Utilizing Cryogenic Ion Vibrational Spectroscopy to Investigate How Side Chains Affect Structure and Solvation of Small Peptides

By

Summer L. Sherman

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN – MADISON

2022

Date of final oral examination: 5/18/2022

The dissertation is approved by the following members of the Final Oral Committee:

Etienne Garand, Professor, Chemistry (Analytical, Physical)
Gilbert Nathanson, Professor, Chemistry (Materials, Physical)
Timothy Bertram, Professor, Chemistry (Analytical, Physical)
Silvia Cavagnero, Professor, Chemistry (Analytical, Chemical Biology, Organic, Physical)

Graduate School
UNIVERSITY OF WISCONSIN-MADISON

Candidate for the degree of PHD

Sherman, Summer Lee
9077929561

Major: *Chemistry PHD*
Minor: *Distributed GMIN*

We, the undersigned, report that as a committee we have examined ***Summer Sherman*** on *Wednesday, May 18, 2022*, and upon the work done in the subjects named and upon the dissertation presented by the candidate we find that the candidate may properly be admitted to the degree of Doctor of Philosophy.

By signing this warrant I am confirming that I have also approved this student's UMI abstract.

Committee Member Name		Committee Member Signature	Date
Garand, Etienne	Advisor		Reader
BERTRAM, TIMOTHY H			Reader
CAVAGNERO, SILVIA			
NATHANSON, GILBERT M			Reader

The following committee members dissent from the report:

Dissenting Member Name	Dissenting Member Signature	Date
Summer Sherman		05/18/22

Dissertation approved by the Graduate School on _____

Utilizing Cryogenic Ion Vibrational Spectroscopy to Investigate How Side Chains Affect Structure and Solvation of Small Peptides

Summer L. Sherman

Under the supervision of Professor Etienne Garand

at the

UNIVERSITY OF WISCONSIN – MADISON

Molecular interactions play a key role in the structure and reactivity of molecules. In biological systems, the major driving force dictating the molecule's structure is determined by intramolecular H-bonding and intermolecular H-bonding with the aqueous environment. In addition to these interactions in biological systems, electronic and steric effects by various amino acid side chains and their placement can compete against these interactions to dictate the final system's overall structure and therefore function. By understanding the fundamentals of each molecular interaction as well as the interplay of these effects it is possible to tune these molecules to achieve a desired structure and function. There is much focus on understanding these effects in their entirety and theoretical work has been increasing complexity and accuracy for these types of systems, however, these computations may be expensive due to their size and types of interactions that they are trying to characterize (intramolecular H-bonding, intermolecular solvent interactions,

solvent-solvent interactions, dispersion, electronic effects). Theoretical work has been increasing complexity and accuracy for these types of systems, however due to the size and types of interactions, these computations may be expensive. Particularly, modeling the non-covalent interactions of intramolecular H-bonding in addition to intermolecular H-bonding with solvent is extremely complicated. Thus, there is a need for experimental work of these systems. Here, we use Cryogenic Ion Vibrational Spectroscopy (CIVS) to study the effects of amino acid side chains and microsolvation on the small amino acids and peptides.

First, a re-examination of the solvation of $GlyH^+(H_2O)_n$ was performed. Two different laser techniques were used (IRMPD vs IRPD) to determine that although there is a different experimental scheme, the resulting population of conformers in the gas-phase remains the same. Rice-Ramsperger-Kassel-Marcus (RRKM) theory of unimolecular reaction rates was used to understand reaction kinetics in the gas-phase. These results demonstrated that for the microsolvation of small amino acids and peptides, the conformer distribution is determined by gas-phase equilibrium energetics unless the interconversion barriers are larger than the H_2O binding energy or the internal energy at room temperature.

Second, the conformational and isomeric population effects were probed for a series of eight tripeptides containing different amounts and orderings of glycine and alanine residues (Gly-Gly-Gly to Ala-Ala-Ala) to sample all permutations of the methyl side-chain position. This was done to provide a comprehensive view of the effects of simple side-chain on the structure of the peptide. IR-IR double resonance spectroscopy was performed to gain experimental conformational data and electronic structure predictions were used to assist in determining the effects of the methyl side chain through proton affinities. The data suggest that there are three main families of

conformations which are defined by protonation site and internal hydrogen bonds, and the relative contributions of each family is highly dependent on the exact amino acid sequence of the tripeptide.

Third, the microsolvation of protonated Gly-Gly-Gly and Ala-Ala-Ala is investigated and compared. The IRPD spectra of $Gly_3H^+(H_2O/D_2O)_2$ and $Ala_3H^+(H_2O/D_2O)_{1-2}$ are obtained and compared to theoretical computations. The conformations for each of the solvated tripeptide clusters are compared against each other as well was to the unsolvated peptide. The data suggest that with the addition of the first water, the extra energy from the binding energy of water brings the initial internal energy of the complex to be above the transition state barrier, allowing each of the bare conformers to form the new lowest energy one water complex. However, a small proportion of the minor conformer determined from the bare species, with the addition of a water molecule, was found for each tripeptide. It was thought that this minor conformer was kinetically trapped, however, with the addition of the second water, it is shown that both tripeptides adopt the same two major conformations found by the addition of the water molecule to the major conformation of the one-water cluster and they also follow their same minor conformations. The population distribution for the minor conformer in the two-water system is much larger, showing that there is interconversion of the species and therefore these species become energetically favorable.

Fourth, the experimental spectra and preliminary calculations and assignments of $Betaine(H_2O)_{0-6}$ are presented. An IR-IR two-color approach was adapted on this instrument and experimental spectra for $Betaine(H_2O)_{2-4}$ are presented with this approach. The solvation effects on a small amino acid with a positive, shielded charge are analyzed and discussed. In particular, the two-color approach is shown to be a valuable asset for delicate hydrogen bonding networks.

Fifth, the experimental spectra of a glycine analog, 1,3-Dimethylhistidine$(H_2O)_{0-12}$ are presented. The overlaid spectra of H_2O and D_2O are shown for 1-2 H_2O cluster sizes and are compared to calculated spectra. Computations were run for the non-ionized and zwitterionic forms of glycine and the difference in energy was taken for the lowest energy species at each cluster size. These calculations suggest that it might be possible to see the zwitterionic form of glycine starting at 5 water molecules and it will form the zwitterion at 7 water molecules and above. More work is needed on this system to see the zwitterionic structure of the glycine analog in the gas-phase to determine how many water molecules is necessary to induce zwitterion formation of the simplest amino acid.

Sixth, a chapter was written in in conjunction with the Wisconsin Initiative for Science Literacy (WISL) to explain the work described in this thesis to a non-scientific audience. This was done in order to bring more understanding and engagement to gas-phase spectroscopy and the work that can be done with this technique, such as monitoring and disentangling molecular interactions in IRPD spectra.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Graduate school, through its many ups and downs, has helped shape me into the person and scientist that I am today. Moving far from home taught me the importance of having a family away from family. Graduate school is not something done alone, and during this process I have been blessed with having many people surround and push me forward. These people have been there to commiserate when things were going wrong in lab, celebrate when things were going well, be there for me when there were family emergencies, stand by my side as I figured out a healthy work-life balance and in general just be great friends, colleagues, and bosses. Though this list is not exhaustive, since there have been so many people who have interacted, supported, and inspired me throughout this process, I do want to thank some people who have been with me along the way. I am deeply grateful for so many of the interactions that I have had in graduate school. Who would have thought that a first-generation college student, raised by three amazingly strong women, would be here writing her dissertation?!

First, I want to thank a couple important teachers from my high school: Mr. Nakamura, Mr. Buck, and Mr. Williamson. Both Mr. Nakamura and Mr. Buck had taught me Chemistry at Chaminade through honors and AP classes. Both of their classes made me fall in love with chemistry; they gave me passion and excitement for the field. Both teachers were extremely invested in me as a student and a person while I was at Chaminade. Most notably, Mr. Buck would chat with me about my strive for perfection, which in hindsight was not completely healthy. He wanted to make sure that when I was out of high school and faced “real” challenges, that I would not destroy myself over not being completely perfect all the time; and this is something that I will keep with me for the rest of my life. I also wanted to thank Mr. Williamson for being a calming

presence in high school and would go the extra mile for his students. I am glad that even after 9 years, I still love going to visit these teachers and talk about my recent accomplishments.

Second, I wanted to thank a few influential people from my undergraduate university. From my first experience with conducting research, I wanted to thank Dr. Robert ("Skip") Pomeroy, Marissa Tessman, Luis Camarda, and Joey Mason. Each of these wonderful individuals guided me in some way on my journey to graduate school, but most importantly showed me what working in a research lab was really like. This is where I formed my first research family and fell in love with conducting experiments and being a part of a group effort. Before my undergraduate career, I believed I was just going to get my bachelor's degree and then get a job, but with the mentorship of Skip and the rest of the people in my lab, I decided to pursue a PhD. I would also like to thank Cole Carter (PhD at Northwestern University) and Josh Cox (PhD at Harvard University) who both embarked on the graduate school journey a year ahead of me, who also gave me guidance on the graduate school application process and were there to answer my calls when I was crying and freaking out about everything.

Third, I am extremely grateful to Etienne for all he has done as a graduate advisor. He has always made himself available for discussions about current projects as well as take time to come into the lab when things were going less than ideally. I am also extremely grateful for Etienne's understanding when family emergencies came up. As someone who stresses and constantly pushes herself, having Etienne say "It's okay, take all the time you need" was needed. I don't know how I would have been able to go through it all without having him respond that way and I am extremely grateful. Even when I was down on myself about my perceived lack of progress, he assured me that I was on the right track. One other thing that I appreciate about Etienne is that he is always

there to encourage the group; though it might not be direct, he will always look for things that he can nominate his students for, or that would fit well with their future goals.

Fourth, I would like to thank past and current Garand Group members who made my time in the group what it was. I was trained by Jon and Kaitlyn and am so grateful for their teachings. It was also entertaining to sit next to Jon in the office and having our late-night chats when we were the last ones in the office. I also worked closely with Kaitlyn on all my projects, and she hosted me when I was visiting graduate schools. I wanted to give Kaitlyn a special shoutout for being the research partner I never thought I would have. From listening to emo music in lab while taking data together all the way to chatting about everything going on in our lives and the department, I am so appreciative to have had someone to be so close and work with so well. Casey was always someone I could be real with and vent about things going on, which was necessary to function in graduate school and I appreciate him so much. In addition, Cole and I joined the group at the same time and have gone through our whole graduate journey together. When we first started in the group, I was too much stress for Cole to handle (!) but I am glad that once we got settled in, we got closer and have been able to share this journey together. I am, also, so appreciative of Gina's constant kindness and checking in on me and my family and will cherish this forever. Kathy has always been there to help me with whatever questions I had and was good in keep me updated on the instrument. Katharina has been instrumental in guiding me in the next coming stages of post-doc life and showing me how exciting it all is. She is also someone I can talk about trash tv with, and I appreciate those talks more than anything. Finally, I wanted to thank Kenny, Grace, and Bri for being such good colleagues and I know that they are going to doing such great work in the future.

Fifth, I would like to thank my first-year graduate school friends who have stuck by me all throughout this journey. Rachel Bergin, Rachel Miller, Madison Fellows, Kendall Kamp, Zac Dyott, Josh Miller, Mary Katherine Andrews have been with me since day one of graduate school and some before day one! Though everyone was on their own path, we were here for each other for all successes and failures. I am so glad that I found such an amazing group of people to surround myself with throughout my time here. Special shout-out to Zac who I taught my first class with as my partner TA and has really been there with me every step of the way.

Sixth, I would like to thank my friends outside of graduate school. Sabrina has been with me and supporting me from UCSD and really has been there for me every step of the way. She pushes me when I need it and lets me sit back when she sees that is best for myself as well. Evey, Alexa, Georgia, and Vanessa have all stuck with me through the good times and bad times and each constantly connected with me throughout the whole journey. My OTF friends have also been essential for me. I never realized how important it was to also have people who are outside of academia to just hang out with. They have helped me on my physical health journey and really are a driving force for me continuing on with it even when I do not want to.

Seventh, I wanted to thank my CHOPs team and all of the faculty that I have been able to work with in the department. CHOPs allowed me to work behind the scenes on something I am extremely passionate about: making the department more diverse and in particular giving opportunities for first-generation students to actually see what graduate school is like. I have got to work with some many amazing people in this department through my time in CHOPs. It also allowed me to network more with the faculty in this department. I can honestly say that I feel that there are many people in the department that are my mentors and they always look out for me. It was wonderful to hear that people on this committee helped Etienne draft a nomination for me to

win the Farrington Daniels Ethical Leadership Award. There have also been many faculty and staff that I have worked with during my time grading for their classes each semester. I have appreciated my time with all of these people and will miss working with them all very dearly.

Eighth, I also would like to thank my family (my mom, aunt, and grandma) for their support and encouragement throughout my life. My mom had sacrificed so much for me to go to private school and get the best education possible. She always said, that though she might not leave me any physical inheritance in life, what she would leave me is my education, and for that I thank her so much. She has given me so much love and support and has been my number one supporter throughout my whole life. My aunt has constantly been there to read my essays, give me corrections, teach me history, and so much more. She has gone above and beyond what any aunt should do for their niece, and I thank her so much for everything she has done. Which brings me to my grandma. Firstly, she had raised my mom and aunt to be the amazing women that they are, who in turn, raised me. She taught me to be positive and happy throughout anything life throws at you. She was one of the strongest women I know, and even with her passing, what she wanted most for me in life was to be happy and finish my PhD. I know how proud she is of me.

Finally, I would like to thank my amazing fiancé, Andrew, and my dog, Triton. Andrew has been with me through so many ups and downs of my own graduate school journey, as well as his own. He has given me so much love and support through all of this and has always pushed me to do my best. He is there for me when I want to talk about science and there for me when I absolutely do not. As I have been finishing up my PhD, he has celebrated with me every milestone that I have come through and I am eager to continue to my post doc and be his support while he finishes his PhD. I also am grateful that he has expanded my family to include his own and for all their constant support as well. Andrew is my best friend and really is my perfect partner. Last but

not least, I wanted to say thank you to my dog, Triton. He came into my life in the beginning of my third year of graduate school and has been one of the best things to happen to me. He is one of the biggest reasons that I have a good work-life balance and appreciate the balance I have. When I am down and don't want to leave the house, he gives me a reason to get up and go out. He is a constant support for me and is always just happy to be around me without expectations.

With all of that said, and yes, I know it was a lot, I am excited to see the current and future students succeed and excel the lab forward. From working with the younger students, I know that Etienne is going to be busy with all the amazing work that they will do.

CHAPTER 1

Introduction

1.1 Overview

Molecular interactions play a crucial role in the structure and reactivity of molecules. For instance, large peptides and proteins are not just defined by the composition or number of their amino acids, but also by the 3-D shape that they form in nature. This 3-D shape is not only controlled by the intramolecular H-bonding forces within the peptides and proteins, but also by the H-bonding intermolecular forces that they can have with their environment, namely, water. The intramolecular forces of the molecule can drastically change the 3-D conformation given the size of the peptide or protein as well as the R-groups of each amino acid that make up the longer structure. Based on the acidity/basicity or sterics of the R-group, each peptide or protein can take on a different conformation, which in turn, changes the function of the species. In addition to amino acid substituents affecting the structure and, therefore, function of the biological molecule, the influence of an aqueous environment must also be considered since this is where many biological molecules exist. Thus, by studying the structural changes induced by R-groups, their placement within peptides, as well as solvation of these peptides, information can be induced about how these factors influence the structure and therefore function of these biomolecules and hopefully give insight to larger species.

Computational chemistry is a strong tool that many use to help study these molecular interactions. However, modeling of these systems suffers two major limitations depending on the model: cost and accuracy. Computational costs can place a limit on the size of the molecule or the accuracy of the method being used. For instance, though ab initio methods such as CCSD and CCSD(T) are extremely accurate,[1] the computational time scales are $O(N^6)$ and $O(N^7)$, respectively, with N representing system size. As the system gets larger, the computational time increases and therefore the cost of each computation also increases. This places a limit on the size

of the system to be computed at extremely accurate computations. In contrast, Molecular Dynamics (MD) methods can be used to compute very large systems including large proteins within aqueous environments, yet they are not very accurate.[2-3] Therefore, there is a compromise to be made to be able to get a reasonable computation approximation within a reasonable computational cost. Within such a compromise, many computational methods also perform poorly while measuring non-covalent interactions, such as any H-bonding interactions. Therefore, it is necessary to be able to experimentally analyze molecules with H-bonding interactions. These experiments will hopefully serve to help benchmarking of theory to be able to full describe these systems with cheap and accurate computations.

Since these biological molecules are in aqueous environments, experimentally, it would be the most beneficial to probe the molecules in the solution phase. However, this has its own set of limitations. Though probing in the solution phase can provide information of the system as a whole, such as in FT-IR and UV-VIS, it does not give information about individual interactions. In addition, large peptides and proteins do not exist as one 3-D structure. Rather, there is a conformer population associated with each one. This conformer population is influenced by its aqueous environment as well as the chain length and R-group of its amino acid substituents and has a distribution which is dependent on the binding energy of each water molecule as well as the energy it takes to rotate in space. The information extracted from solution phase experiments gives a picture of the average of all the conformers in the solution and by itself does not provide conformer-specific structural information. To be able to gather more insight into specific interactions in the molecule, it is necessary to acquire experimental data in the gas phase.

Mass Spectrometry is a useful and versatile technique to study molecular ions in the gas phase. Not only does it allow for the identification of the ion by its mass-to-charge ratio (m/z), but it also

allows for isolation of specific masses. Further advances in mass spectrometry have allowed for activation techniques which fragment ions of interest, giving some structural information such as atom connectivity. Some of these activation techniques include collision induced dissociation (CID),[4-8] surface-induced dissociation (SID),[9-12], electron transfer dissociation (ETD),[13-16] infrared multiphoton dissociation (IRMPD)[17-25] and blackbody infrared radiative dissociation (BIRD).[26-27] The high vacuum environment of mass spectrometry further allows for experiments that are "sensitive" in nature such as probing reaction intermediates, probing and isolating various steps in the catalytic cycle, or producing large water clusters. Though activation techniques allow for more structural information than just m/z of a particular ion, they do not provide information about non-covalent interactions. The results from these methods can also get confusing very quickly with increasing molecule size due to the number of resulting fragments of the molecules. Therefore, scientists have to rely on experimental databases because not all peaks can be assigned directly. However, mass spectrometers can easily be coupled with other techniques to be able to extract this information as long as there is a change in mass involved. By then being able to observe the change of mass as a function of the coupling technique, one can monitor the course and timescale of the reaction. This can help deconvolute data analysis of molecules.

This isolated, high vacuum environment in mass spectrometry makes it ideal to analyze many systems. To be able to gather more direct structural information, it is possible to couple mass spectrometry with spectroscopy. In general, a common method for securing structural information of non-covalent interactions is by absorption spectroscopy. However, this type of spectroscopy requires a high density of sample to produce sufficient absorbance values, and this is not found in mass spectrometry, detailed further in Section 1.3. Action spectroscopy, which does not rely on the amount of light absorbed by the ions as it does in UV/VIS and IR spectroscopies, are

compatible and can provide detailed structural information of the species. These schemes involve the photodissociation of molecular ions, which then causes a change of mass, allowing this change to be monitored by the mass spectrometer. During photodissociation, chemical bonds, molecular clusters, or weakly bonded tags are decomposed as a result of the resonant absorption of light by the molecule. By monitoring the change in mass as a function of laser wavelength, a spectrum can be produced which gives structural information of the system.

Since a change in mass is required, this means that a bond must be dissociated during resonant absorption of the wavelength. The type of bond being broken in a specific local environment will determine the wavelength of light or number of photons required to break the bond. For instance, a covalent O-H bond in H_2O requires 41,550 cm^{-1} of energy to break.[28] Yet, the asymmetric stretch of H_2O appears at 3756 cm^{-1}. This means that in order to fully break the covalent O-H bond in water, it would require 11 IR photons minimum to reach this threshold at this resonant wavelength. In a normal experiment, it is possible to change experimental timings in such a way that an instrument may cycle at 5 Hz while the laser cycles at 10 Hz, leading to 2-3 photons of light hitting the molecule. However, to achieve 11 IR photons minimum, a high-power laser would be required. The IRMPD process, however, can lead to broader, non-linear spectra, is capable of changing the absorption cross-section after initial absorption, and can make the molecules more sensitive towards the higher energy kinetically-trapped conformers that require less photons for dissociation.[29-34] Alternatively, as stated above, UV/VIS photons can be used and provide linear spectra of the molecules. Further, UV/VIS can be coupled with IR photons to do conformational two-laser experiments.[35-40] However, utilizing UV/VIS light can limit the molecules under observation since it requires there to be an adequate chromophore embedded in the molecule.

In order to overcome the limitations addressed above, a tag molecule that weakly attaches itself to the analyte can be used so that one photon of light at a wavelength of resonance is sufficient to see a change of mass, as found in infrared photodissociation (IRPD).[41-45] The weakly bound tag is usually an inert molecule such as H_2, D_2, N_2, among others, and is chosen such that there is a sufficient mass difference, has a low binding energy, and does not perturb the system or its resulting spectrum. When the system is resonant at a specific wavelength, it undergoes intramolecular vibrational energy redistribution which is sufficient to disassociate the tag species off the parent ion. By tagging the system there is less restriction on what can be studied.

In cryogenic ion vibrational spectroscopy (CIVS), ions are introduced into the instrument by electrospray ionization (ESI) and therefore are not tagged upon entry into the system as they are in supersonic expansion. Rather, the ions travel into a cooled ion trap and are held there by voltages for a set amount of time. During this time, a buffer gas that is seeded with the tagged species is introduced to provide collisional cooling to the molecules while attaching the tag. The low binding energy of the tag species chosen (usually ~100's of cm^{-1}) allows the ions to be probed even into the far "fingerprint" region of the IR. The experimental spectrum can be compared direction to theoretical calculated spectrum. Though the tag used is chosen to minimize interactions with the molecular ion, it still might have effects with specific N-H and O-H groups and cause small redshifts in their vibrational bands. Therefore, there still must be awareness of this fact while using a tagged complex.

CIVS provides many advantages to studying molecular ionic species and clusters. Due to its cryogenic temperatures, it can produce well-resolved spectra which can be compared directly, given a linear scaling factor, to theoretical calculated spectra. The following chapters will describe

various systems that have been studied with this technique as well as cares that must be taken when reporting the resulting spectra.

Chapter 3 begins with detailing the cares that one must take when comparing IRPD spectra to IRMPD as well as when assigning specific conformers to IRPD spectra. This chapter is a response to a comment paper published comparing the two types of spectroscopies for the $GlyH^+(H_2O)_n$ system, previous published[46-47] and detailed in Dr. Kaitlyn C. Fischer's Thesis (2019).[48] When comparing the spectra taken by the two methods, certain peaks that were in the IRPD spectra are not in the IRMPD spectra. The group doing the comparison determined that the appearance of these peaks in our IRPD spectra that were not present in their IRMPD spectra must be from kinetic trapping of conformers. In this chapter, IRMPD spectra of these clusters were taken with CIVS and found to be similar to theirs since the peaks that were in the IRPD spectra were due to the D_2 interactions with certain N-H groups. Further, RRKM analysis was used to determine at what barrier height kinetic trapping would occur and bring to light what kinetic trapping means for gas-phase species.

Chapter 4 continues by comparing a series of tripeptides: the protonated forms of Gly-Gly-Gly (GGG), Ala-Gly-Gly (AGG), Gly-Ala-Gly (AGA), Gly-Gly-Ala (GGA), Ala-Ala-Gly (AAG), Ala-Gly-Ala (AGA), Gly-Ala-Ala (GAA), and Ala-Ala-Ala (AAA). The side chain in alanine replaces a hydrogen in glycine with a methyl moiety, thus giving rise to a chiral center in the molecule. Incrementally changing the placement and amount of alanine residues in the protonated tripeptide allows us to draw conclusions on how R-group placement of the electron-donating $-CH_3$ group affects the conformation and population ratios of all species. Since the methyl groups are small enough to not provide steric effects, any changes in structure or conformer population is directly correlated to effects of the electron-donating R-group.

Chapter 5 expands on the results of Chapter 4 by comparing the solvation of Gly_3H^+ to the solvation of Ala_3H^+ with one and two waters. Previously, the conformers of $Gly_3H^+(H_2O)$ were investigated.[49] It was found that the binding energy of one water molecule was enough to overcome the barrier corresponding to a cis to trans amide rotation. The corresponding population conformation was predominately this new trans conformer as well as a minor conformer of the unsolvated triglycine. The addition of a second water molecule, however, does not have the same large binding energy as the first water did to the system. Therefore, by solvating further, we can see how conformers change as a function of solvation. Comparing the microsolvation of Gly_3H^+ to Ala_3H^+, given the electron-donating effects of methyl groups in Ala_3H^+, we can see how these effects may influence solvation and conformer population. Further solvation will reveal that if sterics are not a factor, the conformer population of the two tripeptides may converge to the same conformer, dictated by solvation, or maintain differences due to the electron donating group of the methyl.

Chapter 6 explores the not fully completed work of the microsolvation of a different system, betaine, whose structure is similar to that of glycine yet contains a permanent positive charge induced by trimethylation of the amine. Due to the shielded nature of the charge, we can see a different solvation motif to that of glycine[46-47] in which the water molecules, after solvation of betaine's alcohol group, hydrogen bond to other water molecules. Due to the extensive and sensitive hydrogen bonding network that occurs upon solvation, a two-color approach is used to resolve features that are not attainable with indirect IRMPD. Tentative assignments for each cluster are made, yet without a mass selection step within the traps, conformer specific spectroscopy could not be done. Of interest, one of the lowest energy conformers of DFT and MP2 yielded a structure which contains a H_3O^+ in the solvation network and a neutrally charged betaine lacking a hydrogen

on its alcohol group. With conformer specific spectroscopy utilizing the two-color[50] technique, we would be able to see if this "floating hydrogen" conformer can be seen as early as solvation with three water molecules. This "floating hydrogen" would be the initial framework necessary to induce a zwitterionic structure in the gas-phase. Future work is described in this chapter as well.

Chapter 7 expands on work done in Chapter 6 and aims to answer the question of how many water molecules is necessary to induce a zwitterionic species. Without solvent and within the gas-phase, many amino acids and peptides are known to be in their non-ionized form, however, in solution they are zwitterionic. Glycine is the smallest amino acid and has been shown to follow the trend from above. Thus, work has been done to find a glycine representative which contains a permanent positive charge (necessary for mass spectrometry), yet with the charge shielded such that the glycine base or the added water molecules do not interact with it. Computations have been run to determine at which point we might be able to see this zwitterionic structure. From DFT computations, it has been determined that for glycine, this may occur around seven or eight water molecules, yet CIVS would be able to give us experimental evidence for this. Two glycine analogs have been probed with CIVS and their pros and cons are described. Utilizing the strengths and weaknesses of each molecule, a new glycine analog must be applied to be able to finish this project.

Chapter 8 finishes up this work by containing a chapter which communicates my research to the public. Coming from a family in which I am a first-generation college student, I believe that this chapter is extremely valuable. It is important for scientists to be able to describe what they are doing to a common person who is not an expert in this field to be able to educate the masses rather than gate-keep the knowledge.

1.2 Mass Spectrometry (MS)

Mass spectrometry (MS) is an extremely powerful and versatile technique that uses the mass-to-charge ratio (m/z) of an ion as well as any isotopic patterns that the species might have to ascertain its molecular mass and composition. There are vast MS libraries of molecular species which can be referred to for identification of compounds, and it further can be coupled with other techniques to gain more structural information.[51-53] Because of all of this, it makes sense why mass spectrometry is such a widespread technique in a variety of settings and for many different purposes (e.g. for analyzing everything from an individual atom to whole proteins). With this technique, molecules are ionized and studied in the gas phase under high vacuum. This is extremely useful because it not only allows for less collisions of the molecules than in solution phase, but it makes it possible to easily manipulate ions of interest using electric fields which guide these species through the instrument to be analyzed. Due to less molecular collisions, ions that may form specific conformations due to hydrogen bonding or have solvated structures can be conserved and identified. Each mass spectrometer has in general three main components: the ion source, the mass analyzer, and the detector.[54] There are many different types of mass spectrometry ion sources (electrospray ionization,[55-56] chemical ionization,[57-58] etc[59-60]) and mass analyzers available for use, however, the following sections will describe the ion source (electrospray ionization) and mass analyzer (time-of-flight mass spectrometry coupled with infrared spectroscopy) that is used in the experimental set-up of Cryogenic Ion Vibrational Spectroscopy (CIVS).

1.2.1 *Electrospray Ionization (ESI)*

Electrospray ionization (ESI) is the preferred method for ionizing and getting samples into the gaseous phase for CIVS. It was developed in the late 1980s by Dr. John Fenn who ended up earning the Nobel Prize in Chemistry in 2002.[61-63] ESI is shown to be a "soft" ionization technique, which allows for the gentle transfer of molecules, including large biomolecules, into the gas phase without fragmentation occurring. It therefore preserves the structural integrity of the molecule,[55, 64] unlike other methods (such as Electron Impact[65]) that are much harsher and lead to fragmentation. Because of its success in creating intact gaseous ions, it is a very common ionization source for many instruments including liquid chromatography (LC)[66-67] and ESI-MS.[68-69] It can also create gaseous ions from many different types of solutions: solutions in which the molecules are neutral, solutions where there is a charged species already present in solution, and solutions that have charge induced molecules by the addition of an acid or base.

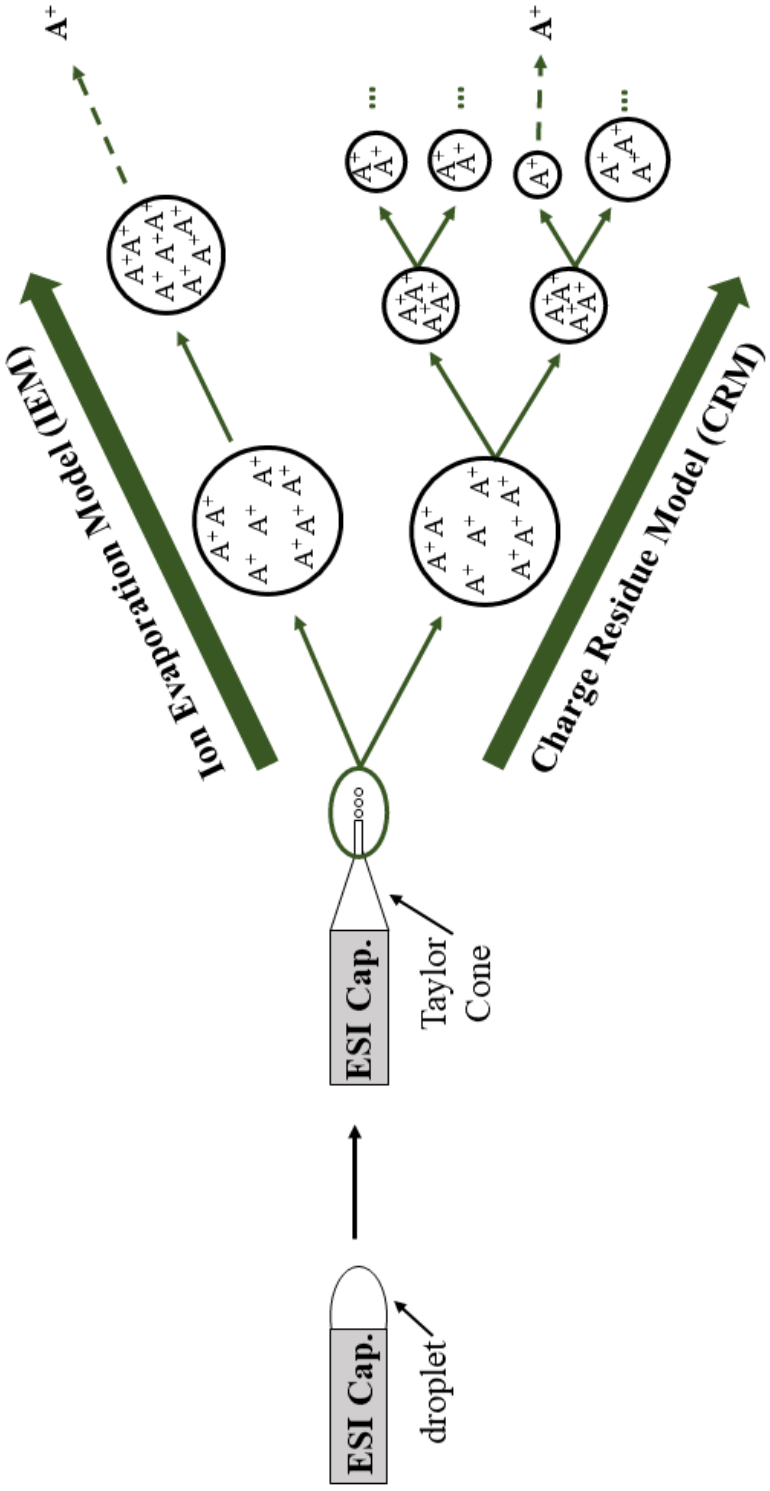
In general, ESI is performed by flowing a dilute solution of the analyte of interest through a small (μm) metal or glass capillary, which has an applied voltage (kV), directly into the mass spectrometer entrance, which also has an applied voltage (V). The large electrical potential difference between the capillary and the MS inlet creates a large electric field around the solution tip. This causes accumulation of like charges in the droplets emerging from the capillary tip. For example, if in positive ion mode, a positive (kV) potential is applied to the capillary and there is a positive (V) potential applied to the inlet of the instrument. When the large positive potential is applied to the capillary, this causes the accumulation of positive charges towards the front of the droplet which is being attracted to the more negative potential on the inlet of the instrument. With sufficiently high capillary potentials, the surface tension of the solution can be overcome, and a Taylor cone will form. From here, a stream of charges droplets will make its way into the

instrument. The droplets will become de-solvated as they travel through the air, and this process can be sped up by heating the capillary or using a nebulizing gas around the ESI set-up. There are two different leading theories of ways that the charged droplets form out of the Taylor Cone: Ion Evaporation Model (IEM) and Charge Residue Model (CRM).[56] A cartoon schematic of bare ion formation by ESI is shown in **Figure 1.1**.

The Ion Evaporation Model (IEM) was first developed by Iribarne and Thomson.[70-71] This model was preferred early on to help explain the generation of large ions from electrosprayed droplets.[61] There are three characteristics for this process: geometric parameter, reaction rate kinetics, and Rayleigh limit.[56] It was found that ion evaporation is played an important role when the size of the droplet fell below 20 nm in diameter (geometric parameter) and that the reaction kinetics depended on the chemical properties of the ion. This model begins with the premise that in every droplet, there are multiple ions. When the droplet gets below the 20 nm diameter limit, the Coulombic repulsions between the ions increase. As the droplet gets smaller, the repulsions continue to increase until it hits the Rayleigh limit. Once the repulsions surpass this threshold, which means that the repulsions are larger than the surface tension of the droplet, an ion is ejected from the droplet. However, Iribarne and Thomson realized that this model could not explain every trend.[70] Some ions with similar solvation energies, had different evaporation rates; and some with very different solvation energies, had the same evaporation rates.

The other leading model is the Charge Residue Model (CRM).[72-73] In this model, it is assumed that in each droplet that is released, there is only one ion, and the ion is released when the solvent evaporates. Therefore, if you start with a larger droplet which might have many ions, as the solvent starts to evaporate, the large droplets divide until only one ion remains. Also, in this model the ionization rate is strongly independent of the identity of the ion.

Figure 1.1. Cartoon depiction of the mechanism of bare ion formation in Electrospray Ionization (ESI) starting from droplet deformation at the capillary.



1.2.2 *Time-of-Flight Mass Spectrometry (TOF-MS)*

Wiley and McLaren[74] were the first to use linear time-of-flight (TOF) mass spectrometry (MS) to determine mass-to-charge ratios based on the time that it takes the ions to travel a known distance. In general, TOF-MS takes ions and imparts the same potential energy (E_p) by an electric field on all ions. The ions, therefore, are ejected into a drift tube at the same time with the same kinetic energy (E_K) which was its original potential energy. Since all ions have the same E_K, this causes species that have different masses (m) to travel at different velocities (v) and therefore arrive at the detector at different times (t). This relationship is expressed as:

$$E_K = \frac{1}{2}mv^2 \qquad \text{(eq 1.2.1)}$$

Since velocity is defined as distance/length (L) traveled per unit time (t), the equation can be re-written as:

$$E_K = \frac{1}{2}m\left(\frac{L}{t}\right)^2 \qquad \text{(eq 1.2.2)}$$

L is defined as distance, or in the case of a TOF tube, it is the length of the tube. Since the electric field that is applied to the ions before ejection imparts potential energy (E_p), E_p can be defined as the product of the electric charge of the ion (q) and the applied potential field (U). This potential energy is converted to kinetic energy as the ion acceleration occurs. This is written as:

$$E_p = qU = E_K \qquad \text{(eq 1.2.3)}$$

Electric charge of an ion (q) can further be defined as the product of the charge of the ion (z) and the elementary charge constant (e). This is written as:

$$q = ze \quad \text{(eq 1.2.4)}$$

By applying Equations 1.2.3 and 1.2.4 into Equation 1.2.2, the following equation relating the mass and charge of an ion to the time-of-flight can be written as:

$$E_K = E_p = \frac{1}{2}m\left(\frac{L}{t}\right)^2 = zeU \quad \text{(eq 1.2.5)}$$

Equation 1.2.5 can be rearranged to show how m/z is related to the square of the time-of-flight multiplied by a constant. This can be written as:

$$\frac{m}{z} = (2eU)\left(\frac{t}{L}\right)^2 \quad \text{(eq 1.2.6)}$$

Throughout the duration of the experiment, variables *e, U,* and *L* are constant which implies that the m/z of the ions is determined by measuring flight time. However, it is still possible that different masses with different charges might have the same m/z. This can occur in a mixture of a peptide and metal, for instance. The metal might have a larger molar mass yet might not be singly charged. Because of this, the metal might have the same m/z as a larger peptide. In this case, spacing of isotopic peaks can be used in order to determine the overall molecular weights of these species.

As it stands, a TOF tube might have limited resolution due to multiple factors, however, the two most important factors are energy resolution and spatial resolution.[74-75] In energy resolution, all the ions ideally start at the same position when they receive the potential energy "kick" from the electric field. However, since ions are not static, they might start at the same position, but their initial velocities might be in different directions. All the ions get the same amount of potential energy and, therefore, kinetic energy. However, since they have different initial velocity directions there might be a turn-around time which causes them to hit the detector later. In spatial resolution, all the ions receive the potential energy "kick" but are not in the same position before entering the TOF. So, ions of the same mass receive a slightly different potential energy "kick." This can be corrected if the detector is placed at the focal point where all masses of the same size with different velocities are at the same distance in time, however, this is usually not the case.

In order to help minimize these resolution fluctuations, a reflectron[76-77] can be used to increase resolution by focusing the ion packet to the detector through applying an electric field (E_r) to slow, stop, and turn around the ions. This reflectron also serves as a secondary mass separation stage without adding too much physical distance as would another TOF tube. The reflectron electric field (E_r) is defined as the electric field potential (U_r) divided by the length of the reflectron (D). Each ion travels a specific depth (d) into the reflectron which is determined by the ion's kinetic energy. The depth is therefore defined as the kinetic energy of the ion (E_K) divided by the product of the electric charge of an ion (q) and the reflectron electric field (E_r). The equation for the depth is described and simplified in Equation 1.2.7.

$$d = \frac{E_K}{qE_R} = \frac{qU}{q\frac{U_R}{D}} = \frac{UD}{U_R} \qquad \text{(eq 1.2.7)}$$

Since the ions are slowed to a stop from their initial velocity, v_0 to 0, over the length (D) of the reflectron, the average velocity is equal to half of the initial velocity, $0.5v_0$. The time (t_D) it takes for the ion to travel by length (D) can be described by Equation 1.2.8.

$$t_D = \frac{D}{\frac{v_0}{2}} = \frac{2D}{v_0} \quad \text{(eq 1.2.8)}$$

However, the reflectron is symmetric. The total time it then takes for the ion to travel through the full reflectron (t_R) can be written as:

$$t_R = 2t_D = 2\left(\frac{2D}{v_0}\right) = \frac{4D}{v_0} \quad \text{(eq 1.2.9)}$$

After the ions exit the reflectron they must travel through the final part of the TOF tube to reach the detector. The TOF tube can therefore be divided into two lengths, the length of the tube before the reflectron (L_1) and the length of the tube after the reflectron (L_2). Therefore, the total time that it takes the ions to travel through the full time-of-flight tube (t_{tube}) can be described as the following:

$$t_{tube} = \frac{L_1 + L_2}{v_0} \quad \text{(eq 1.2.10)}$$

The total time (t_{total}) that it takes the ions to reach the detector is therefore the sum of the time that the ions are in the full TOF tube (t_{tube}) and the time that they are in the reflectron.

$$t_{total} = t_{tube} + t_R = \frac{L_1+L_2+4D}{v_0} \quad \text{(eq 1.2.11)}$$

In order to relate this to mass-to-charge (m/z), Equation 1.2.5 can be rearranged in order to isolate for v_0. Recall from Equation 1.2.2 that the initial velocity (v_0) is equal to the length of the tube (L) traveled per unit time (t).

$$\frac{1}{2}mv_0^2 = zeU \rightarrow v_0 = \sqrt{\frac{2zeU}{m}} \quad \text{(eq 1.2.12)}$$

By substituting v_0 from Equation 1.2.12 into Equation 1.2.11,

$$t_{total} = \frac{L_1+L_2+4D}{\sqrt{\frac{2zeU}{m}}} \quad \text{(eq 1.2.13)}$$

And rearrangement of Equation 1.2.13 can show how m/z can be found in a TOF mass spectrometer. This equation can be written as the following,

$$\frac{m}{z} = \left[\frac{2eU}{(L_1+L_2+4D)^2}\right] t_{total}^2 \quad \text{(eq 1.2.14)}$$

1.3 Infrared (IR) Spectroscopy

Infrared (IR) spectroscopy is commonly used to observe the vibrational motion of atoms in a molecule, which in turn gives structural information for that molecule. The molecular

vibrations occur when the energy of an IR photon is resonant with a vibrational transition of a molecule, such as stretching and bending of bonds and angles. Each vibrational transition occurs at discrete energies, and therefore the frequency at which photons are absorbed gives accurate information about the bond type as well as the local chemical environment surrounding the vibrating atom. Each molecule has a different set of discrete energies that are necessary to induce vibrational motion and by monitoring where these occur, one can obtain a molecular IR spectrum. This spectrum is considered to be "signature" of the molecule and is used to identify and define it.

Since vibrational frequencies are dependent upon the bonds present as well as the local environment that they are found in, the equation for the frequency ($\bar{\upsilon}$) can be written as a product of constants and the square root of the force constant (K), which depends on the bond strength, divided by the reduced mass (μ) of the atoms providing the molecular vibration. The equation can be written as follows,

$$\bar{\upsilon} = \frac{1}{2\pi c}\sqrt{\frac{K}{\mu}} \qquad \text{(eq 1.3.1)}$$

where the reduced mass (μ) of the atoms involved in the molecular vibration is defined as the product of the mass of the first atom (m_1) and the mass of the second atom (m_2) divided by their sum. The equation is defined below,

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \qquad \text{(eq 1.3.2)}$$

As shown by the equations, different vibrational modes will appear at different frequencies. Further, these modes can be perturbed by its local environment. For instance, if the bond is involved in hydrogen bonding (H-bonding), its frequency will become red-shifted.

As stated above, spectroscopy requires the absorption of a photon to see a molecular vibration. Therefore, in conventional IR spectroscopy experiments, the fraction of light that is absorbed by the molecule for a particular vibration is measured. The absorbance for the molecule (*A*) is defined by the ratio of light that is transmitted through the sample (*I*) and the light that the sample receives (I_0). The absorbance equation can be written as the following,

$$\mathrm{A} = log\frac{I}{I_0} \qquad \text{(eq 1.3.3)}$$

Another way to describe absorbance is by Beer's Law,[78] which states that absorbance is equal to the product of the sample's molar absorptivity (ε), the path length that the light travels through the sample (l), and the concentration of the sample (C). Beer's law is written as the following equation

$$\mathrm{A} = \varepsilon lC \qquad \text{(eq 1.3.4)}$$

As the molar absorptivity coefficient is related to the absorption cross section of each vibrational mode ($\sigma(\bar{\upsilon})$) and the number density of the molecules in the sample (ρ), Equation 1.3.4 can be re-written as follows,

$$\mathrm{A} = \sigma(\bar{\upsilon})l\rho \qquad \text{(eq 1.3.5)}$$

Equation 1.3.5 shows that for conventional IR spectroscopy experiments, there must be a large number density of the molecules in the sample (ρ) to cause significant absorbance values. While this is achievable for many experimental set-ups, especially those in the solution phase, this is not suitable for gas-phase measurements of ions. In the gas-phase, though the absorption cross section of each vibrational mode might be large, the number density of molecules is small. Therefore, a different approach to acquire IR spectrum of ions inside of a mass spectrometer is necessary.

1.3.1 *Principles of Infrared Predissociation (IRPD) Spectroscopy*

Since the number density of the molecules in the gas-phase is small, infrared predissociation spectroscopy (IRPD), a type of action spectroscopy, is a technique that can be used to work around this. In conventional spectroscopy, by utilizing Beer's law, Equations 1.3.3 and 1.3.5 can be combined and rearranged to describe and track the intensity of light that passes through the sample at a particular frequency ($I(\bar{\upsilon})$) as the following,

$$I(\bar{\upsilon}) = I_0 e^{-\sigma(\bar{\upsilon}) l \rho} \qquad \text{(eq 1.3.6)}$$

However, in IRPD spectroscopy, the change of intensity of a particular mass peak ($N(\bar{\upsilon})/N_0$), rather than the change of light intensity, is monitored as a function of the frequency. Equation 1.3.6 can be written as follows,

$$N(\bar{\upsilon}) = N_0 e^{-\sigma(\bar{\upsilon}) \phi(\bar{\upsilon})} \qquad \text{(eq 1.3.7)}$$

where N_0 is the initial intensity of the mass peak without laser interaction, $N(v)$ is the intensity of the mass peak after laser interaction, and $\phi(\bar{v})$ is the photon fluence on the sample.

In IRPD, since a mass change is necessary for the generation of an IR spectrum, the parent ion is usually tagged with a weakly bound inert molecule. However, to be able to tag a weakly bound molecule to a parent ion, the system must be cooled to cryogenic temperatures. Each inert tag molecule has a specific temperature range at which it will tag. Ideally, the best candidate for the tag molecule would be one that is inert, has minimal perturbation on the parent ion's vibrational spectrum, has a small binding energy, and can weakly bind to the parent ion in the instruments operating temperature range. Common tag molecules are H_2,[79-80] D_2,[81-82] Ar,[83-84] and N_2.[85-86] In the experiments described in this work, D_2 is the preferred tag of choice for the reasons discussed before. When an IR photon is absorbed by a tagged complex, the molecule undergoes intramolecular vibrational distribution (IVR), which results in the loss of the weakly bound tag.[87-89] Due to the small binding energy of the tag, one photon is enough to induce IVR which results in the loss of the tag. IVR occurs in the order of picoseconds to nanoseconds. The instrumental set-up in the work described here minimizes reattachment of the tag to the parent molecule and therefore the tagged parent and the photofragment are separated in the time-of-flight tube and in the reflectron. Another advantage of IRPD spectroscopy is that it only requires a tabletop laser to obtain spectra, and these are much cheaper and more versatile than the alternative.

D_2 photofragmentation, in the case described above, can be modeled by Rice-Ramsperger-Kassel-Marcus (RRKM) theory.[43, 90-91] There are two main conditions for tag-loss: the binding energy of the tag must be smaller than the wavelength of the resonant laser photon and there must be sufficient energy in the localized vibrational motion corresponding to the tag. However, with IVR the second condition is met because IVR statistically redistributes the energy of the photon

among all the vibrational degrees of freedom. RRKM theory, therefore, computes the probability that this would occur and estimates the timescales for tag loss. The energy dependent rate constant for unimolecular dissociate[92] (*k(E*)*) can be written as the following,

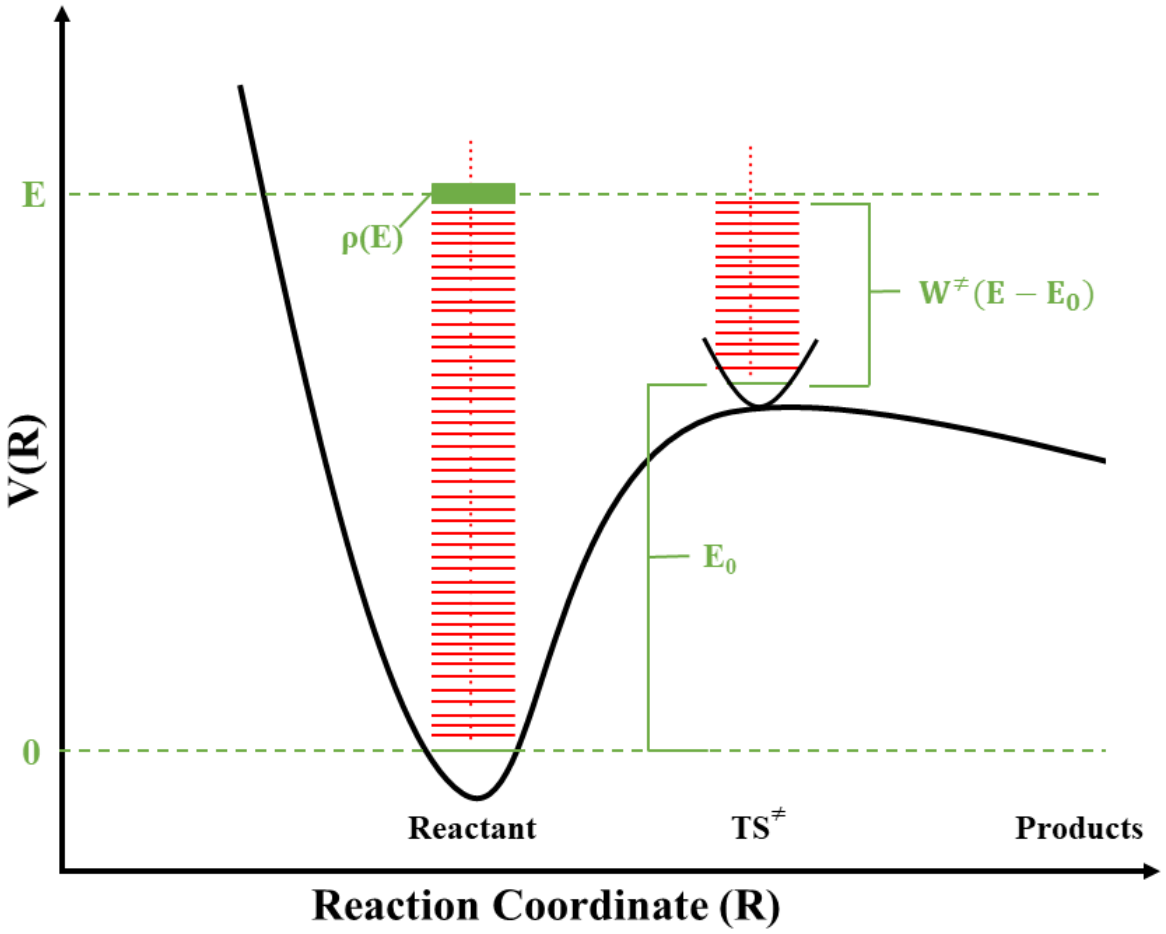
$$k(E^*) = \frac{\sigma W^{\neq}(E-E_0)}{h\rho(E)} \qquad \text{(eq 1.3.8)}$$

where σ is the reaction path degeneracy, $W^{\neq}(E\text{-}E_0)$ represents sum of states between the barrier height of the reaction (or transition state), E_0, and energy E (which usually is defined as the internal energy of the molecule), h is Planck's constant, and $\rho(E)$ which represents the density of states at energy E for the reactant structure. A schematic representation of this is shown in **Figure 1.2**. The Beyer and Swinehart algorithm[93] can be used to determine the sum of states and density of states of the system if the vibrational energy levels of the system are known. When determining the timescale for the disassociation of the tag, the barrier height is assumed to be the binding energy of the tag.

To apply this theory to a molecule discussed in this work, it was found that for Gly_3H^+, the D_2 tag has a binding energy of ~150 cm^{-1}, which was computed at the cam-B3LYP/def2TZVPP level. In the experimental set-up for work done in this thesis, the laser is usually scanned from 1400 to 3800 cm^{-1}. Therefore, RRKM predicts that for this binding energy, the complex dissociates on sub-nanosecond timescales. Since the ion flight time is approximately in the μs range, the photodissociation of the tag is done at a rate that is compatible with the TOF experiment.

However, RRKM not only validates the choice of tag for the species used and the timescale of the instrument, but it can give guidance on whether kinetically trapped species are present in the IRPD spectrum. Since the molecules are cooled quickly through collisional cooling, there is a

Figure 1.2. Schematic of Rice-Ramsperger-Kassel-Marcus (RRKM) Theory for Unimolecular Reaction Rates.



worry that some conformations would not be the lowest energy conformers. When the IRPD spectrum is taken and structural determinations are made, potential energy scans (PES) can be done of moving from one conformation to the other. The energy, *E*, can either be the internal energy of the molecule if there is no solvation, or it can be the binding energy of a water molecule if the species is microsolvated. E_0, would be the largest transition energy of going from one conformer to the other. From there, the sum of states and density of states can be computed and RRKM analysis can take place. It was found that for systems discussed in this work, the minimum energy barrier (E_0) needed to produce conformer half-lives of at least 50 ms at 298 K (the timescale of the work done on this instrument) is approximately the amount of available internal energy of the system (*E*). However, this does not correlate as the chain of the peptide increases. Rather, the minimum energy barrier plateaus for chains longer than 10 monomers. Therefore, for these systems, on the timescale of this instrumentation kinetically trapped species may be produced that have internal energies, *E*, which are much higher than the predicted barrier, E_0. This is discussed and explored further in **Chapter 3**.

1.3.2 *The Harmonic Oscillator Model and Vibrational Selection Rules*

To help interpret experimental spectra, vibrational spectra of various conformers of the species of interest are computed utilizing the harmonic oscillator approximation and then compared to the experimental spectra. In this approximation, atoms are treated as balls connected together by springs. The internuclear potential energy (*V(r)*) between the two atoms at a defined bond distance (*r*), can be written as the following,

$$V(r) = \frac{1}{2}k(r - r_e)^2 + \frac{1}{6}\gamma(r - r_e)^3 + \cdots \qquad \text{(eq 1.3.9)}$$

where k and γ are defined as force constants, r is the bond distance, and r_e is the equilibrium bond distance. If the vibrations being considered have small nuclear displacement, like in fundamental vibrational modes, then higher order terms are negligible. Therefore, for fundamental modes, Equation 1.3.9 can be rewritten as the following:

$$\mathrm{V(r)} = \frac{1}{2}k(r - r_e)^2 \qquad \text{(eq 1.3.10)}$$

Since these vibrations are quantized, the Schrödinger equation can be used to derive the energy and wavefunction for the system. The general form of the Schrödinger equation consists of the Hamiltonian, which is defined as the sum of the kinetic and potential energies, multiplied by a wavefunction ($\psi(r)$) which equals the energy of the system (E) multiplied by the same wavefunction. The general form of the equation is defined as follows,

$$\left(\frac{-\hbar^2}{2m}\nabla^2 + V(r)\right)\psi(r) = E\psi(r) \qquad \text{(eq 1.3.11)}$$

Here, $\hbar$ is defined as Planck's constant divided by 2π and m is defined as the mass. If Equation 1.3.10 is plugged into Equation 1.3.11 and the kinetic energy term is written out,

$$\left(\frac{-\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} + \frac{1}{2}k(r - r_e)^2\right)\psi(r) = E\psi(r) \qquad \text{(eq 1.3.12)}$$

where μ is the reduced mass of the bonded atoms defined in Equation 1.3.2. If Equation 1.3.12 is rearranged to solve for 0,

$$\frac{\partial^2}{\partial r^2}\psi(r) + \frac{2\mu}{\hbar^2}\left(E - \frac{1}{2}k(r - r_e)^2\right)\psi(r) = \ 0 \qquad \text{(eq 1.3.13)}$$

The wavefunction and solution to this equation is only valid for quantized values and therefore the wavefunction and vibrational energy levels are described by the following equations,[94-95]

$$\psi_{\mathrm{v}} = N_{\mathrm{v}} H_{\mathrm{v}} e^{\frac{-\sqrt{\mu k}(r-r_e)^2}{2\hbar}} \qquad \mathrm{v} = 0,1,2,\ldots \qquad \text{(eq 1.3.14)}$$

$$E_{\mathrm{v}} = \frac{h}{2\pi}\left(\frac{k}{\mu}\right)^{\frac{1}{2}}\left(\mathrm{v} + \frac{1}{2}\right) \qquad \mathrm{v} = 0,1,2,\ldots \qquad \text{(eq 1.3.15)}$$

where v is the vibrational quantum number. For Equation 1.3.14, *N*v is the normalization constant for that vibrational level, and *H*v is the Hermite polynomial corresponding to the vibrational level. Equation 1.3.15 can be written more simply as follows:

$$E_{\mathrm{v}} = \hbar\omega\left(\mathrm{v} + \frac{1}{2}\right) = \ h\nu\left(\mathrm{v} + \frac{1}{2}\right) \qquad \mathrm{v} = 0,1,2,\ldots \qquad \text{(eq 1.3.16)}$$

where ω is the angular frequency and ν is the linear oscillator frequency. However, in spectroscopy we are interested when there is an energy difference between two vibrational levels. Therefore, the energy associated with a transition from the 0→1 fundamental level is given by the following equation,

$$\Delta E_{0\rightarrow 1} = \frac{h}{2\pi}\left(\frac{k}{\mu}\right)^{\frac{1}{2}} = h\upsilon_{photon} \quad \text{(eq 1.3.17)}$$

Therefore, if we re-write Equation 1.3.17 to solve for the frequency in wavenumbers ($\bar{\upsilon}$),

$$\bar{\upsilon}_{photon} = \frac{1}{2\pi c}\left(\frac{k}{\mu}\right)^{\frac{1}{2}} \quad \text{(eq 1.3.18)}$$

where *c* is the speed of light.

Though a transition from the fundamental 0→1 level was shown, only vibrational transitions that obey the vibrational selections rules can occur. These selection rules can be derived starting with the transition dipole moment (μ_v).[94-95] This transition dipole moment is correlated to the intensity of the specific allowed transition. This is shown by the following equation,

$$\mu_v = \langle \psi_v | \vec{\mu} | \psi_v' \rangle \quad \text{(eq 1.3.19)}$$

where $\vec{\mu}$ is the molecular dipole moment. The molecular dipole moment can be expanded by a series in terms of (r-r_e) coordinates which results in the following equation,

$$\vec{\mu} = \mu_r = \mu_{r_e} + \left(\frac{\partial \mu_r}{\partial r}\right)(r - r_e) + \left(\frac{\partial^2 \mu_r}{\partial r^2}\right)(r - r_e)^2 + \cdots \quad \text{(eq 1.3.20)}$$

Equation 1.3.20 can be substituted into Equation 1.3.19 to get the following equation,

$$\mu_v = \langle \psi_v | \left[\mu_{r_e} + \left(\frac{\partial \mu_r}{\partial r} \right) (r - r_e) + \left(\frac{\partial^2 \mu_r}{\partial r^2} \right) (r - r_e)^2 + \cdots \right] | \psi_v' \rangle \quad \text{(eq 1.3.21)}$$

The vibrational wavefunction, ψ_v, is defined in Equation 1.3.14.

There are then two conditions in order for a transition to occur. First, the transition dipole must be a non-zero value, and this only can occur when $\Delta v = \pm 1$. Therefore, the second condition is that $\frac{\partial \mu_r}{\partial r} \neq 0$. This implies that there must be a change in the molecular dipole moment in order for absorption to occur.

Since harmonic vibrational computations utilize the harmonic oscillator model to predict molecular vibrations, any anharmonic modes (overtones or combination bands) will not be predicted.[96-98] However, in IRPD spectra of species containing water or species with strong H-bonded moieties, anharmonicities may appear.[99-102] Separate computations could be done, which do not use the harmonic oscillator model, to predict these vibrations.

1.4 References

(1) Harding, M. E.; Metzroth, T.; Gauss, J.; Auer, A. A. Parallel Calculation of CCSD and CCSD(T) Analytic First and Second Derivatives. *J Chem Theory Comput* **2008,** *4* (1), 64-74.

(2) Chmiela, S.; Sauceda, H. E.; Muller, K. R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat Commun* **2018,** *9* (1), 3887.

(3. Heo, L.; Feig, M. Experimental accuracy in protein structure refinement via molecular dynamics simulations. *Proc Natl Acad Sci U S A* **2018,** *115* (52), 13276-13281.

(4) Sleno, L.; Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **2004,** *39* (10), 1091-112.

(5) Johnson, A. R.; Carlson, E. E. Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation. *Anal Chem* **2015,** *87* (21), 10668-78.

(6) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins**2005,** *402*, 148-185.

(7) Williams, B. J.; Kmiec, K. L.; Russell, W. K.; Russell, D. H. Effect of cysteic acid position on the negative ion fragmentation of proteolytic derived peptides. *J Am Soc Mass Spectrom* **2011,** *22* (1), 31-7.

(8) Harvey, D. J. A new charge-associated mechanism to account for the production of fragment ions in the high-energy CID spectra of fatty acids. *J Am Soc Mass Spectrom* **2005,** *16* (2), 280-90.

(9) Zhou, M.; Wysocki, V. H. Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Acc Chem Res* **2014,** *47* (4), 1010-8.

(10) Harvey, S. R.; VanAernum, Z. L.; Wysocki, V. H. Surface-Induced Dissociation of Anionic vs Cationic Native-Like Protein Complexes. *J Am Chem Soc* **2021,** *143* (20), 7698-7706.

(11) Snyder, D. T.; Harvey, S. R.; Wysocki, V. H. Surface-induced Dissociation Mass Spectrometry as a Structural Biology Tool. *Chem Rev* **2022,** *122* (8), 7442-7487.

(12) Harvey, S. R.; O'Neale, C.; Schey, K. L.; Wysocki, V. H. Native Mass Spectrometry and Surface Induced Dissociation Provide Insight into the Post-Translational Modifications of Tetrameric AQP0 Isolated from Bovine Eye Lens. *Anal Chem* **2022,** *94* (3), 1515-1519.

(13) Kim, M. S.; Pandey, A. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics* **2012,** *12* (4-5), 530-42.

(14) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004,** *101* (26), 9528-33.

(15) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* **2006,** *1764* (12), 1811-22.

(16) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal Chem* **2018,** *90* (1), 40-64.

(17) Heine, N.; Asmis, K. R. Cryogenic ion trap vibrational spectroscopy of hydrogen-bonded clusters relevant to atmospheric chemistry. *International Reviews in Physical Chemistry* **2014,** *34* (1), 1-34.

(18) Oomens, J.; Sartakov, B. G.; Meijer, G.; von Helden, G. Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *International Journal of Mass Spectrometry* **2006,** *254* (1-2), 1-19.

(19) Polfer, N. C.; Oomens, J. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom Rev* **2009,** *28* (3), 468-94.

(20) Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; von Helden, G.; Oomens, J. Infrared fingerprint spectroscopy and theoretical studies of potassium ion tagged amino acids and peptides in the gas phase. *J Am Chem Soc* **2005,** *127* (23), 8571-9.

(21) Polfer, N. C. Infrared multiple photon dissociation spectroscopy of trapped ions. *Chem Soc Rev* **2011,** *40* (5), 2211-21.

(22) Shi, Y.; Du, M.; Ren, J.; Zhang, K.; Xu, Y.; Kong, X. Application of Infrared Multiple Photon Dissociation (IRMPD) Spectroscopy in Chiral Analysis. *Molecules* **2020,** *25* (21).

(23) Maitre, P.; Scuderi, D.; Corinti, D.; Chiavarino, B.; Crestoni, M. E.; Fornarini, S. Applications of Infrared Multiple Photon Dissociation (IRMPD) to the Detection of Posttranslational Modifications. *Chem Rev* **2020,** *120* (7), 3261-3295.

(24) Wu, R.; McMahon, T. B. Infrared multiple photon dissociation spectra of proline and glycine proton-bound homodimers. Evidence for zwitterionic structure. *J Am Chem Soc* **2007,** *129* (16), 4864-5.

(25) Scuderi, D.; Correia, C. F.; Balaj, O. P.; Ohanessian, G.; Lemaire, J.; Maitre, P. Structural characterization by IRMPD spectroscopy and DFT calculations of deprotonated phosphorylated amino acids in the gas phase. *Chemphyschem* **2009,** *10* (9-10), 1630-41.

(26) Dunbar, R. C. BIRD (blackbody infrared radiative dissociation): evolution, principles, and applications. *Mass Spectrom Rev* **2004,** *23* (2), 127-58.

(27) Schnier, P. D.; Price, W. D.; Jockusch, R. A.; Williams, E. R. Blackbody infrared radiative dissociation of bradykinin and its analogues: energetics, dynamics, and evidence for salt-bridge structures in the gas phase. *J Am Chem Soc* **1996,** *118* (30), 7178-89.

(28) Luo, Y.-R. Comprehensive Handbook of Chemical Bond Energies. *Boca Raton* **2007,** *1st Edition*.

(29) Cao, W.; Zhang, H.; Yuan, Q.; Zhou, X.; Kass, S. R.; Wang, X. B. Observation of Conformational Simplification upon N-Methylation on Amino Acid Iodide Clusters. *J Phys Chem Lett* **2021,** *12* (11), 2780-2787.

(30) Zhang, H.; Cao, W.; Yuan, Q.; Zhou, X.; Valiev, M.; Kass, S. R.; Wang, X. B. Cryogenic "Iodide-Tagging" Photoelectron Spectroscopy: A Sensitive Probe for Specific Binding Sites of Amino Acids. *J Phys Chem Lett* **2020,** *11* (11), 4346-4352.

(31) Fernandez-Lima, F. A.; Becker, C.; Gillig, K. J.; Russell, W. K.; Tichy, S. E.; Russell, D. H. Ion mobility-mass spectrometer interface for collisional activation of mobility separated ions. *Anal Chem* **2009,** *81* (2), 618-24.

(32) Hoaglund-Hyzer, C. S.; Lee, Y. J.; Counterman, A. E.; Clemmer, D. E. Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures. *Anal Chem* **2002,** *74* (5), 992-1006.

(33) Masson, A.; Kamrath, M. Z.; Perez, M. A.; Glover, M. S.; Rothlisberger, U.; Clemmer, D. E.; Rizzo, T. R. Infrared Spectroscopy of Mobility-Selected H+-Gly-Pro-Gly-Gly (GPGG). *J Am Soc Mass Spectrom* **2015,** *26* (9), 1444-54.

(34) Voronina, L.; Masson, A.; Kamrath, M.; Schubert, F.; Clemmer, D.; Baldauf, C.; Rizzo, T. Conformations of Prolyl-Peptide Bonds in the Bradykinin 1-5 Fragment in Solution and in the Gas Phase. *J Am Chem Soc* **2016,** *138* (29), 9224-33.

(35) de Vries, M. S.; Hobza, P. Gas-phase spectroscopy of biomolecular building blocks. *Annu Rev Phys Chem* **2007,** *58*, 585-612.

(36) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. Guanine tautomerism revealed by UV–UV and IR–UV hole burning spectroscopy. *The Journal of Chemical Physics* **2001,** *115* (10), 4604-4611.

(37) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. Interplay of intra- and intermolecular H-bonding in a progressively solvated macrocyclic peptide. *Science* **2012,** *336* (6079), 320-3.

(38) Rizzo, T. R.; Stearns, J. A.; Boyarkin, O. V. Spectroscopic studies of cold, gas-phase biomolecular ions. *International Reviews in Physical Chemistry* **2009,** *28* (3), 481-515.

(39) Dean, J. C.; Buchanan, E. G.; Zwier, T. S. Mixed 14/16 helices in the gas phase: conformation-specific spectroscopy of Z-(Gly)n, n = 1, 3, 5. *J Am Chem Soc* **2012,** *134* (41), 17186-201.

(40) Gloaguen, E.; Tardivel, B.; Mons, M. Gas phase double-resonance IR/UV spectroscopy of an alanine dipeptide analogue using a non-covalently bound UV-tag: observation of a folded peptide conformation in the Ac-Ala-NH2–toluene complex. *Structural Chemistry* **2015,** *27* (1), 225-230.

(41) Heiles, S.; Cooper, R. J.; DiTucci, M. J.; Williams, E. R. Hydration of guanidinium depends on its local environment. *Chem Sci* **2015,** *6* (6), 3420-3429.

(42) Voss, J. M.; Marsh, B. M.; Zhou, J.; Garand, E. Interaction between ionic liquid cation and water: infrared predissociation study of [bmim](+).(H2O)n clusters. *Phys Chem Chem Phys* **2016,** *18* (28), 18905-13.

(43) Wolk, A. B.; Leavitt, C. M.; Garand, E.; Johnson, M. A. Cryogenic ion chemistry and spectroscopy. *Acc Chem Res* **2014,** *47* (1), 202-10.

(44) Klyne, J.; Schmies, M.; Miyazaki, M.; Fujii, M.; Dopfer, O. Stepwise microhydration of aromatic amide cations: water solvation networks revealed by the infrared spectra of acetanilide(+)-(H2O)n clusters (n </= 3). *Phys Chem Chem Phys* **2018,** *20* (5), 3148-3164.

(45) Suzuki, Y.; Hirata, K.; Lisy, J. M.; Ishiuchi, S. I.; Fujii, M. Double Ion Trap Laser Spectroscopy of Alkali Metal Ion Complexes with a Partial Peptide of the Selectivity Filter in K(+) Channels horizontal line Temperature Effect and Barrier for Conformational Conversions. *J Phys Chem A* **2021,** *125* (44), 9609-9618.

(46) Fischer, K. C.; Sherman, S. L.; Garand, E. Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and beta-Alanine. *J Phys Chem A* **2020,** *124* (8), 1593-1602.

(47) Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J Phys Chem A* **2019,** *123* (15), 3355-3366.

(48) Fischer, K. C. Using Cryogenic Ion Vibrational Spectroscopy to Investigate the Microsolvation of Amino Acids and Small Peptides. *University of Wisconsin - Madison, Ph.D.* **2019**.

(49) Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H(+).(H2O). *J Phys Chem A* **2018,** *122* (41), 8213-8221.

(50) Yang, N.; Duong, C. H.; Kelleher, P. J.; Johnson, M. A.; McCoy, A. B. Isolation of site-specific anharmonicities of individual water molecules in the I−·(H2O)2 complex using tag-free, isotopomer selective IR-IR double resonance. *Chemical Physics Letters* **2017,** *690*, 159-171.

(51) Dubascoux, S.; Le Hécho, I.; Hassellöv, M.; Von Der Kammer, F.; Potin Gautier, M.; Lespes, G. Field-flow fractionation and inductively coupled plasma mass spectrometer coupling: History, development and applications. *Journal of Analytical Atomic Spectrometry* **2010,** *25* (5), 613.

(52) Materazzi, S. Mass Spectrometry Coupled to Thermogravimetry (TG-MS) for Evolved Gas Characterization: A Review. *Applied Spectroscopy Reviews* **1998,** *33* (3), 189-218.

(53) Goebbert, D. J.; Meijer, G.; Asmis, K. R.; Iguchi, T.; Watanabe, K. 10 K Ring Electrode Trap—Tandem Mass Spectrometer for Infrared Spectroscopy of Mass Selected Ions **2009**, 22-29.

(54) Glish, G. L.; Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* **2003,** *2* (2), 140-50.

(55) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization-principles and practice. *Mass Spectrometry Reviews* **1990,** *9* (1), 37-70.

(56) Wilm, M. Principles of electrospray ionization. *Mol Cell Proteomics* **2011,** *10* (7), M111 009407.

(57) Munson, B. Chemical Ionization Mass Spectrometry: Theory and Applications **2006**.

(58) Bertram, T. H.; Kimmel, J. R.; Crisp, T. A.; Ryder, O. S.; Yatavelli, R. L. N.; Thornton, J. A.; Cubison, M. J.; Gonin, M.; Worsnop, D. R. A field-deployable, chemical ionization time-of-flight mass spectrometer. *Atmospheric Measurement Techniques* **2011,** *4* (7), 1471-1479.

(59) O'Connor, P. B.; Costello, C. E. A high pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry ion source for thermal stabilization of labile biomolecules. *Rapid Commun Mass Spectrom* **2001,** *15* (19), 1862-8.

(60) Barber, M.; Bordoli, R. S.; Sedgiyick, R. D.; Tyler, A. N. Fast Atom Bombardment of Solids (F.A.B.) : A New Ion Source for Mass Spectrometry. *J.C.S. Chem. Comm* **1981**.

(61) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

(62) Yamashita, M.; Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **2002,** *88* (20), 4451-4459.

(63) Yamashita, M.; Fenn, J. B. Negative ion production with the electrospray ion source. *The Journal of Physical Chemistry* **2002,** *88* (20), 4671-4675.

(64) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal Chem* **2013,** *85* (1), 2-9.

(65) Orient, O. J.; Chutjian, A. A compact, high-resolution Paul ion trap mass spectrometer with electron-impact ionization. *Review of Scientific Instruments* **2002,** *73* (5), 2157-2160.

(66) Berman, P.; Futoran, K.; Lewitus, G. M.; Mukha, D.; Benami, M.; Shlomi, T.; Meiri, D. A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in Cannabis. *Sci Rep* **2018,** *8* (1), 14280.

(67) Prince, J. T.; Marcotte, E. M. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Anal Chem* **2006,** *78* (17), 6140-52.

(68) Yunker, L. P.; Stoddard, R. L.; McIndoe, J. S. Practical approaches to the ESI-MS analysis of catalytic reactions. *J Mass Spectrom* **2014,** *49* (1), 1-8.

(69) Nguyen, H. P.; Schug, K. A. The advantages of ESI-MS detection in conjunction with HILIC mode separations: Fundamentals and applications. *J Sep Sci* **2008,** *31* (9), 1465-80.

(70) Thomson, B. A.; Iribarne, J. V. Field induced ion evaporation from liquid surfaces at atmospheric pressure. *The Journal of Chemical Physics* **1979,** *71* (11), 4451-4463.

(71) Iribarne, J. V. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **1976,** *64* (6), 2287.

(72) Dole, M.; Hines, R. L.; Mack, L. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Gas Phase Macroions. *Macromolecules* **2002,** *1* (1), 96-97.

(73) Mack, L. L.; Kralik, P.; Rheude, A.; Dole, M. Molecular Beams of Macroions. II. *The Journal of Chemical Physics* **1970,** *52* (10), 4977-4986.

(74) Wiley, W. C.; McLaren, I. H. Time-of-Flight Mass Spectrometer with Improved Resolution. *Review of Scientific Instruments* **1955,** *26* (12), 1150-1157.

(75) Mamyrin, B. A. Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry* **2001,** *206* (3), 251-266.

(76) Mirsaleh-Kohan, N.; Robertson, W. D.; Compton, R. N. Electron ionization time-of-flight mass spectrometry: historical review and current applications. *Mass Spectrom Rev* **2008,** *27* (3), 237-85.

(77) Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. Mass reflection: a new nonmagnetic time-of-flight high resolution mass- spectrometer. *Zh. Eksp. Teor. Fiz.* **1973,** *64* (1), 82-89.

(78) Swineharf, D. F. The Beer-Lambert Law. *Journal of Chemical Education* **1962,** *39* (7).

(79) Bouchet, A.; Klyne, J.; Ishiuchi, S. I.; Fujii, M.; Dopfer, O. Conformation of protonated glutamic acid at room and cryogenic temperatures. *Phys Chem Chem Phys* **2017,** *19* (17), 10767-10776.

(80) Kamrath, M. Z.; Garand, E.; Jordan, P. A.; Leavitt, C. M.; Wolk, A. B.; Van Stipdonk, M. J.; Miller, S. J.; Johnson, M. A. Vibrational characterization of simple peptides using cryogenic infrared photodissociation of H2-tagged, mass-selected ions. *J Am Chem Soc* **2011,** *133* (16), 6440-8.

(81) Leavitt, C. M.; Wolk, A. B.; Fournier, J. A.; Kamrath, M. Z.; Garand, E.; Van Stipdonk, M. J.; Johnson, M. A. Isomer-Specific IR-IR Double Resonance Spectroscopy of D2-Tagged Protonated Dipeptides Prepared in a Cryogenic Ion Trap. *J Phys Chem Lett* **2012,** *3* (9), 1099-105.

(82) Chakraborty, A.; Schmahl, S.; Asmis, K. R. Isomer-Specific Vibrational Spectroscopy of Microhydrated Lithium Dichloride Anions: Spectral Fingerprint of Solvent-Shared Ion Pairs. *Chemphyschem* **2021,** *22* (11), 1036-1041.

(83) Gerardi, H. K.; Gardenier, G. H.; Viswanathan, U.; Auerbach, S. M.; Johnson, M. A. Vibrational predissociation spectroscopy and theory of Ar-tagged, protonated Imidazole (Im) Im1–3H+·Ar clusters. *Chemical Physics Letters* **2011,** *501* (4-6), 172-178.

(84) Patzer, A.; Chakraborty, S.; Solcà, N.; Dopfer, O. IR Spectrum and Structure of the Phenyl Cation. *Angewandte Chemie* **2010,** *122* (52), 10343-10346.

(85. Cismesia, A. P.; Bell, M. R.; Tesler, L. F.; Alves, M.; Polfer, N. C. Infrared ion spectroscopy: an analytical tool for the study of metabolites. *Analyst* **2018,** *143* (7), 1615-1623.

(86) Chatterjee, K.; Dopfer, O. Spectroscopic identification of fragment ions of DNA/RNA building blocks: the case of pyrimidine. *Phys Chem Chem Phys* **2020,** *22* (30), 17275-17290.

(87) Rijs, A. M.; Oomens, J. IR Spectroscopic Techniques to Study Isolated Biomolecules. *Top Curr Chem* **2015,** *364*, 1-42.

(88) Boyall, D.; Reid, K. L. Modern studies of intramolecular vibrational energy redistribution. *Chemical Society Reviews* **1997,** *26* (3), 223.

(89) Stannard, P. R.; Gelbart, W. M. Intramolecular Vibrational Energy Redistribution. *The Journal of Physical Chemistry* **1981,** *85* (24).

(90. Baer, T.; Mayer, P. M. Statistical Rice-Ramsperger-Kassel-Marcus Quasiequilibrium Theory Calculations in Mass Spectrometry. *American Society for Mass Spectrometry* **1996,** *8*, 103-115

(91) Callear, A. B. Chapter 4 Basic RRKM Theory**1983,** *24*, 333-356.

(92) Levine, R. D. Molecular Reaction Dynamics **2009**.

(93) Beyer, T.; Swinehart, D. F. Algorithm 448: number of multiply-restricted partitions. *Communications of the ACM* **1973,** *16* (6), 379.

(94) Levine, I. N. Quantum Chemistry. *Pearson* **2014,** (Seventh Edition ed).

(95) McQuarrie, D. A.; Simon, J. D. Physical Chemistry: A Molecular Approach. *University Science Books* **1997,** (1st edition).

(96) Tabor, D. P.; Kusaka, R.; Walsh, P. S.; Sibert, E. L., 3rd; Zwier, T. S. Isomer-Specific Spectroscopy of Benzene-(H2O)n, n = 6,7: Benzene's Role in Reshaping Water's Three-Dimensional Networks. *J Phys Chem Lett* **2015,** *6* (10), 1989-95.

(97) Carnegie, P. D.; McCoy, A. B.; Duncan, M. A. IR spectroscopy and theory of Cu+(H2O)Ar2 and Cu+(D2O)Ar2 in the O-H (O-D) stretching region: fundamentals and combination bands. *J Phys Chem A* **2009,** *113* (17), 4849-54.

(98) Zabuga, A. V.; Kamrath, M. Z.; Rizzo, T. R. Franck-Condon-like Progressions in Infrared Spectra of Biological Molecules. *J Phys Chem A* **2015,** *119* (42), 10494-501.

(99) Iogansen, A. V. Direct proportionality of the hydrogen bonding energy and the intensification of the stretching ν(XH) vibration in infrared spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **1999,** *55* (7-8), 1585-1612.

(100) Rozenberg, M.; Loewenschuss, A.; Marcus, Y. An empirical correlation between stretching vibration redshift and hydrogen bond length. *Physical Chemistry Chemical Physics* **2000,** *2* (12), 2699-2702.

(101) Rozenberg, M.; Shoham, G.; Reva, I.; Fausto, R. A correlation between the proton stretching vibration red shift and the hydrogen bond length in polycrystalline amino acids and peptides. *Phys Chem Chem Phys* **2005,** *7* (11), 2376-83.

(102) McCoy, A. B. The role of electrical anharmonicity in the association band in the water spectrum. *J Phys Chem B* **2014,** *118* (28), 8286-94.

CHAPTER 2

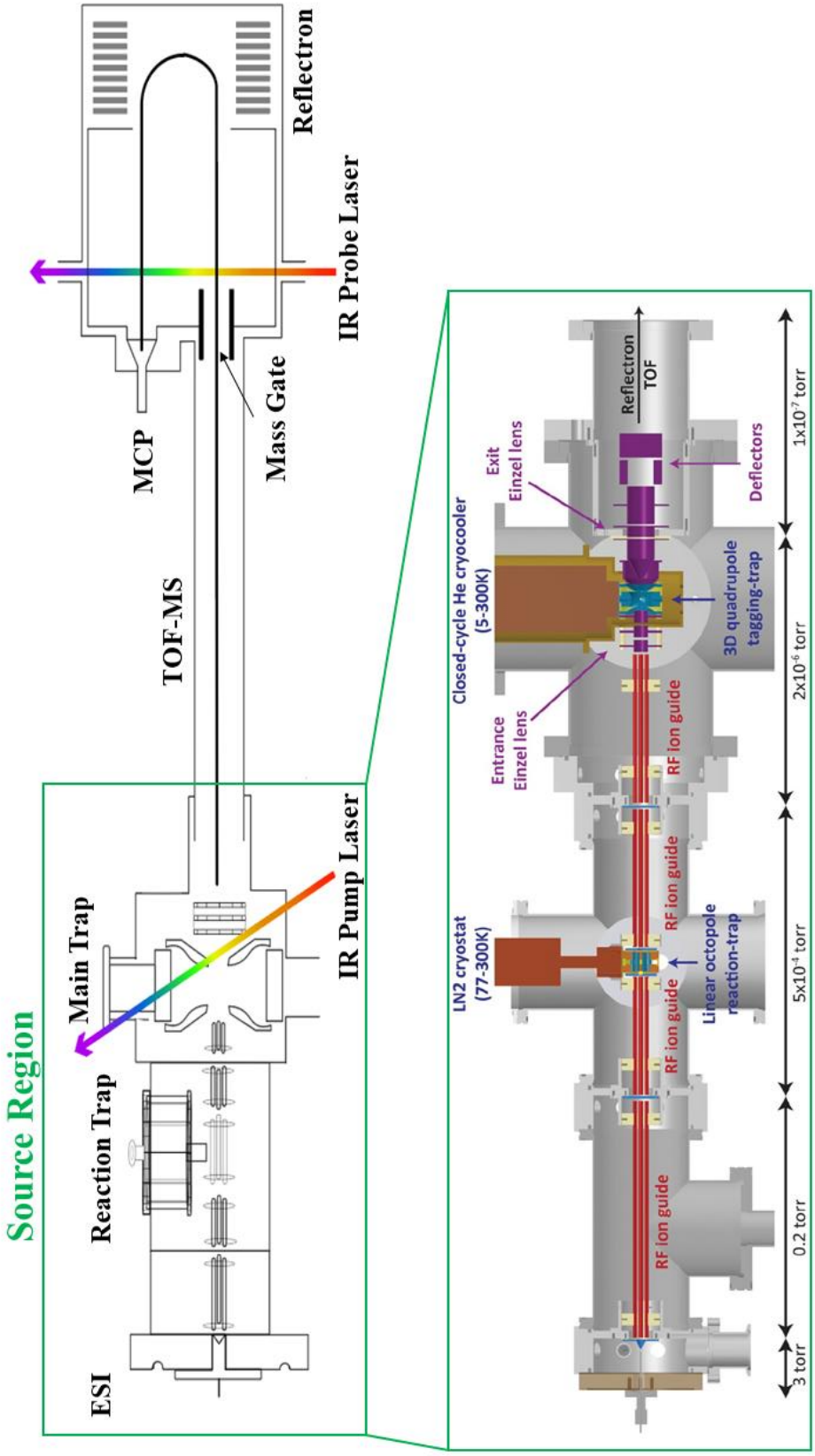
Experimental Details

2.1 Overview

Our home-built dual cryogenic ion trap vibrational spectrometer, shown in **Figure 2.1**, was first built in Spring 2013, based on the design by Mark Johnson.[1-2] This instrument includes an ESI source, reaction trap, main trap, time of flight, reflectron, MPC, and two tunable lasers; all of these components will be discussed later on in this chapter. However, there have been several modifications done to the original design throughout the years to allow for acquisition of better quality of data as well allow as permit for more sophisticated experiments. Firstly, the instrument was upgraded with a longer time-of-flight (TOF) tube to increase mass resolution. Secondly, a linear octupole trap, termed "reaction trap," was built and placed before the main "tagging trap." The reaction trap is cooled with a liquid nitrogen cryostat and can hold the temperature in the trap between 77-300K. This trap serves as an initial trapping stage which allows for more stable signal and greater efficacy of the tagging done in the main tagging trap. It allows for water clustering on the bare ion to create solvated clusters, which is necessary for experiments discussed in Chapters 3-5, and is also useful for procuring reaction intermediates.[3] Finally, the last major modification was the introduction of a second 10 Hz Nd:YAG pumped optical parametric oscillator (OPO)/optical parametric amplifier (OPA) laser, later referred to as the "pump" laser, which was directed through the center ring electrode of the main tagging trap in order to perform IR-IR isomer-specific spectroscopy.[4] Though these major modifications were completed in the beginning years of the laboratory, future work on modifications that will enable the instrument to be more malleable in its design dependent on project have continued as projects for current and future graduate students.

For all experiments discussed in this thesis in Chapters 3-5, protonated ions were generated via electrospray ionization[5-6] (ESI) of an ~1mM solution of the analyte dissolved in methanol with

Figure 2.1. Diagram of the homebuilt dual cryogenic ion trap vibrational spectrometer.



trace amounts of formic acid. The analytes of interest were either obtained through Millipore Sigma or were synthesized according to standard methods (see Chapter 4) by a collaborator in the Gelman Group at UW-Madison. Hexapole ion guides transferred the ions through a series of differentially pumped regions and into the linear octupole ion reaction trap, held at 80 K by the liquid nitrogen cryostat. Collisions of the ions in the trap with the helium buffer gas introduced in an ~1 ms burst at the beginning of the trap sequence induced thermalization of the ions. Additionally, H_2O or D_2O could be seeded into the gas line to allow for cluster formation. These ions were then transferred by another set of hexapole ion guides to a 3D quadrupole ion tagging main trap held at 10 K by a closed-cycle helium cryostat. Here, the ions were further thermalized with helium buffer gas seeded with 10% D_2, allowing the formation of D_2-tagged adducts. These tagged adducts were then extracted into a time-of-flight (TOF) mass spectrometer at 10 Hz. The $[ion^+]\cdot(D_2)_n$ adducts were mass-selected with a gated deflector and intersected with the output of a 10 Hz Nd:YAG pumped optical parametric oscillator (OPO)/optical parametric amplifier (OPA) infrared "probe" laser. Resonant absorption of a single photon was enough to induce the loss of the weakly-bound (~500 cm^{-1} or ~6 kJ/mol) D_2 tag. The photofragment ions were then separated from the parent adducts by a two-stage reflectron and were directed to a multi-channel plate (MCP) detector. The yield of these photofragment ions as a function of laser wavelength thus produced the one-laser IRPD spectra.

Harmonic vibrational frequency calculations are used in order to interpret experimental spectra. A basin hopping program is used to find a collection of conformers for each system (see section 2.14). Relative lowest energy structures are optimized, and harmonic frequencies are calculated at the cam-B3LYP/def2-TZVP/GD3BJ level, which has provided good agreement with experimental spectra for similar systems.[3-4, 7-10] These relative low energy structures obtained with

cam-B3LYP are also optimized at the MP2/def2-TZVP level and corrected by the ZPE from the corresponding DFT calculation. MP2 energetics provide a more accurate energetic landscape for the small peptides and water clusters. Though the DFT and MP2 energetics may disagree at points since DFT overestimates the strength of hydrogen bonds,[11-13] the lowest energy structures from each method respectively can help guide relative stabilities of the various structures in the experimental setup. In this work investigating small peptides and their solvated species, there have been many different isomers that arise. Thus, a basin hopping program, developed in the Hopkins group at the University of Waterloo,[14-18] is used to help identify many possible conformations of each species. The program can randomly move specified parameters such as bond angles, distances, and dihedrals, as well as a combination of all three, to get a large sampling of the conformational landscape. The low energy structures found from this program are subsequently re-optimized at higher levels of theory/basis sets until they have been optimized at cam-B3LYP/def2TZVP and MP2/def2TZVP, which are our highest levels of theory for these systems, and DFT frequencies calculated. This program has been found to be extremely useful in systems in which the molecules are fluxional and thus can adopt many different conformations, as well as when the molecule is solvated, and each solvent molecule can move independently in the network. Though chemical intuition may serve to provide good starting points for each molecule, this program allows for the systematic conformer search of many variations of the same species.

2.2 Ion Source

Electrospray ionization (ESI) was used as the ion source for all experiments performed in this work, with a typical ESI voltage between +1-3kV. Solutions of the analyte were made by diluting the chemical in HPLC grade methanol and adding a couple drops of formic acid to induce

protonation. These samples are approximately 1mM of analyte. The created solution was placed in a syringe (Hamilton) with a 30 μm fused silica capillary tip (New Objective) and pumped at a rate of 1-3 μm/min using a Harvard Apparatus syringe pump. The electrosprayed positively charged ions entered the mass spectrometer through a 10 cm stainless steel capillary with an inner diameter of 0.3 mm which is held in place by a compression fitting. In this first stage of the instrument, there is a 15 L/s mechanical pump that pumps this region to an operating pressure of ~2.5-3 Torr. The ions then pass through a 0.75 mm biased brass skimmer that divides the first and second stages of the instrument. In the second stage of the instrument, the ions continue traveling through a hexapole ion guide, which is comprised of alternating poles with opposite radio frequency (RF) phases of frequency ~1MHz to keep the ions from diverging as they travel through the instrument. The power supplies that drive these hexapole ion guides were built according to the design of O'Connor and coworkers.[19-20] Another 15 L/s mechanical pump pumps this region to an operating pressure of ~300 mTorr.

2.3 Reaction Trap Region

Once the ions have passed through the hexapole ion guides within the second stage of the instrument, they pass through a 1.3 mm diameter biased aperture and continue into the third stage of the instrument. This stage of the instrument is pumped by a 300 L/s turbomolecular pump to ~5 x 10^{-5} Torr. Another set of hexapole ion guides transfer the ions into a linear octupole ion trap. The hexapole ion guides are also powered by a RF power supply identical to those used for the hexapole ion guides in the second stage of the instrument (see section 2.2) The ion trap's entrance and exit each have a 3.8 mm copper aperture which is biased to provide the entrance and exit trapping potentials. The exit potential is pulsed between a high and low DC voltage in order to trap or

extract the ions, respectively. This ion trap is mounted to the base of a liquid nitrogen cryostat (Janis, now known as Lakeshore Cryotronics, model VPF-100), which is electrically insulated by a 1mm thick thermally conductive Shapal ceramic (Precision Ceramics). Indium foil (Millipore Sigma) is used on both sides of the ceramic to provide good thermal contact and conductivity from the cryostat to the trap. The liquid nitrogen cryostat has an operational temperature range of 77-300 K, where the temperature is monitored by a silicone diode and controlled by a 50 W cartridge heater located at the base of the cryostat.

To allow for collisional cooling, either He or N_2 buffer gas is pulsed into the trap using a solenoid pulse valve connected by a short polytetrafluoroethylene (PTFE) tube. Water cluster and reaction intermediates can be formed in this trap. H_2O or D_2O is seeded into the gas line through a valve in which ~10-20 μL of solvent is placed into a reservoir that the gas flows over. Water cluster formation by process of condensation begins to occur at ~220 K, but optimal cluster formation and stability occurs at 80 K which is the temperature used throughout this work. The operational pressure for this reaction trap is ~2.2-5.6 x 10^{-4} Torr with the lower operational pressures better suited for non-solvated ions and the higher operational pressures better suited for larger formation of cluster. To control the extent of clustering, experimental parameters can be adjusted. These parameters include applied electric potentials, reaction trap timings (see section 2.12), buffer gas pressure, as well as buffer gas pulse duration (which is controlled by the width of the pulse).

2.4 Tagging “Main” Trap Region

Once the ions have exited the reaction trap in the third stage, the ions are then transferred into the fourth stage of the instrument by a third set of hexapole ion guides. The fourth stage of the

instrument is pumped by an 800 L/s turbomolecular pump to ~5 x 10^{-8} Torr. The ions enter the fourth stage through a 2 mm biased aperture and are transported by a fourth set of hexapole ion guides, and focused by a set of cylindrical einzel lenses, into a 3D quadrupole ion trap (Jordan TOF). This trap will be referred to as the "main trap" throughout this work. The trap's design is based on that of Wang and Wang[21] and Kamrath et al.[2] This trap is attached to the second stage of a closed-cycle helium cryocooler (Sumitomo RDK-415D, Janis Research Company) with 1.5 W of cooling power at 4.2 K. The closed-cycle helium cryocooler has an operational temperature range of 4.2 - 300 K, where the temperature is monitored, similarly to the reaction trap, by a silicone diode and controlled by a 50 W cartridge heater located at the base of the cryocooler. Indium foil (Millipore Sigma) is used between the trap and second stage of the cryocooler to provide good thermal contact and conductivity from the cryocooler to the trap. The whole cryocooler assembly is enclosed by an aluminum radiation shield that is attached to the first stage of the cryocooler to ensure maximum cooling. The mounted main trap is attached to a linear translation stage (McAllister Technical Services) which is used to align the trap to account for the temperature-induced changes caused by thermal expansion and compression. For typical operation, this linear translation stage, usually has the main tagging trap at a height of 27 mm at 300 K and a height of 29 mm at 10 K.

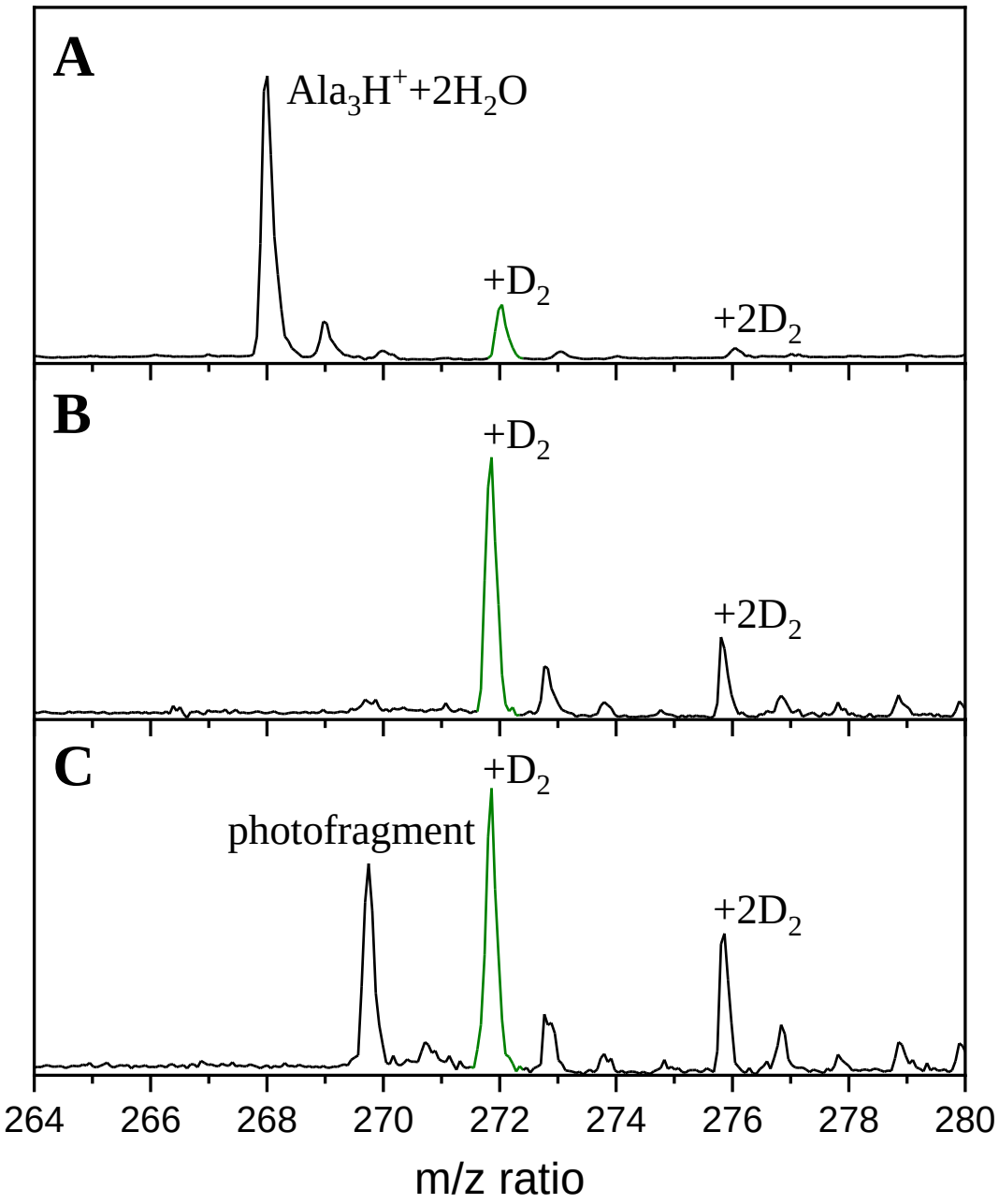
Similarly to the reaction trap, a solenoid pulsed valve is used to introduce buffer gas for collisional cooling to the trap through an attached PTFE tubing. The buffer gas mixture consists of ultrapure 99.999% He with 10% D_2 added for tagging. Though this trap has an operational temperature of 4.2 – 300 K, the trap is usually held at 10 K to induce the formation of D_2-tagging adducts (which can be found to tag between 4 – 50 K). The typical operating pressure with gas pulsed into this trap is ~ 3 x 10^{-6} Torr. After thermalization to 10K, the ions are ejected from the

trap through turning off the trap RF and pulsing the two endcap voltages. These pulsed endcaps act as Wiley-McLaren plates,[22] with the entrance voltage being positive and the exit voltage being negative, to direct the positive ions toward the exit of the trap. This quickly and efficiently extracts ions from the trap and a final grid electrode (TOF liner voltage), located 1 cm from the exit cap, which carries a very large negative voltage, accelerates the ions out of the trap and into the TOF.

2.5 Time-of-Flight (TOF) Region

The TOF region is separated into two stages: the TOF tube (one of the modifications from Mark Johnson's original instrument design, described in Section 2.1) and the photofragment area. The TOF region is pumped by a 300 L/s turbomolecular pump and has an operating pressure of $1x10^{-7}$ Torr. As stated in Section 2.4, once the ions are guided out of the trap by the pulsed Wiley McLaren plates, they pass a third grid electrode which is held at the TOF liner voltage (~-1950V). This helps accelerate the ions into the TOF region through a 3 mm conical aperture. With the initial acceleration kick, all ions receive the same kinetic energy and travel down the first stage of the TOF region which is an ~ 1 m long flight tube, this is where the initial mass separation takes place. The tube, which is held at the liner voltage, contains a set of cylindrical einzel lenses and deflectors at both ends to guide the ions down the tube. The small deflector at the end of this first stage is pulsed to act as a mass gate to isolate the species of interest for IRPD studies at a specific m/z. To be able to select which m/z ions pass through to the second stage of the TOF region, the pulsed deflector is connected to an adjustable timing box which sets the deflector to a high voltage to deflect unwanted ions and, at a specific timing, to a low voltage to allow ions of interest to pass, demonstrated in **Figure 2.2**.

Figure 2.2. Mass spectra of $Ala_3H^+(H_2O)_2 \cdot D_2$ with A) the mass gate and probe laser off, B) the mass gate on and probe laser off or not resonant, and C) the mass gate on and the probe laser resonant with the free carboxylic O—H at 3575 cm^{-1}.



Following the mass gate, the ions enter the second TOF region, the photofragmentation area where they undergo IRPD, or in some cases IRMPD depending on the binding energy of the tag being used. Here, they are intersected by the output of probe laser, discussed in section 2.6 After the laser interaction, all the ions move into a two-stage reflectron (Jordon TOF), which serves as a second mass separation stage and separates the parent ions from its photofragments. The separated ions are directed to a pair of microchannel plate (MCP) detectors, which consists of two 40 mm microchannel plates and is usually held at ~-4.5 kV. The yield of these photofragment ions as a function of laser wavelength thus produces the linear one-laser IR spectra as shown in **Figure 2.3**.

Since ions are differentiated by flight time, it is important to create a calibration to determine the m/z ratio of the new analyte. In general, this is done when starting a new system to make sure the peak observed is the analyte in question. This is important since some analytes might have possible contaminations due to lower purity and therefore there are many peaks within the sample. To begin the calibration, a mass spectrum of an analyte with a known flight time and a similar m/z ratio as the new analyte is taken and is compared to the analyte of interest. Ideally, calibration curves are obtained by taking the mass spectrum of known analytes and enough D_2 tags such that that m/z of the various ions encompasses the m/z of the new analyte. Then the mass spectrum of the new system can be taken under the same TOF/MS settings as the known analytes. This is to make sure that no timings change since flight times can depend on current settings. As derived in **Chapter 1.2.2**, the m/z ratio and flight time are related by Equation 1.2.14. Transformed, this equation used to obtain a linear calibration curve, found by plotting the square root of m/z vs time, is shown by the following equation

Figure 2.3. IRPD spectrum of $Ala_3H^+(H_2O)_2{\cdot}D_2$.

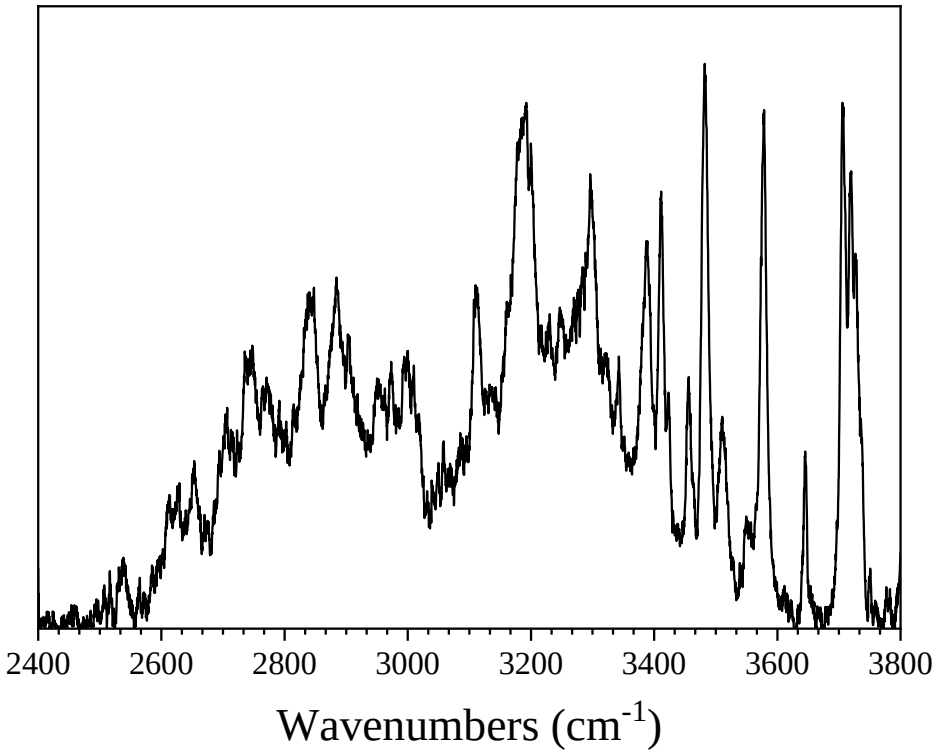
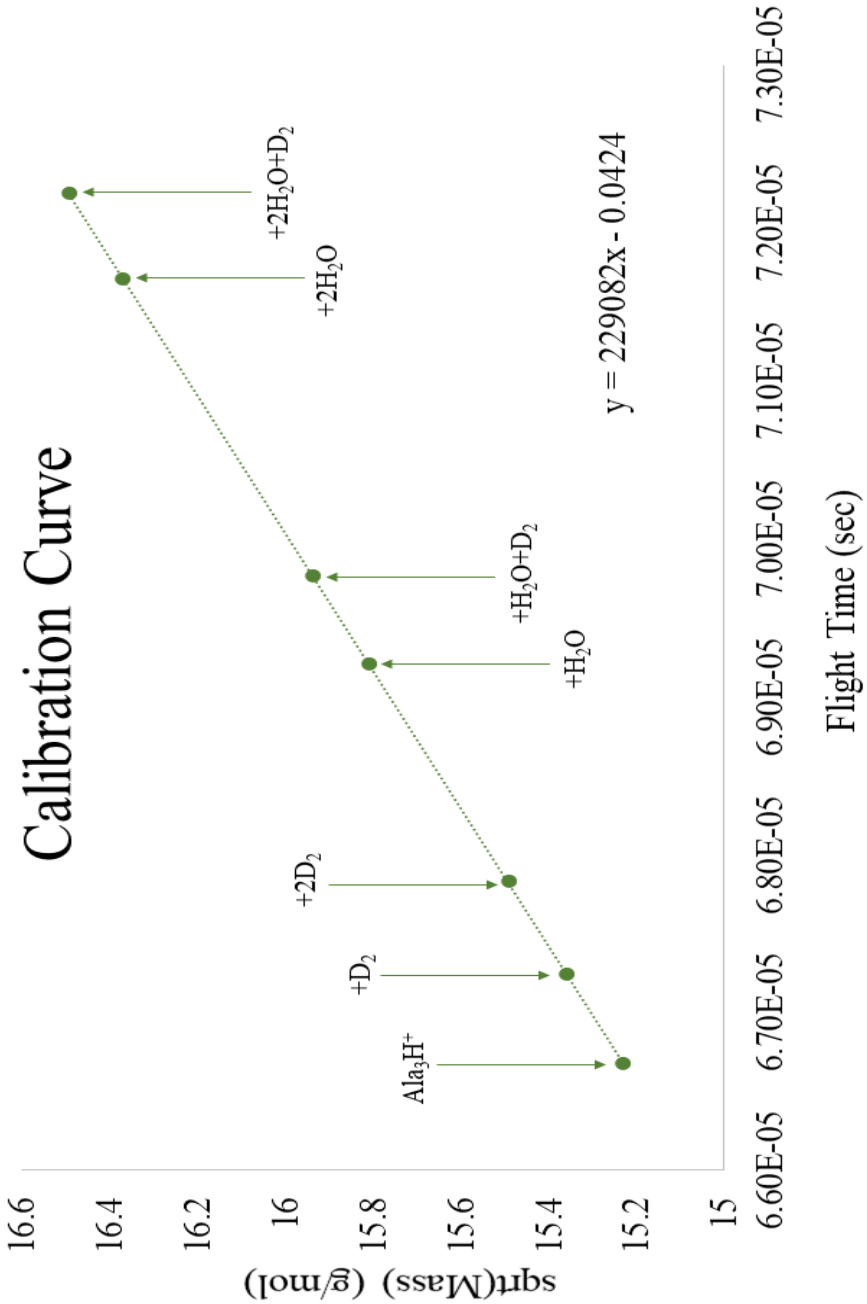


Figure 2.4. Calibration curve for Ala_3H^+ (molar mass: ~232 g/mol) to $Ala_3H^+(H_2O)_2{\cdot}D_2$ (molar mass: ~272 g/mol). Linear regression shown with equation below the line.



$$\sqrt{\frac{m}{z}} = slope * t_{total} + (y - intercept) \quad \text{(eq 2.5.11)}$$

To make a calibration curve, one needs to plot the square root of mass of the known analytes vs their timings. An example of this is shown in **Figure 2.4**. Once a linear curve has been developed, the equation of the line will give the timing of the new analyte.

2.6 Infrared Predissociation Spectroscopy

In the second stage of the TOF region, the ions are intersected by the output of a 1064 nm, 10 Hz Nd:YAG pumped optical parametric oscillator (OPO)/optical parametric amplifier (OPA) infrared "probe" laser. The laser is timed through its Q-switch, such that there is maximal overlap with the ion packet at the focal point. As the laser is scanned, when the laser frequency is resonant with a vibrational mode in the ions, the affected ions undergo intramolecular vibrational energy redistribution which is sufficient to disassociate the tag species off the parent ion, forming the bare ion, termed "enhancement" shown in **Figure 2.2C**. By monitoring the intensity of the photofragment enhancement as a function of wavelength, the linear one-laser IR spectra is generated as shown in **Figure 2.3**.

2.7 IR-IR Double Resonance: Conformer Specific Spectra

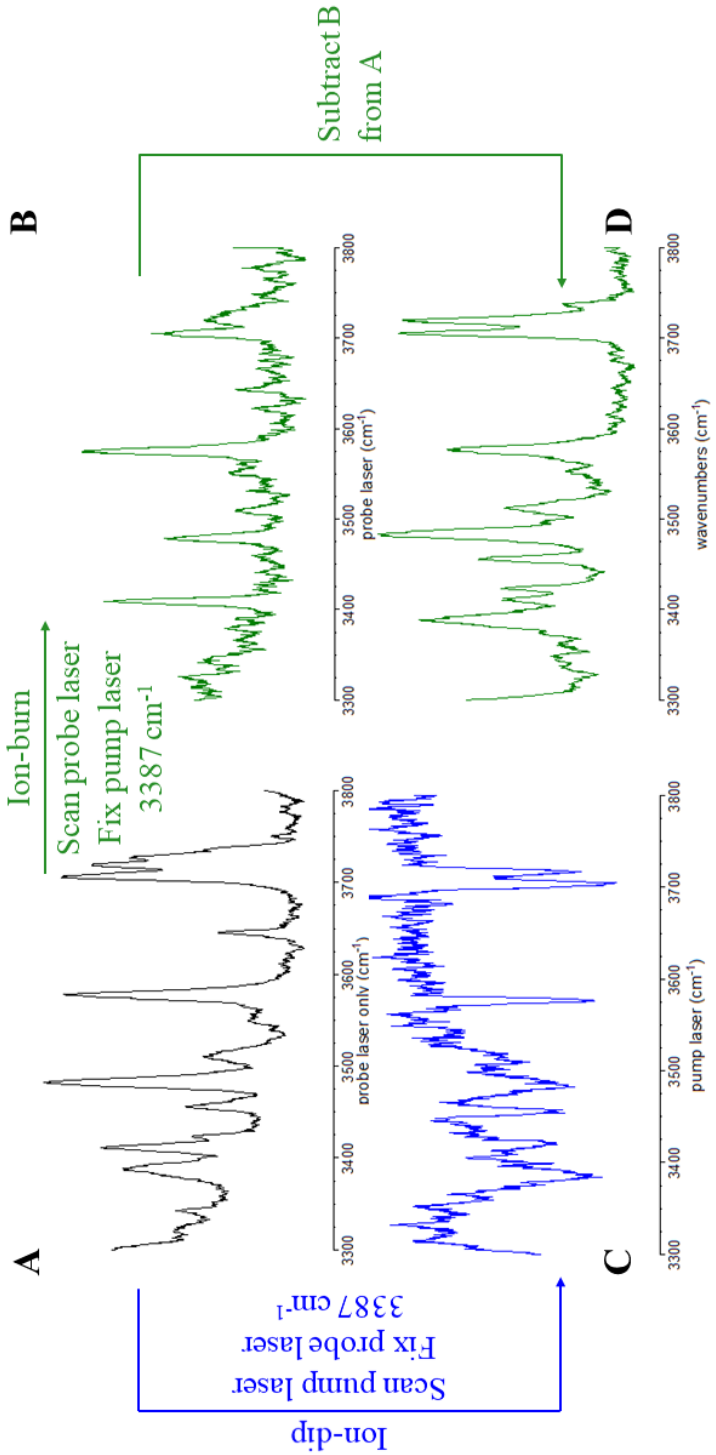
Due to the complicated nature of peptide folding as well as solvent interactions which leads to various peptide conformations, IR-IR double resonance spectroscopy is used to help disentangle these spectra. As described in Section 2.1, the final modification to the instrument was the introduction of a second 10 Hz Nd:YAG pumped optical parametric oscillator (OPO)/optical parametric amplifier (OPA) laser into the main trap. This allows for IR-IR double resonance

experiments. It can isolate one conformer in the population, remove one conformer from the spectra, as well as elucidate conformers which may contain a very sensitive H-bonding network with solvent. Thus, isomer-specific IR-IR spectra[4, 7-10] can be obtained by focusing the output of a pump laser directly into the 10 K tagging main trap, ~94 ms after the introduction of the buffer gas and 1 ms before the ion extraction and focusing the output of the probe laser in the second stage of the TOF region. This timing scheme for the pump laser limits reformation of the tagged parent adducts before ejection into the TOF. There are two main methods for taking IR-IR double resonance conformer spectra: "ion-burn" method or "ion-dip" method, shown in **Figure 2.5**.

In the "ion-burn" method, the laser in the main trap is placed at a resonant frequency of one conformer while the laser in the TOF is scanned. Therefore, all conformers that contain this vibrational frequency will lose their D_2 tag within the main trap. Since the ions are in the main trap for such a short amount of time, the D_2 tag does not have time to reattach with the bare ion. Thus, when passed through the mass gate, the conformer(s) resonant with the wavelength of the trap laser already lost their tag and therefore are filtered out by the mass. The resulting spectrum, therefore, is a collection of all conformers that do not contain that specific frequency. An example of this can be seen in **Figure 2.5 A-B**. The main advantage of this method is that it is "background-free" in nature and thus leads to a higher signal-to-noise spectra. However, though it might give a cleaner spectrum, it cannot yield one specific conformer, unless the system is only two conformers, and doing spectra subtraction requires a good reference spectrum, shown in **Figure 2.5 D**.

In the "ion-dip" method, the laser in the main trap is scanned, while the laser in the TOF is placed at a resonant frequency of one conformer. When the wavelength of the main trap laser is not resonant with the conformer being probed in the TOF region, all conformers pass through the

Figure 2.5. (A) One-laser IR predissociation spectrum of $Ala_3H^+(H_2O)_2{\cdot}D_2$ utilizing the TOF probe laser. (B) IR-IR spectrum employing the ion-dip method. (C) IR-IR spectrum employing the isomer-burning method. (D) Spectrum generated by subtracting the spectrum in (B) from the spectrum in (A)



mass gate and into the TOF region where they are then intercepted by the output of the TOF laser. There, the conformer being probed will lose its D_2 tag, leading to a constant photofragment yield. When the laser in the main trap is resonant with the frequency of the same conformer being probed in the TOF region, it will lose its D_2 tag in the trap, and will therefore be mass gated out, and the photofragment yield dips as a function of wavelength. Thus, this method generates a spectrum of only the conformer being probed and is most useful when there are more than two conformers present in a sample. An example of this can be seen in **Figure 2.5 A,C**. However, this method is not "background-free" in nature and is prone to instabilities from the electrospray and shot-to-shot fluctuations in the overall ion signal which therefore yields a lower signal-to-noise spectrum than the "ion-burn" method.

2.8 IR-IR Double Resonance: Two-Color

IR-IR double resonance can also be utilized for a "two-color approach" to be able to obtain conformers of systems with very sensitive H-bonding networks in which the direct laser interaction may perturb the system. [23-24] This approach takes advantages of the disadvantages of the one laser IRPD where one photon is does not have enough energy to perturb the molecule, hence the requirement for the weakly bound tag. In general, this approach looks at the bare molecular complexes without any tag involved. The first laser is scanned through the series of wavelengths, warming up the molecular complex of interest, yet not breaking the complex apart. These warm molecules can then be excited by the probe laser which is set to a wavelength that is resonant with a transition of the warm complex but not of the cold complex. This therefore leads to dissociation of the molecule and provides a linear IRPD spectrum. Hence, it is possible to achieve tag-free spectra of molecules in which H-bonding networks are sensitive. To determine the wavelength the

probe laser needs to be placed at, one can either take a spectrum of the molecule at temperatures < 160 K or another method is to place the pump laser at a resonant transition of the cold molecule with the tag such that $h\nu_1$ is less than the binding energy of the complex (can turn power of laser down to help this happen) and then scan the probe $h\nu_2$ to find the warm spectrum of the molecules.

2.9 Laser System

The two laser setups used to photodissociate ions in the main tagging trap and the TOF region are identical. They both contain a Continuum Powerlite neodymium: yttrium aluminum garnet (Nd:YAG) pump laser and a LaserVision optical parametric oscillator and optical parametric amplifier (OPO/OPA). The Nd:YAG laser pumps 1064 nm light at 10 Hz into the OPO/OPA. The OPO/OPA is tunable between 600-4500 cm^{-1} and the output of the Nd:YAG laser is restricted to 6W by the laser Q-switch delay timing and the laser flash lamp delay timings. Additionally, to minimize absorption of ambient H_2O and CO_2, all enclosures containing the beam path are purged with N_2 gas.

A schematic of the laser setup is shown in **Figure 2.6**. To generate the IR light by the OPO/OPA used for the experiments, first, a 1064 nm (ω_0) light produced by the Nd:YAG pump lasers is split into two beams. One of the beams passes through a $\lambda/4$ waveplate and then into a β-barium borate (BBO) crystal, where it undergoes second harmonic generation (SHG) to generate 532 nm ($2\ \omega_0$) light. This beam is then directed into the OPO which is composed of two potassium titanyl phosphate (KTP) crystals within an optical cavity. Within the OPO, the 532 nm beam functions as the pump frequency (ω_p) and by second-order nonlinear optical processes undergoes sum frequency generation (SFG) and is converted into two lower energy waves according to the equation

$$\omega_p = \omega_s + \omega_i \quad \text{(eq 2.9.1)}$$

, where ω_s is the signal (higher) frequency and ω_i is idler (lower) frequency; ω_s can range from 710-880 nm and ω_i from 1350-2120 nm. The beams are then directed into the OPA, which is composed of four KTP crystals within an optical cavity. In the OPA, the signal (ω_s) and the original 1064 nm (ω_0) light undergo difference frequency generation (DFG) according to the equation

$$\omega_{\text{mid}} = \omega_s - \omega_0 \quad \text{(eq 2.9.2)}$$

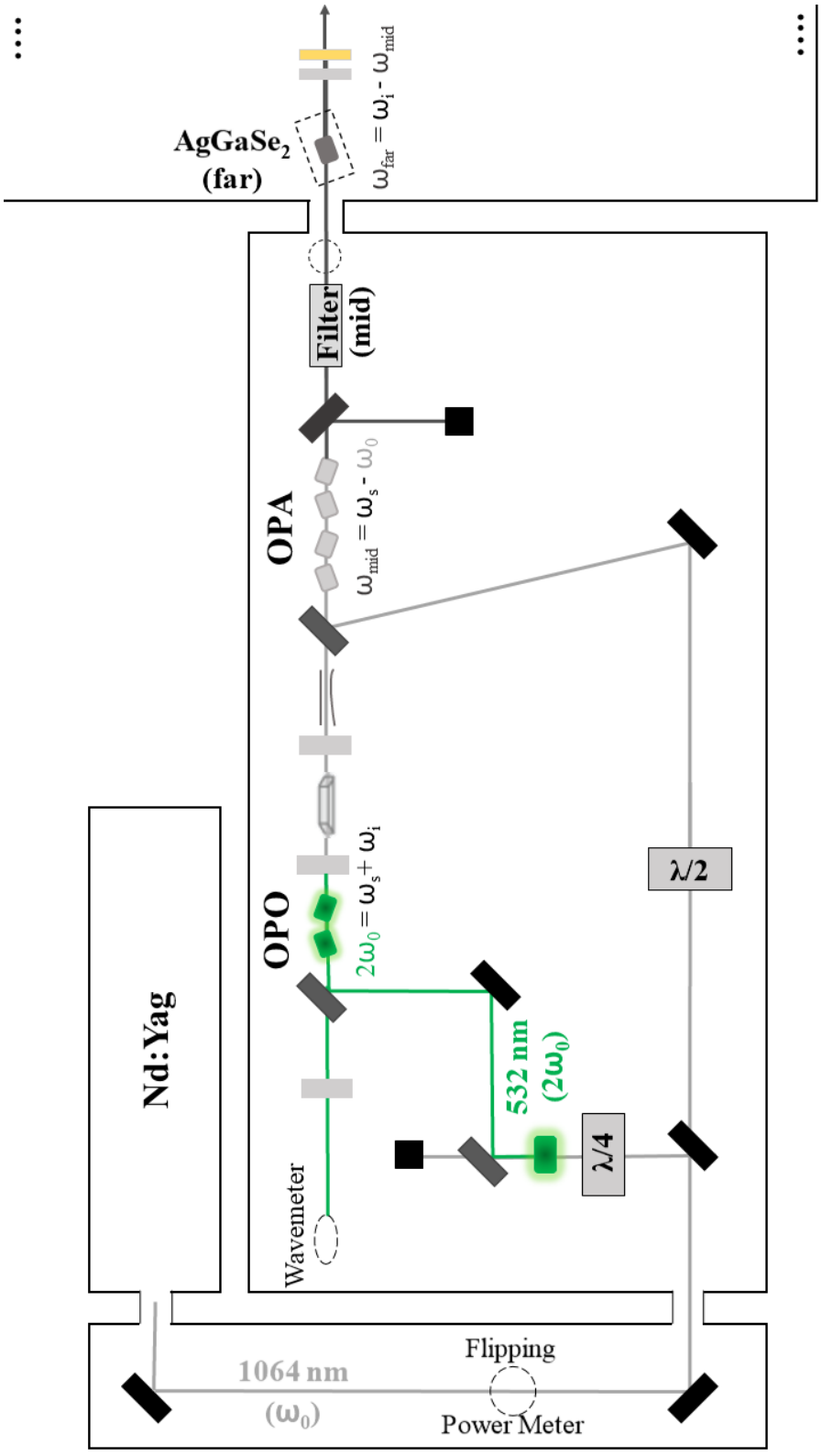
where ω_{mid} ranges from 2000-4500 cm^{-1}. If this wavelength range is the final goal (mid IR light), then the idler (ω_i) frequency is removed by a polarizer filter and any residual 1064 nm light is dumped. Power is measured at 3900 cm^{-1} to avoid power losses and fluctuations from residual ambient H_2O and is typically around 25-30 mJ/pulse.

However, if a lower frequency range is required, the polarized filter placed after the OPA stage must be removed. A $AgGaSe_2$ crystal is added after the OPO/OPA so that the idler (ω_i) beam and ω_{mid} can mix and undergo difference frequency generation (DFG) within the $AgGaSe_2$ crystal according to the equation

$$\omega_{far} = \omega_i - \omega_{mid} \quad \text{(equation 2.4)}$$

where ω_{far} ranges between 600-2400 cm^{-1}. Unwanted residual light from this mixing process is removed after the $AgGaSe_2$ crystal either by a ZnSe filter for the 1400-2400 cm^{-1} range or a 5.5 micron filter for the 600-1800 cm^{-1} range. Typical powers at 1594 cm^{-1} are 1.5-2.0 mJ/pulse and

Figure 2.6. Schematic of pump laser and OPO/OPA setup.



1.0-1.5 mJ/pulse for the ZeSe filter or 5.5 micron filter, respectively, which are much lower than in the mid spectral region. Finally, since there is no compensator crystal after the $AgGaSe_2$ crystal, the beam moves substantially in space between the highest and lowest wavelengths of this region. This can be compensated for by adjusting the Q-switch timings depending on which part of the spectral range the experiment is being measured in.

After passing through the crystals, the beam of the lasers is directed into the instrument by a series of spherical mirrors and a periscope. There is a final spherical mirror that focuses the beam to maximize overlap with the ion pack right before entering the instrument through a KBr window. The output of the beam for the "pump" laser requires an extra step of passing through the middle RF electrode of the 3D quadrupole trap in order to intersect the ions.

2.10 Data Acquisition

The ion signal is detected using a microchannel plate detector (MCP). This signal is then amplified and measured by a 100 MHz digital oscilloscope (PicoScope) which is capable of recording mass spectra on the 10 Hz instrument timescale. Signal can be viewed in real time on an oscilloscope on the instrument (Tetronix) or be routed to the computer for data acquisition using a custom written Labview program. On the instrument, the signal can be viewed and recorded, either MS or IR data, by a homemade LabView (National Instruments) data acquisition (DAQ) program, which was first written by Brett M. Marsh (PhD 2015) and later modified by Steven J. Kregel (PhD 2018). For MS data, a specified number of MS traces are averaged together for one MS. For IR data, the area of the photofragment peak is monitored as a function of laser wavelength and is normalized to a "control" peak to account for signal instability. Thus, it is possible to view in real-time what the spectrum might look like before data processing.

To be able to get the program to acquire a specific set of data, many parameters in the program are adjustable. For starters, there is a toggle for which mode that the user would like to be in, either MS or IR. If in MS, a specified number of MS scans that the user would want to be averaged, usually around 500 MS. Also, there is the trigger delay (μs), which allows the user to begin looking at data from a certain time and onwards; the sample window (μs), which is usually smaller for a narrower view of the data while in IR mode and larger for MS mode; and range (V) of the data, which allows the user to zoom in or out of the peaks. For the IR mode of the data acquisition, there are additional parameters that can be adjusted such as the IR Range, most commonly changed between far and mid IR; scan rate which is 3 cm^{-1}/s for IR data taken in the mid and 0.75 cm^{-1}/s for IR data taken in the far; the start wavenumber and end wavenumber for the IR scan that the user would like to take, specified in the near IR; and finally the toggle switch for which laser, main trap or TOF, the user would like the program to scan. The frequencies given to the program need to be given in near-IR wavenumbers so that the program can communicate correctly with the OPO/OPA software on the laser computers. This software controls the crystal positions (6 or 7 crystals depending if in the mid or far, respectively) and will automatically move the crystals to the correct positions corresponding to the frequencies desired. These positions are precalculated when calibrating the laser and creating the correct fit for each crystal. Conversion from the near-IR to the mid and far IR is given by the equations

$$\tilde{\nu}_{mid} = \tilde{\nu}_{near} - 9394\ cm^{-1} \qquad \text{(equation 2.5)}$$

$$\tilde{\nu}_{far} = 28182\ cm^{-1} - 2\tilde{\nu}_{near} \qquad \text{(equation 2.6)}$$

Finally, the program allows the user to place cursors around a detagged peak and a static peak. This is important for the IR mode of the data acquisition. The detagged peak is the peak area being monitored as a function of the wavelength and the static peak is the peak area which will normalize the detagged peak to account for signal instability.

There is also an option to select how many times the user would like to program to run in a row. Once the program has run once, if in the IR mode, the OPO/OPA computer brings the laser system back to the starting position as specified on the program. No matter the mode, if the number of scans collected is less than initially requested, a new file is created and another scan begins again.

2.11 Data Processing

In the data acquisition stage (section 2.10), though the user is able to view an IR spectrum in "real-time," the data being acquired is just a series of mass spectra as a function of wavelength. Once the mass spectra are acquired, they are analyzed in a separate LabView program which was written by Brett M. Marsh (PhD 2015) and is called "fullanalysis.vi". For MS data, the program takes the average of all the mass scans which was specified in the data acquisition program and saves it as a ".dat" file. For IR data, additional parameters must be specified before the data can be analyzed: 1) the frequency range of the data, which must be specified in near-IR wavenumbers and must be placed in numerical order in the start and end respectively; 2) the interpolation value, which is 1/10 of the scan rate; 3) the number of shots-per-point, which is 15; 4) the integration range that encompasses the area of the photofragment, which a good starting point can be found on the data acquisition program for the range of the "untagged peak;" 5) the display frequency, which is the range of frequencies that the data was taken in, either the far or mid. A smoothing

"spline" is applied to the averaged scan data to smooth the spectra. Once the parameters have been set, the program can run, and a smoothed plot of intensity vs laser frequency is generated. In general, the signal-to-noise ratio of one scan is usually insufficient, therefore multiple scans need to be averaged. The program has a parameter of the number of scans to be inputted for analysis and allows the user to select which files need to be averaged together. The number of scans is highly dependent on the system and which type of experiment is being run. For instance, the number of scans to acquire a one laser IRPD < number of scans to acquire an IR-IR burn spectra ≤ number of scans to acquire an IR-IR dip spectra. Typically, ~10-30 scans are averaged. After ~40 scans, each additional scan will have little effect on the overall scan due to scan saturation in the signal-to-noise.

2.12 Experimental Timings

The instrument and both lasers are on a 10 Hz cycle and each event within this cycle is precisely coordinated to allow for maximum efficiency and ion overlap. The experimental timing scheme is depicted in **Figure 2.7** and a sample timing scheme for the one laser setup is shown in **Table 2.1**. It is important to note that this timing scheme allows for IRPD in which one photon is used to dissociate the tag off of the molecular complex of interest. To be able to achieve IRMPD in the main trap, the scheme changes by a little.

The main timing sequence begins when the pulse valve of the main tagging trap and reaction trap is triggered by the T_0 output of a delay generator (DG1). All other experimental timings are referenced to the T_0 from DG1. DG1 also triggers two other delay generators (DG 2 and 4) such that all timings are referenced to the same value. The timing scheme will be discussed below in the order in which the ions travel through the instrument.

Figure 2.7. Schematic of sample timings for IR-IR double resonance experiment in positive ion mode.

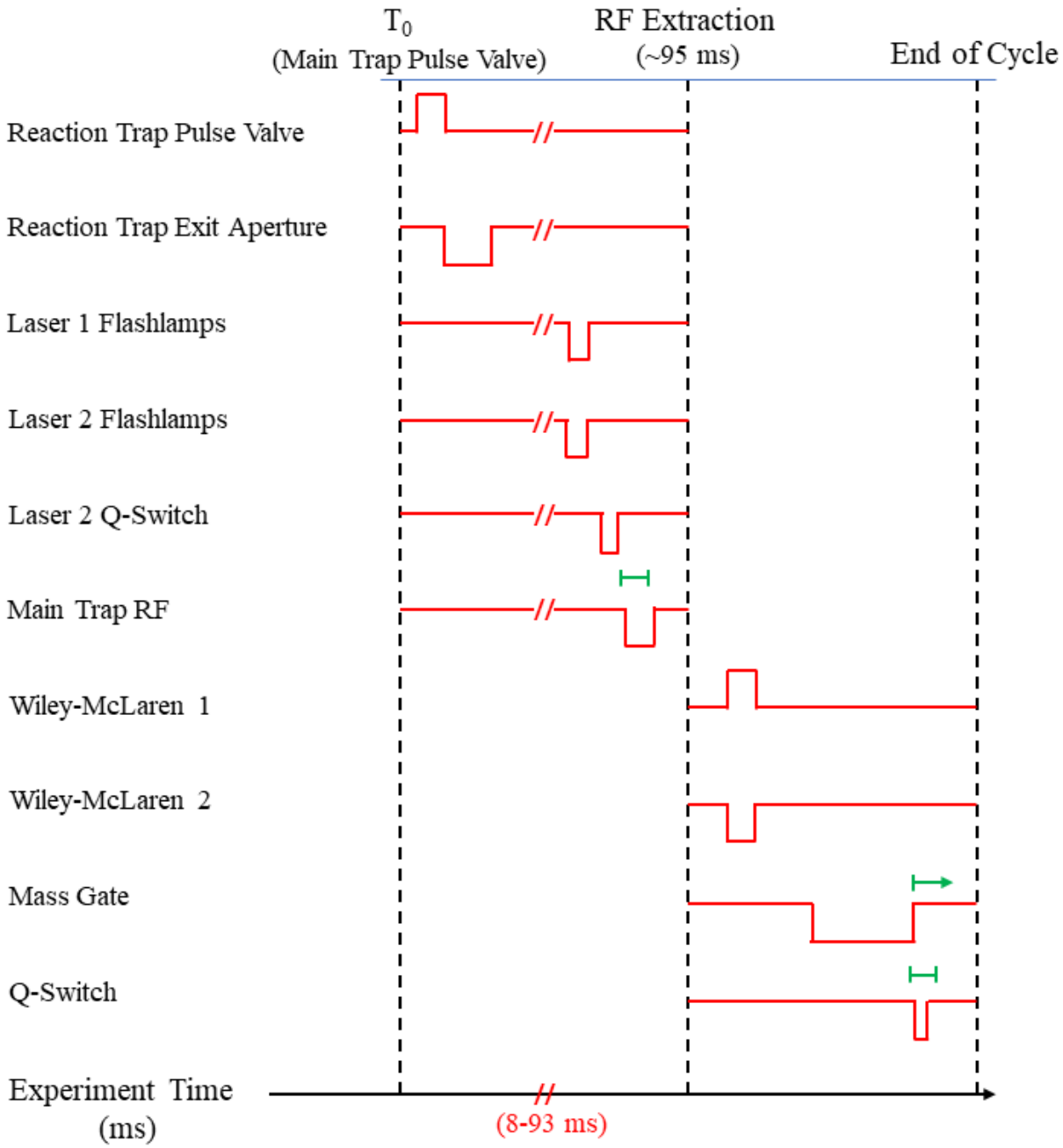


Table 2.1. Sample timings for IR-IR double resonance experiments. Mass Gate (High) and Laser 1 (TOF) Q-switch (Delay) are samples for $Ala_3H^+(H_2O)_2{\cdot}D_2$.

Delay Generator 1 (DG1)	**Channel**	**Timing**
Main Trap Pulse Valve	T_0	-
Main Trap RF	A	T_0 + 95 ms
Laser 1 (TOF) Flash Lamps (Delay)	C	A - 263 us
Laser 1 (TOF) Flash Lamps (Width)	D	C + 10 us
DG2 (Triggered by T_0)	**Channel**	**Timing**
Reaction Trap High	A	7.0 ms
Reaction Trap Low	B	A + 0.900 ms
Reaction Trap Pulse Valve	T_0	-
DG3 (Triggered by Main Trap RF)	**Channel**	**Timing**
Wiley-McLaren (WMC) 1 (Low)	A	0 + 500 ns
WMC 1 (High)	B	A + 500 ns
WMC 2 (Low)	C	0 + 500 ns
WMC 2 (High)	D	C + 500 ns
Mass Gate (High)	E	0 + 35.45 us
Mass Gate (Low)	F	E + 200 us
Laser 1 (TOF) Q-Switch (Delay)	G	0 + 36.76 us
Laser 1 (TOF) Q-Switch (Width)	H	G + 10 us
DG4 (Triggered by T_0)	**Channel**	**Timing**
Laser 2 (MT) Q-Switch Delay	A	T_0 + 94 ms
Laser 2 (MT) Q-Switch Width	B	A + 10 us
Laser 2 (MT) Flashlamp Delay	C	A - 284 us
Laser 2 (MT) Flashlamp Width	D	C + 10 us

Once the pulse valve of the reaction trap is triggered by the T_0 output of DG1, DG1 triggers DG2 which controls the timings associated with the reaction trap. The reaction traps exit aperture is pulsed between two voltages (high and low) which allows ions to either stay in the trap or be extracted. DG2 controls the timing between the high and low voltages which dictates how long ions stay trapped and then finally how long they are extracted for. The main trap pump laser (Laser 2) Q-switch (Channel A of DG4) is fixed at 94 ms. Since the laser is aligned into the main trap, ion overlap at any timepoint in the main trap is maximized. Thus, a 94 ms delay (~1 ms before ion extraction) was chosen since it allows for maximum de-tagging while minimizing retagging of the de-tagged species. Finally, the main trap pump laser which is controlled by Channel C of DG4 traditionally stays constant, yet can be changed to make sure that the power output of the Nd:YAG does not exceed 6 W.

Once the ions have left of the reaction trap, they enter the main tagging trap. The duration of time the ions spend in the main tagging trap (normally ~95 ms) is determined by channel A of DG1. Once it is time for ion extraction, the RF voltages in the main tagging trap are turned off and the RF power supply sends a TTL signal to DG3. DG3 controls the Wiley-McLaren plates, the mass gate, and the Q-switch on the TOF probe laser. Once DG3 is triggered, it sends a signal to the main trap end caps (Wiley McLaren 1 and 2) to switch on and guide/focus the ions out of the trap and into the TOF. As the ions fly down the TOF, the mass gate timings are adjusted to select the desired tagged species. Prior to the selected time, the mass gate voltage is high which deflects unwanted ions. Then at the specified voltage the mass gate voltage drops low and allows all other ions to pass. The mass gate timing is system dependent, with heavier systems requiring a longer time to get to the mass gate. For the typical systems studied throughout this work, the mass gate timing is usually between 20-40 μs. The Q-switch timing is then adjusted to occur after the mass

gate timing such that it allows for the laser to achieve the best temporal overlap with the tagged complex. Therefore, the Q-switch timing is usually set to occur 0.4-1.5 μs after the mass gate timing. Again, the larger the system, the larger the difference between the timing of the mass gate and Q-switch. Finally, the flashlamp timing for the TOF probe laser on DG1 is adjusted such that the power out of the Nd:YAG does not exceed 6W.

As mentioned above, there is a timing set up in case a two-photon process of the main tagging trap pump laser is required. Mainly, it has been found to be required when experimentally determining conformer population ratios, rather than using a comparison of experimental peak areas and computational peak intensities to determine these values. In this setup, the pump main trap laser is on its 10 Hz cycle, however the instrument is now on a 5 Hz cycle. Though much of the timings are similar, there are a couple things that differ. In the one-photon process of the main trap pump laser, the experiments are triggered by the T_0 output of DG4 (instead of DG1). All other experimental timings are referenced to the T_0 from DG4. DG4 also triggers two other delay generators (DG 1 and 2) such that all timings are referenced to the same value. The timing scheme will be discussed below in the order in which the ions travel through the instrument.

This main timing sequence begins when the pulse valve of the main tagging trap is triggered by the T_0 output of a delay generator (DG4). Once DG4 is triggered, it triggers the reaction trap timings (as before) and the reaction trap pulse valve. The reaction trap exit aperture is pulsed as it was before, however, the reaction trap pulse valve is now fired 15 ms after the pulse valve for the main trap is fired. Thus, it holds the gas in the reaction trap for a second cycle since the gas was not introduced before the exit was supposed to happen. Once the next cycle of the instrument starts, the original ions are transferred to the main trap where they are held for 195 ms (double the time than before). This turns DG1 into a 5 Hz cycle box. Since the TOF probe laser

timings are on DG1 which was turned into a 5 Hz cycle they should not be used since they need to run properly at 10 Hz. Once two cycles are completed, the RF voltages in the main tagging trap are turned off and the RF power supply sends a TTL signal to DG3, just as before. Everything else following occurs the same way as before. Sample timings for the two-photon main trap pump laser is shown in **Table 2.2**.

2.13 Voltage Settings

The voltages applied at the various stages of the instrument can vary from day to day and are dependent on the system being studied. Traditionally, peptides follow very similar settings, with the main differences being in the RF1 voltage. However, the voltages can vary a bit when switching from a peptide system to a catalyst or metal system. Regardless, the voltages typically follow trend where they gradually decrease as the ions move through the various stages of the instrument. Typical voltages for both the front half of the instrument and the TOF region can be found in **Table 2.3.** The settings shown are fairly "soft" settings which prevents fragmentation, promotes the formation of water clusters in the reaction trap, and promotes the formation of D_2-tagged adducts in the main trap. These voltages can be adjusted to "hard" settings to induce fragmentation, not have as much tagging, or not form clusters. Specifically, peptide ion fragmentation is sensitive to the ESI voltage and RF1 voltage: the higher the voltage, with respect to the rest of the voltages, the higher the fragmentation yield. Likewise, clustering in the reaction trap is sensitive to settings such as RT Entrance, RT Bias, RT Exit High, and RT Exit Low. Finally, tagging in the main trap is very sensitive to Einzel lens 1/3, Enizel lens 2, Main Trap Bias, and Wiley-McLaren (WMC) 1 and WMC2. The more drastic the difference between previous voltage

Table 2.2. Sample timings for two-photon experiments utilizing Laser 2 (MT). The Laser 1 (TOF) Delays and Widths are grayed out since during these experiments, Laser 1 cannot be turned on. The difference between the two-photon MT experiment and normal IR-IR experiments is that here, DG4 is internally triggered rather than DG1, and the Main Trap RF timing is 195 ms rather than 95 ms, making it 5 Hz.

DG1 (Triggered by T_0)	**Channel**	**Timing**
Main Trap Pulse Valve	T_0	-
Main Trap RF	A	T_0 + 195 ms
Laser 1 (TOF) Flash Lamps (Delay)	C	A - 263 μs
Laser 1 (TOF) Flash Lamps (Width)	D	C + 10 μs
DG2 (Triggered by T_0)	**Channel**	**Timing**
Reaction Trap High	A	5.6 ms
Reaction Trap Low	B	A + 1.800 ms
Reaction Trap Pulse Valve	C	15.0 ms
DG3 (Triggered by Main Trap RF)	**Channel**	**Timing**
Wiley-McLaren (WMC) 1 (Low)	A	0 + 500 ns
WMC 1 (High)	B	A + 500 ns
WMC 2 (Low)	C	0 + 500 ns
WMC 2 (High)	D	C + 500 ns
Mass Gate (High)	E	0 + 35.45 μs
Mass Gate (Low)	F	E + 200 μs
Laser 1 (TOF) Q-Switch (Delay)	G	0 + 36.76 μs
Laser 1 (TOF) Q-Switch (Width)	H	G + 10 μs
Delay Generator 4 (DG4)	**Channel**	**Timing**
Laser 2 (MT) Q-Switch Delay	A	T_0 + 94 ms
Laser 2 (MT) Q-Switch Width	B	A + 10 μs
Laser 2 (MT) Flashlamp Delay	C	A - 284 μs
Laser 2 (MT) Flashlamp Width	D	C + 10 μs

Table 2.3. Typical voltages used for water clustering. In this case, these voltages are for the optimization of $Ala_3H^+(H_2O)_2{\cdot}D_2$. Acronyms used: RF = Radio Frequency, RT = Reaction Trap, MT = Main Trap.

Region	Power Supply	Applied Voltage
Ion Source	ESI	+2.2 kV
	Capillary	+146.0 V
	Skimmer	+141.0 V
	RF1	+102.5 V
Reaction Trap	Aperture	+89.1 V
	RF2	+87.1 V
	RT Entrance	+86.4 V
	RT Bias	+82.8 V
	RT Exit High	+90.2 V
	RT Exit Low	+78.6 V
	RF 3 Bias	+77.5 V
Main Trap	Aperture 2	-7.5 V
	RF4 Bias	+58.0 V
	Einzel 1 & 3	-81.2 V
	Einzel 2	+40.4 V
	Main Trap Bias	+54.6 V
TOF/Reflectron	HV1 (WMBC 1)	+75.0 V
	HV 2 (WMBC 2)	-308.0 V
	Liner	-1981 V
	Focus	0.0 V
	Deflector Left/Right 1	-1990 V
	Deflector Up/Down	-1951 V
	Einzel 2	-2270 V
	Deflector Left/Right 2	-2170 V
	Reflectron Front	-930.0 V
	Reflectron Back	+150.0 V
	Mass Gate Low (-)	-2098 V
	Mass Gate High (+)	+1004 V
	MCP	-4500 V
Amplitude	RF 1 Amplitude	+393.0 V
	RF 2 Amplitude	+456.0 V
	RF 3 Amplitude	+409.0 V
	RF 4 Amplitude	+388.0 V
	RT RF Amplitude	+333.0 V

and the voltage that follows, the "harder" the settings and the less likely it will be to create clustered and tagged species.

2.14 Computations

Unless otherwise specified, all calculations were performed using Gaussian 16.[25] For the bulk of our calculations, DFT is utilized largely due to the size of the systems we are calculating. However, ab initio methods, such as MP2, are used on a small subset of computations once the lowest energy conformers have been found with DFT. In general, computational time scales as $O(N^X)$ where N depends on the size of the molecule that the computations are being done on and X depends on which method is being utilized. DFT scales as $\sim O(N^3)$ whereas MP2 scales as $\sim O(N^5)$.[26] Hence, DFT is more feasible for a larger amount of computations for these larger systems.

To begin our computations, a basin hopping program developed by the Hopkins group[14-17] at the University of Waterloo is employed. This program can probe a system's conformational landscape utilizing AMBER parameters. It has proven to be very useful for fluxional molecules that can adopt various conformations, such as for a variety of protonated tripeptides, or for probing various solvent arrangements around such fluxional molecules.[7-9, 27] The program can iteratively and randomly move specified parameters such as bond angles, distances, and dihedrals, as well as a combination of all three, within a specified range and then reoptimize after each step for a specified number of steps. Dihedral rotations are most commonly probed for the various angles within peptides, and translational and rotational parameters are commonly probed for each individual water molecule. This allows for a large sampling of the conformational landscape and gives a range of conformations in which the lowest energy unique conformers can be selected for

further analysis utilizing Gaussian 16. Multiple iterations of the program can be run at the same time with the same and different starting structures to thoroughly find the maximum number of unique conformers that each system may have.

Once the basin hopping program has given as many unique conformers as it can, these conformers are sequentially optimized in DFT, increasing in both functional accuracy and basis set size until the final optimization utilizing the cam-B3LYP functional, def2-TZVP basis set, and Grimme's dispersion correction with Becke-Johnson Damping (GD3BJ). Harmonic frequencies to obtain vibrational spectra are also calculated at the cam-B3LYP/def2-TZVP/GD3BJ level. This functional and basis set has been shown to provide good agreement with our experimental spectra and is found to be a good balance for accuracy and computational time.[3-4, 7-10] Once the energetic ordering is obtained through the DFT calculations, MP2/def2-TZVP is utilized to provide a more accurate energetic comparison between conformers. Due to its larger computational time scale, it is usually not feasible to calculate optimizations of all conformers at the level. Thus, MP2 is used only on the lowest 10-20 isomers. Though MP2, in general, gives more accurate energetics, for molecular systems with H-bonding systems, there seems to be a discrepancy between MP2 and DFT (see Chapter 4). Therefore, the lowest energy structures found from DFT and MP2 respectively are used to help identify the conformers within the system.

2.15 References

(1) Kamrath, M. Z.; Garand, E.; Jordan, P. A.; Leavitt, C. M.; Wolk, A. B.; Van Stipdonk, M. J.; Miller, S. J.; Johnson, M. A. Vibrational characterization of simple peptides using cryogenic infrared photodissociation of H2-tagged, mass-selected ions. *J Am Chem Soc* **2011,** *133* (16), 6440-8.

(2) Kamrath, M. Z.; Relph, R. A.; Guasco, T. L.; Leavitt, C. M.; Johnson, M. A. Vibrational predissociation spectroscopy of the H2-tagged mono- and dicarboxylate anions of dodecanedioic acid. *International Journal of Mass Spectrometry* **2011,** *300* (2-3), 91-98.

(3) Marsh, B. M.; Voss, J. M.; Garand, E. A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *J Chem Phys* **2015,** *143* (20), 204201.

(4) Voss, J. M.; Kregel, S. J.; Fischer, K. C.; Garand, E. IR-IR Conformation Specific Spectroscopy of Na(+)(Glucose) Adducts. *J Am Soc Mass Spectrom* **2018,** *29* (1), 42-50.

(5) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

(6) Yamashita, M.; Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **2002,** *88* (20), 4451-4459.

(7) Fischer, K. C.; Sherman, S. L.; Garand, E. Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and beta-Alanine. *J Phys Chem A* **2020,** *124* (8), 1593-1602.

(8) Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J Phys Chem A* **2019,** *123* (15), 3355-3366.

(9) Voss, J. M.; Fischer, K. C.; Garand, E. Revealing the structure of isolated peptides: IR-IR predissociation spectroscopy of protonated triglycine isomers. *Journal of Molecular Spectroscopy* **2018,** *347*, 28-34.

(10) Voss, J. M.; Fischer, K. C.; Garand, E. Accessing the Vibrational Signatures of Amino Acid Ions Embedded in Water Clusters. *J Phys Chem Lett* **2018,** *9* (9), 2246-2250.

(11) Gillan, M. J.; Alfe, D.; Michaelides, A. Perspective: How good is DFT for water? *J Chem Phys* **2016,** *144* (13), 130901.

(12) Willow, S. Y.; Zeng, X. C.; Xantheas, S. S.; Kim, K. S.; Hirata, S. Why Is MP2-Water "Cooler" and "Denser" than DFT-Water? *J Phys Chem Lett* **2016,** *7* (4), 680-4.

(13) Yoo, S.; Xantheas, S. S. Communication: The effect of dispersion corrections on the melting temperature of liquid water. *J Chem Phys* **2011,** *134* (12), 121105.

(14) Campbell, J. L.; Yang, A. M.; Melo, L. R.; Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J Am Soc Mass Spectrom* **2016,** *27* (7), 1277-84.

(15) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: lessons learned from tetraalkylammonium cations and their isomers. *J Am Soc Mass Spectrom* **2014,** *25* (9), 1583-91.

(16) Lecours, M. J.; Chow, W. C.; Hopkins, W. S. Density functional theory study of Rh(n)S(0,+/-) and Rh(n+1)(0,+/-) (n = 1-9). *J Phys Chem A* **2014,** *118* (24), 4278-87.

(17) Liu, C.; Le Blanc, J. C.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using differential mobility spectrometry to measure ion solvation: an examination of the roles of solvents and ionic structures in separating quinoline-based drugs. *Analyst* **2015,** *140* (20), 6897-903.

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

(19) Mathur, R.; O'Connor, P. B. Design and implementation of a high power rf oscillator on a printed circuit board for multipole ion guides. *Review of Scientific Instruments* **2006,** *77* (11), 114101.

(20) O'Connor, P. B.; Costello, C. E.; Earle, W. E. A high voltage RF oscillator for driving multipole ion guides. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (12), 1370-1375.

(21) Wang, X. B.; Wang, L. S. Development of a low-temperature photoelectron spectroscopy instrument using an electrospray ion source and a cryogenically controlled ion trap. *Rev Sci Instrum* **2008,** *79* (7), 073108.

(22) Wiley, W. C.; McLaren, I. H. Time-of-Flight Mass Spectrometer with Improved Resolution. *Review of Scientific Instruments* **1955,** *26* (12), 1150-1157.

(23) Yang, N.; Duong, C. H.; Kelleher, P. J.; Johnson, M. A.; McCoy, A. B. Isolation of site-specific anharmonicities of individual water molecules in the I−·(H2O)2 complex using tag-free, isotopomer selective IR-IR double resonance. *Chemical Physics Letters* **2017,** *690*, 159-171.

(24) Yeh, L. I.; Okumura, M.; Myers, J. D.; Price, J. M.; Lee, Y. T. Vibrational spectroscopy of the hydrated hydronium cluster ions H3O+·(H2O)n (n=1, 2, 3). *The Journal of Chemical Physics* **1989,** *91* (12), 7319-7330.

(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.;

Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

(26) Matczak, P.; Wojtulewski, S. Performance of Moller-Plesset second-order perturbation theory and density functional theory in predicting the interaction between stannylenes and aromatic molecules. *J Mol Model* **2015,** *21* (3), 41.

(27) Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H(+).(H2O). *J Phys Chem A* **2018,** *122* (41), 8213-8221.

CHAPTER 3

Comment on “Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy”

Published: Sherman, S. L.; Nickson, K. A.; Garand, E. *The Journal of Physical Chemistry Letters* **2022,** *13* (8), 2046-2050.

Abstract

This Viewpoint presents a re-examination of the conclusions of a study reported in *The Journal of Physical Chemistry Letters* (Saparbaev, *et al.* 2021, 12, 907) that compared the structure of microsolvated ions formed by electrospray ionization to those formed in the gas-phase via a previously published cryogenic ion trap approach. We conducted additional experiments that clearly show that most of the observed differences in the IR spectra can be accounted for by considering the different spectroscopic action schemes used to obtain them. In particular, the presence of the D_2-tag induces shifts in some of the N—H and O—H peaks which need to be carefully considered before comparing spectra. Once these spectral effects are taken into account, we show that both clustering approaches yield similar cluster structures for the small $GlyH^+(H_2O)_n$ species. Using unimolecular reaction rate theory, we also show that for the small complexes considered here, only the gas-phase equilibrium distribution of conformers should be expected in both experimental approaches. In addition, the barrier heights necessary to kinetically trap high-energy conformers at 298 K is explored using a series of model polyalanine chains.

Author Contributions

SLS and KAN acquired the additional experimental IRMPD spectra. SLS performed data analyses. SLS and EG wrote the manuscript.

3.1 Viewpoint

The Letter by Saparbaev *et al.*[1] aimed to compare the structure of microhydrated protonated glycine [$GlyH^+(H_2O)_n$] and triglycine [$Gly_3H^+(H_2O)$] clusters produced by two different approaches. In their experiment, clusters were formed directly by soft electrospray ionization ("soft-ESI" hereafter) and thermalized in a room temperature storage ion trap for 45 ms before they were cooled to 6 K in a cryogenic ion trap prior to spectroscopic interrogation. The infrared spectra of these clusters were directly compared to those previously obtained by our group,[2-5] in which the same species were formed by clustering H_2O onto the bare $GlyH^+$ or Gly_3H^+ ions in an 80K ion trap ("clustering approach" hereafter).[6] The clusters were then cooled to 10 K in a cryogenic ion trap and D_2-tagged prior to spectroscopic interrogation. The observed differences in the $GlyH^+(H_2O)_{1-3}$ spectra of the species produced by either method led Saparbaev *et al.* to conclude that some of the conformers identified in the clustering approach were kinetically trapped "intrinsic gas-phase conformers", while all of those observed directly from soft-ESI must be the "native solution structures". Such kinetic trapping of solution-phase structures, or high-energy gas-phase structures, via soft-ESI has been observed numerous times previously.[7-13]

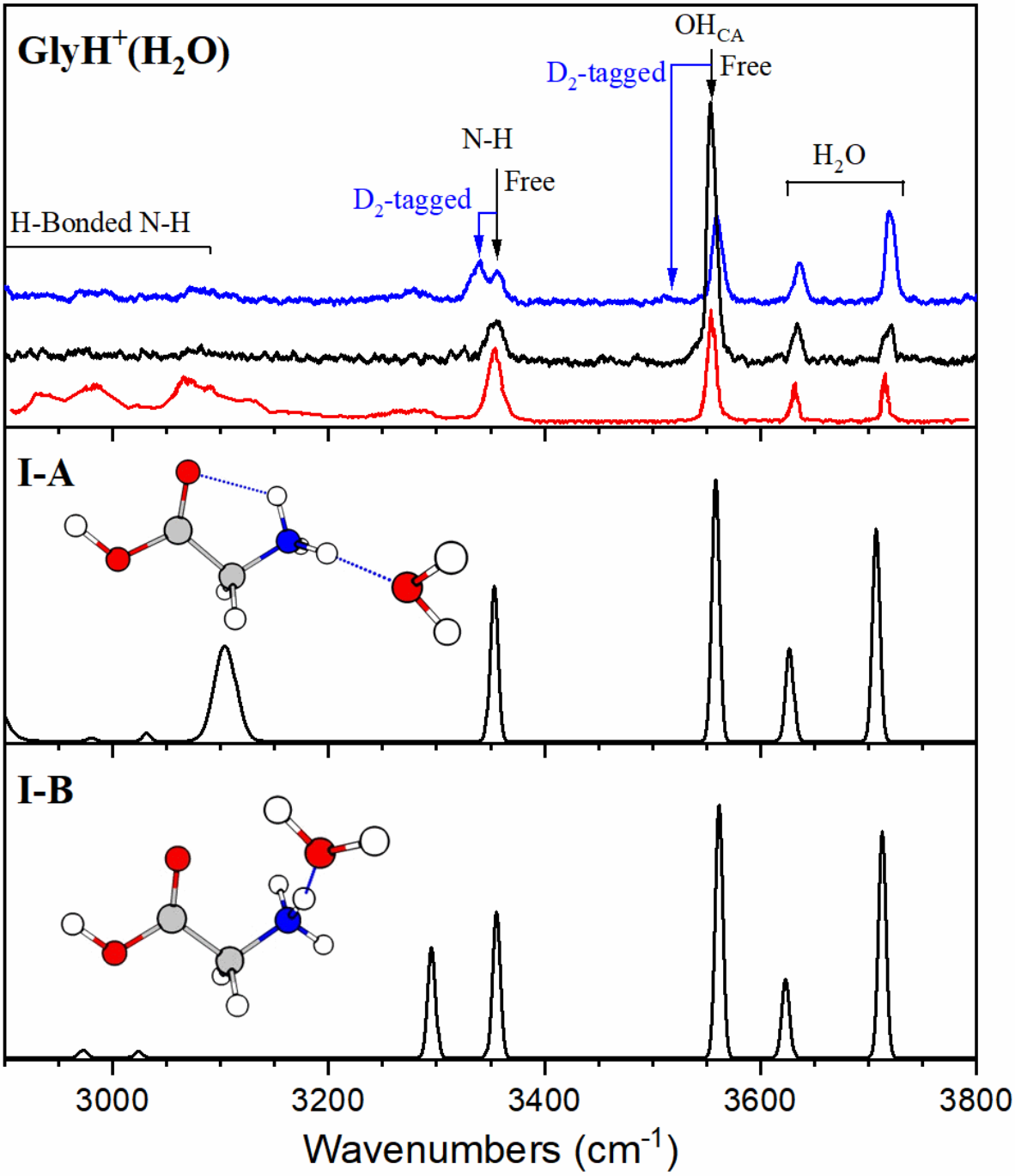
Herein, we show that most of the observed differences in the IR spectra can be accounted for by considering the different spectroscopic action schemes used to obtain them. The species formed by the clustering approach were probed using infrared predissociation (IRPD) of the D_2-tagged species[14-15] whereas the soft-ESI clusters were probed by infrared multiple photon dissociation (IRMPD). Both schemes are indirect action spectroscopies that require careful considerations when analyzing the resulting spectra. For the one-photon IRPD experiment, while the effects of the D_2 presence are usually quite small, its interaction with specific N—H and O—H groups can nonetheless cause a ~10-50 cm^{-1} redshift in their infrared vibrational bands.[14,

[16-17] The H_2O binding energy in $GlyH^+(H_2O)$ is calculated (cam-B3LYP/def2TZVP) to be around 6500 cm^{-1} (77.8 kJ/mol), and therefore the IRMPD scheme requires the absorption of 2-3 photons in the 2900-3800 cm^{-1} range studied by Saparbaev *et al*. The potential changes in absorption cross-section after the absorption of the first photons, or variation in the rate of vibrational energy redistribution, may cause the intensity of features in IRPD and IRMPD spectra to differ.[18-19] The multiphoton nature of the IRMPD process can also make it more sensitive toward the higher-energy kinetically trapped conformers that require fewer photons for dissociation.

We have reacquired the IR spectra of the $GlyH^+(H_2O)_{1-3}$ clusters, using the exact same 80 K ion trap clustering approach used in Refs 2 and 3, but this time we probed them using the same IRMPD scheme as Saparbaev *et al*. The resulting IR spectra are shown in the top panel of **Figures 3.1-3.3** (black trace) and compared to those of Saparbaev *et al*. (red trace). The D_2-tagged IRPD spectra from Refs 2 and 3 (blue trace) are also included for comparison. All spectra in the figures have been normalized to the common H_2O sym. stretch around 3630-3650 cm^{-1} peak to show differences in intensities. Some of the differences between the spectra of the ion trap clustering species acquired with IRPD and IRMPD highlights the perturbations induced by the D_2-tag. We recognize that our previous work did not adequately describe the inferred effects of the D_2 tag on the spectra and that this might have affected the interpretation of Saparbaev et al. Herein, we will use the D_2-tagged IRPD and IRMPD spectra acquired under the same conditions to highlight these perturbations.

In **Figure 3.1**, the peak at 3340 cm^{-1} in the $GlyH^+(H_2O)$ IRPD spectrum which is absent in the IRMPD spectrum, can be attributed to a redshift of one of the N—H stretches resulting from direct interactions with the D_2 tag. In our previous work, we assigned the observed spectrum to the presence of two conformers, I-A and I-B, whose structures, and calculated spectra (without D_2

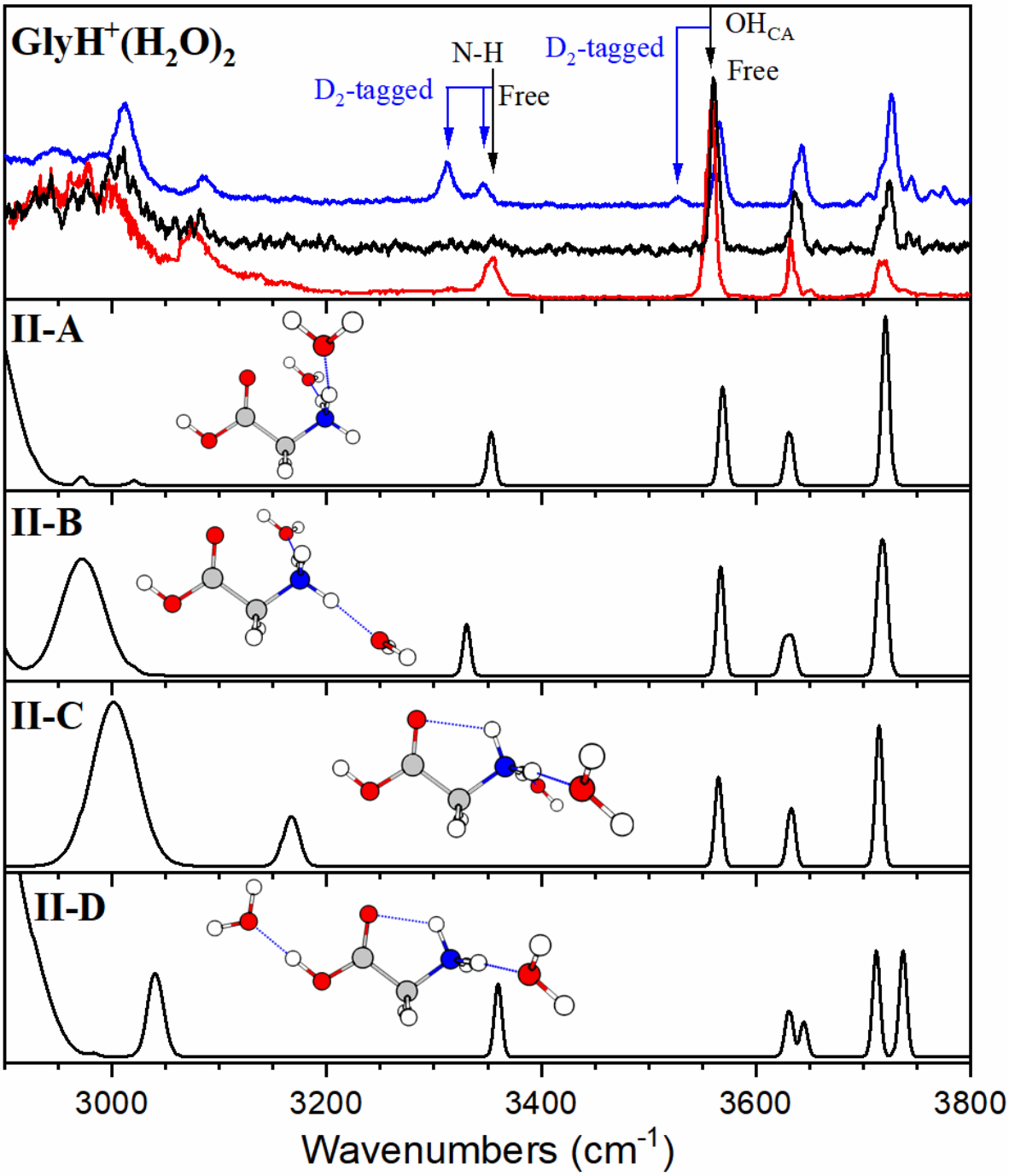
Figure 3.1. (Top panel) IRMPD spectra of $GlyH^+(H_2O)$, produced through clustering approach (black trace; this work) and soft-ESI (red trace; Saparbaev et al.[1]). IRPD spectra of $GlyH^+{\cdot}D_2(H_2O)$, produced via clustering approach (blue trace; previous work[2-3]). All traces are normalized to their respective free O—H water sym. stretch between 3630 and 3650 cm^{-1}. (Bottom panels) Harmonic spectra calculated at the cam-B3LYP/def2TZVP/GD3BJ level.



tag) are shown in **Figure 3.1**. These two conformers are necessary to explain the presence of small peaks at 3080 and 3277 cm^{-1}. Because conformer I-A has only one free N—H, the addition of a D_2-tag is expected to redshift it and therefore, the peak at 3339 cm^{-1} was explicitly assigned to I-A. Conformer I-B has two free N—H and therefore should exhibit at least one mostly unperturbed N—H, which led us to assign the highest N—H feature at 3357 cm^{-1} to this conformer. In the absence of D_2-tag, the calculation predicts that these two features should be overlapping, consistent with the single peak at 3360 cm^{-1} in both IRMPD shown in **Figure 3.1**. We note that the IRMPD of Saparbaev *et al*. also exhibits peaks at 3080 and 3277 cm^{-1}, consistent with the presence of the same I-A and I-B conformers in their experiment.

The assignment is not as straightforward for the $GlyH^+(H_2O)_2$ spectrum. In our previous work, we identified the presence of 3 conformers (II-A, II-B and II-D) which are shown in **Figure 3.2**. The presence of ~15% of II-D, in which the water interacts with the carboxylic acid O—H, was confirmed and quantified with the observation of its unique C=O stretch at 1757 cm^{-1} (not shown here; see Ref 3). The remaining ~85% conformer population was assigned to conformers II-A and II-B. In Ref 3, isomer-specific IR-IR double resonance was used to confirm that the two peaks at 3313 and 3344 cm^{-1} belong to separate conformers. Because both II-A and II-B have only one free N—H, we would expect them to be red-shifted by the presence of a D_2-tag. Unfortunately, the signal/noise ratio in our IRMPD spectra is not sufficient to clearly identify the position of these N—H stretches without the D_2-tag. However, the IRMPD of Saparbaev *et al.* clearly shows a feature at 3375 cm^{-1} which is consistent with the calculated spectra of II-A, II-B, and II-D. Here the presence of multiple conformers with similar IR spectra makes it difficult to conclusively compare the conformers produced by both clustering methods.

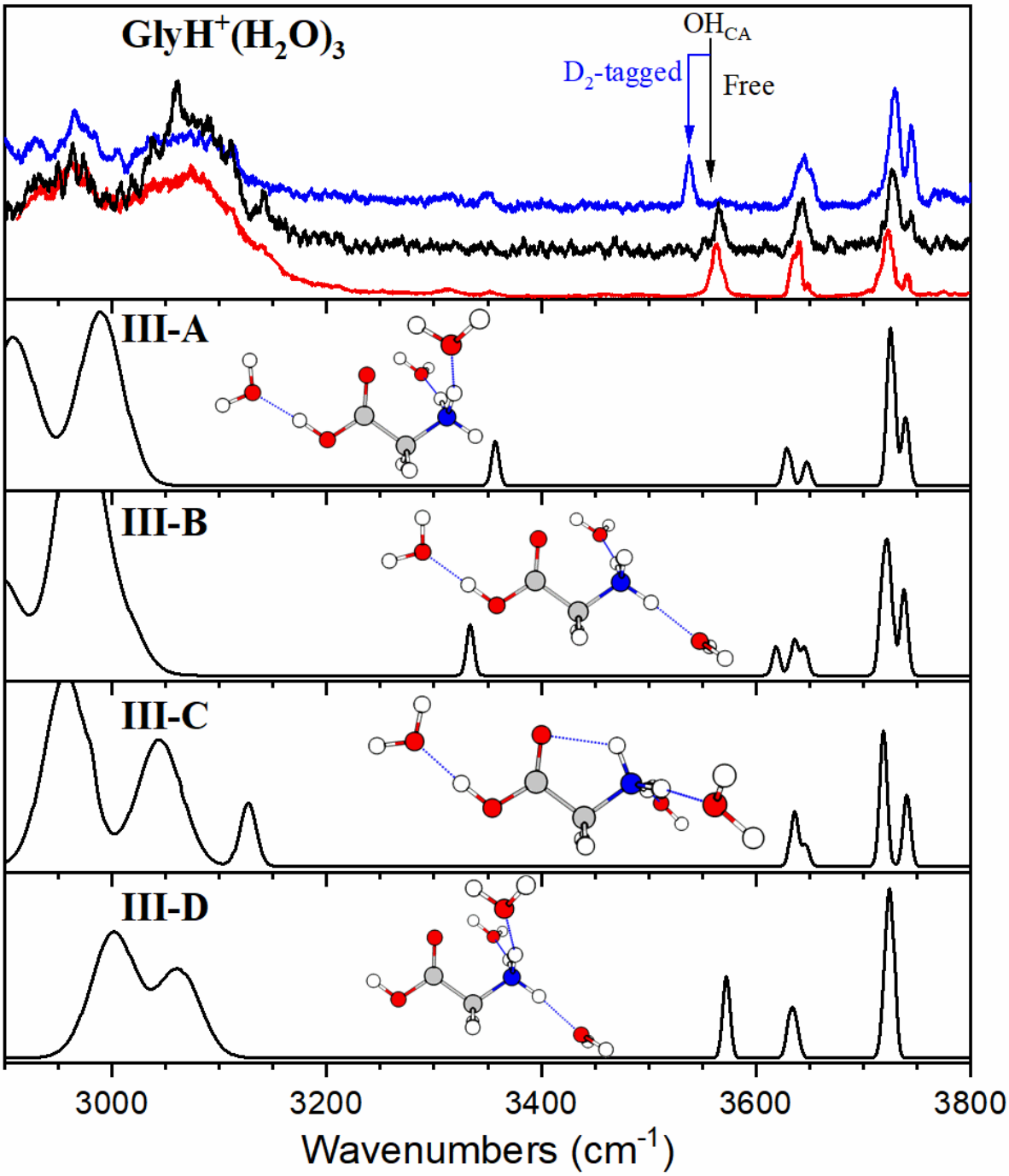
Figure 3.2. (Top panel) IRMPD spectra of $GlyH^+(H_2O)_2$, produced through clustering approach (black trace; this work) and soft-ESI (red trace; Saparbaev et al.[1]). IRPD spectra of $GlyH^+{\cdot}D_2(H_2O)_2$, produced via clustering approach (blue trace; previous work[2-3]). All traces are normalized to their respective free O—H water sym. stretch between 3630 and 3650 cm^{-1}. (Bottom panels) Harmonic spectra calculated at the cam-B3LYP/def2TZVP/GD3BJ level.



Finally, in the $GlyH^+(H_2O)_3$ spectra shown in **Figure 3.3**, the presence of the D_2-tag redshifts the carboxylic O—H feature from 3560 cm^{-1} in the IRMPD spectrum to 3538 cm^{-1} in the D_2-tagged spectrum. In our previous work, we identify the presence of 4 conformers for $GlyH^+(H_2O)_3$, which are shown in **Figure 3.3**. Conformer III-D, in which all three waters are bound to the NH_3^+ group, is the only one that can give rise to the observed carboxylic acid O—H near 3550 cm^{-1}. Because the N—H groups are completely occupied by the water molecules, the most likely D_2-tagging site is the carboxylic acid O—H, consistent with the shifted frequency of this feature in the D_2-tagged IRPD spectrum. In our previous work, the small peaks at 3312 and 3349 cm^{-1} were assigned to the presence of conformers III-B and III-A, respectively. Finally, the presence of conformer III-C, which does not exhibit any unique features, could not be ruled out. Using the calculated relative intensities of all the conformers, we estimated that the contribution of III-C is less than 10%. The IRMPD spectra of Saparbaev *et al*. clearly exhibit features at ~3315, ~3355 and ~3560 cm^{-1} which are consistent with the presence of the same III-A, III-B and III-D major conformers.

Overall, careful analysis of the spectra shows that similar $GlyH^+(H_2O)_n$ cluster structures are generated by both clustering methods. This conclusion is also supported by the unimolecular dynamic analysis presented in the next section. Another important conclusion is that careful analysis is always required in order to directly compare action IR spectra produced by different schemes. The spectral features should not be directly compared, but rather, the resulting structures obtained from properly analyzing them within their respective framework should be considered. We also note that the IRMPD of the same species acquired in each lab (black and red traces) show some marked difference in peak intensities. For example, in $GlyH^+(H_2O)$ the carboxylic O—H peak at 3550 cm^{-1} is much more intense in the black trace. In $GlyH^+(H_2O)_2$ and $GlyH^+(H_2O)_3$, the

Figure 3.3. (Top panel) IRMPD spectra of $GlyH^+(H_2O)_3$, produced through clustering approach (black trace; this work) and soft-ESI (red trace; Saparbaev *et al.*[1]). IRPD spectra of $GlyH^+{\cdot}D_2(H_2O)_3$, produced via clustering approach (blue trace; previous work[2-3]). All traces are normalized to their respective free O—H water sym. stretch between 3630 and 3650 cm^{-1}. (Bottom panels) Harmonic spectra calculated at the cam-B3LYP/def2TZVP/GD3BJ level.



N—H peak between 3300 and 3400 cm^{-1} is a lot weaker and sometime barely visible in the black traces. These could be due to saturation effects, different signal-to-noise ratios, and mostly different laser fluence between the two experiments. It highlights once again that direct comparison between action spectra must always be done carefully. While experimentally more demanding, the IR-IR two-color approach pioneered by Lee and co-workers[20] and recently adapted by Yang *et. al.*[21] might provide vibrational spectra of mass-selected ions which are free tag perturbation and non-inear effects and closer to the true linear absorption spectra.

The results that both cluster formation schemes essentially yield similar conformers for the $GlyH^+(H_2O)_n$ clusters is not too surprising given the very small barriers between the structures labeled as "native solution structure" and "intrinsic gas-phase conformer" by Saparbaev *et al*. (see **Figure 3.4**), which were computed and presented in Ref 3. Conversion between these structures involve a simple C—N bond rotation with calculated barriers of ~1.2 kJ/mol for $GlyH^+(H_2O)_1$ and ~9.0 kJ/mol for $GlyH^+(H_2O)_2$. For reference, the average vibrational internal energies, calculated using harmonic frequencies, is 15.0 and 24.5 kJ/mol for $GlyH^+(H_2O)_1$ and $GlyH^+(H_2O)_2$ at 298 K, respectively. Using the Rice-Ramsperger-Kassel-Marcus (RRKM) theory[22] of unimolecular reaction rates, we can roughly estimate that the half-life for the conversion between the "native structure" and "intrinsic gas-phase conformer" would be less than 100 ps at 298 K. These barriers are simply too small to prevent interconversion at room temperature for 45 ms, as in the experiment of Saparbaev *et al*.

For relatively small species such as $GlyH^+(H_2O)_2$, the fast unimolecular dynamics means that kinetic trapping of conformers is unlikely to occur on the millisecond experimental time scale unless the activation barrier is very close to, or larger than, the vibrational internal energy. This is illustrated in **Figure 3.5**, which shows the half-life of a $GlyH^+(H_2O)_2$ conformer at room

Figure 3.4. $GlyH^+(H_2O)$ labeled as (A) "intrinsic gas-phase" and (B) "native solution-phase" form as labeled by Saparbaev et al.[1] $GlyH^+(H_2O)_2$ in its (C) "intrinsic gas-phase" and (D) "native" form.

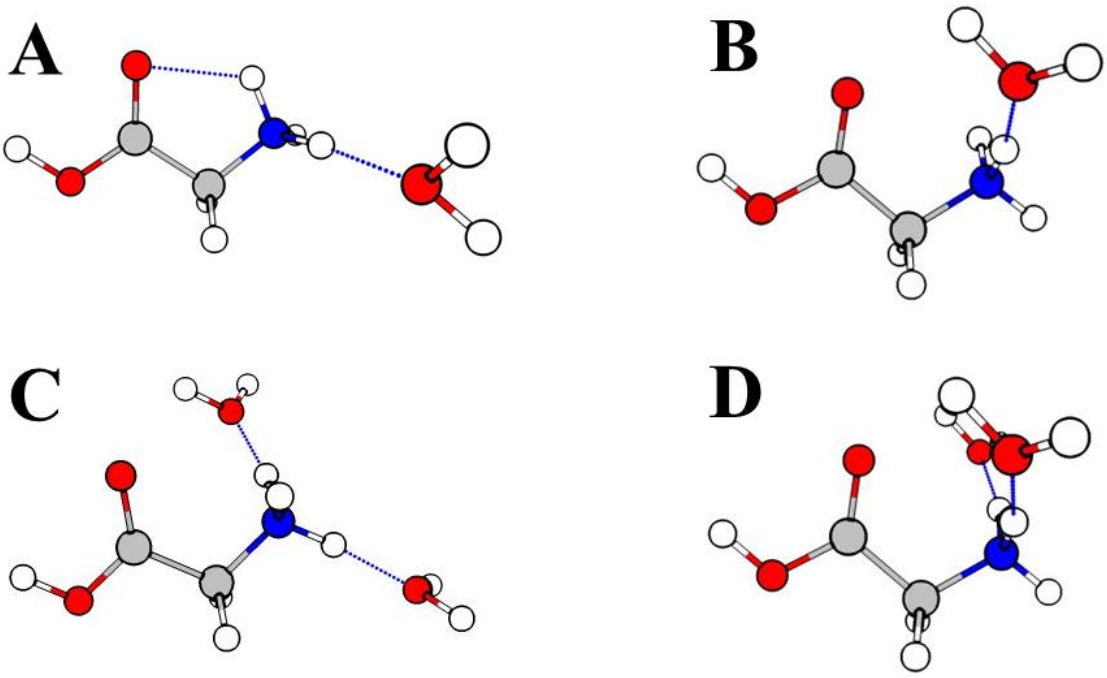
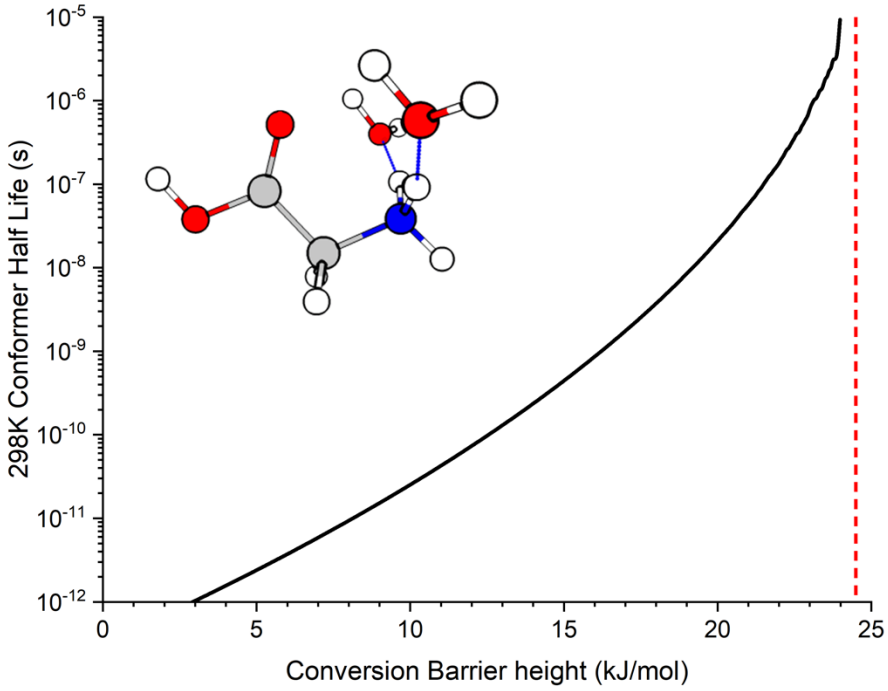
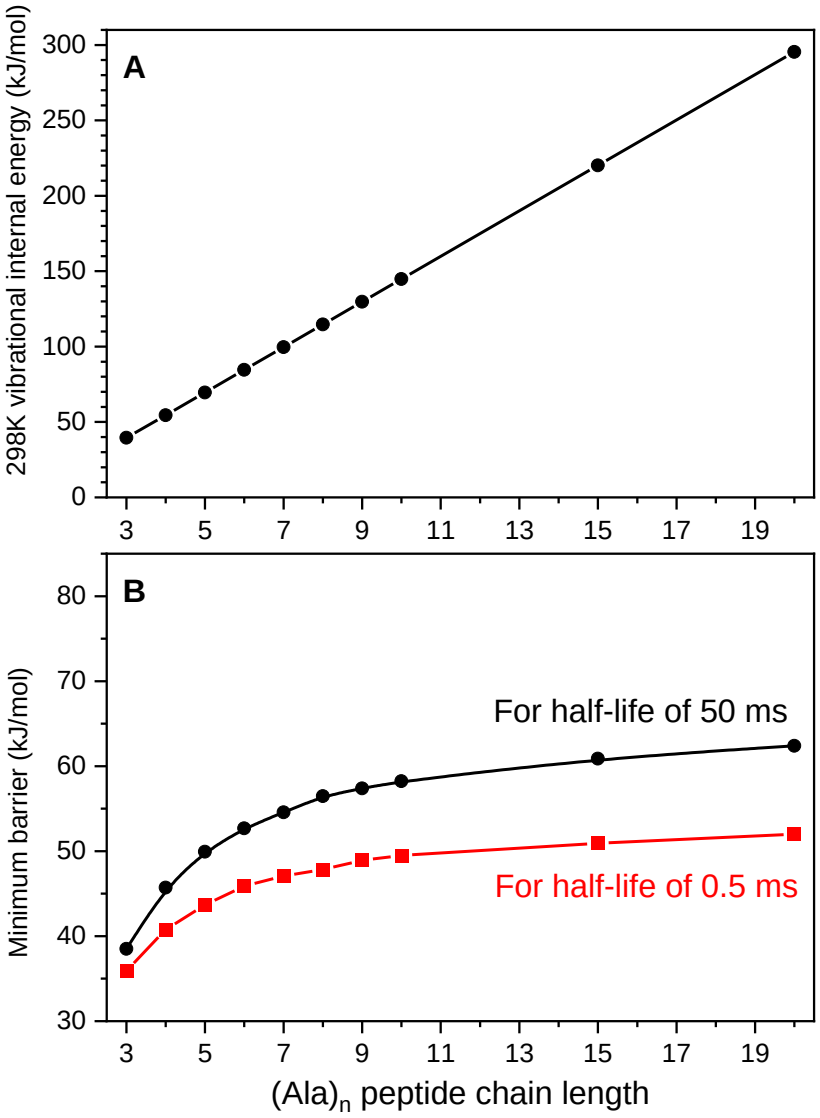


Figure 3.5. Calculated half-life of a $GlyH^+(H_2O)_2$ conformer at room temperature as a function of the conversion barrier height. The red dotted line marks the internal energy of $GlyH^+(H_2O)_2$ at 298 K.



temperature as a function of a hypothetical interconversion barrier height. For an internal energy of 24.5 kJ/mol, half-lives shorter than 1 μs are found for all barrier heights less than ~23 kJ/mol. For increasingly more complex species or larger clusters, the rate of unimolecular reaction would decrease because of the large increase in vibrational state density. However, the vibrational internal energy available at room temperature will also increase with vibrational state density. To look at these competing effects, we computed the harmonic vibrational state density and unimolecular reaction rate of simple linear polyalanine peptide chains with 3-20 monomers. The distribution of vibrational states of the hypothetical conformer conversion transition state was approximated to be the same as the reactant structure but shifted by the height of the arbitrary barrier. The average vibrational internal energy at 298 K, shown in **Figure 3.6a**, scales almost linearly from 40 to 295 kJ/mol when the chain length is increased from 3 to 20 monomers. **Figure 3.6b** (black trace) shows the minimum conversion barrier height necessary to produce conformer half-life of at least 50 ms at 298 K as a function of the number of monomers. For the smallest $(Ala)_3$ species, the minimum barrier essentially corresponds to the available internal energy. However, for the larger species, the necessary barrier does not scale linearly with internal energy but instead plateaus slightly above 60 kJ/mol for chains longer than 10 monomers. This implies that for the larger species, the conversion rate remains slow enough to produce kinetic trapping even at internal energies well above the barrier. We note that Zwier and co-workers[23] reached similar conclusions by applying similar RRKM calculations to the conversion of the protonated pentapeptide YGPAA in a cryogenic ion trap. The 50 ms time scale is representative of many experiments in which ions are stored at room temperature between each laser spectroscopy or ion mobility spectrometry characterization steps. An alternative approach that favors kinetic trapping in soft-ESI would be to cryogenically cool the ions immediately after the ESI step. To explore this possibility, we

Figure 3.6. (A) Calculated average vibrational internal energy at 298 K for linear polyalanine peptide chains with 3-20 monomers. (B) Minimum conversion barrier height necessary to produce a conformer half-life of at least 50 ms (black trace) or 0.5 ms (red trace) at 298 K as a function of the number of monomers.



calculated the necessary barriers for producing a conformer half-life of 0.5 ms at 298 K, as shown in **Figure 3.6b** (red trace). Despite a 100 x shorter time scale, the required barrier heights are reduced by less than only 20%. This shows that the combination of internal energies and unimolecular reaction rates are the dominant factors in producing kinetically trapped structures. While the RRKM calculations presented here are only crude estimates, they nevertheless provide a rough guide of when kinetic trapping should or should not be expected in soft-ESI.

A similar kinetic analysis can be made for the ion trap clustering approach. At first glance, it would appear to favor the formation of kinetically trapped structures because of its lower ~80 K working temperature. However, we must consider that during the clustering process, the binding energy between the clustering H_2O molecule and the ion will be available as internal energy of the newly formed cluster. For the $GlyH^+ + H_2O$ complex, this binding energy is calculated to be ~77.8 kJ/mol, which is 5 times larger than the internal energy of the 298 K cluster. The formation of kinetically trapped structures would require cooling of this excess energy on unrealistically short picosecond time scale. Using the inelastic collision model of Drahos and Vekey,[24] we estimate that the time scale for cooling at 1×10^{-4} Torr to be on the order of milliseconds, which leaves ample time for conformer conversion. We have demonstrated previously[4] that for $Gly_3H^+(H_2O)$, a ~68 kJ/mol barrier between conformers could be overcome by the ~70 kJ/mol water binding energy during the clustering process at 80 K, yielding predominantly the same lowest energy gas-phase conformer as the soft-ESI approach used by Saparbaev *et al.* Under the specific clustering conditions used in Ref 4, only 10% of the conformer population remained kinetically trapped. Once again, for the larger species, the slower unimolecular dynamics would increase the probability of kinetic trapping. Although we have not explored this possibility in detail yet, the

fraction of kinetically trapped conformers could potentially be controlled by manipulating the ion trap buffer gas pressure and thus the cooling rate of the binding energy.

In summary, the analysis presented here suggests that for relatively small species, both soft-ESI and ion trap clustering approach are likely to yield a conformer distribution determined by gas-phase equilibrium energetics, unless the interconversion barriers are larger than the H_2O binding energy (clustering approach) or the internal energy at room temperature (soft-ESI). The slower unimolecular dynamics in the larger species and clusters would favor the formation of kinetically trapped species in both cases. Finally, we note that one largely unexplored aspect of cryogenic ion trap spectroscopy, common to both approaches discussed here, is the effect of buffer gas cooling on conformer distributions. The fact that multiple conformers are routinely observed after cooling to below 10 K suggests that the collisional cooling is sufficiently fast for preserving some of the more entropically favored conformations. However, it is unclear whether the final observed conformer distribution is representative of the room-temperature equilibrium or an intermediate temperature. Moreover, the extent of observed kinetic trapping is likely to be highly dependent on cooling rate and specific experimental conditions, which will all need to be characterized when comparing results from different laboratories.

3.2 References

(1) Saparbaev, E.; Aladinskaia, V.; Zviagin, A.; Boyarkin, O. V. Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy. *J Phys Chem Lett* **2021,** *12* (2), 907-911.

(2) Fischer, K. C.; Sherman, S. L.; Garand, E. Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and beta-Alanine. *J Phys Chem A* **2020,** *124* (8), 1593-1602.

(3) Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J Phys Chem A* **2019,** *123* (15), 3355-3366.

(4) Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H(+).(H2O). *J Phys Chem A* **2018,** *122* (41), 8213-8221.

(5) Voss, J. M.; Fischer, K. C.; Garand, E. Revealing the structure of isolated peptides: IR-IR predissociation spectroscopy of protonated triglycine isomers. *J. Mol. Spectrosc* **2018,** *347*, 28-34.

(6) Marsh, B. M.; Voss, J. M.; Garand, E. A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *J Chem Phys* **2015,** *143* (20), 204201.

(7) Cao, W.; Zhang, H.; Yuan, Q.; Zhou, X.; Kass, S. R.; Wang, X. B. Observation of Conformational Simplification upon N-Methylation on Amino Acid Iodide Clusters. *J Phys Chem Lett* **2021,** *12* (11), 2780-2787.

(8) Zhang, H.; Cao, W.; Yuan, Q.; Zhou, X.; Valiev, M.; Kass, S. R.; Wang, X. B. Cryogenic "Iodide-Tagging" Photoelectron Spectroscopy: A Sensitive Probe for Specific Binding Sites of Amino Acids. *J Phys Chem Lett* **2020,** *11* (11), 4346-4352.

(9) Fernandez-Lima, F. A.; Becker, C.; Gillig, K. J.; Russell, W. K.; Tichy, S. E.; Russell, D. H. Ion mobility-mass spectrometer interface for collisional activation of mobility separated ions. *Anal Chem* **2009,** *81* (2), 618-24.

(10) Hoaglund-Hyzer, C. S.; Lee, Y. J.; Counterman, A. E.; Clemmer, D. E. Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures. *Anal Chem* **2002,** *74* (5), 992-1006.

(11) Masson, A.; Kamrath, M. Z.; Perez, M. A.; Glover, M. S.; Rothlisberger, U.; Clemmer, D. E.; Rizzo, T. R. Infrared Spectroscopy of Mobility-Selected H+-Gly-Pro-Gly-Gly (GPGG). *J Am Soc Mass Spectrom* **2015,** *26* (9), 1444-54.

(12) Voronina, L.; Masson, A.; Kamrath, M.; Schubert, F.; Clemmer, D.; Baldauf, C.; Rizzo, T. Conformations of Prolyl-Peptide Bonds in the Bradykinin 1-5 Fragment in Solution and in the Gas Phase. *J Am Chem Soc* **2016,** *138* (29), 9224-33.

(13) Voronina, L.; Rizzo, T. R. Spectroscopic studies of kinetically trapped conformations in the gas phase: the case of triply protonated bradykinin. *Phys Chem Chem Phys* **2015,** *17* (39), 25828-36.

(14) Kamrath, M. Z.; Garand, E.; Jordan, P. A.; Leavitt, C. M.; Wolk, A. B.; Van Stipdonk, M. J.; Miller, S. J.; Johnson, M. A. Vibrational characterization of simple peptides using cryogenic infrared photodissociation of H2-tagged, mass-selected ions. *J Am Chem Soc* **2011,** *133* (16), 6440-8.

(15) Wolk, A. B.; Leavitt, C. M.; Garand, E.; Johnson, M. A. Cryogenic ion chemistry and spectroscopy. *Acc Chem Res* **2014,** *47* (1), 202-10.

(16) Tsybizova, A.; Paenurk, E.; Gorbachev, V.; Chen, P. Perturbation of Pyridinium CIVP Spectra by N2 and H2 Tags: An Experimental and BOMD Study. *J Phys Chem A* **2020,** *124* (41), 8519-8528.

(17) Masson, A.; Williams, E. R.; Rizzo, T. R. Molecular hydrogen messengers can lead to structural infidelity: A cautionary tale of protonated glycine. *J Chem Phys* **2015,** *143* (10), 104313.

(18) Eyler, J. R. Infrared multiple photon dissociation spectroscopy of ions in Penning traps. *Mass Spectrom Rev* **2009,** *28* (3), 448-67.

(19) Oomens, J.; Sartakov, B. G.; Meijer, G.; von Helden, G. Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *Int. J. of Mass Spectrom.* **2006,** *254* (1-2), 1-19.

(20) Yeh, L. I.; Okumura, M.; Myers, J. D.; Price, J. M.; Lee, Y. T. Vibrational spectroscopy of the hydrated hydronium cluster ions H3O+·(H2O)n (n=1, 2, 3). *J. Chem. Phys.* **1989,** *91* (12), 7319-7330.

(21) Yang, N.; Duong, C. H.; Kelleher, P. J.; Johnson, M. A.; McCoy, A. B. Isolation of site-specific anharmonicities of individual water molecules in the I−·(H2O)2 complex using tag-free, isotopomer selective IR-IR double resonance. *Chem. Phys. Lett.* **2017,** *690*, 159-171.

(22) Callear, A. B. Chapter 4 Basic RRKM Theory**1983,** *24*, 333-356.

(23) Harrilal, C. P.; DeBlase, A. F.; Fischer, J. L.; Lawler, J. T.; McLuckey, S. A.; Zwier, T. S. Infrared Population Transfer Spectroscopy of Cryo-Cooled Ions: Quantitative Tests of the Effects of Collisional Cooling on the Room Temperature Conformer Populations. *J Phys Chem A* **2018,** *122* (8), 2096-2107.

(24) Drahos, L.; Vekey, K. MassKinetics: a theoretical model of mass spectra incorporating physical processes, reaction kinetics and mathematical descriptions. *J Mass Spectrom* **2001,** *36* (3), 237-63.

CHAPTER 4

Conformational Changes Induced by Methyl Side-Chains in Protonated Tripeptides Containing Glycine and Alanine Residues

Submitted: Sherman, S. L.; Fischer, K. C.; Garand, E. *The Journal of Physical Chemistry A.* **2022.**

Abstract

We present a systematic study of the conformational and isomeric populations in gas-phase protonated tripeptides containing glycine and alanine residues using infrared predissociation spectroscopy of cryogenically cooled ions. Specifically, the protonated forms of Gly-Gly-Gly, Ala-Gly-Gly, Gly-Ala-Gly, Gly-Gly-Ala, Ala-Ala-Gly, Ala-Gly-Ala, Gly-Ala-Ala, and Ala-Ala-Ala allow us to sample all permutations of the methyl side-chain position, providing a comprehensive view of the effects of this simple side-chain on the 3-D structure of the peptide. The individual structural populations for all but one of these peptide species are determined via conformer-specific IR-IR double resonance spectroscopy and comparison with electronic structure predictions. The observed structures can be classified into three main families defined by the protonation site and number of internal hydrogen-bonds. The relative contribution of each structural family is highly dependent on the exact amino acid sequence of the tripeptide. These observed changes in structural population can be rationalized in terms of the electron-donating effect of the methyl side-chain modulating the local proton affinities of the amine and various carbonyl groups in the tripeptide.

Author Contributions

SLS and KCF acquired experimental spectra. SLS performed theoretical computations and data analyses. SLS and EG wrote the manuscript.

4.1 Introduction

Hydrogen bonding (H-bonding), both inter- and intramolecular, plays a vital role in the conformational and folding patterns of peptides and proteins. The overall 3-D structure of peptides and proteins is therefore directly dependent on factors that can modulate the relative strengths of H-bonds. The simple alkane side-chains on amino acid residues have obvious steric effects on H-bonding, e.g. their steric bulk can hinder H-bond formation. However, previous work[1] in our group has shown that the electron donating nature of a methyl side-chain can also affect the strength of nearby H-bonds. Specifically, the methyl side-chain changes the strength of the intramolecular H-bond in protonated alanine compared to protonated glycine, which in turn influences the number of water molecules necessary to break the intramolecular H-bond, leading to the two amino acids having different micro-solvated cluster structures. We rationalized this electronic effect of the methyl group by correlating the observed intramolecular H-bond strengths with the measured proton affinities of the glycine and alanine amino acids. Here, we extend this work by systematically studying the structural landscape of protonated tripeptides containing glycine and alanine residues and analyzing the results in terms of local proton affinities.

The increased chain length in these tripeptides gives rise to many more possible conformations due to their flexibility and ability to fold to create a network of intramolecular H-bonds. In fact, triglycine and trialanine have been shown to be the simplest and smallest peptides that can adopt structural and conformational features of the larger peptides.[2-3] This increasing flexibility and the additional stabilizing effect of internal H-bonds is also reflected in the increasing gas-phase basicity[4-7] and acidity[8] of the peptides with increasing chain length. The electronic effect of the side-chain on the gas-phase basicity of peptides has also been studied.[9-10] Carr *et al.* found that protonated Gly-Gly-Ala and Gly-Ala-Gly featured similar proton transfer reaction efficacies,

but it was slower in Ala-Gly-Gly.[9] This highlighted the effect of the methyl side-chain presence near the N-terminus, the expected protonation site. Bokatzian-Johnson *et al.* determined that the gas-phase acidity of Gly-Ala-Gly, Ala-Gly-Ala, and Ala-Ala-Ala to be slightly smaller than that of Gly-Gly-Gly. This was consistent with electronic structure calculations showing that the methyl substituent had little effect on the structure of the neutral and anionic form of these peptides.[8]

Vibrational spectroscopy of mass-selected ions is a powerful technique for studying 3-D structures of isolated flexible peptides. The frequencies of the N-H, O-H and C=O modes are particularly sensitive to the presence and strength of H-bond interactions, and thus serve as excellent in situ indicators of the overall 3-D structure. Peptide structures in the gas-phase have been widely studied using IR multiple photon dissociation (IRMPD),[11-17] IR-UV hole burning[18-23] and IR predissociation (IRPD) of messenger-tagged ions.[24-29] IRPD can be further enhanced via conformer-specific IR-IR double resonance approaches,[30-37] which provide the vibrational signatures of individual isomers or conformers contributing to the overall spectrum. This is particularly useful in the case of flexible peptides which can adopt a wide range of conformations. For example, in an earlier study, we used IR-IR double-resonance spectroscopy to study the structure of protonated triglycine,[33] and found two isomers, differing by the protonation site, contributing to the overall spectrum.

In this paper, we present a systematic study of the effect of methyl side-chain on the structural populations in protonated tripeptides containing glycine and alanine residues. The protonated forms of Gly-Gly-Gly (**GGG**), Ala-Gly-Gly (**AGG**), Gly-Ala-Gly (**GAG**), Gly-Gly-Ala (**GGA**), Ala-Ala-Gly (**AAG**), Ala-Gly-Ala (**AGA**), Gly-Ala-Ala (**GAA**), and Ala-Ala-Ala (**AAA**) allow us to explore all possibilities of positioning the methyl groups within the tripeptide and provide more meaningful interpretation of the experimental results. The structural population

for each tripeptide species is determined with the help of conformer-specific IR-IR double resonance spectroscopy and electronic structure computations. Comparisons between experimental IRPD and calculated spectra provided structural assignments for 7 of the 8 species. Overall, we found 3 structural families, differing primarily by protonation site and number of internal H-bonds. We use these results to draw conclusions about how the methyl side-chain can influence the structural population ratios. Calculated local proton affinities, modulated by the presence of the methyl side-chain, are used to rationalize these changes in population.

4.2 Experimental and Computational Details

The IRPD spectra presented here were obtained using our home-built dual cryogenic ion trap vibrational spectrometer described in detail previously.[38] Protonated tripeptide ions were generated via electrospray ionization of an ~1 mM tripeptide solution in methanol with trace amounts of formic acid. The GGG and AAA peptides were obtained from Millipore Sigma while the other mixed peptides (AGG, GAG, GGA, AAG, AGA, GAA) were synthesized according to standard methods which are detailed in the **Section 4.3**. Hexapole ion guides transferred the ions through a series of differentially pumped regions and into a linear octupole ion reaction trap, held at 80 K by a liquid nitrogen cryostat. Collisions of the ions in the trap with the helium buffer gas introduced in an ~1 ms burst at the beginning of the trap sequence induced thermalization of the ions. These ions were then transferred to a 3D quadrupole ion tagging main trap held at 10 K by a closed-cycle helium cryostat. Here, the ions were further thermalized with helium buffer gas seeded with 10% D_2, allowing the formation of D_2-tagged adducts. These tagged adducts were then extracted into a time-of-flight (TOF) mass spectrometer. The [protonated tripeptide]$^+ \cdot D_2$ adducts were mass-selected and intersected with the output of a 10 Hz Nd:YAG pumped optical parametric oscillator (OPO)/optical parametric amplifier (OPA) infrared “probe” laser. Resonant

absorption of a single photon was sufficient to induce the loss of the weakly-bound D_2 tag. The photofragment ions were then separated from the parent adducts by a two-stage reflectron. The yield of these photofragment ions as a function of laser wavelength produced the one-laser IRPD spectra.

Conformer-specific IR-IR spectra[30-31] were obtained by focusing the output of a second Nd:YAG pumped tunable OPO/OPA infrared "pump" laser directly into the 10 K tagging main trap, ~94 ms after the introduction of the buffer gas and 1 ms before the ion extraction. This timing scheme limits re-formation of the tagged parent adducts before ejection into the TOF. To acquire an "ion-dip" spectrum, the TOF probe laser was fixed at a particular resonant frequency such that monitoring decreases in the photofragment yield as a function of pump laser wavelength results in an IR-IR spectrum that only contains contributions from the probed vibration. To acquire an "ion-burn" spectrum, the pump laser was fixed at a resonant frequency, thus removing those conformer(s) from the mass selected ion packet interrogated by the TOF probe laser.

The IRPD spectra presented here were normalized to the most intense feature in each spectral region; that is, the 1400-1900 cm^{-1} region was normalized to the C=O stretch, while the 2600-3800 cm^{-1} region was normalized to the OH stretch.

The lowest energy structures of protonated GGG had been identified previously.[10, 33, 39-41] However, to insure that all low energy structures of the other seven species are considered, we performed systematic computational structure searches using the molecular mechanics basin hopping program developed by Hopkins and co-workers.[42-46] A linear chain structure was the starting point and the dihedral angles were iteratively stepped 7,000 times by a random value between ±5°. After each iteration, the resulting structure was optimized using the AMBER force field. The Gaussian 16 program[47] was used for the subsequent DFT and MP2 calculations. The

AMBER structures were optimized at the BLYP/3-21g level, and all duplicate structures were rejected. Structures within 140 kJ/mol of the lowest energy structure were further optimized at the cam-B3LYP/6-311g level and duplicate structures were again rejected. Finally, all structures with a relative energy less than 50 kJ/mol above the lowest energy structure found were optimized at the cam-B3LYP/def2-TZVP/GD3BJ level. Harmonic frequency calculations were then performed on structures with relative energies less than 25 kJ/mol above the lowest energy structure found to determine zero-point energy (ZPE) corrections. This level of theory has previously been shown to yield good agreements with experimental spectra of similar species.[1, 31-33, 38] These lowest energy structures were then optimized at the MP2/def2-TZVP level. The ZPE corrections from the corresponding DFT calculations are applied to the MP2/def2-TZVP energies. Though the MP2 energies have a different ordering than the cam-B3LYP energies, it can help guide relative stabilities of the various structures. Finally, DLPNO-CCSD single point energies are found for the DFT structures at the def2-TZVP def2/J level.[48-49] The relevant calculated results are summarized in **Table 4.1** and the full results are included in **Table 4.2**.

The structure naming scheme used here is similar to Voss *et al.*[33] and Fischer *et al.*[32] It is based on the protonation site (**N** for the amine, and **O** for the amide carbonyl), configuration of the first amide bond (**c** for *cis*, and **t** for *trans* amide), and final letter denoting the conformational family. All the structures discussed here fall into three distinct structure families which are shown in **Figure 4.1**: 1) Amide carbonyl protonated (**O^{t}A**), 2) Amine protonated with all three carbonyls H-bonded to the amine (**N^{c}A**), and 3) Amine protonated with only two carbonyls H-bonded to the amine (**N^{t}B** and **N^{c}B**).

The calculated harmonic IR spectra presented here were scaled in the 2400-3800 cm^{-1} region using the experimental position of each species' respective carboxyl OH stretch, and in the

Table 4.1. Calculated relative energies of tripeptide structures in kJ/mol. DFT energies are from optimized structures at the cam-B3LYP/def2-TZVP/GD3BJ level and include ZPE correction. MP2 energies are from optimized structures at the MP2/def2-TZVP level and include ZPE correction from the DFT structures. DLPNO-CCSD are single point DLPNO-CCSD/def2-TZVP def2/J energies for the DFT structures.

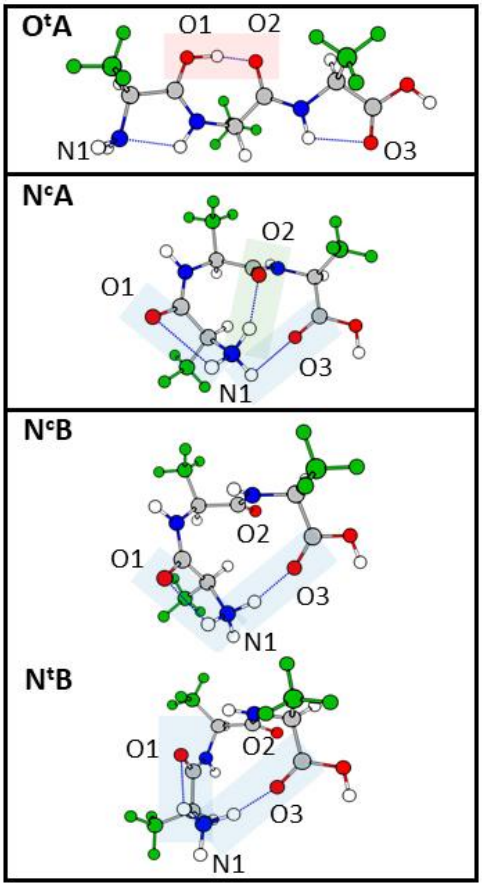
		DFT	MP2	DLPNO-CCSD
GGG	**O^tA**	0.00	7.61	14.61
	N^cA	7.29	0.00	0.00
	N^cB	8.22	4.46	6.40
AGG	**O^tA**	2.05	16.57	21.87
	N^cA	0.00	0.00	0.00
	N^tB	6.26	9.11	6.85
GAG	**O^tA**	0.00	4.55	11.12
	N^cA	9.43	0.00	0.00
	N^tB	8.42	2.29	3.98
GGA	**O^tA**	0.00	5.57	13.03
	N^cA	8.32	0.00	0.00
	N^cB	8.94	3.19	5.35
	N^tB	12.21	4.96	5.43
AAG	**O^tA**	0.75	13.54	18.24
	N^cA	1.02	0.00	0.00
	N^tB	0.00	2.80	3.53
AGA	**O^tA**	1.04	14.61	20.18
	N^cA	0.00	0.00	0.00
	N^cB	1.88	3.84	5.47
GAA	**O^tA**	0.00	2.05	6.59
	N^cB	8.34	0.00	0.00
AAA	**O^tA**	0.00	9.61	14.27
	N^cA	7.54	5.06	2.68
	N^cB	0.16	0.00	0.00

Table 4.2. Calculated relative energies (kJ/mol) of lowest energy isomers for all species. The cam-B3LYP energies include ZPE correction. MP2 include the cam-B3LYP ZPE correction. DLPNO-CCSD and LCCSD(T) are single point energies at the cam-B3LYP structures. LLCSD(T) include ZPE scaling corrections. Those structures that are bolded are found within the experiment.

		cam-B3LYP/ def2-TZVP/GD3BJ	MP2/ def2-TZVP	DLPNO-CCSD/ def2-TZVP	LCCSD(T)/ def2-TZVP
	OtA	0.00	7.61	14.69	1.15
	NcA	7.29	0.00	0.07	1.63
GGG	NcA*	7.29	0.00	0.00	0.00
	NcB	8.22	4.46	6.47	2.69
	NtB	11.35	6.27	8.10	8.70
	NtB*	11.35	4.46	8.09	2.16
	OtA	2.05	16.57	22.02	19.66
	NcA	0.00	0.00	0.15	0.00
AGG	NcA*	0.00	0.00	0.00	11.31
	NcB	2.17	5.49	7.01	5.31
	NtB	3.92	6.66	7.52	7.87
	NtB*	6.26	9.11	7.00	8.53
	OtA	0.00	4.55	11.12	0.00
	NcA	9.43	0.00	0.00	0.01
GAG	NcA*	17.49	8.87	8.46	31.68
	NcB	7.75	1.97	4.44	2.09
	NtB	23.13	17.13	18.18	57.95
	NtB*	8.42	2.29	3.98	2.55
	OtA	0.00	5.57	13.03	0.76
	NcA	14.66	6.02	5.32	6.07
GGA	**NcA***	8.32	0.00	0.00	0.00
	NcB	8.94	3.19	5.35	3.89
	NtB	12.21	4.96	5.43	7.71
	NtB*	20.91	13.23	13.21	62.70
	OtA	0.75	13.54	18.24	18.90
	NcA	1.02	0.00	0.00	0.00
AAG	NcA*	8.79	8.28	8.28	38.02
	NcB	0.63	2.92	5.08	3.03
	NtB	14.71	17.70	17.82	55.15
	NtB*	0.00	2.80	3.53	1.96
	OtA	1.04	14.61	20.18	14.51
	NcA	6.48	6.20	5.58	1.55
AGA	**NcA***	0.00	0.00	0.00	3.87
	NcB	1.88	3.84	5.47	0.00
	NtB	3.58	5.31	4.96	0.35

	NtB*	12.17	13.48	12.59	55.45
GAA	**OtA**	0.00	2.05	6.59	0.00
	NcA	17.01	5.70	2.85	8.22
	NcA*	18.64	8.40	5.72	28.70
	NcB	8.34	0.00	0.00	3.28
	NtB	23.45	15.12	12.92	56.66
	NtB*	17.95	8.80	7.09	63.42
AAA	**OtA**	0.00	9.61	14.27	2.44
	NcA	7.54	5.06	2.68	4.06
	NcA*	8.73	6.93	5.33	15.37
	NcB	0.16	0.00	0.00	0.00
	NtB	13.83	14.91	12.39	23.46
	NtB*	8.06	8.35	6.22	51.33

Figure 4.1. Representative structures of the low-energy structure families. **O^{t}A** is protonated at the O1 site; **N^{c}A** is protonated at the N1 amine with three internal H-bonds; **N^{x}B** is also protonated at the N1 amine but only has two internal H-bonds. The methyl side-chain groups are in green.



1400-1900 cm^{-1} based on each species' respective C=O stretches. The calculated spectra are Gaussian broadened with areas corresponding to the calculated intensities to facilitate comparisons to the experimental spectra.

Furthermore, in order to help rationalize the experimental observations, the site-specific proton affinities of the tripeptides were computed at the camB3LYP/def2-TZVP level. This was achieved using a rigid linear peptide chain structure with fixed backbone dihedral angles and placing a proton at the terminal amine (N1), the carbonyl closest to the amine terminal (O1), the central carbonyl (O2), and the carboxyl terminal (O3), shown in **Figure 4.2**. The proton affinity was then computed using a standard method.[50] The calculated results are summarized in **Table 4.3**.

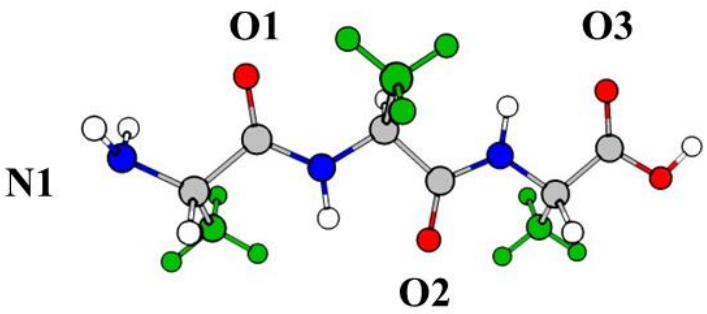
4.3 Synthesis of Tripeptides

The synthesis method detailed here is a standard method for synthesizing peptides in the Gellman Group at the University of Wisconsin, Madison. Mary Katherine Andrews in the Gellman Group synthesized all non-commercialized tripeptides with the following method.

4.3.1 General Procedure

Peptide Synthesis and Purification NovaPEG Wangresin (200 μmol by acid loading [0.9 mmol/g], 222 mg) is added to a Torviq solid-phase peptide synthesis vessel along with a micro stir bar. Resin is swelled in DCM for 30 minutes before beginning synthesis. Fmoc amino acid (3.5 eq., 0.7 mmol) and HATU (3.45 eq., 0.69 mmol) are weighed into 24 mL glass vials. Amino acid and coupling reagent are pre-activated by adding 12 mL biotech grade DMF, DIEA (8 eq., 280 μL, 1.6 mmol), and vortexing. The pre-activated amino acid sits at room temperature for 5 minutes.

Figure 4.2. Structure of tripeptides used for calculating the proton affinities at various positions. Each protonated structure energy was compared to that of the top linear neutral structure. N1 has the excess proton located on the amine, O1 has the excess proton located on the carbonyl at the N-terminus, O2 has the excess proton located on the central carbonyl, and O3 has the excess proton located on the carboxylic acid.



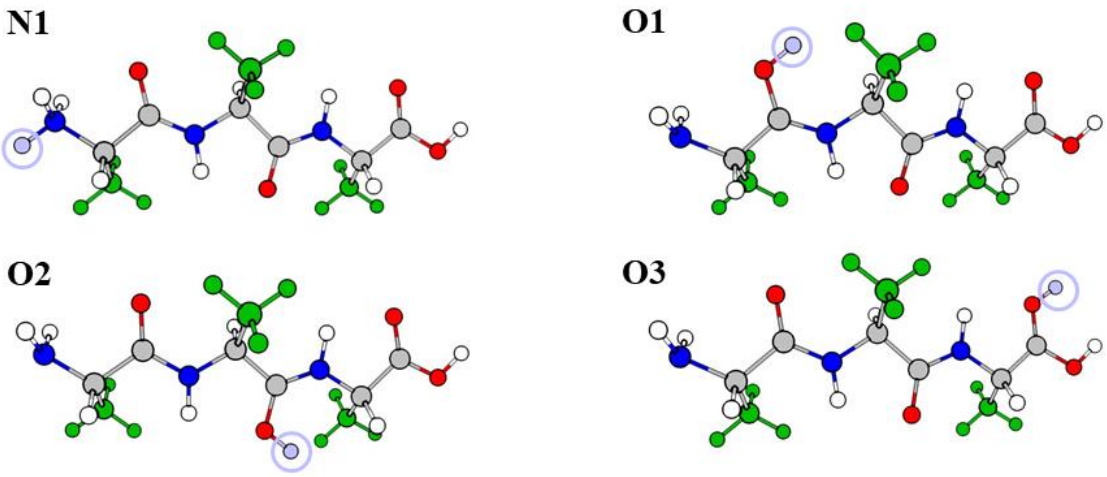


Table 4.3. Calculated proton affinities of each linear tripeptide at each possible site of protonation. All proton affinity values are in kJ/mol. Proton affinities are found at the cam-B3LYP/def2-TZVP level with ZPE correction and temperature corrections to the vibrational enthalpy.

	Position	PA (kJ/mol)
GGG	N1	888.5
	O1	855.4
	O2	803.5
	O3	638.0
AGG	N1	895.6
	O1	857.5
	O2	800.4
	O3	634.2
GAG	N1	889.0
	O1	860.0
	O2	807.8
	O3	640.9
GGA	N1	888.3
	O1	858.8
	O2	810.7
	O3	650.3
AAG	N1	898.3
	O1	862.1
	O2	804.0
	O3	636.2
AGA	N1	897.9
	O1	862.1
	O2	811.2
	O3	643.4
GAA	N1	887.1
	O1	860.3
	O2	808.0
	O3	644.2
AAA	N1	900.6
	O1	866.6
	O2	812.0
	O3	645.1

DCM used to swell the resin is aspirated from the vessel, the resin is washed 5X with DMF and then capped at the bottom. The pre-activated amino acid solution is added to the vessel.

4.3.2 *Coupling of amino acids*

The reaction vessel containing the resin and pre-activated amino acid is transferred to the microwave reactor, and the temperature sensor is placed in the vessel. The microwave coupling program is initiated which S3 ramps to 70ºC over two minutes, and holds at 70ºC for thirteen minutes. After the coupling, the vessel is taken out of the microwave and aspirated. The resin is then washed 5X with ACS grade DMF. The vessel is then capped.

4.3.3 *Deprotection of amino acids*

To the reaction vessel is added 12 mL of 20 %v/v piperidine in ACS DMF, and the vessel is moved to the microwave reactor. The temperature probe is placed in the vessel. The microwave deprotection program is initiated, which ramps to 80ºC over two minutes and holds at 80ºC for four minutes. After deprotection, the vessel is aspirated and the resin is washed 5X with ACS DMF. Subsequent coupling and deprotection steps are carried out until the final residue has been coupled.

4.3.4 *Cleavage of peptides from resin*

Cleavage is performed by withdrawing 12 mL of 2.5 % TIPS, 2.5 % Water, 95% TFA into the reaction vessel using the syringe plunger and rocking the vessel for 3 hour at room temperature. Crude peptide solution is expunged into a 50-mL centrifuge tube, resin is washed 3X with TFA, and TFA is blown off under a stream of N^2. Once most of the TFA is removed, the crude peptide

is precipitated by addition of 40 mL cold diethyl ether and pelleted using a centrifuge at 4,300 rpm for 5 minutes. The supernatant ether is decanted, and the crude peptide solid is dried under a stream of N_2.

4.4 Calculating Conformer/Isomer Population Ratios

4.4.1 Isomer-specific Parent Ion Depletion Experiment

Isomer depletion experiments in the cryogenic ion trap were performed for GGG and AAA at three different wavelengths: at the carboxyl O-H (common to all isomers) and at wavelengths unique to the O^tA, and the N^cA/N^cB conformer. The laser pulse energy was increased in small steps, and the remaining parent ion mass intensity was recorded. Between each data point recording, a "blank" of the full signal was recorded to account for ion signal fluctuations due to the signal over time. The normalized parent ion intensity was then plotted against the laser pulse energy and fitted with an exponential decay function. The carboxylic acid O-H, common to all isomers/conformers, was used to ensure that the infinite laser pulse energy extrapolation was close to ~100% depletion of the parent signal. A similar extrapolation for the wavelengths unique to each isomer/conformer was used to determine their relative population. The peaks chosen for GGG were 3573 cm^{-1} (free carboxylic O-H), 3480 cm^{-1} (N^cA), and 3398 cm^{-1} (O^tA). At infinite laser pulse energy, the depletion were found to be 95(3)% at 3573 cm^{-1}, 26(5)% at 3480 cm^{-1}, and 66(11)% at 3398 cm^{-1}. The peaks chosen for AAA were 3573 cm^{-1} (free O-H), 3445 cm^{-1} (O^tA), and 3419 cm^{-1} (N^cB). At infinite laser pulse energy, the depletion were found to be 94(1)% at 3573 cm^{-1}, 62(11)% at 3445 cm^{-1}, and 34(5)% at 3419 cm^{-1}.

4.4.2 Calculated Intensity Comparison Method

The second method used to determine the conformer population uses the calculated intensities of unique bands. Gaussian curves were fitted to peak unique to individual conformers in the experimental spectra and used to determine their relative intensities. The experimental peak area was then divided by the normalized computed absorption cross-section at the cam-B3LYP/def2-TZVP/GD3BJ level of theory.

4.5 Results

The IRPD overview (1600-3600 cm^{-1}) spectra of all 8 tripeptides considered here are shown in **Figure 4.3**. Four spectral regions of interest are highlighted: 1) The free (i.e. non H-bonded) carboxyl O-H stretch around 3575 cm^{-1} in green, 2) the free amine/amide N-H stretches between 3380-3500 cm^{-1} in blue, 3) the broad and redshifted H-bond donor N-H and O-H stretches between 2900-3400 cm^{-1} highlighted in purple, and 4) the carbonyl C=O and amide I modes between 1650-1800 cm^{-1} in orange. We note that the various C-H modes appear between 2900 cm^{-1} and 3050 cm^{-1} but are expected to be quite weak compared to the O-H and N-H modes. A summary of all peak assignments is shown in **Table 4.4**.

Notable variations in the IRPD spectra can be clearly seen for this series of tripeptides. They are most striking for the N-H and C=O stretches, whose positions are highly sensitive to intramolecular H-bonds and, in turn, to the overall peptide conformation. These overview spectra clearly show that the introduction of a simple methyl side-chain can heavily modulate the structural populations of the tripeptide. The use of systematic Gly/Ala substitutions here can therefore inform on the mechanism by which this modulation occurs. In order to perform this analysis, the isomeric structure and population for each tripeptide must first be determined.

Figure 4.3. Overview of the IRPD spectra of all 8 peptides considered here in the 1600-3600 cm^{-1} range. The free O-H region is highlighted in green, the free N-H in blue, H-bonded N-H & O-H in purple, and the C=O region in orange.

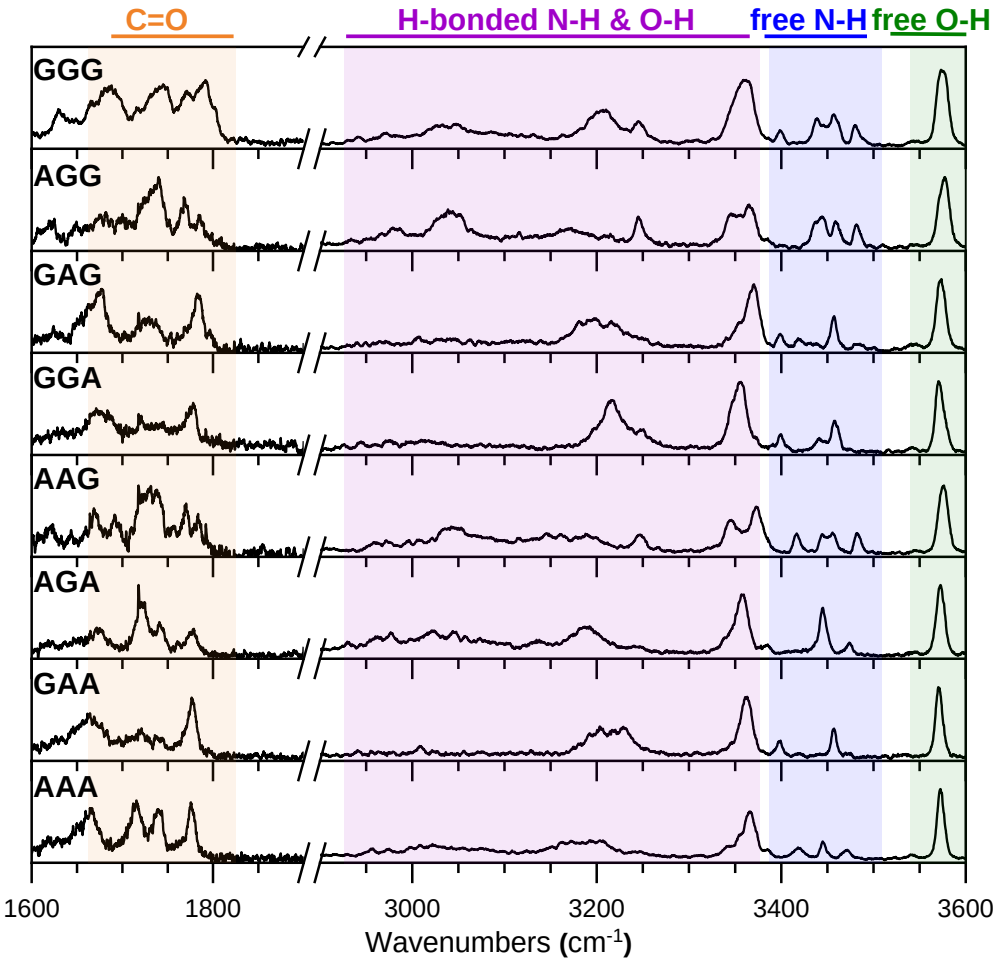


Table 4.4. IRPD peak positions and assignments. All values are in cm^{-1}.

Peptide	Assignment	Exp. Freq.	Isomer Specific features		
	OH_{CA} str	3575			
	$NH_{2,\ asym.}$	3457 **(O^tA)**	3347 **(N^cA)**		
	$NH_{2,\ sym.}$	3397 **(O^tA)**	3246 **(N^cA)**		
	N^1H		3035 **(N^cA)**		
GGG	N^2H	3205 **(O^tA)**	3438 **(N^cA)**		
	N^3H	3358 **(O^tA)**	3481 **(N^cA)**		
	CH	2900-3000			
	C=O str	1790 **(O^tA)**	1770 **(N^cA)**		
	OH_{CA} str	3575			
	$NH_{2,\ asym.}$	3446 **(O^tA)**	3352 **(N^cA)**		3206 **(N^tB*)**
	$NH_{2,\ sym.}$	3385 **(O^tA)**	3247 **(N^cA)**		3350 **(N^tB*)**
	N^1H		3042 **(N^cA)**		
AGG	N^2H	3170 **(O^tA)**	3437 **(N^cA)**		3460 **(N^tB*)**
	N^3H	3366 **(O^tA)**	3483 **(N^cA)**		3446 **(N^tB*)**
	CH	2880-3025			
	C=O str	1784 **(O^tA)**	1769 **(N^cA)**		1743 **(N^tB*)**
	OH_{CA} str	3575			
	$NH_{2,\ asym.}$	3456 **(O^tA)**			3359 **(N^tB*)**
	$NH_{2,\ sym.}$	3398 **(O^tA)**			3189 **(N^tB*)**
	N^1H				3022 **(N^tB*)**
GAG	N^2H	3216 **(O^tA)**			3418 **(N^tB*)**
	N^3H	3370 **(O^tA)**			3456 **(N^tB*)**
	CH	2890-3125			
	C=O str	1785 **(O^tA)**			1742 **(N^tB*)**
	OH_{CA} str	3573			
	$NH_{2,\ asym.}$	3458 **(O^tA)**	3350 **(N^cA*)**	3090 **(N^cB)**	3214 **(N^tB)**
	$NH_{2,\ sym.}$	3400 **(O^tA)**	3251 **(N^cA*)**	3052 **(N^cB)**	2976 **(N^tB)**
	N^1H		3027 **(N^cA*)**	3361 **(N^cB)**	3450 asy, 3359 sym **(N^tB)**
GGA	N^2H	3219 **(O^tA)**	3440 **(N^cA*)**	3442 **(N^cB)**	3450 asy, 3359 sym **(N^tB)**
	N^3H	3356 **(O^tA)**	3462 **(N^cA*)**	3476 **(N^cB)**	3450 **(N^tB)**
	CH	2911-3031			
	C=O str	1779 **(O^tA)**	1750 **(N^cA*)**	1737 **(N^cB)**	1737 **(N^tB)**
	OH_{CA} str	3574			
	$NH_{2,\ asym.}$	3457 **(O^tA)**	3349 **(N^cA)**		3166 **(N^tB*)**
	$NH_{2,\ sym.}$	3388 **(O^tA)**	3148 **(N^cA)**		3070 **(N^tB*)**
	N^1H		3045 **(N^cA)**		3351 **(N^tB*)**
AAG	N^2H	3192 **(O^tA)**	3421 **(N^cA)**		3444 **(N^tB*)**

	N^3H	3374 **(O^tA)**	3482 **(N^cA)**		3454 **(N^tB*)**
	CH	2870-3040			
	C=O str	1777 **(O^tA)**	1739 **(N^cA)**		1744 **(N^tB*)**
	OH_{CA} str	3574			
	$NH_{2,\ asym.}$	3474 **(O^tA)**		3140 **(N^cB)**	3237 **(N^tB)**
	$NH_{2,\ sym.}$	3388 **(O^tA)**		3023 **(N^cB)**	2975 **(N^tB)**
	N^1H			3349 **(N^cB)**	
AGA	N^2H	3190 **(O^tA)**		3441 **(N^cB)**	3446 asy, 3346 sym **(N^tB)**
	N^3H	3360 **(O^tA)**		3474 **(N^cB)**	3446 asy, 3346 sym **(N^tB)**
	CH	2880-3070			
	C=O str	1780 **(O^tA)**		1742 **(N^cB)**	1742 **(N^tB)**
	OH_{CA} str	3547			
	$NH_{2,\ asym.}$	3457 **(O^tA)**		3010 **(N^cB)**	
	$NH_{2,\ sym.}$	3400 **(O^tA)**		2953 **(N^cB)**	
	N^1H			3361 **(N^cB)**	
GAA	N^2H	3228 **(O^tA)**		3418 **(N^cB)**	
	N^3H	3362 **(O^tA)**		3472 **(N^cB)**	
	CH	2860-3050			
	C=O str	1778 **(O^tA)**		1737 **(N^cB)**	
	OH_{CA} str	3574			
	$NH_{2,\ asym.}$	3447 **(O^tA)**		3171 **(N^cB)**	
	$NH_{2,\ sym.}$	3382 **(O^tA)**		3023 **(N^cB)**	
	N^1H			3355 **(N^cB)**	
AAA	N^2H	3195 **(O^tA)**		3420 **(N^cB)**	
	N^3H	3367 **(O^tA)**		3473 **(N^cB)**	
	CH	2920-3080			
	C=O str	1776 **(O^tA)**		1742 **(N^cB)**	

Figure 4.4 and **Figure 4.5** shows the spectra of GGG and AAA in the 1400-1900 cm^{-1} and 2800-3800 cm^{-1} regions. For both species, the number of peaks within the C=O stretch and N-H stretch regions is more than what would be expected for a single tripeptide structure, indicating the presence of more than one isomer and/or conformer. The spectral deconvolution, assignment, and quantification for GGG was described in detail previously[33] and the same procedure is used here. Briefly, using an IR-IR ion-dip scheme with the probe laser fixed at 3480 cm^{-1} and 3200 cm^{-1}, two distinct spectra were obtained, as shown in grey in **Figure 4.4**. Comparing these spectra with the calculated spectra of some of the lowest energy structures found in the computational search yields good agreement with the calculated spectra of $\mathbf{N^cA}$ and $\mathbf{O^tA}$, as shown. These two structures are sufficient for making assignments to all the peaks in the experimental spectrum and a third potential structure based on relative energies, $\mathbf{N^cB}$, can be ruled out. The isomer population distribution can be estimated by comparing experimental peak areas and calculated intensities of peaks unique to each isomer. In GGG, we chose the symmetric and asymmetric stretches of the free NH_2 (3400 cm^{-1} and 3457 cm^{-1} respectively) of $\mathbf{O^tA}$ and the amide stretches (3440 cm^{-1} and 3482 cm^{-1}) of $\mathbf{N^cA}$. This analysis yields $\mathbf{O^tA}$ and $\mathbf{N^cA}$ contributions of 65(18)% and 35(3)%, respectively.

To verify the validity of this method for the tripeptide species, we also carried out an isomer depletion experiment. The fraction of ion depletion resulting from the IR absorption at a wavelength unique to each isomer was measured as a function of laser pulse intensity.[51-52] The resulting data was fitted with an exponential decay function and extrapolated to infinite laser pulse intensity. The isomer ratio obtained from the isomer depletion experiment was found to be consistent within error bars to the ratio found via intensity comparison method described in the previous paragraph. Therefore, we opt to use the easier intensity comparison method for all the

Figure 4.4. IRPD spectra of Gly-Gly-Gly H^+ compared to harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level. The isomer specific IR-IR spectra are overlaid on the calculated spectra of the structure that they correspond. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.

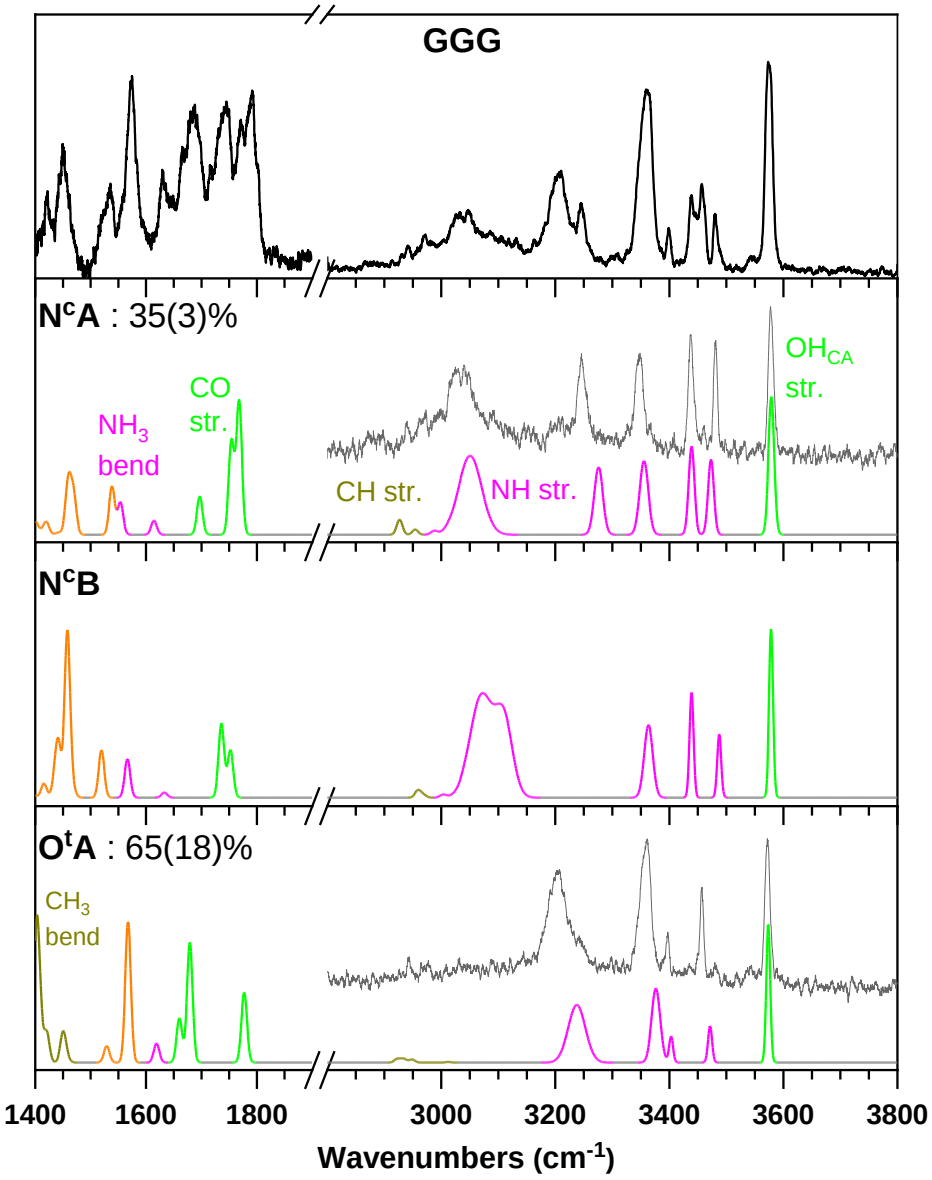
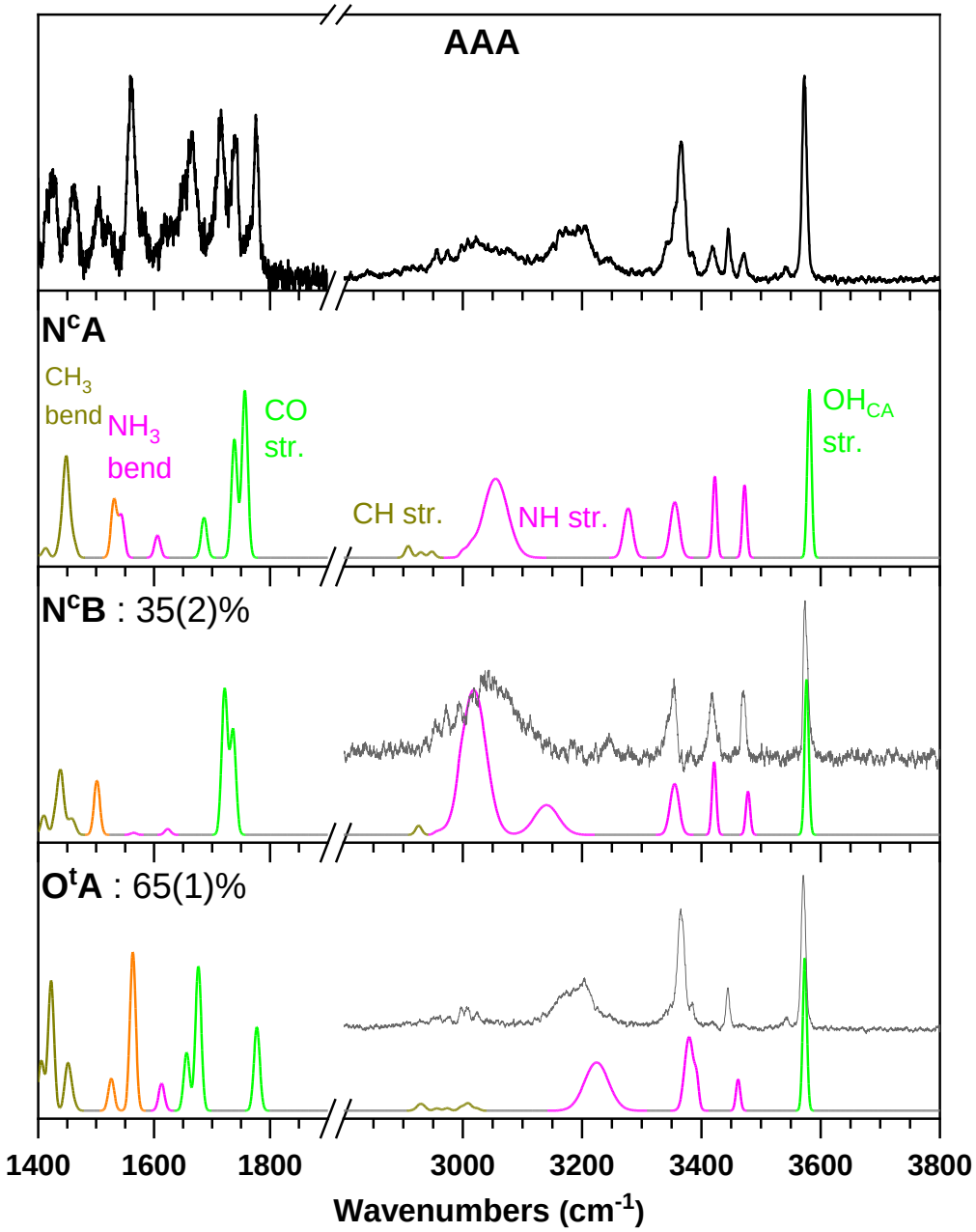


Figure 4.5. IRPD spectra of Ala-Ala-Ala H^+ compared to the lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the isomer specific ion dip spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.



other species presented in this paper. More details on both isomer quantification methods can be found in **Section 4.4**.

A similar approach is applied to the analysis of AAA IRPD spectra, shown in **Figure 4.5**. An IR-IR ion-burn with the pump laser fixed at 3471 cm^{-1} produces a spectrum in good agreement with the calculated spectrum of $\mathbf{O^tA}$. By subtracting this IR-IR spectrum from the ion-burn spectrum acquired with laser fixed at 3448 cm^{-1}, we obtained a spectrum which matches the calculated spectrum of $\mathbf{N^cB}$. Significant contribution from $\mathbf{N^cA}$ can be eliminated based on the absence of a peak around 3275 cm^{-1}. Using the experimental peak areas and calculated intensities of the symmetric and asymmetric stretches of the free NH_2 (3380 cm^{-1} and 3444 cm^{-1} respectively) of $\mathbf{O^tA}$ as well as the amide stretches (3420 cm^{-1} and 3472 cm^{-1}) of $\mathbf{N^cB}$ indicate an $\mathbf{O^tA}$ and $\mathbf{N^cB}$ contributions of 65(1)% and 35(2)%, respectively.

The structure identification and quantification for all other tripeptides follow the same methods as described above. Therefore, only the relevant details will be highlighted in the next two sub-sections.

4.5.1 *One Alanine substitution, (AGG, GAG, GGA)*

The IRPD spectra of AGG is shown in the top panel in **Figure 4.6** and **Figure 4.7**. It contains six features in the free amine NH and carboxyl OH stretch regions, which again indicate the presence of more than one structure. Using the IR-IR ion-dip approach with probe laser fixed at 3482 cm^{-1}, 3459 cm^{-1}, and 3370 cm^{-1} allowed us to isolate the spectral signatures of three distinct structures. These isomer-specific spectra, shown in the three lower panels of **Figure 4.6**, are in excellent agreement with the calculated spectra of $\mathbf{N^cA}$, $\mathbf{N^tB}$, and $\mathbf{O^tA}$, respectively. All three structures were found to have nearly equal population of 30(1)%, 40(1)%, and 30(2)%,

Figure 4.6. IRPD spectra of Ala-Gly-Gly H^+ compared to the lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the isomer specific ion dip spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.

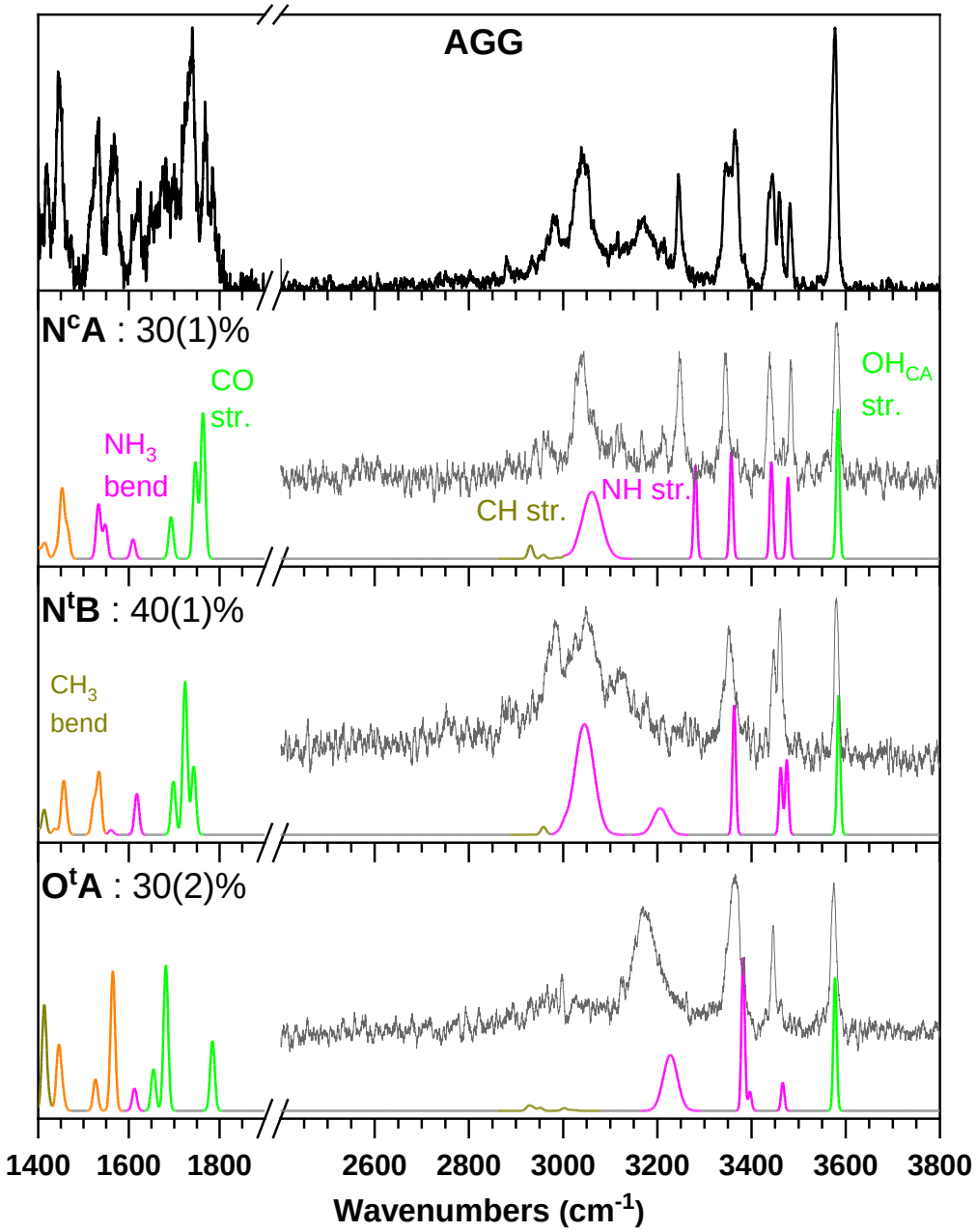
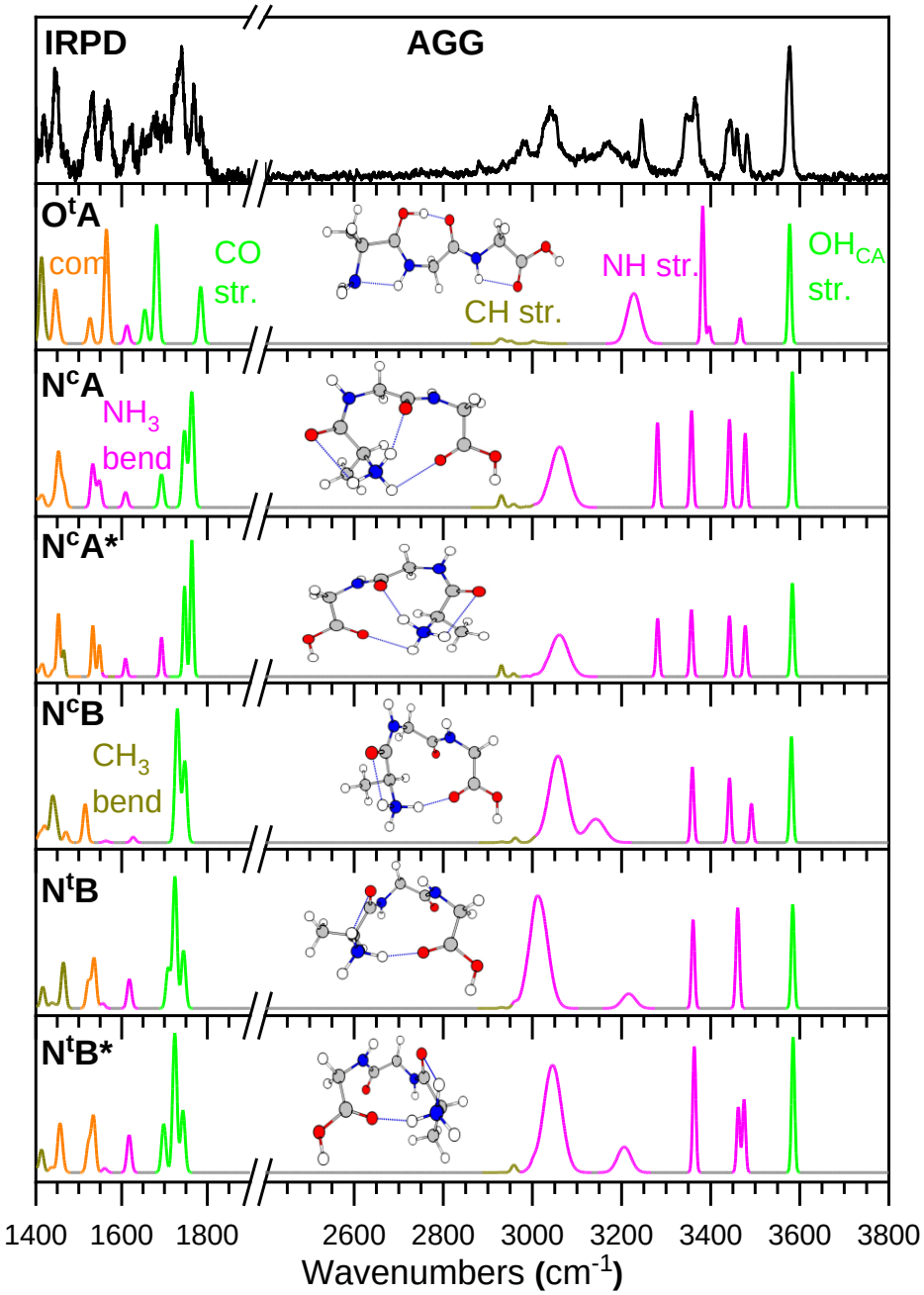


Figure 4.7. IRPD spectra of AGG and calculated harmonic spectra of the six lowest energy structures at the cam-B3LYP/def2-TZVP/GD3BJ level of theory. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. For relative energies of these structures, see **Table 4.2**.



respectively. The IRPD spectra of AGG and calculated harmonic spectra of the six lowest energy structures is shown in **Figure 4.7**.

The one-laser IRPD spectrum of GAG is shown in **Figure 4.8** and **Figure 4.9**. The presence of the distinctive symmetric (3399 cm^{-1}) and asymmetric (3460 cm^{-1}) NH_2 stretches suggest contribution from **O^tA**. To verify this, we carried out IR-IR ion-burn with the pump laser fixed at 3424 cm^{-1}, corresponding to a free amide stretch of a different isomer. This spectrum is overlaid on the calculated spectra of **O^tA** on the bottom panel of **Figure 4.8** and displays excellent agreement. The ion-burn spectrum is then subtracted from the one-laser IRPD spectra, yielding a spectrum in agreement with **N^tB** (middle panel, **Figure 4.8**). Hence, these two isomers account for all the observed peaks in the IRPD spectrum. The symmetric stretch of the free NH_2 (3399 cm^{-1}) of **O^tA** and free amide N-H amide stretch (3419 cm^{-1}) yield an **O^tA/N^tB** ratio of 70(2)%/30(2)%. The IRPD spectra of GAG and calculated harmonic spectra of the six lowest energy structures is shown in **Figure 4.9**.

The IRPD spectrum of GGA is shown in **Figure 4.10** and **Figure 4.11**. The general appearance of the spectrum suggest that **O^tA** is the most abundant structure. The **O^tA** spectrum was isolated through IR-IR ion-dip with the probe laser fixed at 3399 cm^{-1} (bottom panel, **Figure 4.10**), yielding a spectrum in excellent agreement with the calculated one. Almost all peaks in the one-laser spectrum are accounted for by **O^tA**, except shoulder peaks at 3247 cm^{-1} and 3441 cm^{-1}. Another IR-IR ion dip with the probe laser fixed on peak 3441 cm^{-1} was acquired (second panel, **Figure 4.10**). It displays at least three partially resolved peaks between 3400 cm^{-1} and 3500 cm^{-1}, indicating the presence of at least two of the **N^cA**, **N^cB and N^tB** conformers. Unfortunately, these peaks are not resolved enough to be separated by IR-IR methods. Although it is clear that **O^tA** is the dominant isomer, we are not able to confidently assign structure ratios for GGA due to the

Figure 4.8. IRPD spectra of Gly-Ala-Gly H^+ compared to lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the conformer-specific IR-IR spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.

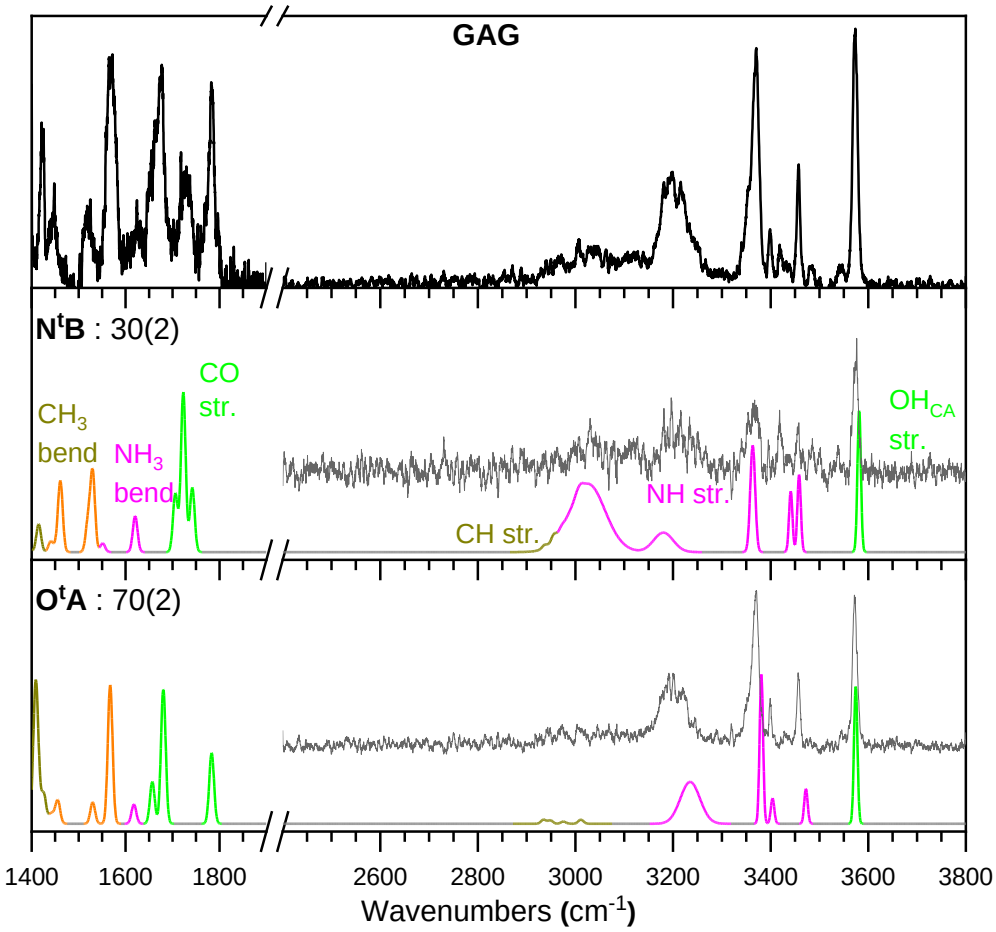


Figure 4.9. IRPD spectra of GAG and calculated harmonic spectra of the six lowest energy structures at the cam-B3LYP/def2-TZVP/GD3BJ level of theory. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. For relative energies of these structures, see **Table 4.2**.

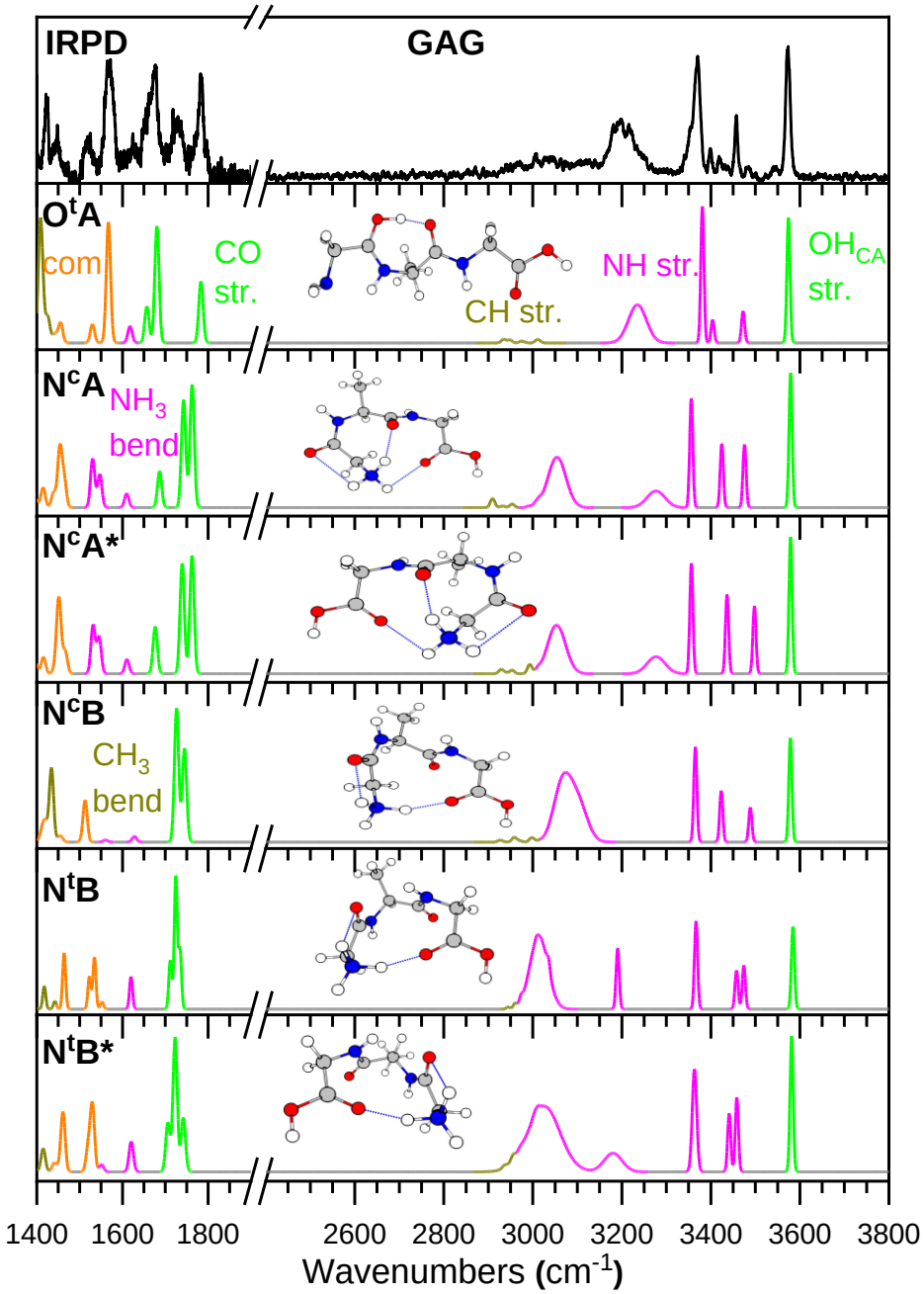


Figure 4.10. IRPD spectra of Gly-Gly-Ala H^+ compared to lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the conformer-specific IR-IR spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra.

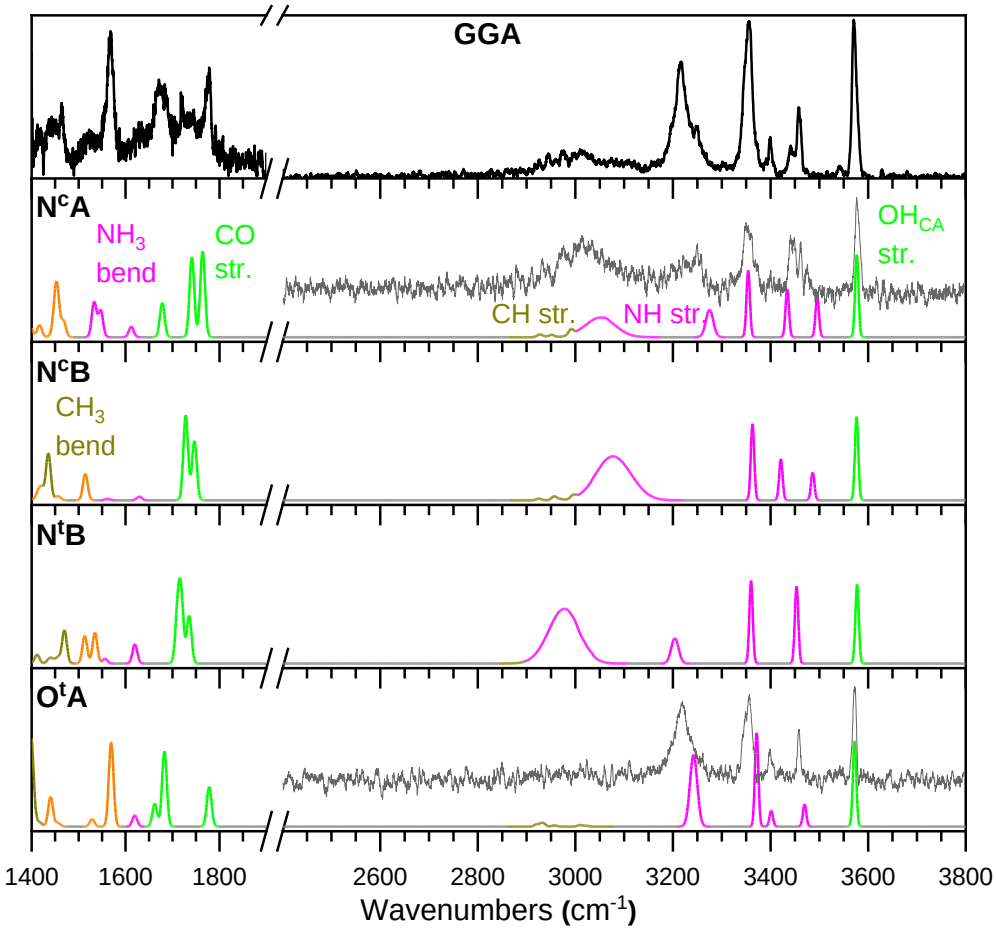
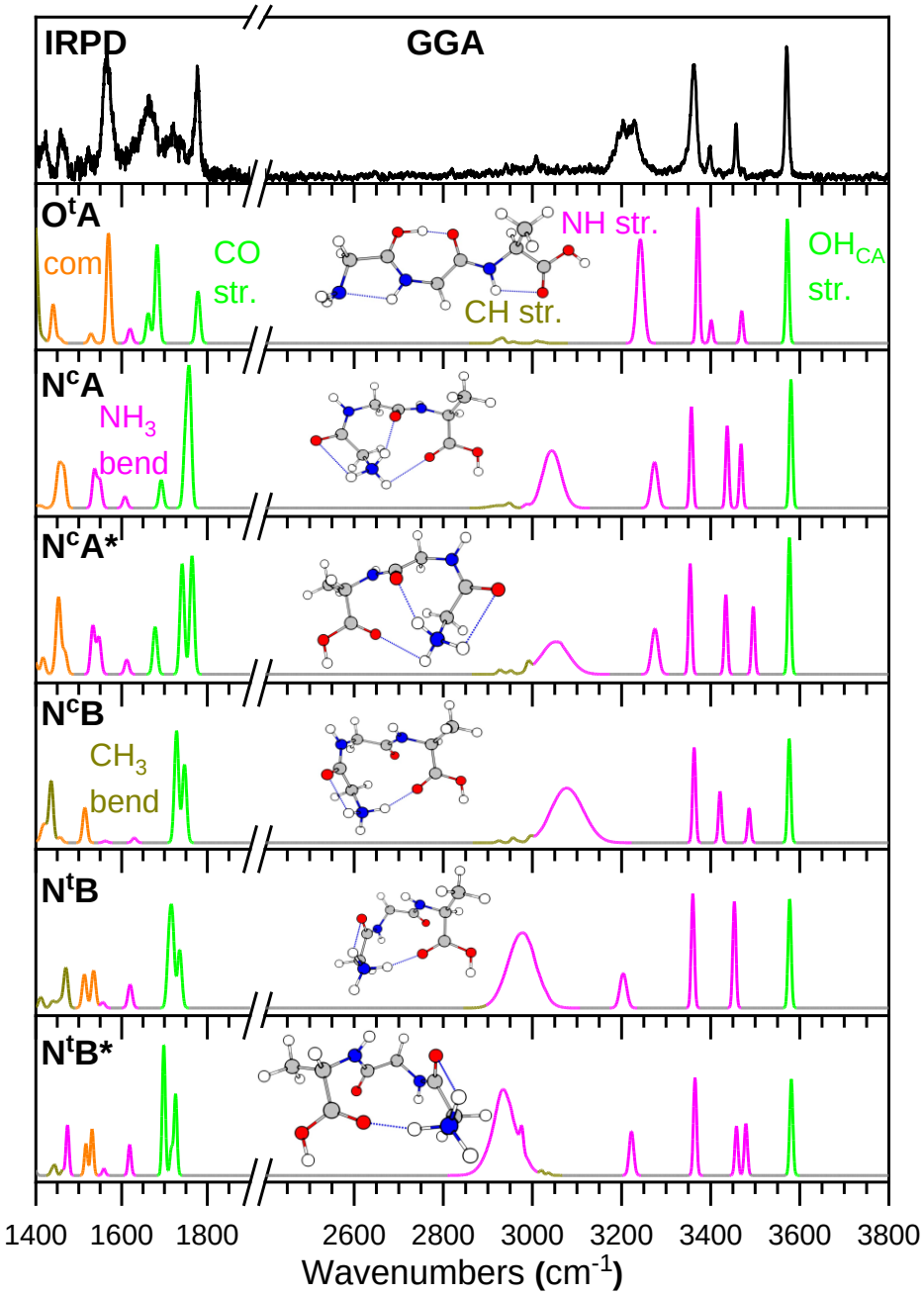


Figure 4.11. IRPD spectra of GGA calculated harmonic spectra of the six lowest energy structures at the cam-B3LYP/def2-TZVP/GD3BJ level of theory. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. For relative energies of these structures, see **Table 4.2**.



inability to fully identify and quantify all the structures present. The IRPD spectra of GGA and calculated harmonic spectra of the six lowest energy structures is shown in **Figure 4.11**.

4.5.2 *Two Alanine substitution, (AAG, AGA, and GAA)*

The IRPD spectrum of AAG as well as the three lowest energy structures, **O^{t}A**, **N^{c}A**, and **N^{t}B**, are shown in **Figure 4.12** and **Figure 4.13**. Once again, there are numerous features in the 3100-3500 cm^{-1} region, indicating the presence of multiple structures. IR-IR ion-dip with the probe laser fixed at 3482 cm^{-1} (second panel, **Figure 4.12**), yielded a spectrum in good agreement with the calculated spectrum of **N^{c}A**. The probe laser was then set to 3445 cm^{-1} and we were able to obtain a spectrum that shows good agreement with the calculated spectrum of **N^{t}B**. These two conformers cover majority of the peaks within the overall IRPD spectrum, with the exception of a peak at 3376 cm^{-1}. Thus, setting the probe laser at that wavelength yielded a spectrum in good agreement with the calculated **O^{t}A** spectrum. Utilizing the N-H free amine and amide stretches from **N^{c}A** (3419 cm^{-1} and 3483 cm^{-1}) and **N^{t}B** (3458 cm^{-1}) as well as H-bonded amide stretch from **O^{t}A** (3373 cm^{-1}) we obtain a **O^{t}A/N^{c}A/N^{t}B** ratio of 45(1)%/30(1)%/25(1)%. The IRPD spectra of AAG and calculated harmonic spectra of the six lowest energy structures is shown in **Figure 4.13**.

The IRPD spectrum of AGA is shown in **Figure 4.14** and **Figure 4.15**. The main N-H peaks at 3188, 3356, and 3447 cm^{-1} and the minor N-H peak at 3384 are consistent with **O^{t}A** being the main isomer. This was confirmed using IR-IR ion dip at 3188 cm^{-1}. This assignment leaves only the small peak at 3473 cm^{-1} unexplained. Subtracting the **O^{t}A** IR-IR spectrum from the one-laser IRPD gives us the signature of the minor isomer which matches nicely with **N^{c}B**. Given this assignment, we obtained a **O^{t}A/N^{c}B** population ratio of 85(6)%/15(7)%, using unique peaks

Figure 4.12. IRPD spectra of Ala-Ala-Gly H^+ compared to lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the conformer-specific IR-IR spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.

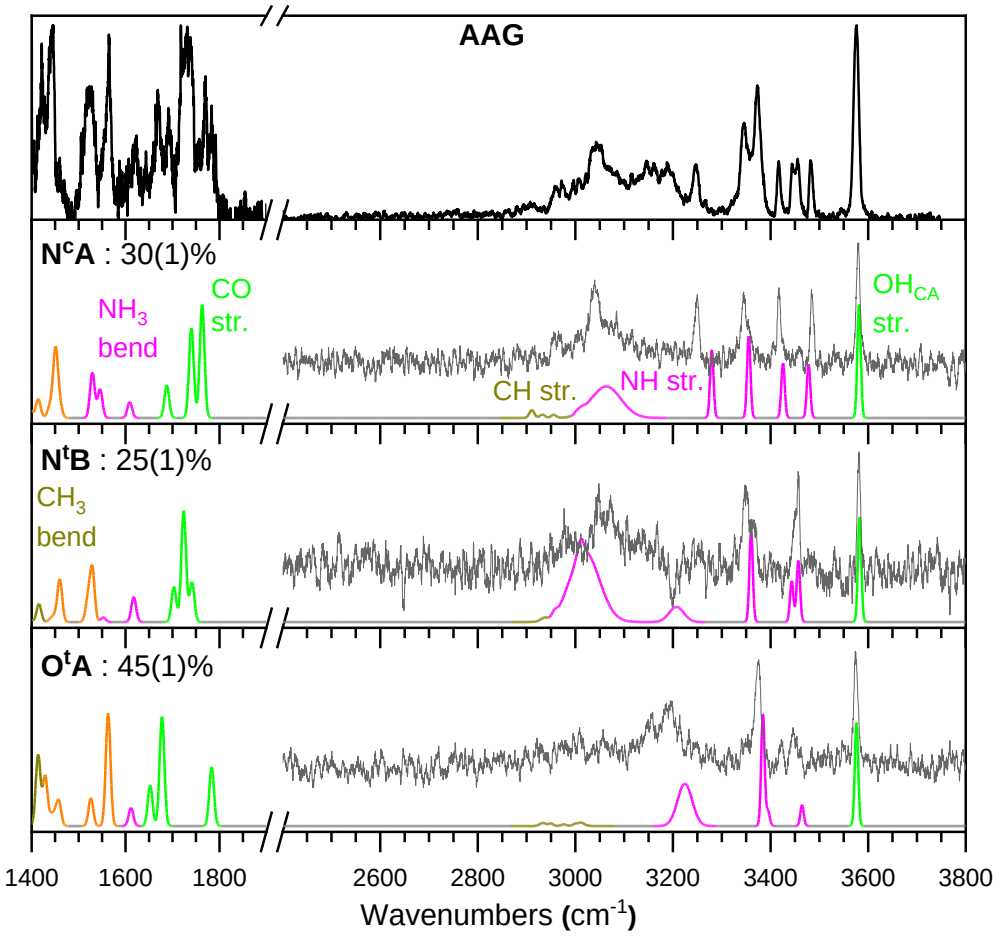


Figure 4.13. IRPD spectra of AAG and calculated harmonic spectra of the six lowest energy structures at the cam-B3LYP/def2-TZVP/GD3BJ level of theory. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. For relative energies of these structures, see **Table 4.2**.

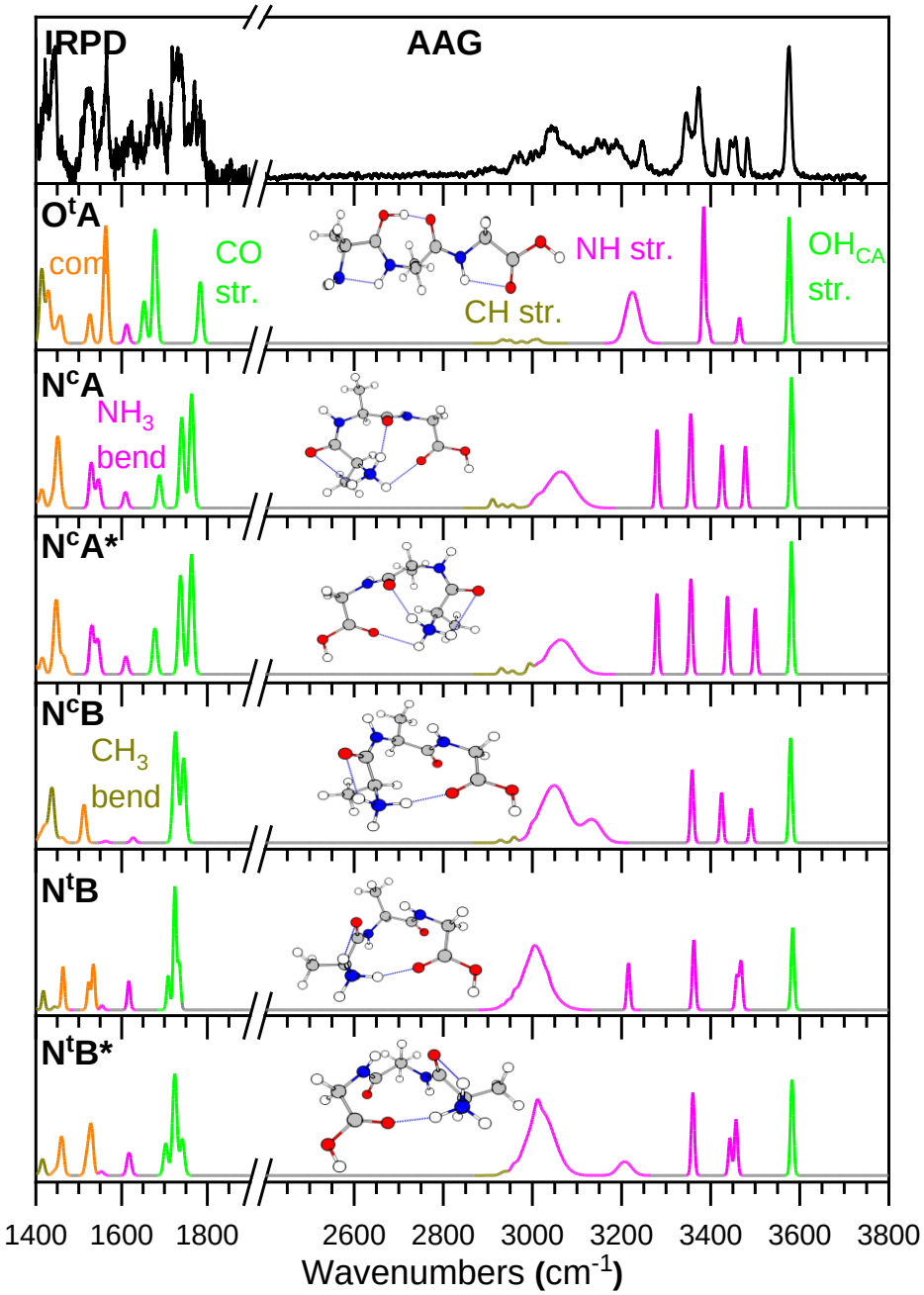


Figure 4.14. IRPD spectra of Ala-Gly-Ala H^+ compared to lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the conformer-specific IR-IR spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.

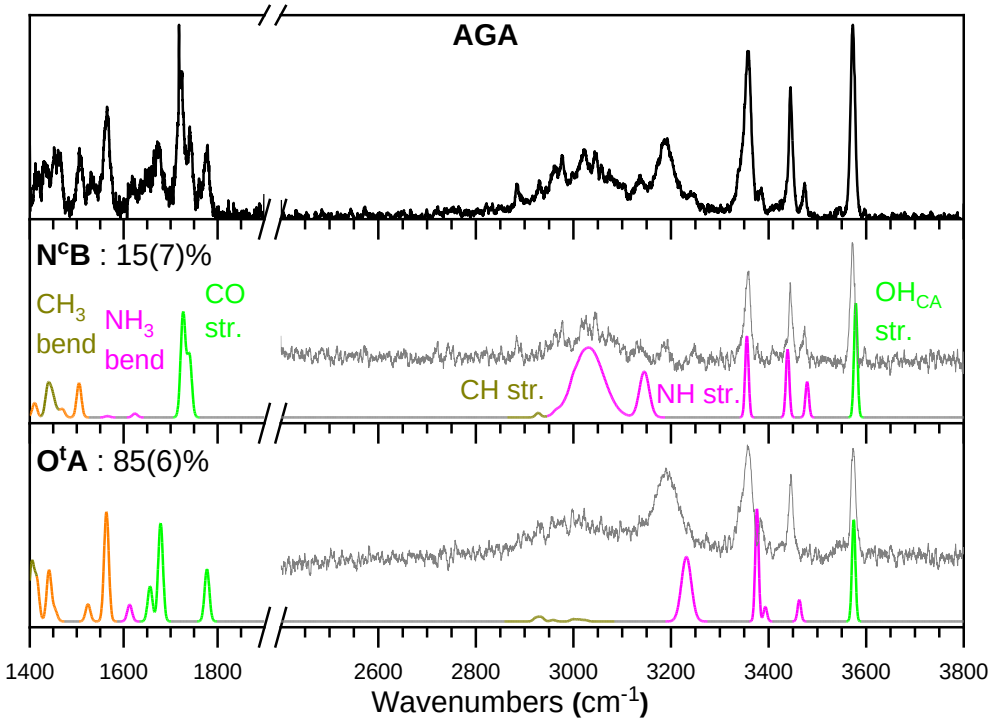
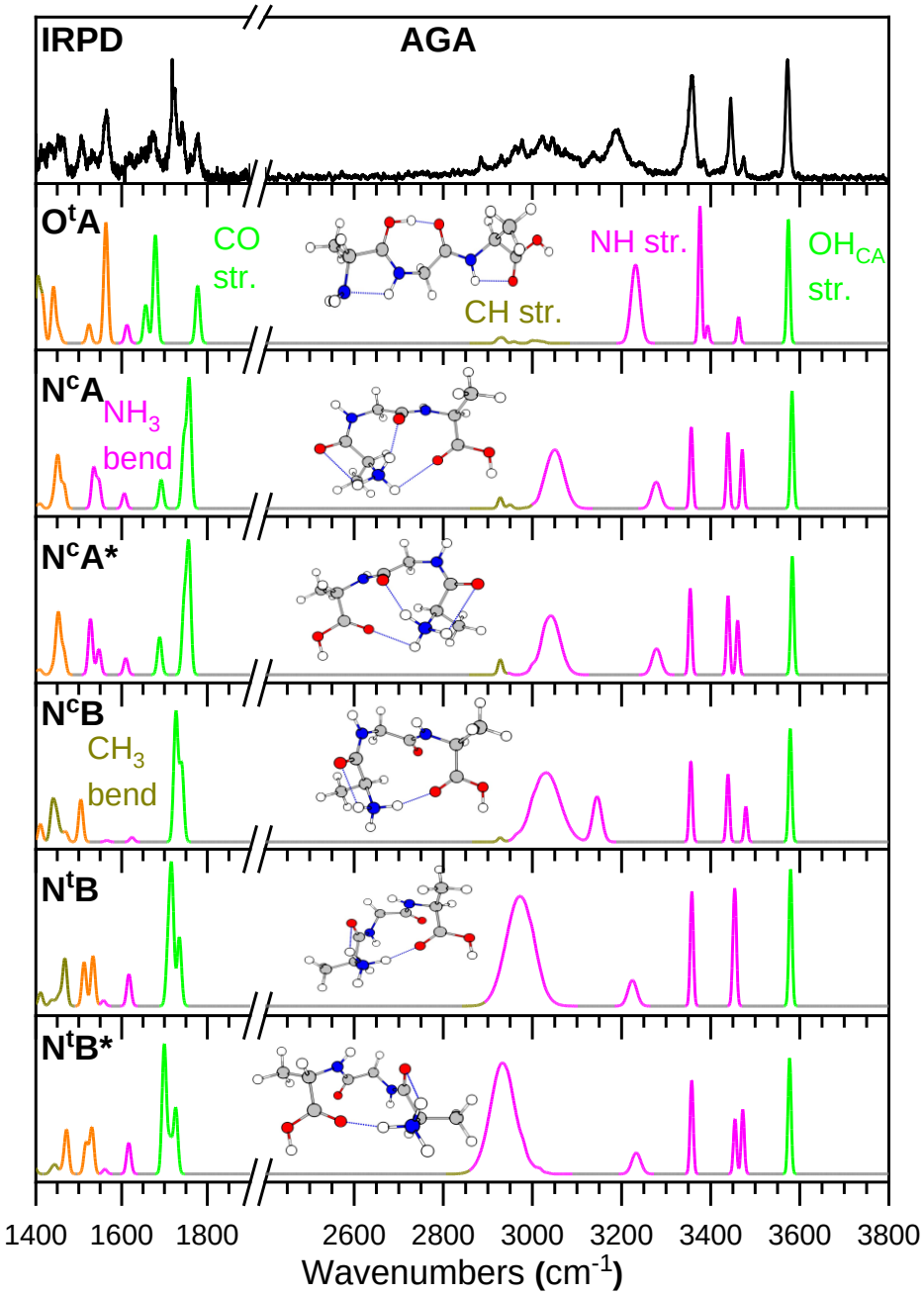


Figure 4.15. IRPD spectra of AGA and calculated harmonic spectra of the six lowest energy structures at the cam-B3LYP/def2-TZVP/GD3BJ level of theory. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. For relative energies of these structures, see **Table 4.2**.



discussed above. The IRPD spectra of AGA and calculated harmonic spectra of the six lowest energy structures is shown in **Figure 4.15**.

Finally, the IRPD spectrum of GAA is shown in **Figure 4.16** and **Figure 4.17**. Looking at the 2400-3800 cm^{-1} region, it appears that only one structure is present since there are only 5 major peaks. These peaks again display the signature NH_2 modes of $\mathbf{O^tA}$. IR-IR ion-dip with the probe laser fixed at 3457 cm^{-1} yields a spectrum in excellent agreement with the calculated spectrum of $\mathbf{O^tA}$. However, looking at the 1400-1900 cm^{-1} region, there is one major peak at 1720 cm^{-1} that cannot be accounted for by $\mathbf{O^tA}$. This feature, as well as some very small peaks in the N-H region, are consistent with $\mathbf{N^cB}$. Therefore, we can assign $\mathbf{N^cB}$ as the minor isomer. Utilizing the peaks in the N-H stretch region we have obtained a ratio of $\mathbf{O^tA/N^cB}$ as 90(2)%/10(15)%. The IRPD spectra of GAA and calculated harmonic spectra of the six lowest energy structures is shown in **Figure 4.17**.

4.6 Analysis and Discussion

The structures found for all the tripeptides considered here fall in three main families: $\mathbf{O^tA}$, $\mathbf{N^cA}$, $\mathbf{N^xB}$, as shown in **Figure 4.1**. In $\mathbf{O^tA}$, protonation occurs at O1 site and is directly stabilized by a strong H-bond to O2. In $\mathbf{N^cA}$ and $\mathbf{N^xB}$, protonation occurs at the amine terminus and is stabilized by either two or three intramolecular H-bonds. An additional structural variation within these isomers involves the first amide groups adopting either the *trans* or *cis* configuration. The calculated relative energies presented in **Table 4.1** show a dependence on the amino acid sequence. However, as discussed in our earlier work on GGG,[33] the calculated relative energies cannot easily be corelated with the observed population ratios. First, there is a large difference between the DFT and MP2 results. The DFT method tends to strongly favor the $\mathbf{O^tA}$ isomers while MP2 favors the

Figure 4.16. IRPD spectra of Gly-Ala-Ala H^+ compared to lowest energy harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ level and the conformer-specific IR-IR spectra for each structure. See **Figure 4.4** caption for color code of calculated spectra. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.

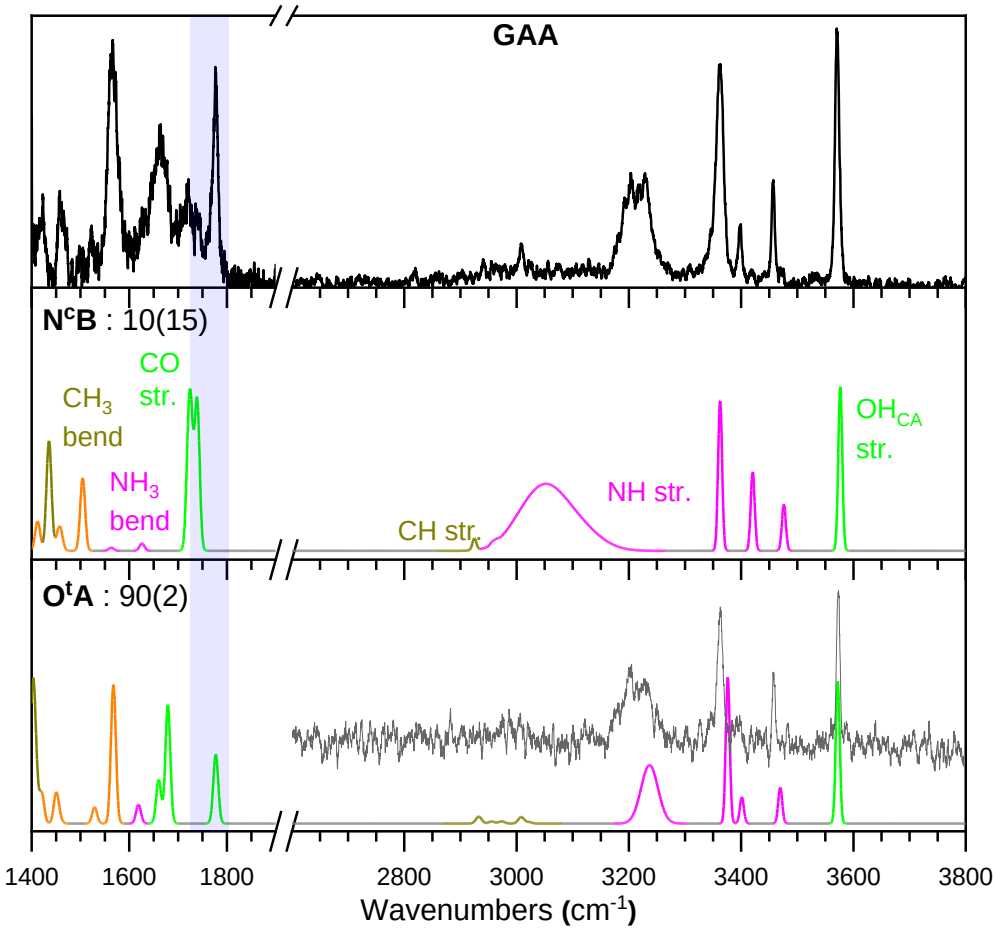
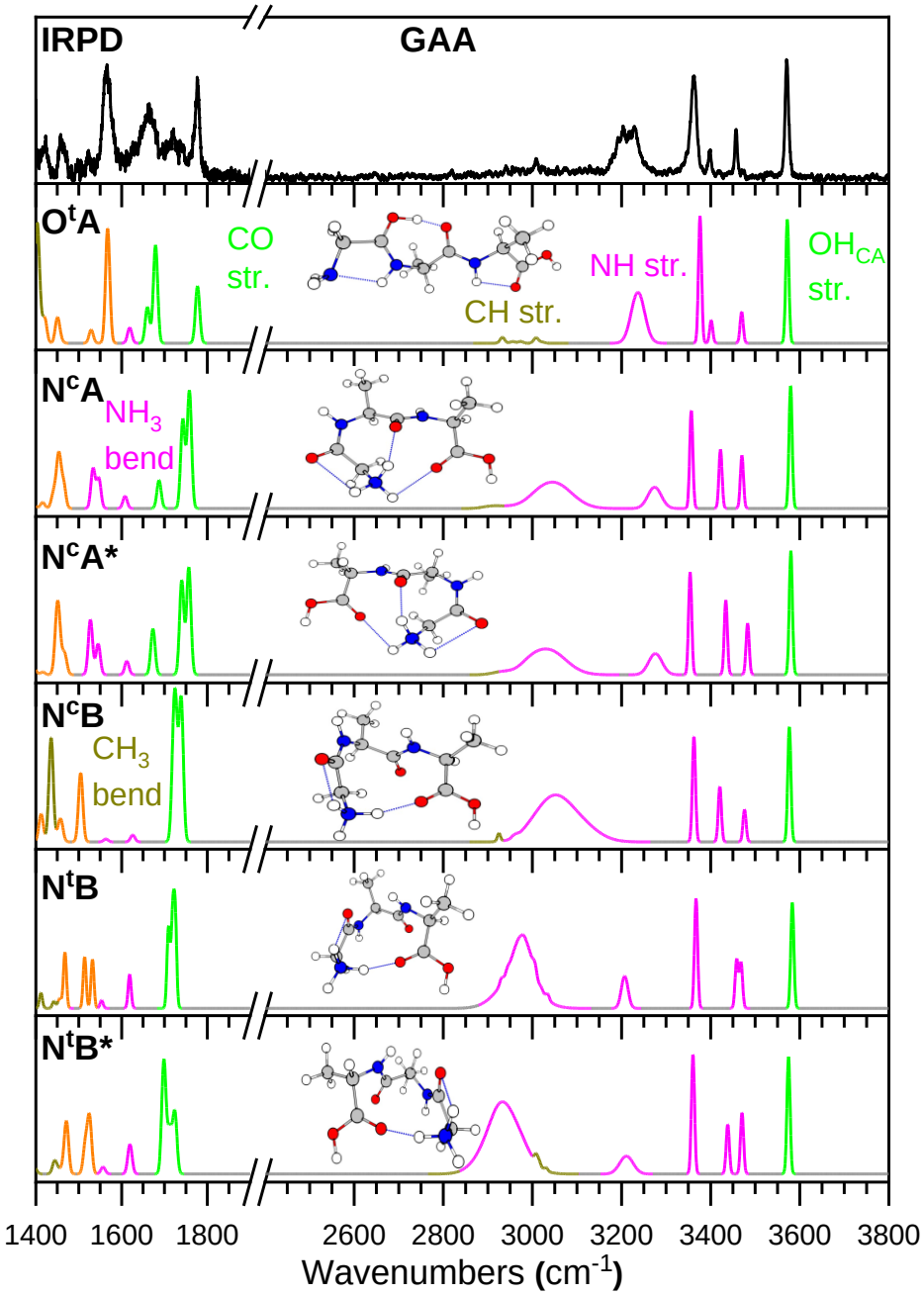


Figure 4.17. IRPD spectra of GAA and calculated harmonic spectra of the six lowest energy structures at the cam-B3LYP/def2-TZVP/GD3BJ level of theory. The calculated carboxyl group vibrations are green, amine group vibrations are magenta, CH_2 group vibrations are tan, and combination bands are denoted in an orange color. For relative energies of these structures, see **Table 4.2**.



N^{c}A and **N^{x}B** isomers. The DPLNO-CCSD favors the **N^{c}A** and **N^{x}B** isomers even more strongly than MP2. The source of this discrepancy is unclear but is probably related to the strong H-bonds formed at the protonation site. Moreover, for each method, the calculated relative energies are generally not consistent with the experimental observation of multiple contributing structures, even if thermal and entropic contributions are taken into account. We note that kinetic trapping of solution-structures is unlikely in our experiment. This is supported by the observation of structures with *cis* amide group or protonation at the carbonyl, both of which are not favored in solution. In our previous work,[32] we have shown that collision with water in the gas-phase can provide sufficient energy to overcome the ~70 kJ/mol barrier for amide bond rotation and ~45 kJ/mol barrier for proton transfer between the N1 and O1 site. Therefore, such collisions in the high-pressure region of our instrument are likely to create conformer population representative of gas-phase energetics, despite disagreements with calculated energies.

Figure 4.1 also shows that the influence of the methyl side-chain on the tripeptide 3-D structure is not a steric effect. The methyl group(s) are all located on the outside perimeter and are not hindering the formation of any of the three structural families. Moreover, all three structural families are found to have similar structural parameters, independent of the presence of the methyl side-chains.[10] Therefore, we turn to the other potential effect of the side-chain, the inductive electronic effects that can modulate local proton affinities. For example, the presence of the methyl side-chain makes the proton affinity of alanine (901.6 kJ/mol)[11, 14-15] slightly larger than that of glycine (886.5 kJ/mol)[11-12, 14]. This change in proton affinities also translates into changes in the strength of H-bond formed in these species. This is exemplified in protonated glycine and alanine, where the frequency shift of the H-bonded N-H, and thus the H-bond strength, are different.[1, 13] For the tripeptides considered here, with multiple sites available for protonation and

H-bonding, local modulation of the proton affinity by the methyl side-chain can energetically favor a specific structure depending on the position of the Ala substitution.

Therefore, to analyze the structure population in terms of local proton affinities, we performed electronic structure calculations starting from a fully extended neutral tripeptide chain shown in **Figure 4.2**. Geometry optimizations and frequency calculations were performed for structures corresponding to protonation at the amine (N1), each of the carbonyl oxygen (O1 and O2), and carboxyl oxygen (O3) sites. For all protonated structures, all dihedral angles were constrained to avoid folding, which allowed computation of local proton affinities representative of the interaction of a proton at a single specific site, without contributions/complications from intramolecular H-bonds. The local proton affinities were calculated using a standard method[53] and summarized in **Table 4.3**. These local proton affinities are smaller than the overall proton affinity of the molecule because they don't include the inevitable formation of H-bonds following protonation. For example, the highest calculated local proton affinity in GGG is 888.5 kJ/mol at the N1 site while the experimental proton affinity of GGG is 966.8 kJ/mol.[14] However, these local proton affinities provide a number that can be more easily compared to show how the methyl side-chain affect the proton affinity of different sites. For example, the local proton affinities of GGG can be subtracted from those of AGG to show the effect of a Gly/Ala substitution at the N-terminus. These differences upon substitution are collected in **Table 4.5**.

In order to discuss the electronic effects of methyl side-chain substitution on the tripeptide structure distribution more effectively, we will focus on two distinct molecular changes: 1) The location of the extra proton, i.e., on the amide carbonyl or terminal amine, and 2) the presence of two or three H-bonds in the case of the amine protonation. In order to elucidate the origin of these effects, it is useful to express the structure populations in two different ways that directly correlate

Table 4.5. Differences in the calculated proton affinities to determine how the change in the methyl placement changes the proton affinity of the molecule per proton placement. Proton affinities are found at the cam-B3LYP/def2-TZVP level with ZPE correction and temperature corrections to the vibrational enthalpy.

	ΔPA (kJ/mol)		
N-terminus	**A**GG-**G**GG	**A**AG-**G**AG	**A**AA-**G**AA
N1	7.18	9.33	13.51
O1	2.11	2.08	6.33
O2	-3.15	-3.81	3.97
O3	-3.79	-4.73	0.96
Central	G**A**G-G**G**G	A**A**G-A**G**G	A**A**A-A**G**A
N1	0.55	2.70	2.70
O1	4.62	4.59	4.52
O2	4.26	3.59	0.80
O3	2.97	2.04	1.72
C-terminus	AG**A**-AG**G**	AA**A**-AA**G**	GA**A**-GA**G**
N1	2.31	2.31	-1.87
O1	4.59	4.52	0.27
O2	10.81	8.03	0.24
O3	9.25	8.94	3.25

Table 4.6. Experimentally determined structural population ratios with their respective errors in parenthesis. These total population ratios are then split into population ratios of the N isomers (N^cA +N^xB) and the O isomer (O^tA). And then finally within the N isomers, the change between N^cA and N^xB conformers.

	N^cA	**N^xB**	**O^tA**	**N%**	**O%**	**A%**	**B%**
GGG	35(3)	0	65(18)	35	65	100	0
AGG	30(1)	40(1)	30(2)	70	30	45	55
GAG	0	30(2)	70(2)	30	70	0	100
AAG	30(1)	25(1)	45(1)	55	45	55	45
AGA	0	15(7)	85(6)	15	85	0	100
GAA	0	10(15)	90(2)	10	90	0	100
AAA	0	35(2)	65(1)	35	65	0	100

with the observed changes: 1) The O% and N% protonation percentages defined as (**O^tA/total)*100** and ((**N^cA+N^xB)/total)*100** , and 2) The A% and B% defined as (**N^cA/(N^cA + N^xB))*100** and (**N^xB/(N^cA + N^xB))*100**. These ratios are summarized in **Table 4.6** for each species.

4.6.1 *Protonation at Amine Vs Carbonyls*

To see the effect of the methyl side-chain on the preferred protonation site, we can look at pairs of tripeptides which only differ by a single Gly/Ala substitution at one specific position. For example, AGG vs. GGG, AAG vs. GAG, and AAA vs. GAA all highlight the effect of the Gly-to-Ala substitution at the N-terminus. Looking at the differences in the local proton affinities for these pairings (**Table 4.5**), it is clear that this substitution substantially increases the N1 proton affinity more than any other site. This is consistent with the experimental observation that this Ala substitution increases the contribution of the N-protonated isomers, by 25-35%, for all three pairs (**Table 4.6**). However, it is important to note that there are changes in the proton affinity at other positions in the tripeptide as well. These more remote changes can also be relevant because they can influence the strength of their respective local H-bond interactions within the structure. This is particularly relevant when trying to understand the changes induced by Ala substitution at the central amino acid.

The proton affinity differences between GAG and GGG, AAG and AGG, and AAA and AGA highlight the effect of the Gly-to-Ala substitution at the central position. For GAG vs. GGG and AAG vs. AGG, the largest differences in local proton affinities are located at the O1 and O2 sites. Although it should be noted that the difference in proton affinity here is only about half of that in the N-terminus substitution. Overall this is consistent with these Ala substitutions favoring,

by 5-15%, the O1 protonation site, in which the proton is also forming a strong H-bond with O2. The situation is different for AGA vs. AAA, where the central Ala substitution led to a 20% increase for N protonated isomers. Unlike the two previous pairs, the proton affinity changes related to this Ala substitution is predominantly on the N1 and O1 sites. The mechanism for this different behavior is unclear. However, the proton affinity change is consistent with the observed increased preference for N-protonation since these structures have a strong H-bond between protonated amine and O1.

For Ala substitution at the C-terminus, there are again three different pairs of peptides that can be compared: AGA vs. AGG, AAA vs. AAG, and GAG vs. GAA. For AGA vs. AGG and AAA vs. AAG, the Ala substitution led to the largest positive proton affinity difference at the O2 and O3 positions. Both N- and O-protonated isomers have H-bonds that involve these sites. However, the substitution also gave rise to large positive proton affinity difference at the O1 site and a smaller one at the N1. Therefore, the combined strong increase of the O1 and O2 proton affinity is consistent with the observed 20-55% increase in preference for the O protonation site upon Ala substitution. For the last GAG vs. GAA pair, the changes in proton affinities are rather small but still favor slightly the O1 and O2 position over the N1, again consistent with an increase in fraction of O-protonated structure.

The overall effect we observe here is that the differences in local proton affinities arising from the methyl group's electron donating characteristic can provide valuable insights regarding the preferential protonation site, a key characteristic that can significantly influence a peptide's 3D structure.

4.6.2 *Changes in the N^cA and N^xB conformer populations*

We also examined the changes in population between different anime protonated conformers: **N^cA** vs. **N^xB**. As shown in **Figure 4.1**, the main difference between these two groups of conformers is that in **N^cA**, O1, O2 and O3 are all involved in H-bonding with the protonated amine while in **N^xB**, only O1 and O3 are H-bonded to the protonated amine. However, these conformers also differ in the relative strength of the various H-bonds, which can be deduced from their calculated lengths. For the **N^cA** conformer of GGG, the N1H-O2 H-bond is the shortest at 1.8 Å while the N1H-O1 and N1H-O3 H-bonds have similar calculated lengths of 2.4 Å. In the **N^xB** conformers, the length of the N1H-O1 and N1H-O3 H-bonds are reduced to 2.1 Å and 1.8 Å, respectively. Using the empirical relationship between H-bond length and bond energy,[54] we can estimate that the total H-bonding energy in all of N-protonated conformers are similar, despite the different number of individual interactions.

The population ratios of these two groups of conformers were determined for seven of the tripeptides as listed in **Table 4.6**. The GGG, AGG, and AAG peptides are the only species that contain the **N^cA** conformer. All the N1-protonated GGG are in the **N^cA** conformation, while about half of them are **N^cA** for AGG and AAG. For all the other peptides considered here, only the **N^xB** conformations are observed. Therefore, starting from GGG, any Ala substitution generally favors the formation of **N^xB** over **N^cA**, i.e. favoring formation of stronger but fewer intramolecular H-bonds.

Analyzing the **N^cA/N^xB** population ratio in terms of local proton affinity changes is more difficult because of the multiple competing factors involved in the relative H-bond strengths**.** However, using the difference in H-bond highlighted above, we can infer that the increase in local proton affinities at the O1 and O3 sites, relative to that of the O2 site, should favor the **N^xB**

conformations by increasing the strength of the N1H-O1 and N1H-O3 H-bonds relative to the N1H-O2 H-bond. Looking at the effect of Ala substitutions in **Table 4.5**, we see that the total difference at the O1 and O3 sites is always more positive than that of the O2 site for an Ala substitution at any position within the peptide. The calculated proton affinity changes are therefore generally consistent with the experimental observations.

4.7 Conclusions

In this paper, we report the experimental IR-IR double resonance spectra and computational efforts that allowed us to determine how Gly-to-Ala substitution influences the gas-phase structural population of model protonated tripeptides. The structures present fall into three structural families that differ mainly by protonation site and number of internal H-bonds. The 3-D geometry of each isomer and conformer is such that steric effect of the methyl side chain cannot explain the observed variations in population upon Ala substitution. Through this systematic study, we found that the presence of the methyl side-chain modulates the local proton affinities of the functional groups that control the preferred protonation site and the relative strength of intramolecular H-bonds. We demonstrated that the calculated local proton affinities can be used to qualitatively explain the changes in structural populations of the tripeptides. Overall, our results show that, in addition to the well-known steric effect of a methyl group, its electronic inductive effect can also significantly alter the 3-D geometry of a peptide. For example, in going from AGG to AGA, half of the population changed from the N-protonated isomers to the O-protonated isomer.

4.8 References

(1) Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J Phys Chem A* **2019,** *123* (15), 3355-3366.

(2) Guo, M.; Rosbottom, I.; Zhou, L.; Yong, C. W.; Zhou, L.; Yin, Q.; Todorov, I. T.; Errington, E.; Heng, J. Y. Y. Triglycine (GGG) Adopts a Polyproline II (pPII) Conformation in Its Hydrated Crystal Form: Revealing the Role of Water in Peptide Crystallization. *J Phys Chem Lett* **2021,** *12* (34), 8416-8422.

(3) Tashiro, S.; Kobayashi, M.; Fujita, M. Folding of an Ala-Ala-Ala tripeptide into a beta-turn via hydrophobic encapsulation. *J Am Chem Soc* **2006,** *128* (29), 9280-1.

(4) Zhang, K.; Cassady, C. J.; Chung-Phillips, A. Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity. *Journal of the American Chemical Society* **1994,** *116* (25), 11512-11521.

(5) Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. Experimental and ab initio studies of the gas-phase basicities of polyglycines. *Journal of the American Chemical Society* **1993,** *115* (23), 10812-10822.

(6) Carr, S. R.; Cassady, C. J. Gas-phase basicities of histidine and lysine and their selected di- and tripeptides. *Journal of the American Society for Mass Spectrometry* **1996,** *7* (12), 1203-1210.

(7) Wu, Z.; Fenselau, C. Structural Determinants of Gas Phase Basicities of Peptides. *Tetrahedron* **1993,** *49* (41), 9197-9206.

(8) Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. Gas-phase deprotonation of the peptide backbone for tripeptides and their methyl esters with hydrogen and methyl side chains. *J Phys Chem B* **2012,** *116* (51), 14844-58.

(9) Cassady, C. J.; Carr, S. R.; Zhang, K.; Chung-Phillips, A. Experimental and Ab Initio Studies on Protonations of Alanine and Small Peptides of Alanine and Glycine. *The Journal of Organic Chemistry* **2002,** *60* (6), 1704-1712.

(10) Wu, J.; Lebrilla, C. B. Intrinsic basicity of oligomeric peptides that contain glycine, alanine, and valine—The effects of the alkyl side chain on proton transfer reactions. *Journal of the American Society for Mass Spectrometry* **1995,** *6* (2), 91-101.

(11) Heine, N.; Asmis, K. R. Cryogenic ion trap vibrational spectroscopy of hydrogen-bonded clusters relevant to atmospheric chemistry. *International Reviews in Physical Chemistry* **2014,** *34* (1), 1-34.

(12) Oomens, J.; Sartakov, B. G.; Meijer, G.; von Helden, G. Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *International Journal of Mass Spectrometry* **2006,** *254* (1-2), 1-19.

(13) Polfer, N. C. Infrared multiple photon dissociation spectroscopy of trapped ions. *Chem Soc Rev* **2011,** *40* (5), 2211-21.

(14) Polfer, N. C.; Oomens, J. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom Rev* **2009,** *28* (3), 468-94.

(15) Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; von Helden, G.; Oomens, J. Infrared fingerprint spectroscopy and theoretical studies of potassium ion tagged amino acids and peptides in the gas phase. *J Am Chem Soc* **2005,** *127* (23), 8571-9.

(16) Shi, Y.; Du, M.; Ren, J.; Zhang, K.; Xu, Y.; Kong, X. Application of Infrared Multiple Photon Dissociation (IRMPD) Spectroscopy in Chiral Analysis. *Molecules* **2020,** *25* (21).

(17) Maitre, P.; Scuderi, D.; Corinti, D.; Chiavarino, B.; Crestoni, M. E.; Fornarini, S. Applications of Infrared Multiple Photon Dissociation (IRMPD) to the Detection of Posttranslational Modifications. *Chem Rev* **2020,** *120* (7), 3261-3295.

(18) de Vries, M. S.; Hobza, P. Gas-phase spectroscopy of biomolecular building blocks. *Annu Rev Phys Chem* **2007,** *58*, 585-612.

(19) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. Guanine tautomerism revealed by UV–UV and IR–UV hole burning spectroscopy. *The Journal of Chemical Physics* **2001,** *115* (10), 4604-4611.

(20) Nagornova, N. S.; Rizzo, T. R.; Boyarkin, O. V. Interplay of intra- and intermolecular H-bonding in a progressively solvated macrocyclic peptide. *Science* **2012,** *336* (6079), 320-3.

(21) Rizzo, T. R.; Stearns, J. A.; Boyarkin, O. V. Spectroscopic studies of cold, gas-phase biomolecular ions. *International Reviews in Physical Chemistry* **2009,** *28* (3), 481-515.

(22) Dean, J. C.; Buchanan, E. G.; Zwier, T. S. Mixed 14/16 helices in the gas phase: conformation-specific spectroscopy of Z-(Gly)n, n = 1, 3, 5. *J Am Chem Soc* **2012,** *134* (41), 17186-201.

(23) Gloaguen, E.; Tardivel, B.; Mons, M. Gas phase double-resonance IR/UV spectroscopy of an alanine dipeptide analogue using a non-covalently bound UV-tag: observation of a folded peptide conformation in the Ac-Ala-NH2–toluene complex. *Structural Chemistry* **2015,** *27* (1), 225-230.

(24) Heiles, S.; Cooper, R. J.; DiTucci, M. J.; Williams, E. R. Hydration of guanidinium depends on its local environment. *Chem Sci* **2015,** *6* (6), 3420-3429.

(25) Voss, J. M.; Marsh, B. M.; Zhou, J.; Garand, E. Interaction between ionic liquid cation and water: infrared predissociation study of [bmim](+).(H2O)n clusters. *Phys Chem Chem Phys* **2016,** *18* (28), 18905-13.

(26) Wolk, A. B.; Leavitt, C. M.; Garand, E.; Johnson, M. A. Cryogenic ion chemistry and spectroscopy. *Acc Chem Res* **2014,** *47* (1), 202-10.

(27) Klyne, J.; Schmies, M.; Fujii, M.; Dopfer, O. Stepwise microhydration of aromatic amide cations: formation of water solvation network revealed by infrared spectra of formanilide(+)-(H2O)(n) clusters (n </= 5). *J Phys Chem B* **2015,** *119* (4), 1388-406.

(28) Klyne, J.; Schmies, M.; Miyazaki, M.; Fujii, M.; Dopfer, O. Stepwise microhydration of aromatic amide cations: water solvation networks revealed by the infrared spectra of acetanilide(+)-(H2O)n clusters (n </= 3). *Phys Chem Chem Phys* **2018,** *20* (5), 3148-3164.

(29) Suzuki, Y.; Hirata, K.; Lisy, J. M.; Ishiuchi, S. I.; Fujii, M. Double Ion Trap Laser Spectroscopy of Alkali Metal Ion Complexes with a Partial Peptide of the Selectivity Filter in K(+) Channels horizontal line Temperature Effect and Barrier for Conformational Conversions. *J Phys Chem A* **2021,** *125* (44), 9609-9618.

(30) Leavitt, C. M.; Wolk, A. B.; Fournier, J. A.; Kamrath, M. Z.; Garand, E.; Van Stipdonk, M. J.; Johnson, M. A. Isomer-Specific IR-IR Double Resonance Spectroscopy of D2-Tagged Protonated Dipeptides Prepared in a Cryogenic Ion Trap. *J Phys Chem Lett* **2012,** *3* (9), 1099-105.

(31) Voss, J. M.; Kregel, S. J.; Fischer, K. C.; Garand, E. IR-IR Conformation Specific Spectroscopy of Na+(Glucose) Adducts. *Journal of The American Society for Mass Spectrometry* **2018,** *29* (1), 42-50.

(32) Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H(+).(H2O). *J Phys Chem A* **2018,** *122* (41), 8213-8221.

(33) Voss, J. M.; Fischer, K. C.; Garand, E. Revealing the structure of isolated peptides: IR-IR predissociation spectroscopy of protonated triglycine isomers. *Journal of Molecular Spectroscopy* **2018,** *347*, 28-34.

(34) Fischer, K. C.; Sherman, S. L.; Garand, E. Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and beta-Alanine. *J Phys Chem A* **2020,** *124* (8), 1593-1602.

(35) Khuu, T.; Yang, N.; Johnson, M. A. Vibrational spectroscopy of the cryogenically cooled O- and N-protomers of 4-Aminobenzoic acid: Tag effects, isotopic labels, and identification of the E,Z isomer of the O-protomer. *Int J Mass Spectrom* **2020,** *457*.

(36) Chin, W.; Piuzzi, F.; Dognon, J. P.; Dimicoli, I.; Mons, M. Gas-phase models of gamma turns: effect of side-chain/backbone interactions investigated by IR/UV spectroscopy and quantum chemistry. *J Chem Phys* **2005,** *123* (8), 084301.

(37) Chiba, T.; Okuyama, K.; Fujii, A. Observation of Evidence for the pi*-sigma* Hyperconjugation in the S1 State of o-, m-, and p-Fluorotoluenes by Double-Resonance Infrared Spectroscopy. *J Phys Chem A* **2016,** *120* (28), 5573-80.

(38) Marsh, B. M.; Voss, J. M.; Garand, E. A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *J Chem Phys* **2015,** *143* (20), 204201.

(39) Mookherjee, A.; Van Stipdonk, M. J.; Armentrout, P. B. Thermodynamics and Reaction Mechanisms of Decomposition of the Simplest Protonated Tripeptide, Triglycine: A Guided Ion Beam and Computational Study. *J Am Soc Mass Spectrom* **2017,** *28* (4), 739-757.

(40) Wu, R.; McMahon, T. B. Infrared multiple photon dissociation spectroscopy as structural confirmation for GlyGlyGlyH+ and AlaAlaAlaH+ in the gas phase. Evidence for amide oxygen as the protonation site. *J Am Chem Soc* **2007,** *129* (37), 11312-3.

(41) Wu, R.; McMahon, T. B. Protonation sites and conformations of peptides of glycine (Gly(1-5)H(+)) by IRMPD spectroscopy. *J Phys Chem B* **2009,** *113* (25), 8767-75.

(42) Campbell, J. L.; Yang, A. M.; Melo, L. R.; Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J Am Soc Mass Spectrom* **2016,** *27* (7), 1277-84.

(43) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: lessons learned from tetraalkylammonium cations and their isomers. *J Am Soc Mass Spectrom* **2014,** *25* (9), 1583-91.

(44) Lecours, M. J.; Chow, W. C.; Hopkins, W. S. Density functional theory study of Rh(n)S(0,+/-) and Rh(n+1)(0,+/-) (n = 1-9). *J Phys Chem A* **2014,** *118* (24), 4278-87.

(45) Liu, C.; Le Blanc, J. C.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using differential mobility spectrometry to measure ion solvation: an examination of the roles of solvents and ionic structures in separating quinoline-based drugs. *Analyst* **2015,** *140* (20), 6897-903.

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

(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith,

T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

(48) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2*, 73-78.

(49) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017,** *8*, e1327.

(50) Tajkhorshid, E.; Suhai, S. Influence of the Methyl Groups on the Structure, Charge Distribution, and Proton Affinity of the Retinal Schiff Base. *The Journal of Physical Chemistry B* **1999,** *103* (26), 5581-5590.

(51) Harrilal, C. P.; DeBlase, A. F.; Fischer, J. L.; Lawler, J. T.; McLuckey, S. A.; Zwier, T. S. Infrared Population Transfer Spectroscopy of Cryo-Cooled Ions: Quantitative Tests of the Effects of Collisional Cooling on the Room Temperature Conformer Populations. *J Phys Chem A* **2018,** *122* (8), 2096-2107.

(52) Seaiby, C.; Zabuga, A. V.; Svendsen, A.; Rizzo, T. R. IR-induced conformational isomerization of a helical peptide in a cold ion trap. *J Chem Phys* **2016,** *144* (1), 014304.

(53) Maksic, Z. B.; Kovacevic, B.; Vianello, R. Advances in determining the absolute proton affinities of neutral organic molecules in the gas phase and their interpretation: a theoretical account. *Chem Rev* **2012,** *112* (10), 5240-70.

(54) Rozenberg, M.; Loewenschuss, A.; Marcus, Y. An empirical correlation between stretching vibration redshift and hydrogen bond length. *Physical Chemistry Chemical Physics* **2000,** *2* (12), 2699-2702.

CHAPTER 5

Effect of Methyl Side-Chains on the Solvation Induced Changes in Flexible Protonated Tripeptides

In Preparation: Sherman, S. L.; Fischer, K. C.; Garand, E. *"*Effect of Methyl Side-Chains on the Solvation Induced Changes in Flexible Protonated Tripeptides*"* **2022**.

Abstract

We present a study comparing and analyzing the infrared predissociation (IRPD) spectra of the $Gly_3H^+(H_2O)_{1\text{-}2}$ and $Ala_3H^+(H_2O)_{1\text{-}2}$ clusters with the goal of determining the effect of the methyl side-chains on the microsolvated structures of these flexible peptides. It was shown in our past work that the amount and position of methyl moieties in a tripeptide, and the addition of one water molecule in Gly_3H^+ can induce a drastic change to the conformer population. Therefore, it raises the question of whether the electron donating effect of the methyl side-chain is sufficient to compete with that of the water binding energy. Upon microsolvation, we showed that the two tripeptides had similarities in the major conformer solvation structures which are all trans species. However, we found that the two tripeptides differed in their minor cis conformer. As solvation size increases, due to the population ratios of all conformers, it can be shown that the minor cis conformers in both tripeptides move from being possibly kinetically trapped to energetically competitive structures.

Author Contributions

SLS and KCF acquired experimental spectra. KCF performed theoretical computations and preliminary data analyses for $Gly_3H^+(H_2O)_2$. SLS performed theoretical computations and data analyses for $Ala_3H^+(H_2O)_{1\text{-}2}$. SLS performed final data analyses for $Gly_3H^+(H_2O)_2$. SLS and EG wrote the manuscript.

5.1 Introduction

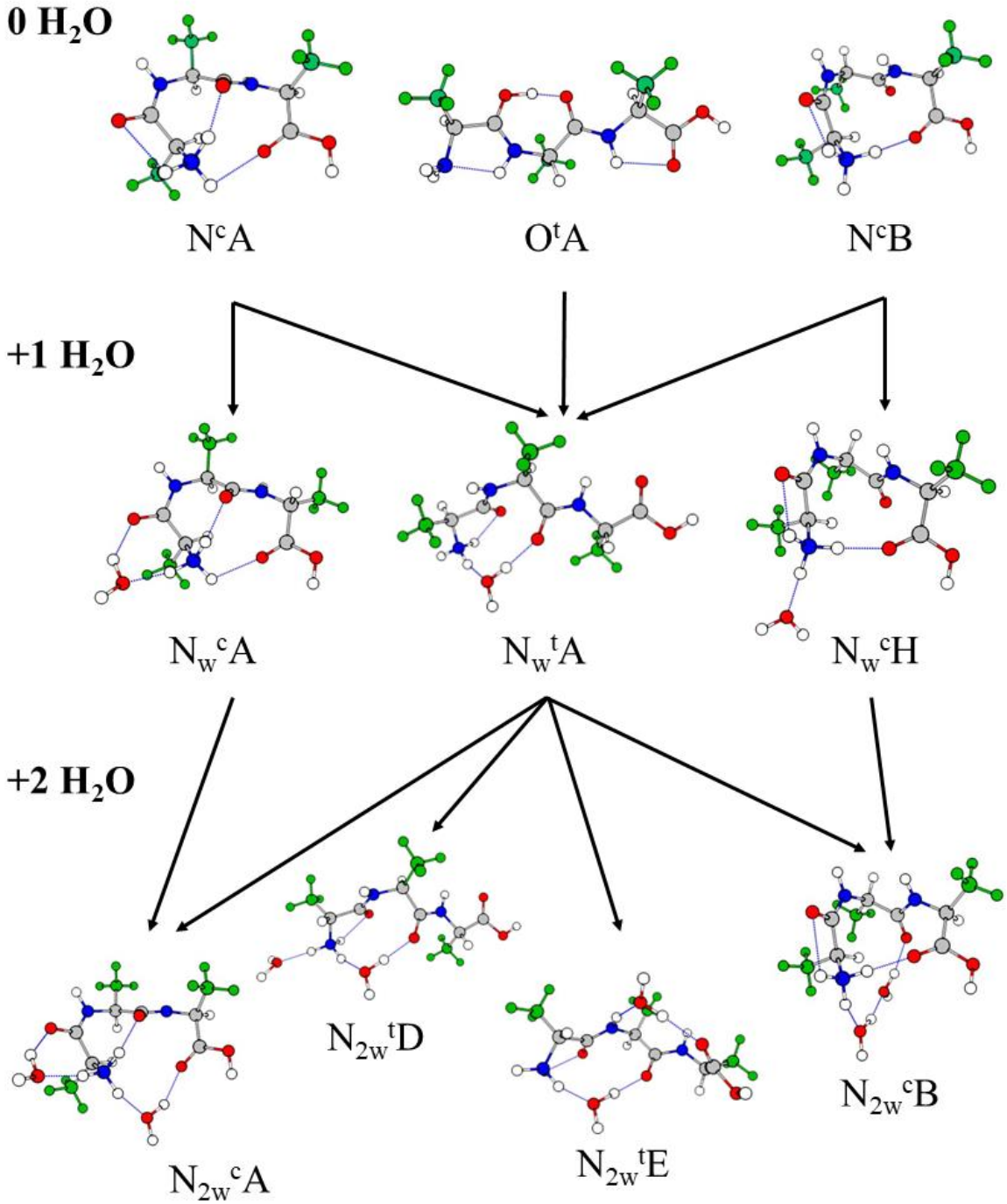
Peptides and large biomolecules can adopt a wide range of conformations due to their high fluxional nature. The factors that modulate the conformational structure of peptides includes sterics[1-2] and intramolecular hydrogen-bond interactions[3-6] which depends on the exact sequence of amino acids.[7-8] However, the presence of solvent, particularly water can dramatically change the conformational landscape through the formation of additional hydrogen-bond interactions.[9-14] These new intermolecular interactions compete with the intramolecular hydrogen bonding. For larger peptides, the sums of all these interactions combined to define the secondary and tertiary structures of proteins and therefore dictate their function. Hence, it is important to fundamentally understand how different amino acids and their placement within the peptide chain can affect conformations and the response to solvent interactions. However, probing these structures and interactions with traditional techniques is difficult due to the multiple conformers which complicate the spectra and therefore the relevant information that can be extracted from them. Here we use conformer specific infrared spectroscopy[7, 10, 15-20] of microsolvated clusters of protonated triglycine and trialanine to study the influence of the methyl side-chain on the solvent and peptide chain interactions.

Previously we have probed the microsolvation structure of the protonated Glycine and Alanine amino acids.[21] We have shown that the presence of the methyl side-chain in Alanine modulate the strength of the intramolecular H-bond between the protonated amine and the carboxylic acid carbonyl. This leads to different structures for the small $GlyH^+(H_2O)_n$ and $Ala^+(H_2O)_n$ clusters because more water solvent molecules are required to break the intramolecular interaction in $AlaH^+$. Similar effects were found[22] in β-$AlaH^+$ in which feature a stronger

intramolecular H-bond than $AlaH^+$ and $GlyH^+$. These structural differences vanish in the larger clusters with sufficient number of water molecule to outcompete the intramolecular H-bond.[22] We have also probed the structures of the bare Gly_3H^+ and Ala_3H^+ flexible peptides and found that once again the presence of the methyl side-chains modulate the strength of intramolecular interactions and therefore affect the relative stability of different conformations. Three main conformer families were found for these species and are shown in **Figure 5.1**: Amide carbonyl protonated (**O^tA**), 2) Amine protonated with three internal H-bonds (**N^cA**), and 3) Amine protonated with only two internal H-bonds (**N^tB** and **N^cB**). The relative contribution of these structure highly depends on the exact sequence of Gly and Ala in the tripeptide.[7] They also contain structural motifs such as carbonyl protonation and cis amide bonds which are atypical of solution-phase peptides. However, for Gly_3H^+, these unusual motifs disappear upon the addition of a single H_2O molecule, and the most stable $Gly_3H^+(H_2O)$ complex involves amine protonation and all trans amide bonds (**Figure 5.1**).[10]

Here we present, analyze, and compare the infrared predissociation (IRPD) spectra of the $Gly_3H^+(H_2O)_{1-2}$ and $Ala_3H^+(H_2O)_{1-2}$ clusters with the goal of determining the effect of the methyl side-chains on the microsolvated structures of these flexible peptides. Considering the differences in conformer population of each bare ion and the drastic changes induced with the addition of one water molecule, it raises the question of whether the electron donating effect of the methyl side-chain is sufficient to compete with that of the water binding energy. The large conformational space afforded by the peptide chain flexibility and variety of possible water positions also poses a challenge to the analysis of the spectra. Thus, we have utilized H_2O/D_2O isotopic substitution as well as conformer-specific IR-IR double resonance schemes to assist in structural analysis. We found that for both tripeptides, the dominant carbonyl protonated isomer

Figure 5.1. Structures of Gly_3H^+ and $Gly_3H^+(H_2O)$ identified in previous studies[10, 17] along with their respective conformer population ratios as well as pathway of conformational change upon addition of one water molecule. Structures of Ala_3H^+ identified in a previous study.[7] Structures of $Gly_3H^+(H_2O)_2$ and $Ala_3H^+(H_2O)_{1-2}$ determined in their respective IRPD spectra.



in the bare protonated ion completely disappears upon the addition of water and that the same conformers with amine protonation and all trans amide bond conformers are formed. However, difference in the conformers involving cis amide bond are found to persist with the addition of one or two water molecules.

5.2 Experimental Details

All the IRPD spectra presented were obtained with our home-built dual cryogenic trap vibrational spectrometer described in detail previously.[23] The bare protonated ions were generated via electrospray ionization (ESI) of an ~1 mM solution of either Gly_3 or Ala_3 dissolved in methanol with trace amounts of formic acid. The ions were guided through hexapole ion guides and through differentially pumped regions and into a linear octupole ion trap mounted to a liquid nitrogen cryostat and held at 80K. Nitrogen buffer gas seeded with either H_2O or D_2O was introduced into the trap in an ~1 ms burst to induce collisional thermalization and the formation of microsolvated clusters. We have shown previously that under these clustering conditions, H/D exchange between the ion and the D_2O molecules is almost completely quenched. This leads to the formation of mainly [tripeptideH^+*(all-H)*]·$(D_2O)_n$ clusters with minimal presence of isotopomers containing a partially deuterated peptide and HDO.[24] The microsolvated ion clusters were then gently transferred through additional hexapole ion guides a 3D quadrupole ion trap held at 10 K by a closed-cycle helium cryostat. Here, the helium buffer gas seeded with 10% D_2 allowed for further thermalization to 10 K and the formation of weakly-bound D_2 tagged-adducts.[16, 23, 25] The tagged adducts were extracted into a time-of-flight (TOF) mass spectrometer. The [tripeptide$^+$]·$(H_2O/D_2O)_{1\text{-}2}$·D_2 adducts were mass-selected by a deflector and intersected with the output of a 10 Hz Nd:YAG pumped tunable optical parametric oscillator (OPO)/optical parametric

amplified (OPA) infrared "probe" laser. Resonant absorption of a single photon was enough to induce the loss of the weakly-bound D_2 tag. The resulting photofragment and parent ions were then temporally separated by a two-stage reflectron. The one-laser IRPD spectra was produced by monitoring the yield of the photofragment ions as a function of laser wavelength.

Isomer-specific IR-IR spectra[15-16] were obtained by focusing the output of an additional Nd:YAG pumped tunable OPO/OPA infrared "pump" laser directly into the 10 K tagging main trap ~94.5 ms after the introduction of the buffer gas and ~0.5 ms before ion extraction.[18] To acquire an "ion-burn" spectrum, the main trap pump laser was fixed at a particular resonant frequency of one or more conformers such that all conformers containing that frequency dissociate from the tag in the main trap and therefore get mass-gated out. The probe laser was then scanned to obtain the IR-IR spectra of all other conformers not containing the resonant frequency of the main trap pump laser. To acquire an "ion-dip" spectrum, the TOF probe laser was fixed at a frequency unique to one conformer such that monitoring decreases in the photofragment yield as a function of main trap pump laser wavelength. This results in an IR-IR spectrum that contains only contributions from the probed vibration of the conformer of interest.

5.3 Computational Details

The lowest energy conformers of Gly_3H^+,[17] Ala_3H^+,[7] and $Gly_3H^+(H_2O)$[10] ion have been identified previously. Here we use a similar approach to find the lowest energy structures of $Ala_3H^+(H_2O)_{1-2}$ and $Gly_3H^+(H_2O)_2$. A systematic computation isomer structure search was done using the molecular mechanics basin hopping program developed by Hopkins and coworkers.[26-29] The lowest energy structures found for Ala_3H^+ were the starting base structures for $Ala_3H^+(H_2O)_1$ and the water was placed on the protonated amine and hydrogen bonded to one of the C=O of an

amide. The lowest energy structures for $Ala_3H^+(H_2O)_1$ and $Gly_3H^+(H_2O)_1$ were used as the starting base for the $Ala_3H^+(H_2O)_2$ and $Gly_3H^+(H_2O)_2$ clusters. The dihedral angles were iteratively stepped 7,000 times by a random value between ±5° and the water was iteratively rotated and translated by a random value between ±5° and 0.2 Å, respectively. After each iteration, the resulting structure was optimized with the AMBER force field. Each unique structure was subsequently optimized using the Gaussian 16[30] program. First, each structure was optimized at the BLYP/3-21g level, and all duplicate structures and structures above 140 kJ/mol of the lowest energy structure were rejected. All remaining structures were further optimized at the cam-B3LYP/6-311g level and duplicate structures and structures with relative energies less than 50 kJ/mol to the lowest energy structure were again rejected. All remaining structures were further optimized at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level, including correction for basis set superposition error.[31-33] Harmonic frequency calculations were then performed on structures with relative energies less than 25 kJ/mol to the lowest energy structure found to determine zero-point energy (ZPE) corrections. This level of theory has previously been shown to yield good agreements with experimental IR spectra of similar species.[7, 10, 17, 21-24] These low energy structures were also optimized at the MP2/def2-TZVP level and the ZPE correction of the corresponding cam-B3LYP calculation were applied to the MP2 energetics. The lowest energy structures for $Gly_3H^+(H_2O)_2$ and $Ala_3H^+(H_2O)_{1\text{-}2}$ are shown in **Figure(s) 5.2**, **5.3**, and **5.4**, respectively.

The calculated harmonic peptide modes were scaled using the experimental position of the free carboxylic acid OH stretch. The calculated harmonic water modes were scaled with a linear scaling factor 0.7829*(water frequency)+669.76 which was determined though experimental spectra and calculated structures from previous experiments.[10, 21-22, 34] The calculated spectra were Gaussian broadened with areas corresponding to the calculated intensities to facilitate comparisons

Figure 5.2. Structures of the lowest energy conformers of $Gly_3H^+(H_2O)_2$.

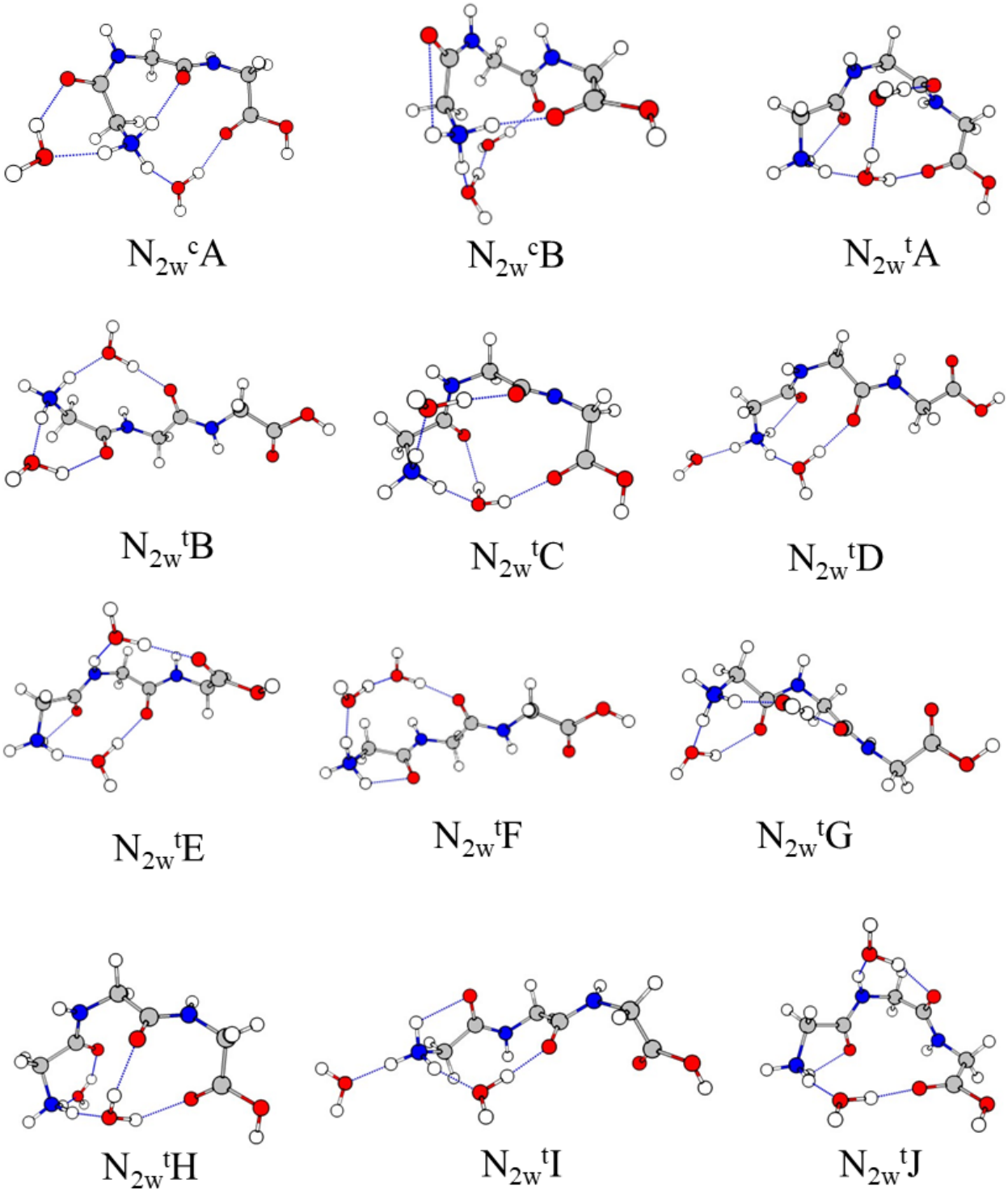


Figure 5.3. Structures of the lowest energy conformers of $Ala_3H^+(H_2O)$ with $Gly_3H^+(H_2O)$ naming. $Ala_3H^+(H_2O)$ energetic naming is denoted with a subscript a before the name.

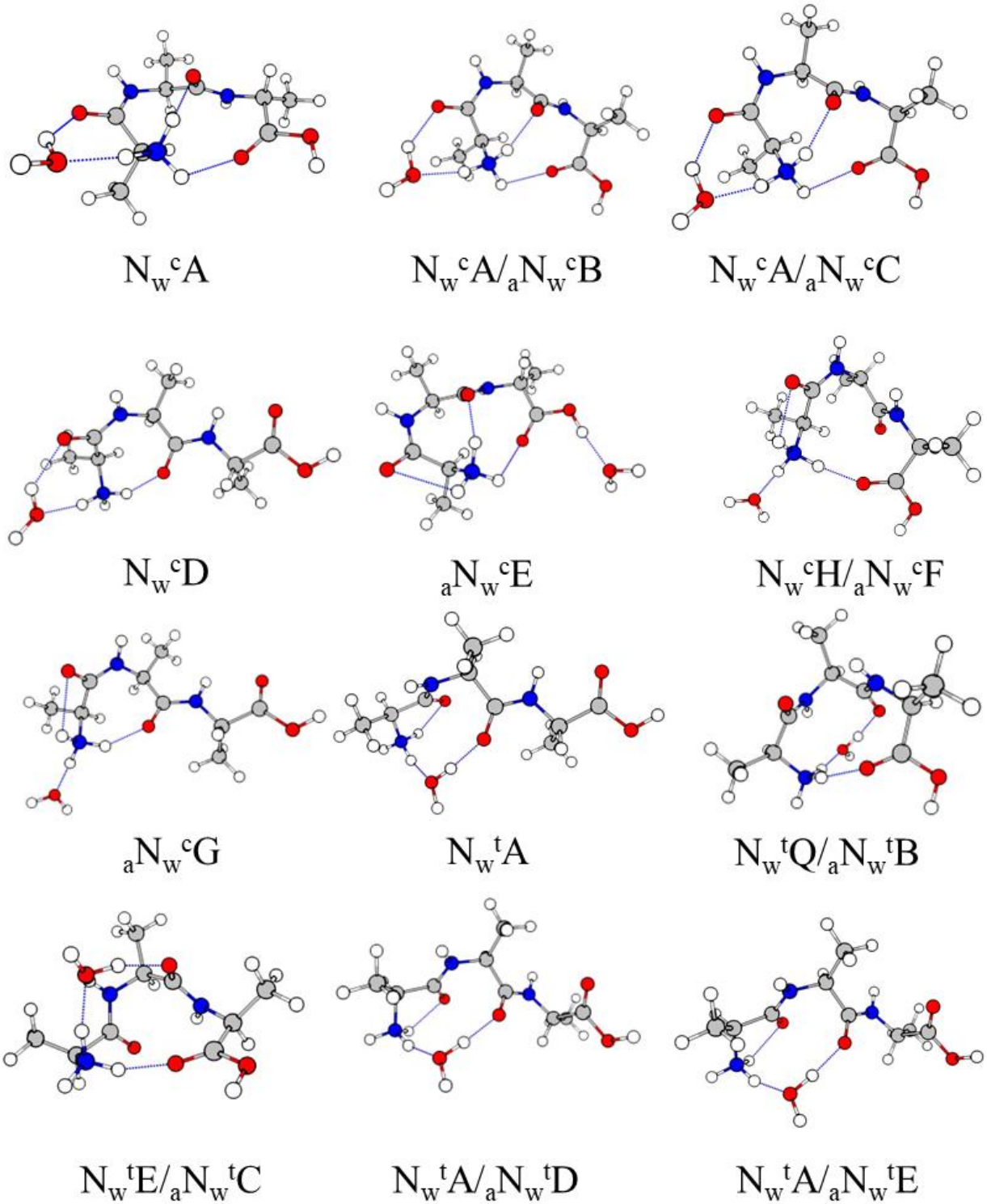
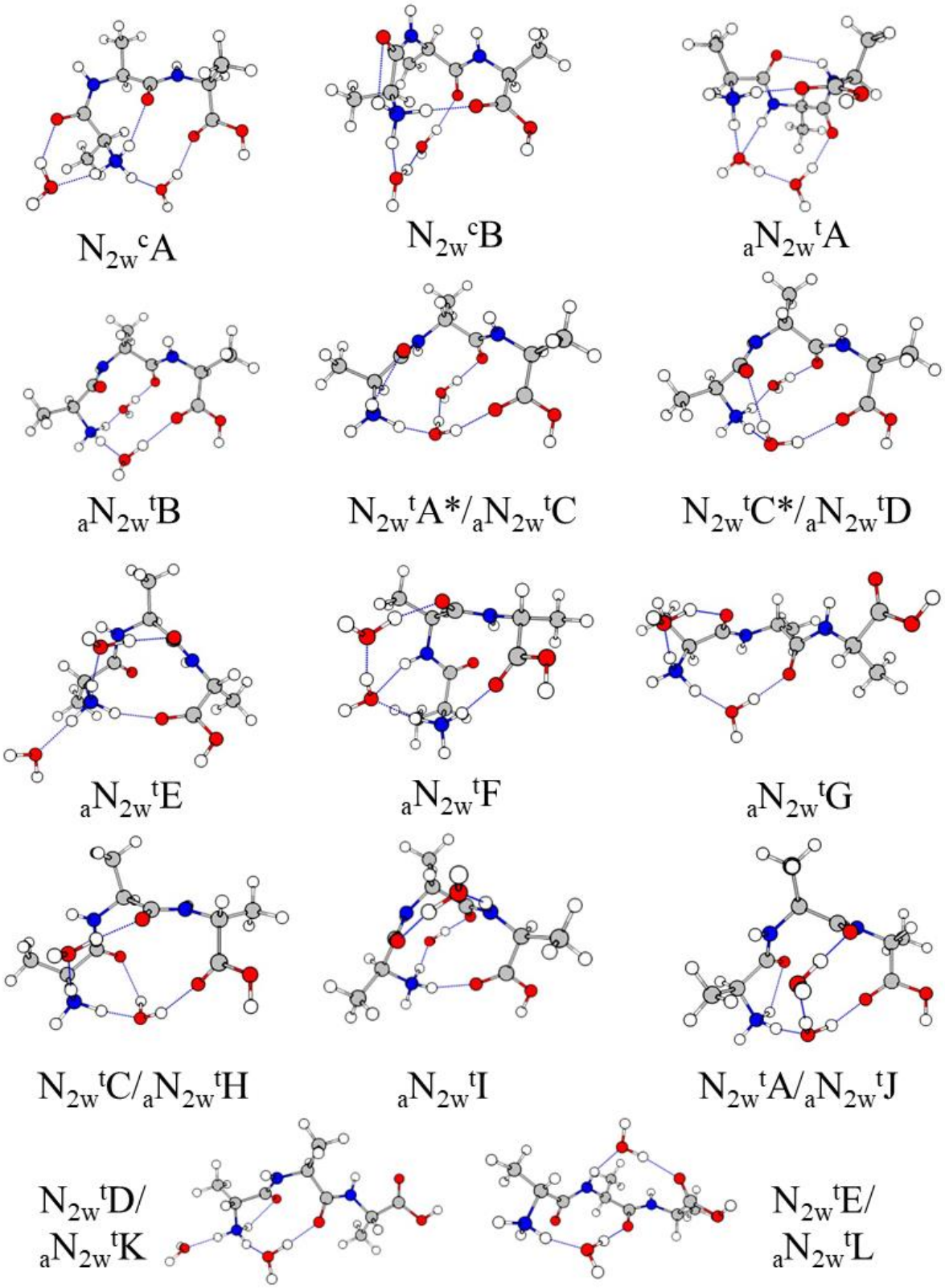


Figure 5.4. Lowest energy conformers of $Ala_3H^+(H_2O)_2$ with $Gly_3H^+(H_2O)$ naming. $Ala_3H^+(H_2O)_2$ energetic naming is denoted with a subscript a before the name.



to the experimental spectra. The isomer naming scheme used here is similar to previous papers of the bare structures,[7, 10, 17] where the first letter is based on the protonation site (**N** for the amine, and **O** for the amide carbonyl), configuration of the first amide bond (**c** for cis, and **t** for trans amide), and letter denoting the energetic ordering for the Gly_3H^+ species. The energetics for the lowest energy conformers for $Gly_3H^+(H_2O)_{1-2}$ and $Ala_3H^+(H_2O)_{1-2}$ are shown in **Table 5.1** and **Table 5.2**, respectively. The $Ala_3H^+(H_2O)_{1-2}$ energetic ordering is denoted by the same naming scheme, however, has a subscript (a) in front of the name. The corresponding $Gly_3H^+(H_2O)_{1-2}$ structure name is also denoted. The ten lowest energy calculated harmonic spectra for $Gly_3H^+(H_2O)_2$ are shown in **Figure(s) 5.5** and **5.6**. The fifteen lowest energy calculated harmonic spectra for $Ala_3H^+(H_2O)$ are shown in **Figure(s) 5.7**, **5.8**, and **5.9**. The fifteen lowest energy calculated harmonic spectra for $Ala_3H^+(H_2O)_2$ are shown in **Figure(s) 5.10**, **5.11**, and **5.12**.

5.4 Results

The IRPD spectra of $Gly_3H^+(H_2O/D_2O)\cdot D_2$[10] and $Gly_3H^+(H_2O/D_2O)_2\cdot D_2$ are shown in the top panels of **Figure 5.13** and **Figure 5.15**, respectively, and those of $Ala_3H^+(H_2O/D_2O)\cdot D_2$, and $Ala_3H^+(H_2O/D_2O)_2\cdot D_2$ are shown in the top panels of **Figure 5.14** and **Figure 5.16**, respectively. The spectra of the H_2O species are shown in blue and the corresponding D_2O spectrum is overlaid in red. The band associated with the water molecules are shifted by the isotopic substitution and therefore appear in blue (O-H) and red (O-D) in the figure. The features which are common to both spectra, and appear in purple, therefore belong to the peptide ion.[10, 21-22] This allows the separation of the peptide and solvent bands, which provide an important additional comparison criterion with the calculated spectra. We note that there are also some features which appear in

Table 5.1. Lowest energy conformers of $Gly_3H^+(H_2O)_{1-2}$.

	Conformer	DFT	MP2
Gly_3H^+ + 1 H_2O	**N_w^tA**	0.0	0.0
	N_w^tB	6.3	2.4
	N_w^tC	7.3	1.5
	N_w^tD	9.6	-
	N_w^tE	10.0	1.0
	N_w^cA	7.8	0.1
	N_w^cB	10.5	12.0
	N_w^cC	11.3	4.9
	N_w^cD	11.6	13.4
	N_w^cE	12.2	3.8
Gly_3H^+ + 2 H_2O	N_{2w}^tA	0.0	0.0
	N_{2w}^tB	0.1	3.6
	N_{2w}^cA	0.6	5.5
	N_{2w}^tC	1.4	2.1
	N_{2w}^tD	1.5	3.4
	N_{2w}^tE	2.9	4.2
	N_{2w}^tF	4.7	7.8
	N_{2w}^tG	5.1	6.1
	N_{2w}^tH	5.5	4.6
	N_{2w}^tI	6.5	6.0
	N_{2w}^ttJ	6.9	9.6
	N_{2w}^tK	7.9	5.6
	N_{2w}^tL	8.6	9.7
	N_{2w}^tM	9.2	10.0

Table 5.2. Lowest energy conformers of $Ala_3H^+(H_2O)_{1-2}$.

	Conformer	Conformer (Ala Energetic Ordering)	DFT	MP2
Ala_3H^+ + 1 H_2O	N_w^cA	$_aN_w^cA$	0.0	0.0
	N_w^tA	**$_aN_w^tA$**	4.6	5.1
	N_w^tQ	$_aN_w^tB$	4.9	6.3
	N_w^tE	$_aN_w^tC$	6.2	7.3
	N_w^cA	$_aN_w^cB$	7.0	6.8
	N_w^cA	$_aN_w^cC$	7.0	6.8
	N_w^tA	$_aN_w^tD$	8.4	7.3
	N_w^tA	$_aN_w^tE$	8.4	7.3
	-	$_aN_w^tF$	8.7	8.3
	N_w^cD	$_aN_w^cD$	12.2	16.6
	-	$_aN_w^cE$	14.4	11.8
	-	$_aN_w^tG$	14.6	11.6
	-	$_aN_w^tH$	14.6	11.6
	N_w^cH	**$_aN_w^cF$**	15.9	15.1
	N_w^cL	$_aN_w^cG$	20.4	21.4
Ala_3H^+ + 2 H_2O	-	$_aN_{2w}^tA$	0.0	6.1
	-	$_aN_{2w}^tB$	0.2	0.0
	N_{2w}^tA*	$_aN_{2w}^tC$	0.5	0.5
	N_{2w}^tC*	$_aN_{2w}^tD$	0.8	1.5
	-	$_aN_{2w}^tE$	2.3	3.7
	-	$_aN_{2w}^tF$	3.5	9.4
	-	$_aN_{2w}^tG$	4.6	6.8
	N_{2w}^tC	$_aN_{2w}^tH$	4.9	6.4
	-	$_aN_{2w}^tI$	5.5	10.0
	N_{2w}^tA	$_aN_{2w}^tJ$	6.5	7.5
	N_{2w}^tD	**$_aN2wtK$**	7.3	7.8
	N_{2w}^tE	**$_aN2wtL$**	8.5	11.2
	-	$_aN_{2w}^tM$	8.6	10.2
	-	$_aN_{2w}^tN$	9.1	12.1
	-	$_aN_{2w}^tO$	9.1	12.1
	-	$_aN_{2w}^tP$	10.7	12.1
	-	$_aN_{2w}^tQ$	11.3	16.7
	-	$_aN_{2w}^tR$	11.6	13.4
	-	$_aN_{2w}^tS$	12.9	17.9
	-	$_aN_{2w}^tT$	13.0	13.0
	-	$_aN_{2w}^tU$	13.2	15.1
	-	$_aN_{2w}^tV$	13.4	14.7

-	$_{a}N_{2w}{}^{t}W$	15.4	21.1
-	$_{a}N_{2w}{}^{t}X$	15.7	20.3
$N_{2w}{}^{c}A$	$_{a}N_{2w}{}^{c}A$	15.7	18.9
-	$_{a}N_{2w}{}^{t}Z$	16.1	19.8
$N_{2w}{}^{c}B$	**$_{a}N_{2w}{}^{c}B$**	19.7	

Figure 5.5. IRPD spectra of $Gly_3H^+(H_2O)_2$ compared to harmonic spectra of lowest energy conformers (#1-5) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Gly_3H^+ + 2H_2O$ (blue) overlaid with $Gly_3H^+ + 2D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

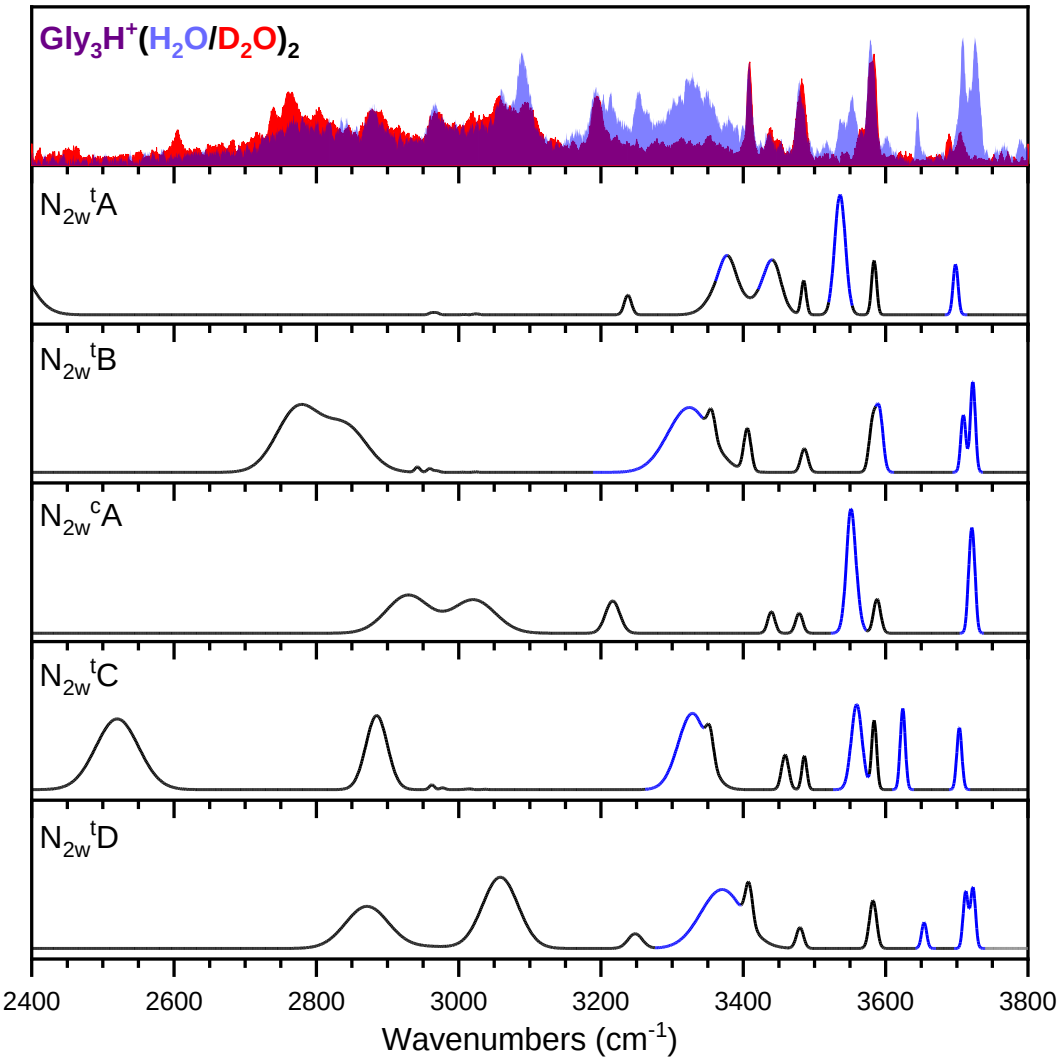


Figure 5.6. IRPD spectra of $Gly_3H^+(H_2O)_2$ compared to harmonic spectra of lowest energy conformers (#6-10) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Gly_3H^+ + 2H_2O$ (blue) overlaid with $Gly_3H^+ + 2D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

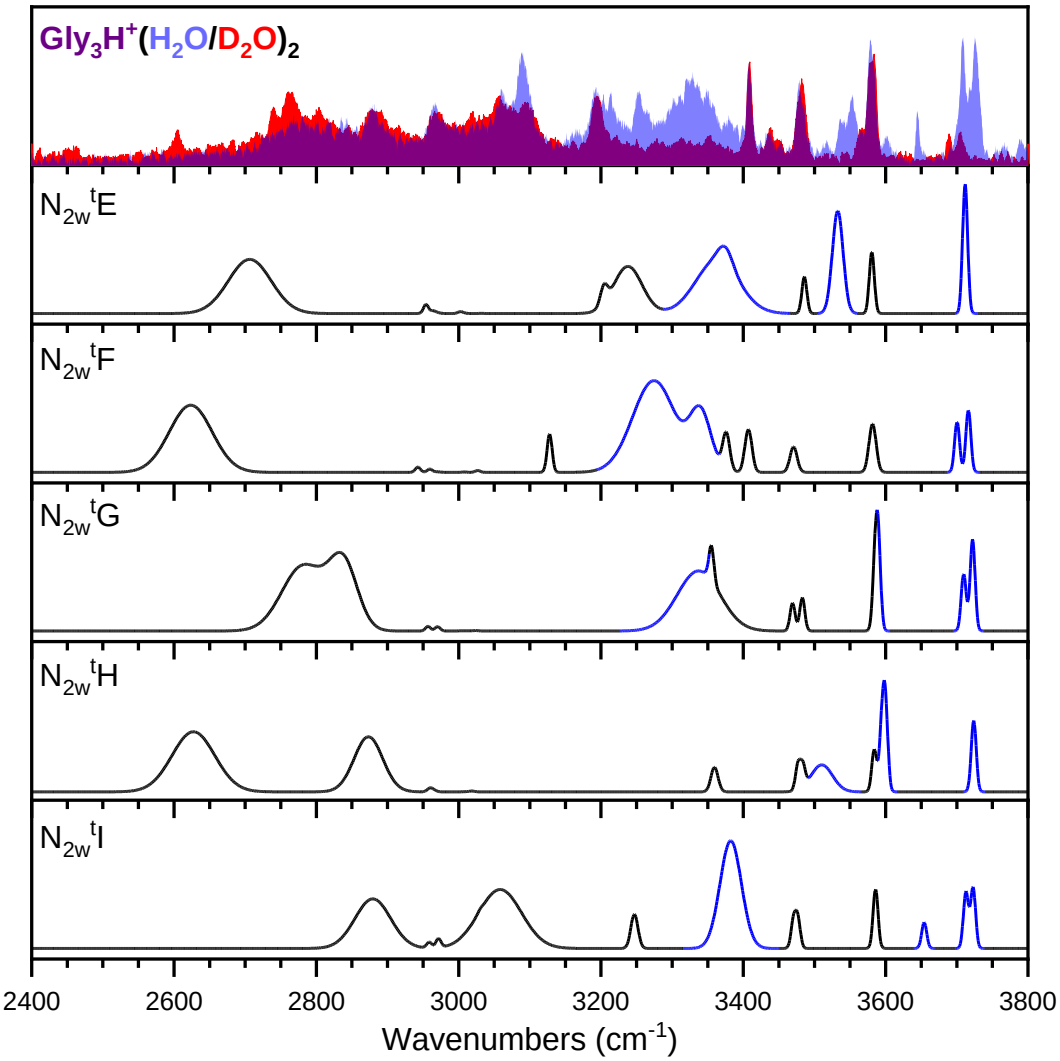


Figure 5.7. IRPD spectra of $Ala_3H^+(H_2O)$ compared to harmonic spectra of lowest energy conformers (#1-5) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Ala_3H^+ + H_2O$ (blue) overlaid with $Ala_3H^+ + D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

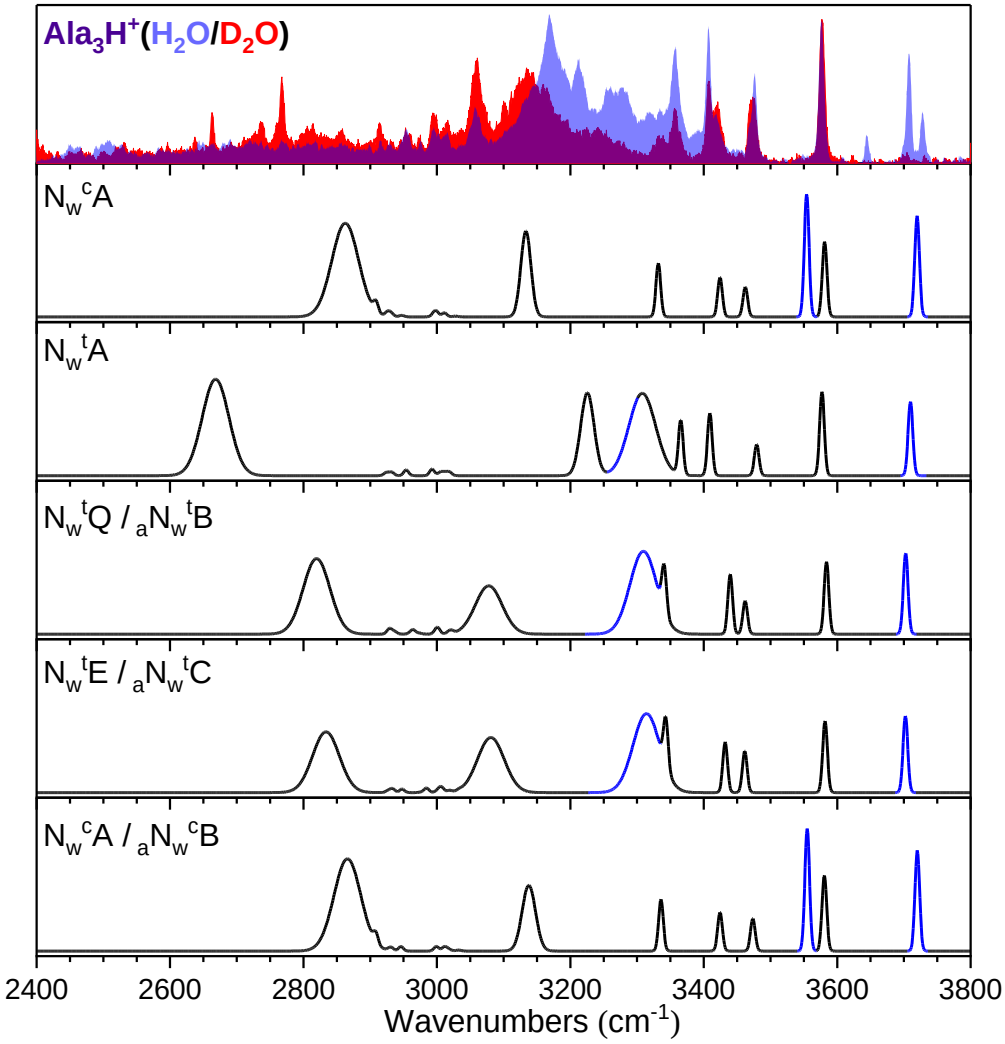


Figure 5.8. IRPD spectra of $Ala_3H^+(H_2O)$ compared to harmonic spectra of lowest energy conformers (#6-10) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Ala_3H^+ + H_2O$ (blue) overlaid with $Ala_3H^+ + D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

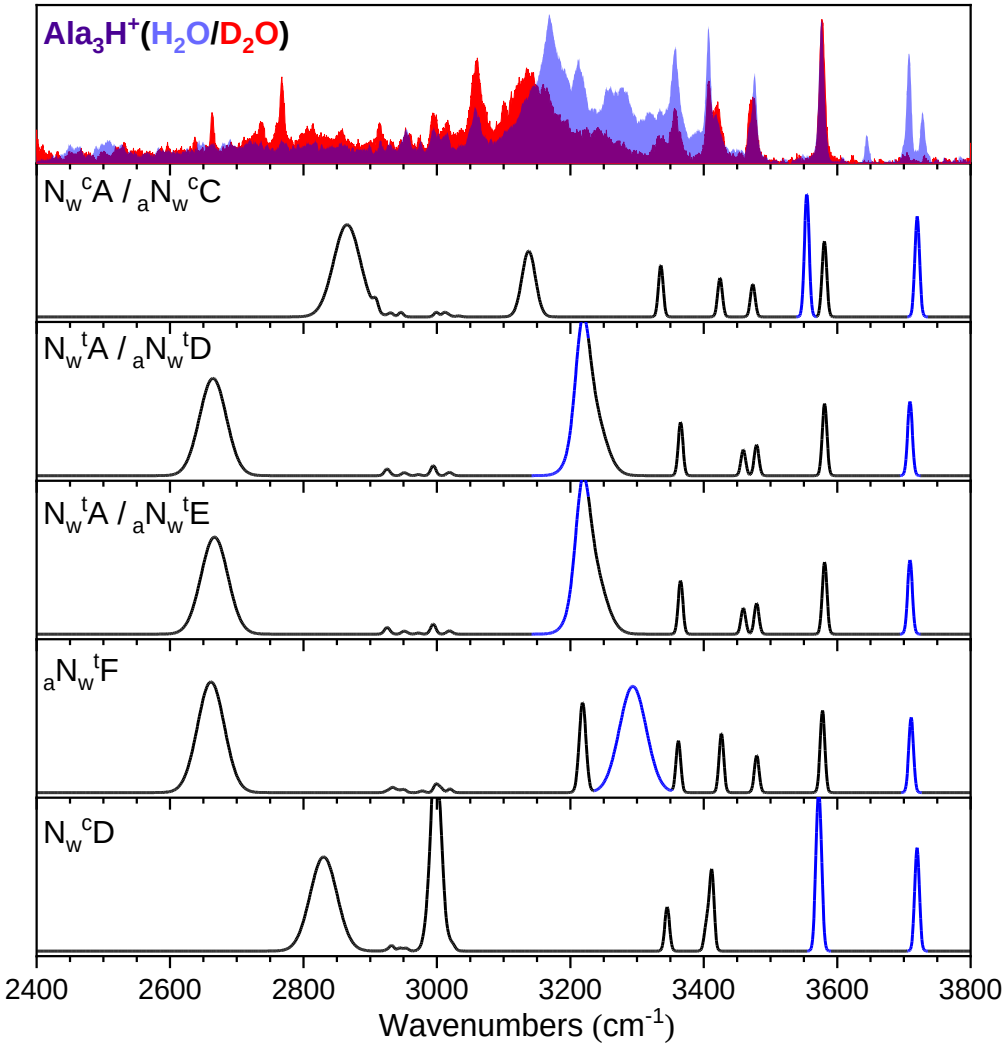


Figure 5.9. IRPD spectra of $Ala_3H^+(H_2O)$ compared to harmonic spectra of lowest energy conformers (#11-15) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Ala_3H^+ + H_2O$ (blue) overlaid with $Ala_3H^+ + D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

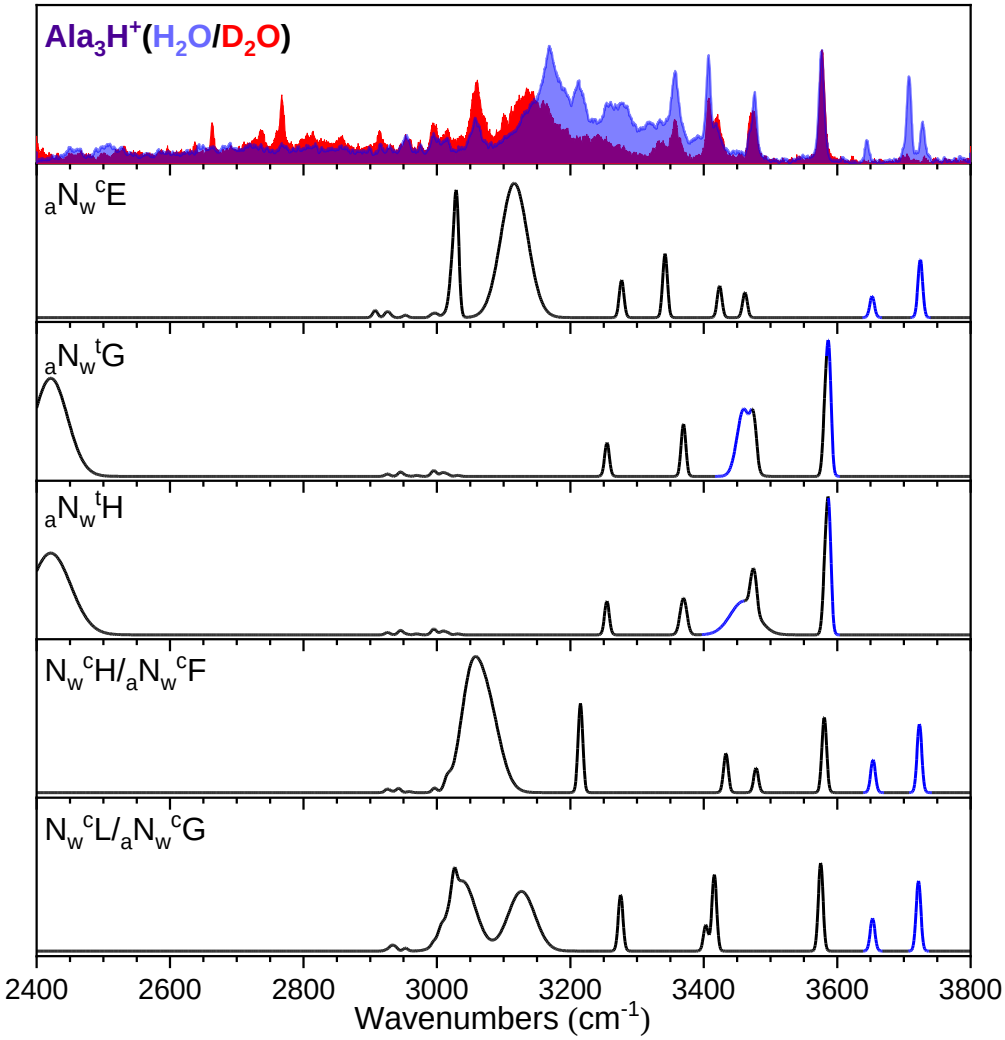


Figure 5.10. IRPD spectra of $Ala_3H^+(H_2O)_2$ compared to harmonic spectra of lowest energy conformers (#1-5) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Ala_3H^+ + 2H_2O$ (blue) overlaid with $Ala_3H^+ + 2D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

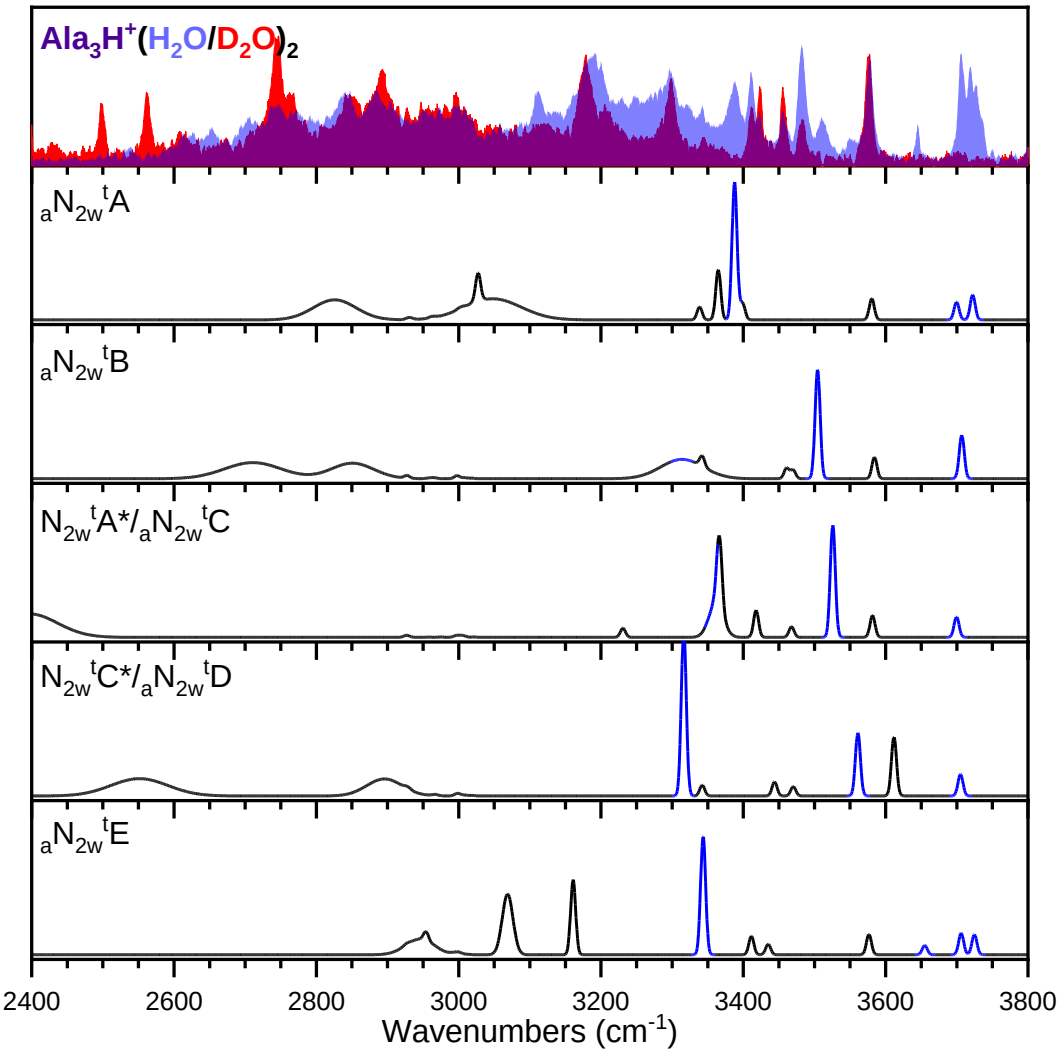


Figure 5.11. IRPD spectra of $Ala_3H^+(H_2O)_2$ compared to harmonic spectra of lowest energy conformers (#6-10) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Ala_3H^+ + 2H_2O$ (blue) overlaid with $Ala_3H^+ + 2D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

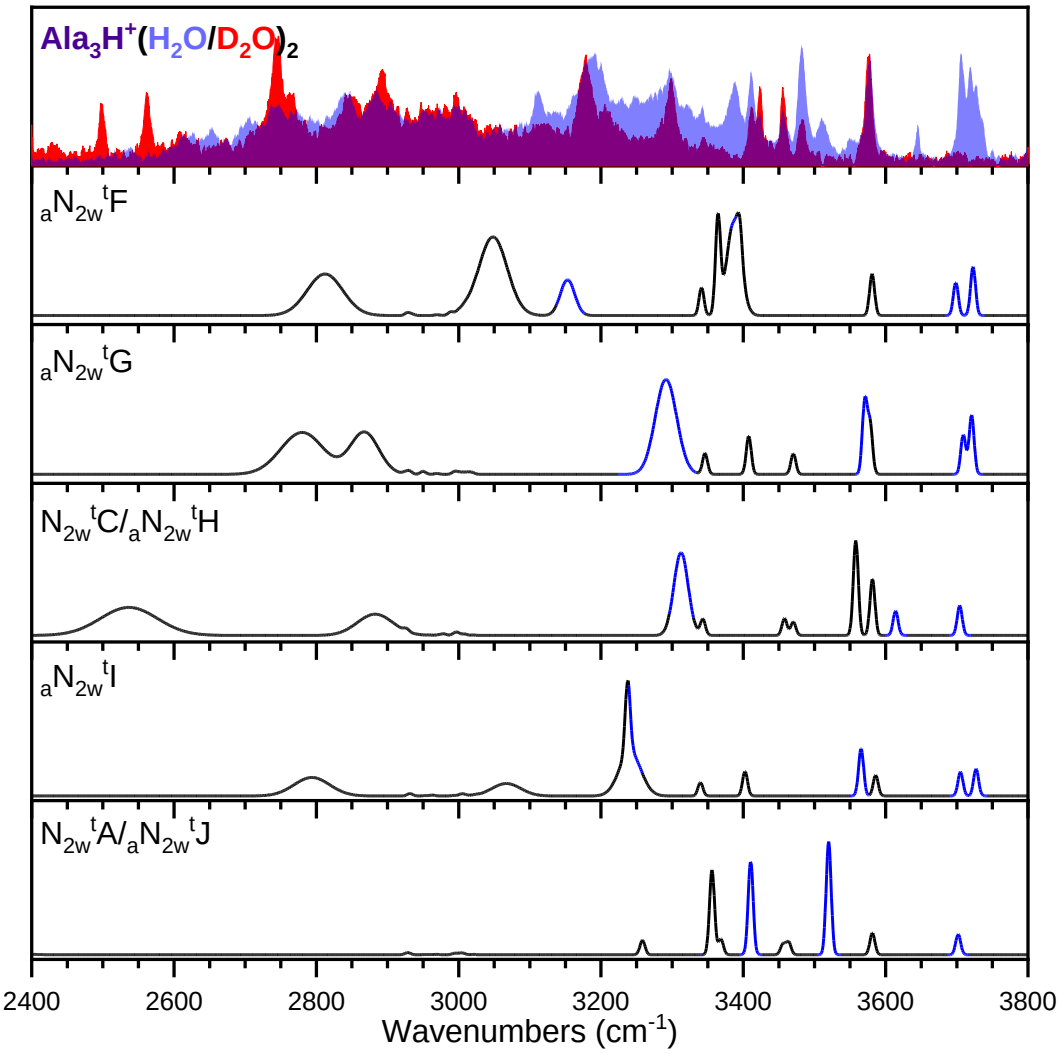
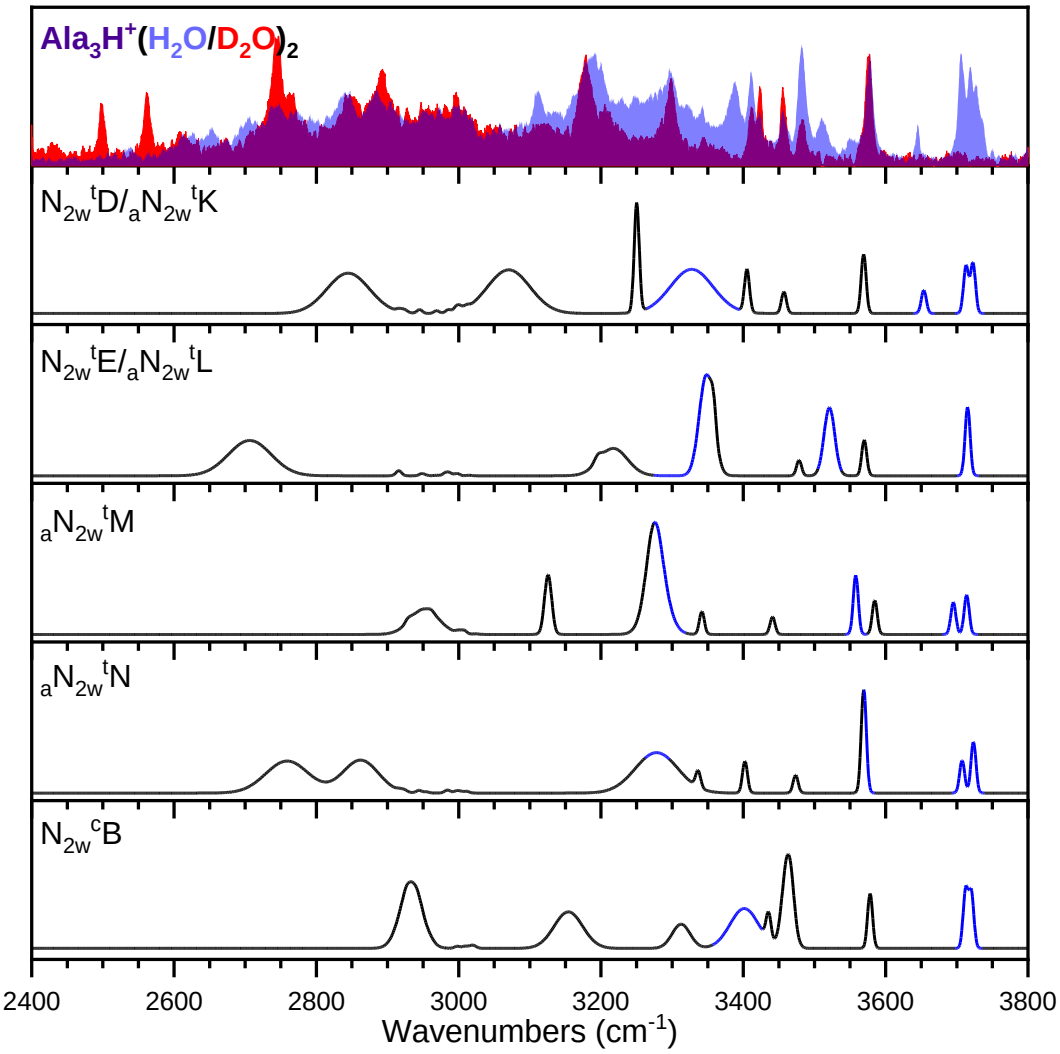


Figure 5.12. IRPD spectra of $Ala_3H^+(H_2O)_2$ compared to harmonic spectra of lowest energy conformers (#11-15) in order calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The one-laser IRPD spectrum of $Ala_3H^+ + 2H_2O$ (blue) overlaid with $Ala_3H^+ + 2D_2O$ (red) is shown on top. All features pertaining to an H_2O stretch are shown in blue.

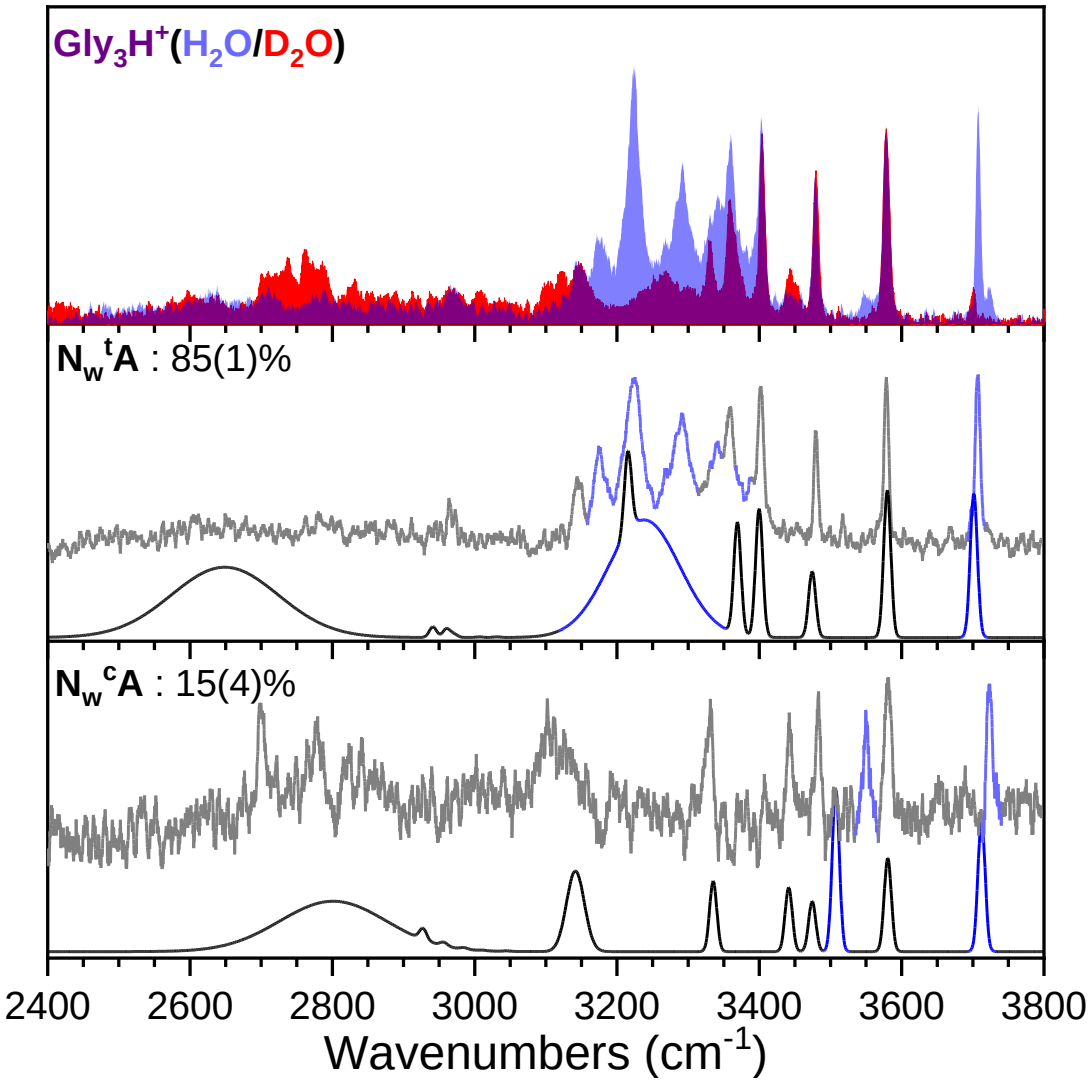


red above 2800 cm^{-1} even though they belong to D_2O. The explanation for these features is that they are peptide modes which are strongly coupled to the H_2O modes. When the H_2O is substituted for a D_2O and this coupling disappear, these peptide features slightly change position and intensity.

Each experimental spectrum contains more features than what would be expected if there was only one conformer in each population. For the amine protonated (tripeptideH^+)·H_2O system, we would expect 2 O-H stretches for the water, a carboxylic acid O-H stretch, and 5 N-H stretches. In the $Ala_3H^+(H_2O)$ spectrum, there are three features in the 3600-3800 cm^{-1} free O-H region which belong to the H_2O molecules. Multiple additional H_2O bands in the H-bonded O-H region between 3100 and 3400 cm^{-1} are also visible. A similar number of H_2O O-H stretches are found in the $Gly_3H^+(H_2O)$ spectra. Therefore, simply by looking at the number of water O-H stretches, it is possible to determine that there must be at least two conformers present for both singly solvated tripeptide. A similar analysis can be done for the (tripeptide$^+$)$(H_2O)_2$ spectra in which we would expect 4 O-H stretches from H_2O. A larger number of H_2O features are found for both $Gly_3H^+(H_2O)_2$ and $Ala_3H^+(H_2O)_2$. Thus, having determined that each cluster are composed of multiple conformers, we use IR-IR double resonance spectroscopy to obtain conformer-specific spectra.

The IRPD and IR-IR spectra of $Gly_3H^+(H_2O)$ are shown in in **Figure 5.13**. The isomer deconvolution, assignment, and quantification for $Gly_3H^+(H_2O)$ was described in detail previously.[10] Briefly, using an IR-IR ion-dipping scheme with the probe laser fixed at 3225 cm^{-1} and 3546 cm^{-1} produces two distinct spectra shown in grey in **Figure 5.13**. Comparing these with the calculated spectra of some of the lowest energy structures found in the computational search yields good agreement with $\mathbf{N_w^tA}$ and $\mathbf{N_w^cA}$, respectively. The geometries and energies of these

Figure 5.13. IRPD spectra of $Gly_3H^+(H_2O/D_2O)\cdot D_2$[10] compared to harmonic spectra calculated at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE level. The isomer specific IR-IR spectra are overlaid on the calculated spectra that they correspond to. All features highlighted in blue are from H_2O stretching.

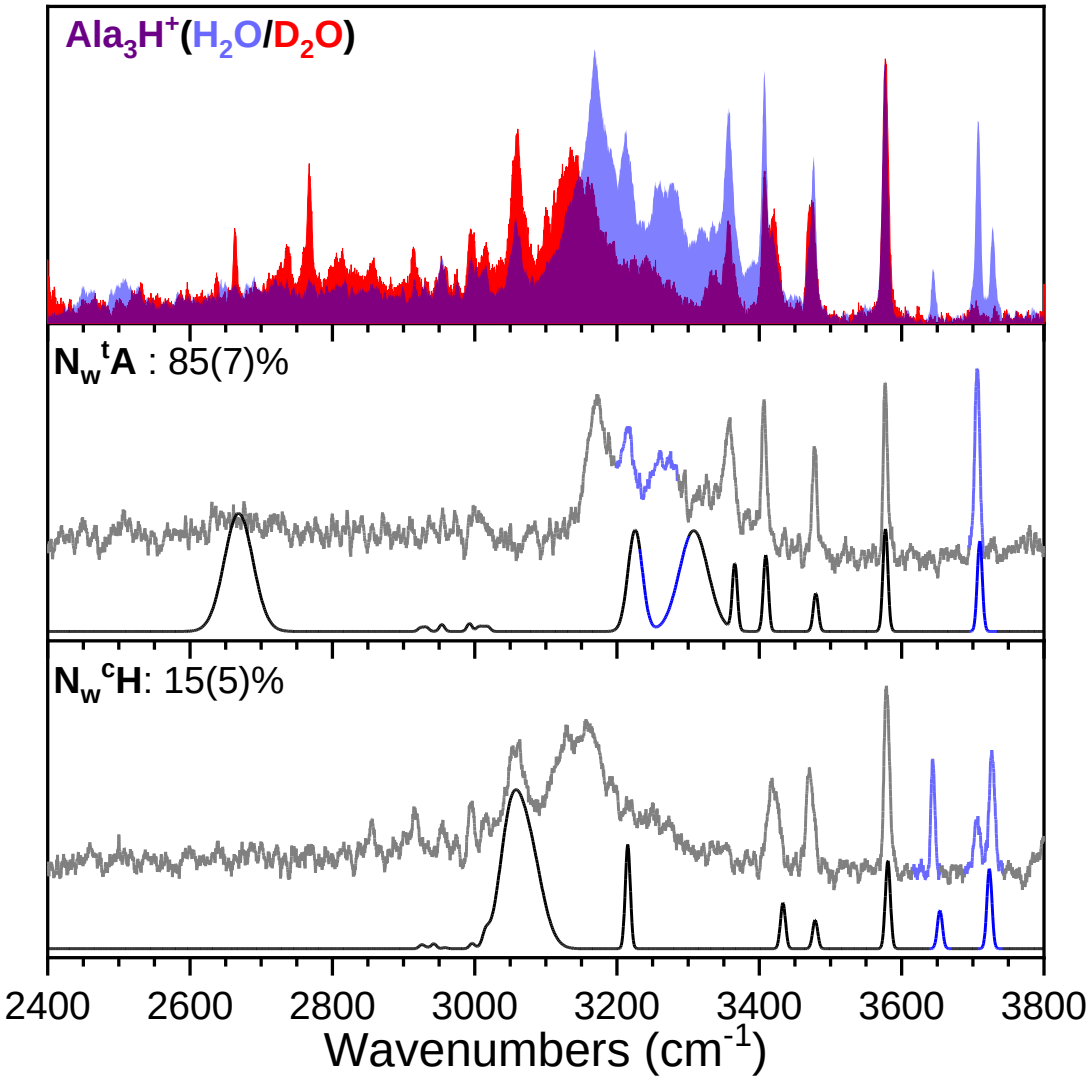


structures are shown in **Figure 5.1**. The $\mathbf{N_w^tA}$ conformer is protonated at the amine and has all the amide in the *trans* configuration. The water molecule is bridging the protonated amine and one of the amide carbonyls in a single H-bond acceptor/single H-bond donor (AD) motif. This gives rise to a strongly H-bond interaction with the corresponding water O-H stretch around 3200 cm^{-1}. The $\mathbf{N_w^cA}$ conformer is also protonated at the amine, but the first amide bond is in the cis configuration, similarly to the bare N^cA conformer. The water molecule is occupying the site between the protonated amine and the first amide carbonyl in a bridging AD configuration. The water in this conformer is engaging in a much weaker H-bond interaction with a predicted O-H stretch at 3551 cm^{-1}.

These two conformer structures are sufficient to account for all the peptide peaks and all the H_2O peaks above 3500 cm^{-1}. However, the one-laser IRPD spectrum display multiple H-bonded O-H peaks from H_2O between 3100 cm^{-1} and 3400 cm^{-1} and the corresponding $\mathbf{N_w^tA}$ structure only has one. We have shown previously, using systematic IR-IR spectroscopy, that all these H-bond O-H features belong to the $\mathbf{N_w^cA}$ conformer.[10] They originate from anharmonic effects, most probably a coupling between the O-H stretch and a low frequency peptide backbone mode.[35-39] The relative $\mathbf{N_w^tA}$ and $\mathbf{N_w^cA}$, conformer population can be estimated by comparing the calculated peak intensities to some of the experimental features unique to a single conformer. Using the amine peaks at 3404 ($\mathbf{N_w^tA}$) and 3448 ($\mathbf{N_w^cA}$) this analysis yields contributions of 85(1)% and 15(4)% for $\mathbf{N_w^tA}$ and $\mathbf{N_w^cA}$, respectively.

A similar approach is taken for $Ala_3H^+(H_2O)$ and using the IR-IR ion-dipping scheme with the probe laser fixed at 3358 and 3645 cm^{-1} yields the conformer specific spectra shown in grey in **Figure 5.14**. These spectra are found to be in good agreement with the $\mathbf{N_w^tA}$ and $\mathbf{N_w^cH}$ structures.

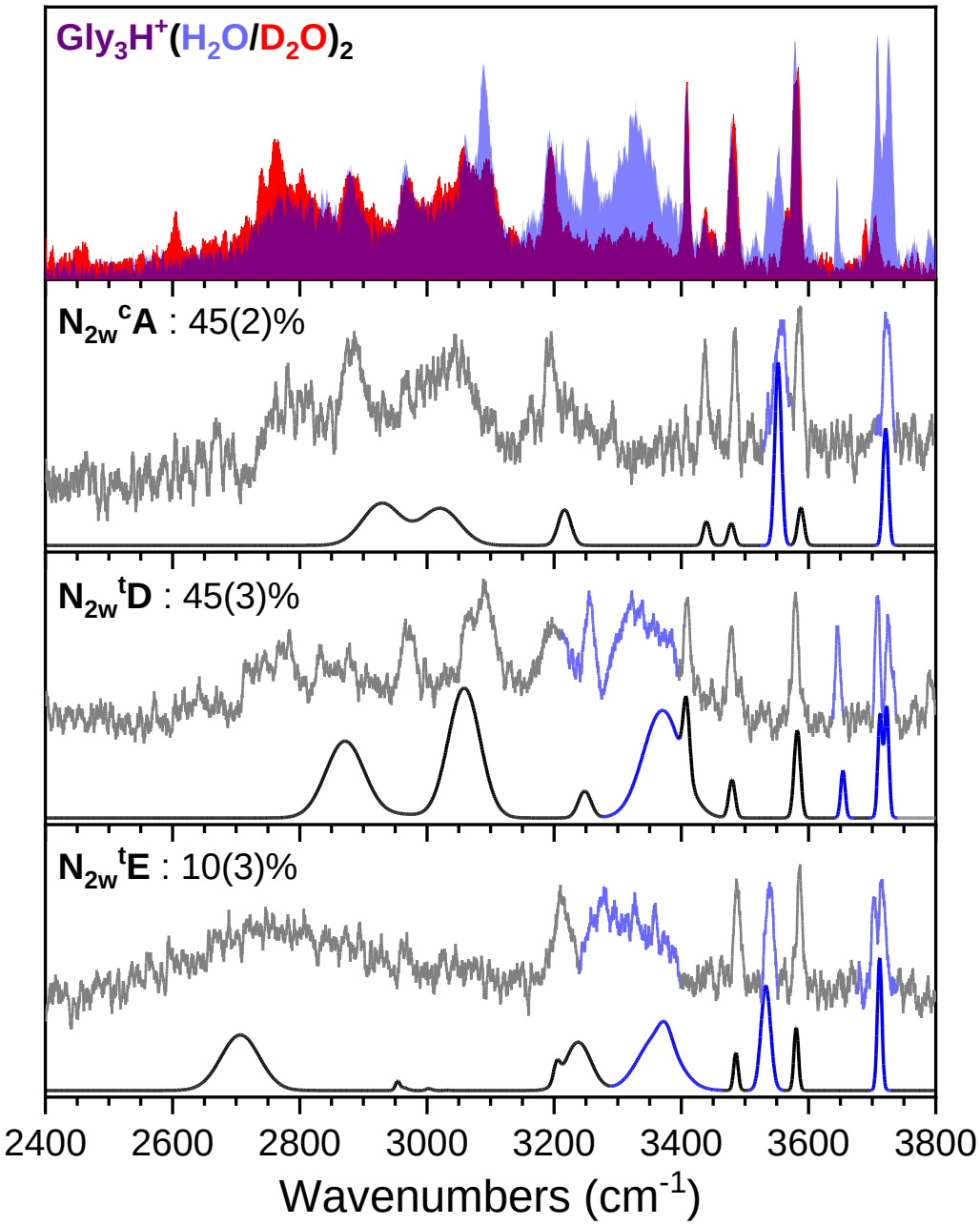
Figure 5.14. IRPD spectra of $Ala_3H^+(H_2O/D_2O)\cdot D_2$ compared to harmonic spectra calculated at the cam-B3LYP/def2-TZVP/BSSE level. The isomer specific IR-IR spectra are overlaid on the calculated spectra that they correspond to.



The $Ala_3H^+(H_2O)$ **N_w^tA** conformer is identical to that of $Gly_3H^+(H_2O)$. The **N_w^cH** structure is protonated at the amine and has the first amide in the cis configuration. The water is interacting with the protonated amine group with a single H-bond acceptor motif which give rise to two free water O-H stretches. The peptide chain has a conformation which is similar to that of the bare minor conformer, **N^cB**. Similarly to $Gly_3H^+(H_2O)$, these two conformers do not account for the multiple H-bonded O-H peaks from H_2O between 3100 cm^{-1} and 3400 cm^{-1}. Because the **N_w^tA** for structure for $Ala_3H^+(H_2O)$ is the same as $Gly_3H^+(H_2O)$, we assign these extra numerous peaks to similar anharmonic effects. The relative conformer population was determined using the amine peak at 3357 cm^{-1} (**N_w^tA**) and the symmetric free water stretch at 3645 cm^{-1} (**N_w^cH**) and was found to be 85(7):15(5), respectively.

For $Gly_3H^+(H_2O)_2$, the IR-IR ion-dipping scheme was used with the probe laser fixed at 3536 cm^{-1}, 3552 cm^{-1}, and 3645 cm^{-1}, yielding the three conformer specific spectra shown in grey in **Figure 5.15**. Comparison of these spectra to those calculated for the low energy structures yielded good agreement with for the **N_{2w}^cA**, **N_{2w}^tD**, and **N_{2w}^tE** structures. **N_{2w}^cA** has a similar peptide conformation as **N_w^cA** with second water molecule inserted between the protonated amine and the carboxylic acid C=O. Both water molecules are engaged in H-bonding interactions with very similar strength and therefore give rises to overlapping O-H stretches at 3557 cm^{-1} and 3722 cm^{-1}. Both **N_{2w}^tD** and **N_{2w}^tE** have a similar structure as **N_w^tA**. **N_{2w}^tD** has the additional water H-bonded on the protonated amine in a single H-bond acceptor configuration. On the other hand, **N_{2w}^tE** has the second water molecule H-bonded between the carboxylic acid C=O and the central amide N-H. The isomer population was determined using the O-H stretch peak at 3557 cm^{-1} (**N_{2w}^cA**), the coupled N3 amide N-H and water O-H stretch peak at 3537 cm^{-1} (**N_{2w}^tE**) and the

Figure 5.15. IRPD spectra of $Gly_3H^+(H_2O/D_2O)_2 \cdot D_2$ compared to harmonic spectra calculated at the cam-B3LYP/def2-TZVP/BSSE level. The isomer specific IR-IR spectra are overlaid on the calculated spectra that they correspond to.



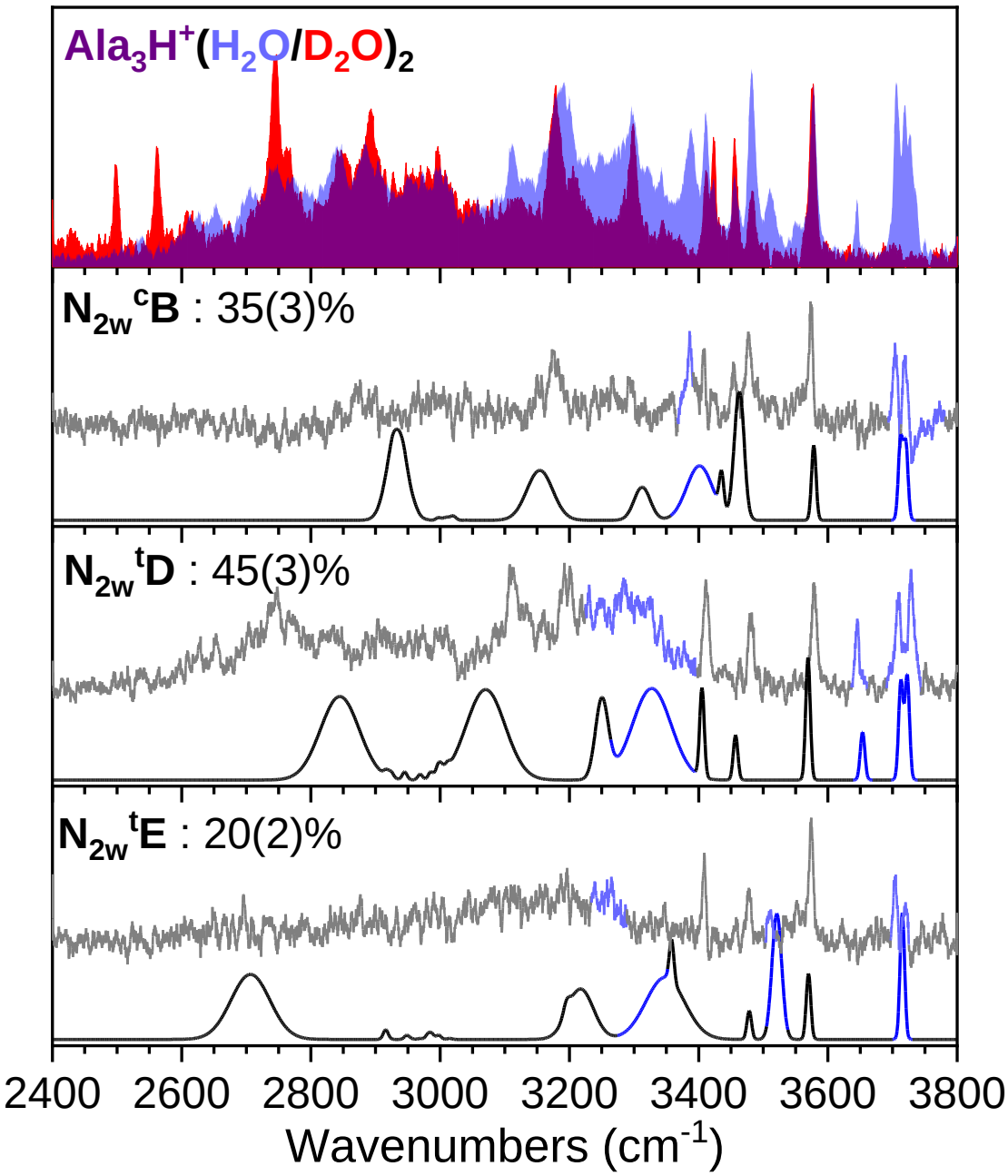
symmetric free water stretch peak at 3646 cm^{-1} (**$N_{2w}{}^{t}$D**) and was found to be 45(2):10(3):45(3), respectively.

Finally, the conformer specific spectra For $Ala_3H^+(H_2O)_2$ are shown in grey in **Figure 5.16**. Comparing these spectra to the lowest energy structure spectra found through DFT and MP2 calculations yielded good agreement with structures **$N_{2w}{}^{c}$B**, **$N_{2w}{}^{t}$D**, and **$N_{2w}{}^{t}$E**. **$N_{2w}{}^{c}$B** is similar to the minor conformers **$N_w{}^{c}$H,** with the second water molecule the first water molecule and to the central amide carbonyl. The other two conformers have a similar conformation as **$N_w{}^{t}$A,** the major conformer in both $Ala_3H^+(H_2O)$ and $Gly_3H^+(H_2O)$, but differ in the placement of the second water molecule. The isomer population was estimated utilizing the hydrogen bonded water O-H stretch peak at 3457 cm^{-1} (**$N_{2w}{}^{c}$B**), the coupled N3 amide N-H and water O-H stretch peak at 3513 cm^{-1} (**$N_{2w}{}^{t}$E**) and the symmetric free water stretch peak at 3645 cm^{-1} (**$N_{2w}{}^{t}$D**) and was found to be 35(3):20(2):45(3), respectively.

5.5 Discussion

As discussed in our previous work,[10] the addition of the first water molecule to Gly_3H^+ induced large conformational changes. First, protonation at the amide carbonyls (**O^{t}A**), which represent 65% of the isomeric population in the bare Gly_3H^+ completely disappears and all the observed $Gly_3H^+(H_2O)$ conformer involve protonation at the amine site. Second, the 35% of Gly_3H^+ with amine protonation and a cis amide bond (**N^{c}A**) is reduced to about 15% in $Gly_3H^+(H_2O)$, in the form of the **$N_w{}^{c}$A** conformer. The remaining 85% of the conformer population is in the form of **$N_w{}^{t}$A** lowest energy conformer, with all trans amide. This is quite remarkable since the conversion of cis/trans amide and protonation site involve significant barriers. However, we have shown that the water molecular binding energy, coupled with the slow cooling

Figure 5.16. IRPD spectra of $Ala_3H^+(H_2O/D_2O)_2 \cdot D_2$ compared to harmonic spectra calculated at the cam-B3LYP/def2-TZVP/BSSE level. The isomer specific IR-IR spectra are overlaid on the calculated spectra that they correspond to.



of the buffer gas, provide sufficient time and energy for most of the clusters to overcome these barriers.[10, 40] The presence of the slightly higher energy $\mathbf{N_w^cA}$ conformer can thus be seen as the small fraction of kinetically trapped conformer who couldn't overcome the largest cis/trans isomerization barrier. However, the MP2 calculations predict that the $\mathbf{N_w^cA}$ and $\mathbf{N_w^tA}$ conformers are nearly isoenergetic and it is possible that the presence of 15% of $\mathbf{N_w^cA}$ to be due to an equilibrium distribution before the conformers get trapped by the large conversion barrier.

We find a very similar situation for the addition of the first water onto Ala_3H^+. Similarly to Gly_3H^+, the main $\mathbf{O^tA}$ conformer of the bare Ala_3H^+ involves protonation at the amide carbonyl which is not present in any the observed $Ala_3H^+(H_2O)$ conformers. The minor amine protonated isomer of the bare Ala_3H^+ is slightly different than in Gly_3H^+ and differ by the number of hydrogen bonds between the protonated amine and the different carbonyl groups. These differences are due to the electron donating effects from the methyl side-chains which change the proton affinities of the amine and carbonyl groups.[7] However, the main conformer found for $Ala_3H^+(H_2O)$ is $\mathbf{N_w^tA}$ which is exactly the same as in the case of $Gly_3H^+(H_2O)$. This shows that effects of the methyl-side chains don't affect structure formed by the addition of the first water molecule. However, we note that this is conformer is predicted to be ~5 kJ/mol higher in energy than the lowest energy conformer for $Ala_3H^+(H_2O)$. The conformation of the peptide chain in the minor conformer of $Ala_3H^+(H_2O)$ is the same as in the bare Ala_3H^+ $\mathbf{N^cB}$ conformer. This structure is also predicted to be ~10 kJ/mol higher in energy than $\mathbf{N_w^tA}$ by both DFT and MP2 methods. Therefore, its presence in the spectra can also be explained by the fraction of $\mathbf{N^cB}$ that couldn't overcome the amide isomerization barrier and got kinetically trapped. We find a very similar ratio of minor/major conformer for both $Gly_3H^+(H_2O)$ and $Ala_3H^+(H_2O)$, which support similar formation mechanisms.

The addition of the second water molecule onto $Gly_3H^+(H_2O)$ generates three conformers for $Gly_3H^+(H_2O)_2$. Two of them, $\mathbf{N_{2w}^tD}$ and $\mathbf{N_{2w}^tE}$, are related to the addition of a water molecule onto $\mathbf{N_w^tA}$, the most abundant conformer in $Gly_3H^+(H_2O)$. The third one, $\mathbf{N_{2w}^cA}$, with a cis amide bond can be thought has the addition of a water molecule onto $\mathbf{N_w^cA}$, the minor conformer in $Gly_3H^+(H_2O)$. However, the population ratio suggests that the presence of this cis amide conformer here is not due to kinetic trapping. It represents about 40% of $Gly_3H^+(H_2O)_2$ while the corresponding $\mathbf{N_w^cA}$ it was only 15% of $Gly_3H^+(H_2O)$. Therefore, it is clear that the binding energy from the addition of the second water molecule is also sufficient to induce amide isomerization. The large population of the cis amide conformer must therefore be due to its competitive energetics. This is consistent with the computational results which indicate that the three conformers found for $Gly_3H^+(H_2O)$ are all calculated to be relatively close in energy and amongst the lowest ones. However, other conformations with trans amide bond such as $\mathbf{N_{2w}^tA}$, $\mathbf{N_{2w}^tB}$, and $\mathbf{N_{2w}^tC}$ are also calculated to be energetically competitive and are not detected in the experiment. Because we already established that the clustering process for the second water molecule can induce isomerization of the amide bond, it is unlikely the absence of $\mathbf{N_{2w}^tA}$, $\mathbf{N_{2w}^tB}$ and $\mathbf{N_{2w}^tC}$ stems from kinetic trapping. More likely, they are higher in energy than what the calculations suggest.

For $Ala_3H^+(H_2O)_2$ we find the exact same two conformers with trans amide bond, $\mathbf{N_{2w}^tD}$ and $\mathbf{N_{2w}^tE}$, which both stem from the main $\mathbf{N_w^tA}$ conformer in $Ala_3H^+(H_2O)$. However, the third conformer, $\mathbf{N_{2w}^cB}$, is different than in Gly_3H^+, but closely related to the minor $\mathbf{N_w^cH}$ conformer in $Ala_3H^+(H_2O)$. Once again, the ~35% population of this conformer in $Ala_3H^+(H_2O)_2$, compared to the ~15% parent in $Ala_3H^+(H_2O)$, also indicates that this conformer appears because it is energetically competitive rather than kinetically trapped. However, for $Ala_3H^+(H_2O)_2$, the three

detected conformers are calculated to be relatively high in energy compared to other structures. Because of the similarity with $Gly_3H^+(H_2O)_2$, this most likely represents shortcoming of the computations. We have observed similar discrepancies between the computed and observed structure for the bare tripeptides.[7]

Overall, we find the following similarities and differences between the microsolvation structure of Ala_3H^+ and Gly_3H^+. First, in both cases, the isomer with protonation at the amide carbonyl disappear in favor of the amine protonation site once water molecules are added. All the microsolvated structures found here involve at least one water molecule forming a strong H-bond with the charged protonated amine site. This strong favorable interaction with the solvent is not possible for the **O^{t}A** structure in which the proton is shared between two carbonyl groups and therefore independent of the presence of methyl side-chains. Second, the addition of water molecules induced the formation of the same tripeptide conformations with trans amide bonds and one water H-bonded between the protonated amine and the carboxylic acid carbonyl. However, the conformers involving a cis amide bond remains distinct upon the addition of water molecule. For Gly_3H^+, the internal H-bond between the protonated amine and the second amide carbonyl remain intact while in Ala_3H^+, the H-bond between the protonated amine and the carboxylic acid carbonyl survive the addition of water molecule. These differences highlight the difference in internal H-bond strength induced by the presence of electron donating methyl[21] side-chains found for the bare tripeptides.[7, 41-45]

5.6 Conclusions

In this paper, we presented the experimental IR-IR double resonance spectra and computational efforts that allowed us to determine the conformers in the microsolvated $Gly_3H^+(H_2O)_{1-2}$ and $Ala_3H^+(H_2O)_{1-2}$ clusters. We found similarity in the solvation structure of

these two tripeptides: 1) The dominant carbonyl protonated isomer in the bare protonated ion completely disappears upon the addition of water. 2) The same microsolvated clusters conformers with amine protonation and all trans amide bond conformers are formed. However, the two peptides produce different conformers that involve amine protonation and one cis amide bond. These have a rather small contribution in the $Gly_3H^+(H_2O)$ and $Ala_3H^+(H_2O)$ clusters which is probably due to kinetic trapping. However, they constitute a larger fraction $Gly_3H^+(H_2O)_2$ and $Ala_3H^+(H_2O)_2$ clusters, suggesting that they are again energetically competitive structures.

5.7 References

(1) Herzberg, O.; Moult, J. Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* **1991,** *11* (3), 223-9.

(2) Pandey, A. K.; Naduthambi, D.; Thomas, K. M.; Zondlo, N. J. Proline editing: a general and practical approach to the synthesis of functionally and structurally diverse peptides. Analysis of steric versus stereoelectronic effects of 4-substituted prolines on conformation within peptides. *J Am Chem Soc* **2013,** *135* (11), 4333-63.

(3) de Groot, B. L.; Daura, X.; Mark, A. E.; Grubmuller, H. Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. *J Mol Biol* **2001,** *309* (1), 299-313.

(4) Toniolo, C. Intramolecularly hydrogen-bonded peptide conformations. *CRC Crit Rev Biochem* **1980,** *9* (1), 1-44.

(5) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012,** *338* (6110), 1042-6.

(6) Pauling, L.; Corey, R. B.; Branson, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* **1951,** *37* (4), 205-11.

(7) Sherman, S. L.; Fischer, K. C.; Garand, E. Conformational Changes Induced by Methyl Side-Chains in Protonated Tripeptides Containing Glycine and Alanine Residues. *(submitted)* **2022**.

(8) Qi, X.; Kondoh, K.; Krusling, D.; Kelso, G. J.; Leonova, T.; Grabowski, G. A. Conformational and amino acid residue requirements for the saposin C neuritogenic effect. *Biochemistry* **1999,** *38* (19), 6284-91.

(9) Lucent, D.; Vishal, V.; Pande, V. S. Protein folding under confinement: a role for solvent. *Proc Natl Acad Sci U S A* **2007,** *104* (25), 10430-4.

(10) Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H(+).(H2O). *J Phys Chem A* **2018,** *122* (41), 8213-8221.

(11) Frauenfelder, H.; Fenimore, P. W.; Chen, G.; McMahon, B. H. Protein folding is slaved to solvent motions. *Proc Natl Acad Sci U S A* **2006,** *103* (42), 15469-72.

(12) Yu, Y.; Wang, J.; Shao, Q.; Shi, J.; Zhu, W. The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: a microscopic view. *Sci Rep* **2016,** *6*, 19500.

(13) García, A. E.; Hillson, N.; Onuchic, J. N. Solvent Effects on Protein Folding/Unfolding. *Progress of Theoretical Physics Supplement* **2000,** *138*, 282-291.

(14) Vaiana, S. M.; Manno, M.; Emanuele, A.; Palma-Vittorelli, M. B.; Palma, M. U. The role of solvent in protein folding and in aggregation. *J Biol Phys* **2001,** *27* (2-3), 133-45.

(15) Khuu, T.; Yang, N.; Johnson, M. A. Vibrational spectroscopy of the cryogenically cooled O- and N-protomers of 4-Aminobenzoic acid: Tag effects, isotopic labels, and identification of the E,Z isomer of the O-protomer. *Int J Mass Spectrom* **2020,** *457*.

(16) Leavitt, C. M.; Wolk, A. B.; Fournier, J. A.; Kamrath, M. Z.; Garand, E.; Van Stipdonk, M. J.; Johnson, M. A. Isomer-Specific IR-IR Double Resonance Spectroscopy of D2-Tagged Protonated Dipeptides Prepared in a Cryogenic Ion Trap. *J Phys Chem Lett* **2012,** *3* (9), 1099-105.

(17) Voss, J. M.; Fischer, K. C.; Garand, E. Revealing the structure of isolated peptides: IR-IR predissociation spectroscopy of protonated triglycine isomers. *Journal of Molecular Spectroscopy* **2018,** *347*, 28-34.

(18) Voss, J. M.; Kregel, S. J.; Fischer, K. C.; Garand, E. IR-IR Conformation Specific Spectroscopy of Na(+)(Glucose) Adducts. *J Am Soc Mass Spectrom* **2018,** *29* (1), 42-50.

(19) Chiba, T.; Okuyama, K.; Fujii, A. Observation of Evidence for the pi*-sigma* Hyperconjugation in the S1 State of o-, m-, and p-Fluorotoluenes by Double-Resonance Infrared Spectroscopy. *J Phys Chem A* **2016,** *120* (28), 5573-80.

(20) Chin, W.; Piuzzi, F.; Dognon, J. P.; Dimicoli, I.; Mons, M. Gas-phase models of gamma turns: effect of side-chain/backbone interactions investigated by IR/UV spectroscopy and quantum chemistry. *J Chem Phys* **2005,** *123* (8), 084301.

(21) Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J Phys Chem A* **2019,** *123* (15), 3355-3366.

(22) Fischer, K. C.; Sherman, S. L.; Garand, E. Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and beta-Alanine. *J Phys Chem A* **2020,** *124* (8), 1593-1602.

(23) Marsh, B. M.; Voss, J. M.; Garand, E. A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *J Chem Phys* **2015,** *143* (20), 204201.

(24) Voss, J. M.; Fischer, K. C.; Garand, E. Accessing the Vibrational Signatures of Amino Acid Ions Embedded in Water Clusters. *J Phys Chem Lett* **2018,** *9* (9), 2246-2250.

(25) Wolk, A. B.; Leavitt, C. M.; Garand, E.; Johnson, M. A. Cryogenic ion chemistry and spectroscopy. *Acc Chem Res* **2014,** *47* (1), 202-10.

(26) Campbell, J. L.; Yang, A. M.; Melo, L. R.; Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J Am Soc Mass Spectrom* **2016,** *27* (7), 1277-84.

(27) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: lessons learned from tetraalkylammonium cations and their isomers. *J Am Soc Mass Spectrom* **2014,** *25* (9), 1583-91.

(28) Lecours, M. J.; Chow, W. C.; Hopkins, W. S. Density functional theory study of Rh(n)S(0,+/-) and Rh(n+1)(0,+/-) (n = 1-9). *J Phys Chem A* **2014,** *118* (24), 4278-87.

(29) Liu, C.; Le Blanc, J. C.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using differential mobility spectrometry to measure ion solvation: an examination of the roles of solvents and ionic structures in separating quinoline-based drugs. *Analyst* **2015,** *140* (20), 6897-903.

(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

(31) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **2011,** *32* (7), 1456-65.

(32) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **2006,** *19* (4), 553-566.

(33) Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *The Journal of Chemical Physics* **1996,** *105* (24), 11024-11031.

(34) Voss, J. M.; Marsh, B. M.; Zhou, J.; Garand, E. Interaction between ionic liquid cation and water: infrared predissociation study of [bmim](+).(H2O)n clusters. *Phys Chem Chem Phys* **2016,** *18* (28), 18905-13.

(35) Heine, N.; Kratz, E. G.; Bergmann, R.; Schofield, D. P.; Asmis, K. R.; Jordan, K. D.; McCoy, A. B. Vibrational spectroscopy of the water-nitrate complex in the O-H stretching region. *J Phys Chem A* **2014,** *118* (37), 8188-97.

(36) Myshakin, E. M.; Jordan, K. D.; Sibert, E. L.; Johnson, M. A. Large anharmonic effects in the infrared spectra of the symmetrical CH3NO2−·(H2O) and CH3CO2−·(H2O) complexes. *The Journal of Chemical Physics* **2003,** *119* (19), 10138-10145.

(37) Robertson, W. H.; Price, E. A.; Weber, J. M.; Shin, J.-W.; Weddle, G. H.; Johnson, M. A. Infrared Signatures of a Water Molecule Attached to Triatomic Domains of Molecular Anions: Evolution of the H-bonding Configuration with Domain Length. *The Journal of Physical Chemistry A* **2003,** *107* (34), 6527-6532.

(38) Zabuga, A. V.; Kamrath, M. Z.; Rizzo, T. R. Franck-Condon-like Progressions in Infrared Spectra of Biological Molecules. *J Phys Chem A* **2015,** *119* (42), 10494-501.

(39) Tabor, D. P.; Kusaka, R.; Walsh, P. S.; Sibert, E. L., 3rd; Zwier, T. S. Isomer-Specific Spectroscopy of Benzene-(H2O)n, n = 6,7: Benzene's Role in Reshaping Water's Three-Dimensional Networks. *J Phys Chem Lett* **2015,** *6* (10), 1989-95.

(40) Sherman, S. L.; Nickson, K. A.; Garand, E. Comment on "Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy". *J Phys Chem Lett* **2022,** *13* (8), 2046-2050.

(41) Zhang, K.; Cassady, C. J.; Chung-Phillips, A. Ab Initio Studies of Neutral and Protonated Triglycines: Comparison of Calculated and Experimental Gas-Phase Basicity. *Journal of the American Chemical Society* **1994,** *116* (25), 11512-11521.

(42) Zhang, K.; Zimmerman, D. M.; Chung-Phillips, A.; Cassady, C. J. Experimental and ab initio studies of the gas-phase basicities of polyglycines. *Journal of the American Chemical Society* **1993,** *115* (23), 10812-10822.

(43) Carr, S. R.; Cassady, C. J. Gas-phase basicities of histidine and lysine and their selected di- and tripeptides. *Journal of the American Society for Mass Spectrometry* **1996,** *7* (12), 1203-1210.

(44) Wu, Z.; Fenselau, C. Structural Determinants of Gas Phase Basicities of Peptides. *Tetrahedron* **1993,** *49* (41), 9197-9206.

(45) Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. Gas-phase deprotonation of the peptide backbone for tripeptides and their methyl esters with hydrogen and methyl side chains. *J Phys Chem B* **2012,** *116* (51), 14844-58.

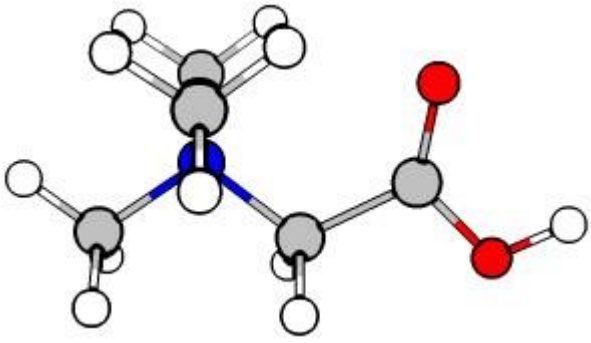
CHAPTER 6

Future Directions: Two-Color Approach in the Microsolvation of Betaine

6.1 Introduction

The experiments presented in **Chapters 3-5** showcase how cryogenic ion vibrational spectroscopy is an exceptional tool for studying amino acids, small peptides, and their solvation in the instrument. Studying small model systems helps identify how fundamental properties change upon solvation, such as intramolecular H-bonds,[1-2] thus enabling scientists to understand how protein folding might change depending on which amino acid is being used and its placement within the system. However, other small systems (and their solvated networks) may be studied, that are important on their own, such as betaine, also called trimethylglycine, shown in **Figure 6.1**.

Figure 6.1. Betaine, also known as trimethylglycine.



As betaine is synthesized as a salt with a counterion, it features the same framework as glycine but the amine group is trimethylated leading to a permanently positive shielded charge when unsolvated rather than a neutral species. Betaine is a substance produced in the body and used in the metabolization of homocysteine.[3-8] Betaine is an important regulator of this amino acid since high levels of homocysteine are often associated with heart disease and stroke.[9-11] Since approximately 70% of the human body (fat-free mass) is water,[12] it is an appropriate assumption

to believe that betaine is solvated in the human body. Solvation of amino acids leads to the formation of a zwitterion by losing the hydrogen on the carboxylic acid to the water around it and betaine can follow this trend. Since this compound is important within the human body, it is necessary to understand how the structure of betaine changes when solvated and the solvation network around it. In the previous chapters as well as past experiments done within the group, all experiments of amino acids[1-2, 13-14] and tripeptides[15-16] (also see **Chapter 4**) have been in their protonated on the N-terminus form rather than their naturally occurring form which is neutrally charged. Since the charge is shielded, this molecule would give a better picture of preferential solvation rather than solvating the pronation site first.

Cyrogenic ion vibrational spectroscopy can be used to probe betaine and its solvated structures. Due to the delicate nature of hydrogen bonding, this technique can be modified to create a linear IR spectrum (two-color spectrum) which is sensitive and nonperturbative to the network, thus giving a clear picture of the solvated structures.

6.2 Experimental Details

All IRPD and IRMPD spectra presented here were obtained using our home-built dual cryogenic ion trap vibrational spectrometer described in detail previously (see **Chapter 2**).[17] Protonated ions were generated via electrospray ionization of an ~1 mM solution of betaine in methanol. Hexapole ion guides gently transferred the ions through a series of differentially pumped regions into a linear octupole ion reaction trap, held at 80 K by a liquid nitrogen cryostat. Thermalization of the ions was induced by an ~1 ms burst of helium buffer gas, which was seeded with H_2O, at the beginning of the trapping sequence. At these sufficiently cold temperatures, water clustering of the ions can occur and will vary depending on reaction trap pressure. The thermalized

and clustered ions were then gently transferred to a 3D quadrupole ion tagging main trap which was held at 10 K by a closed-cycle helium cryostat. Ions and their clusters were further thermalized by helium buffer gas seeded with 10% D_2. Due to the weakly bound nature of the D2-tag, only D2-tagged adducts can form on the smaller cluster sizes (e.g. clusters of 0-1 water molecules). These adducts and clusters were then extracted into the TOF mass spectrometer, mass-selected by a mass gate, and intersected with the output of a 10 Hz Nd:YAG pumped OPO/OPA infrared probe laser. Resonant absorption of a single photon resulted in the loss of the D_2 tag for Betaine$(H_2O)_{0-1}$ and loss of a H_2O molecule for Betaine$(H_2O)_{2-6}$. For larger H_2O clusters utilizing water as a tag, a single resonant photon may dissociate the water when its energy exceeds the binding energy of the water molecule, however, this process typically requires multiple photons for lower energy features (e.g. features below 3100 cm^{-1}).

Usually as solvation increase, rather than the waters interacting with the peptides, they tend to interact with other water molecules. These types of interactions are delicate in nature and can be greatly perturbed by a photon. The photon can interact with the network and change the structure of it. Thus once, the second photon comes in to be absorbed at that same wavelength, since the structure has changed, the second photon does not interact as it should. As water-water interactions begin in betaine clusters with two or more water molecules, the hydrogen bonding interactions are intrinsically delicate, so a two-color process was used to elucidate the H-bonded peaks found below 3300 cm^{-1}. The IR-IR two-color approach was pioneered by Lee and co-workers[18] and was recently adapted by Yang *et. al.*[19] on a similar instrument to our own. On Mark Johnson's instrument,[20] both of the lasers are in the TOF region which allows clusters to already be mass separated and makes sure that clusters cannot be re-cooled after warming. However, on CIVS, the output of a second 10 Hz Nd:YAG pumped OPO/ OPA infrared pump laser is focused

directly into the 10 K tagging main trap. This does not allow the clusters to be mass selected since they are all in the same trap and there is a worry that the clusters will re-cool once hit with a resonant laser. The timing scheme of the laser in the main trap is optimized to minimize re-cooling and performs best when fired at the time of the ion extraction. This thus enables users on this instrument to be able to probe the warm clusters with the TOF later in the instrument to yield a linear two-color spectrum.

6.3 Computational Details

A systematic computation isomer structure search was done utilizing the molecular mechanics basin hopping program developed by Hopkins and coworkers[21-24] to find the lowest energy structures to compare to the experimental IRPD/IRMPD spectra. With increasing solvation, Betaine$(H_2O)_n$, the lowest energy structures from the prior solvation, Betaine$(H_2O)_{n-1}$, was utilized as the starting point for each geometry optimization with the basin hoping program. In the program the dihedral angles were iteratively stepped 7,000 times by a random value between ±5° and the water was iteratively rotated and translated by a random value between ±5° and 0.2 Å, respectively. After each iteration, the resulting structure was optimized utilizing the AMBER force field. After utilizing the program, each unique structure was subsequently optimized in the Gaussian 16[25] program with increasing accuracy of the methods. The final lowest energy structures were optimized at the cam-B3LYP/def2-TZVP/GD3BJ level and counterpoise corrected. Harmonic frequencies were also calculated. These same structures were optimized with MP2/def2-TZVP and the ZPE correction found through the corresponding DFT calculation is applied to the MP2 energetics. All energetics on conformers displayed in this chapter are shown in **Table 6.1**.

Table 6.1. Relative calculated energies (kJ/mol) of various conformers for Betaine(H_2O)$_{0-6}$ clusters at the cam-B3LYP/def2-TZVP/GD3BJ/BSSE (DFT) and MP2/def2TZVP with DFT ZPE (MP2) level.

	DFT	**MP2**
0-A	0.00	0.00
I-A	0.00	0.00
II-A	0.00	0.00
II-B	0.04	2.29
II-C	1.77	3.92
III-A	0.00	0.05
III-B	7.07	5.24
III-C	7.68	6.60
III-D	8.45	3.12
III-E	10.30	10.25
III-F	11.51	0.00
IV-A	0.00	0.00
IV-B	5.34	0.92
IV-C	6.04	2.53
IV-D	7.03	1.85
V-A	0.00	2.05
V-B	5.17	0.78
V-C	7.40	0.00
V-D	8.27	5.63
V-E	8.48	3.92
VI-A	0.00	0.00
VI-B	5.64	7.58
VI-C	6.14	4.30
VI-D	6.63	6.16
VI-E	8.30	7.73

6.4 Results and Discussion

The work presented here is the computational and experimental work performed on this system thus far. For Betaine$(H_2O)_{0-2}$, the work completed is sufficient to confidently assign the spectra. More work is needed for Betaine$(H_2O)_{3-6}$: particularly, IR-IR double resonance spectroscopy under the two-color framework. However, a second reaction trap would need to be implemented first to allow for mass selection of the cluster of interest to be able to do these experiments.

The IRPD/IRMPD spectra of Betaine$(H_2O)_{1-9}$, taken by Brett Marsh (Ph.D. 2015), is shown in **Figure 6.2**. At the time the original data was taken, the instrument had only the TOF laser and no main trap laser. No computations were done by Brett or by any graduate student of the Garand Group for this system prior. As seen in **Figure 6.2**, much of the data was saturated as noted by the similar peak intensity of the most intense peaks within the spectra. As a consequence of the data saturation from this work, much of the data for Betaine$(H_2O)_{1-6}$ was re-collected and the IRPD spectrum of Betaine$\cdot D_2$ was added (shown in **Figure 6.3** as black traces). Due to the extensive hydrogen bonding network of this species, an IR-IR two color approach was used for Betaine$(H_2O)_{2-4}$ and is shown as red traces in **Figure 6.3**. Notably, upon solvation with any amount of water, the preferential first solvation spot is at the carboxylic acid free O-H since no peak appears around the corresponding O-H stretch band position around ~3570 cm^{-1}. Unlike protonated glycine, there are no free N-H stretches to observe in the 3200-3400 cm^{-1} spectral region since betaine contains N-CH_3 bonds rather than N-H bonds. Most of the high intensity peaks are therefore due to stretches in the water network and the hydrogen bonded carboxylic acid.

The one-laser IRPD spectrum of Betaine$\cdot D_2$ is shown in **Figure 6.4** along with the spectrum of its lowest energy conformer. The spectrum in the mid-IR is simple and contains two

Figure 6.2. Overview of the IRPD spectra Betaine$(H_2O)_{1-8}$ taken by Brett Marsh (Ph.D. 2015) in the 2400-3800 cm^{-1} range.

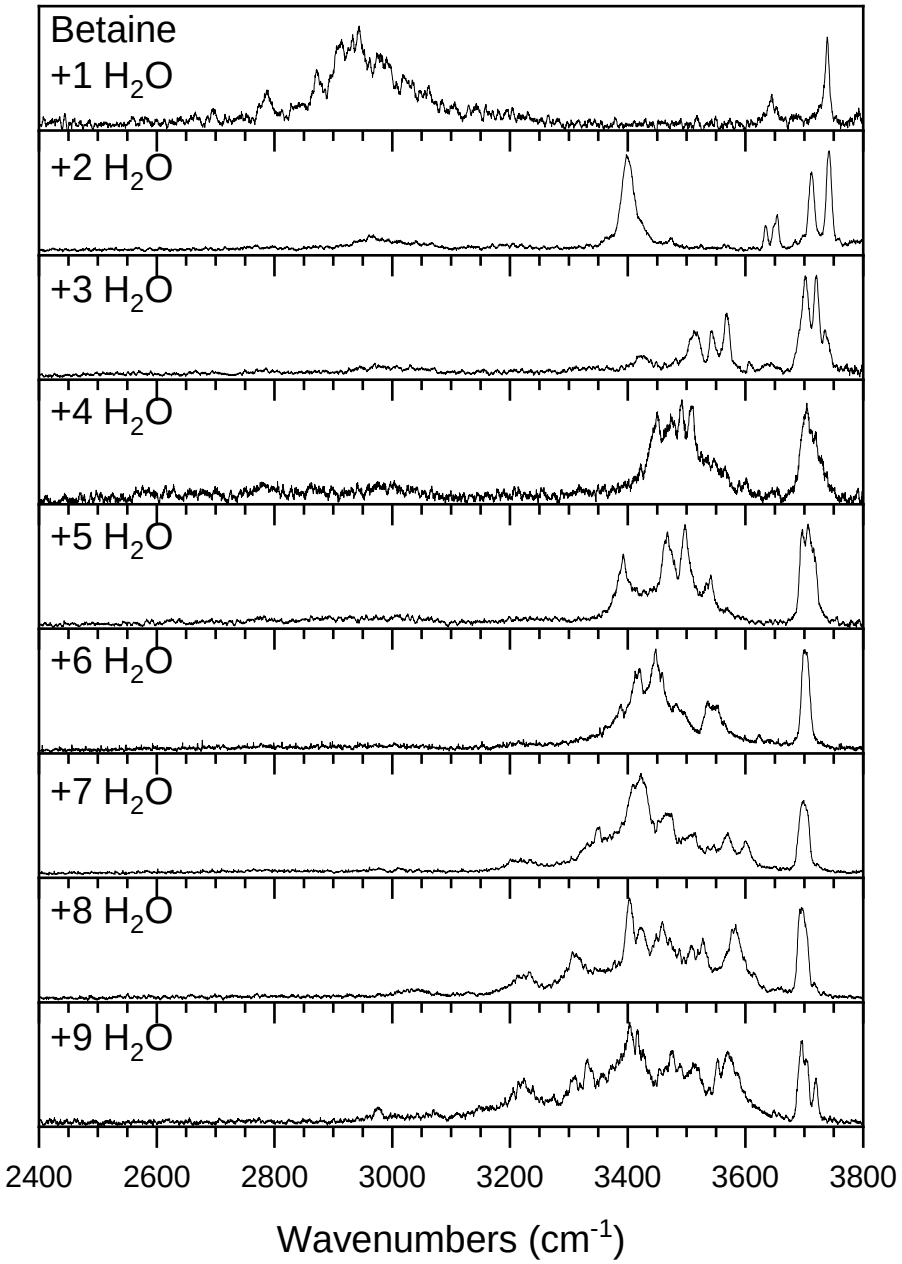
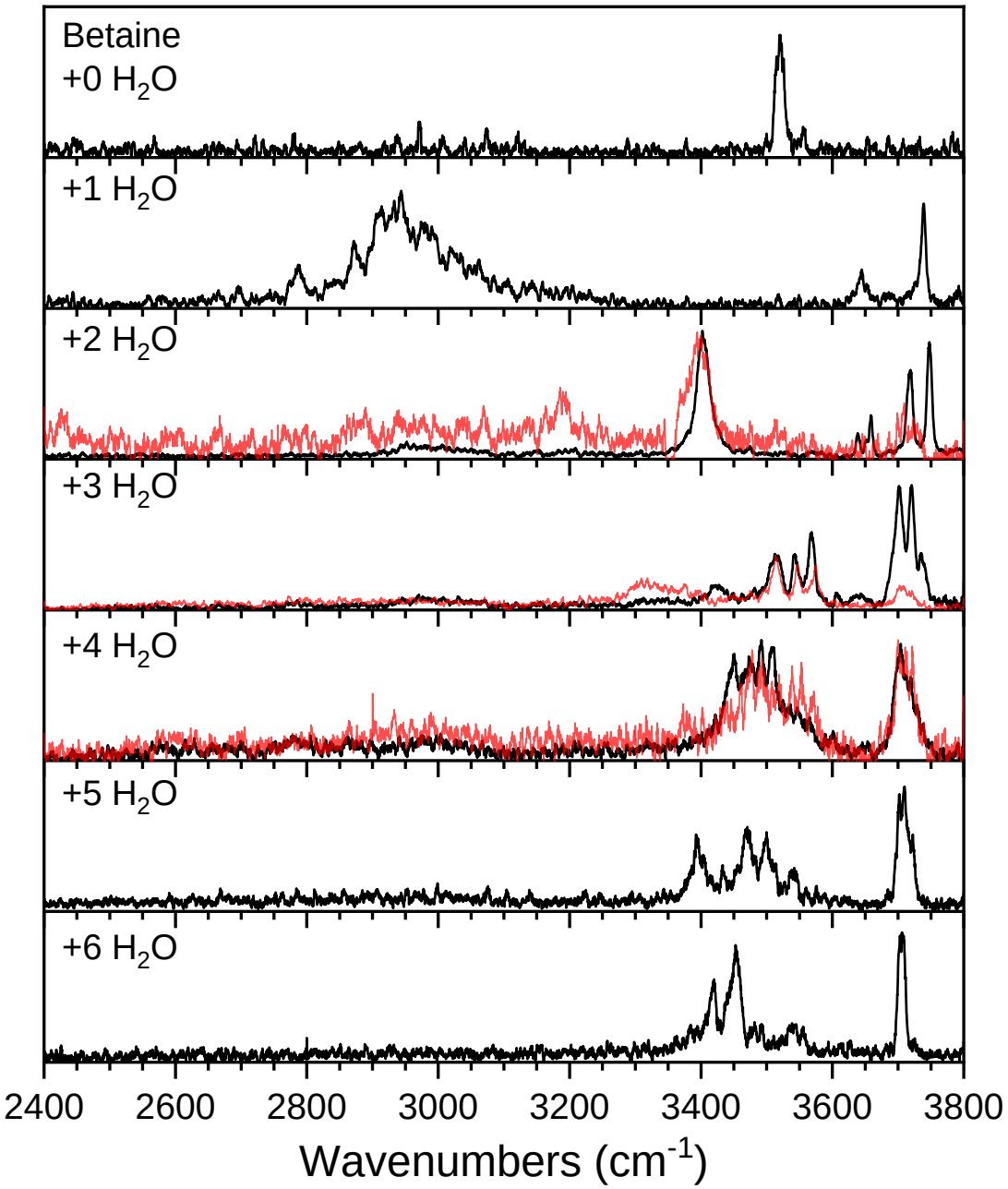


Figure 6.3. (Black Trace) Updated IRPD spectra of Betaine$(H_2O)_{0-6}$ in the 2400-3800 cm^{-1} range. (Red Trace) Linear two-color IRPD spectra of Betaine$(H_2O)_{2-4}$ in the 2400-3800 cm^{-1} range.



carboxylic acid free O-H peaks featured between 3500 and 3580 cm^{-1}, and the weak C-H stretches between 2900 and 3100 cm^{-1}. There are two carboxylic acid free O-H peaks in this spectrum: one weak peak corresponding to a completely unperturbed O-H stretch around 3556 cm^{-1}, and the more intense peak at 3520 cm^{-1} that is red shifted due to the D_2-tag binding to the O-H. The population of this species exists nearly exclusively as the D_2-tag binding to the O-H with a minimal proportion of the population having the tag bind anywhere else on the molecule. Utilizing the peak areas for a rough estimate of population percentages, we see ~90:10 for the D_2-tag binding to the free O-H vs the D_2-tag binding elsewhere. This contrasts to previous studies within the Garand group of amino acids and small peptides which suggest a small perturbation of the D_2-tag on the free O-H, especially on the bare molecule since there is preferential binding to the charged -NH_3 group.[16, 26] Therefore, though the methyl groups on the amine are small, they display a large shielding effect on the charge which results in preferential tagging on the carboxylic acid. Because of this, upon solvation water will bind preferentially to the carboxylic acid free O-H.

The one-laser IRPD spectrum of Betaine(H_2O)·D_2, shown in **Figure 6.5**, features two sharp peaks at 3645 and 3738 cm^{-1} corresponding to the free water symmetric and asymmetric stretch respectively, and a broad feature at 2941 cm^{-1}, corresponding to a hydrogen bonded carboxylic acid O-H stretch. There is only one conformer that this corresponds to, the lowest energy structure shown in **Figure 6.5**, I-A, which is verified by the fact that there is no free carboxylic acid O-H stretch around 3550 cm^{-1} and there is no visible perturbation on the O-H from the D_2-tag, as it must be weakly bound to the shielded quaternary amine.

Upon further solvation, thus making Betaine(H_2O)$_2$, it was appropriate to use chemical intuition to determine where the second water would go. The second water could either hydrogen bond directly to the first water and have both its O-H bonds free, it may hydrogen bond directly to

Figure 6.4. IRPD spectra of Betaine compared to lowest energy harmonic spectra calculated without D_2-tag as well as D_2-tag binding at three different positions around betaine at the cam-B3LYP/def2-TZVP/GD3BJ level.

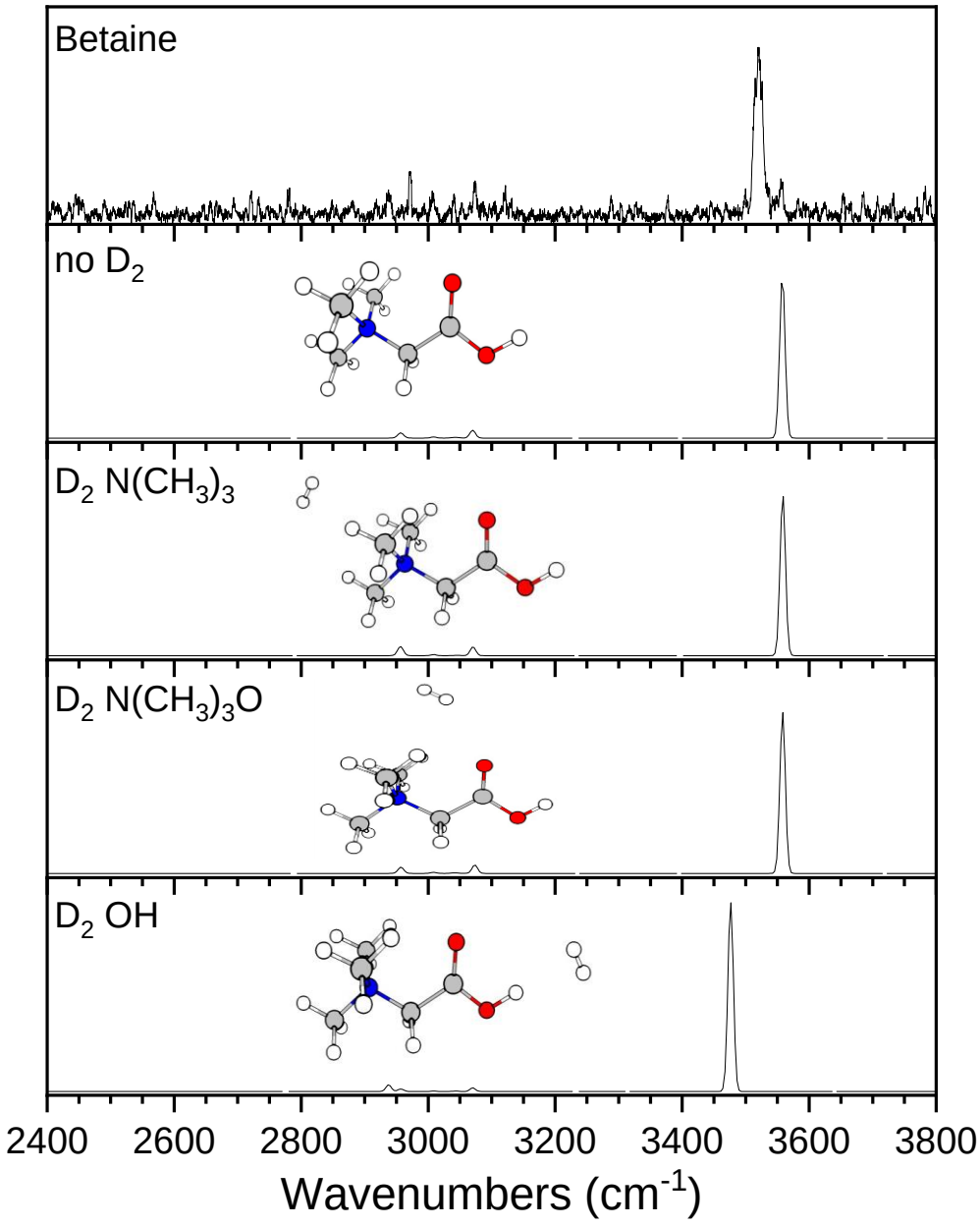
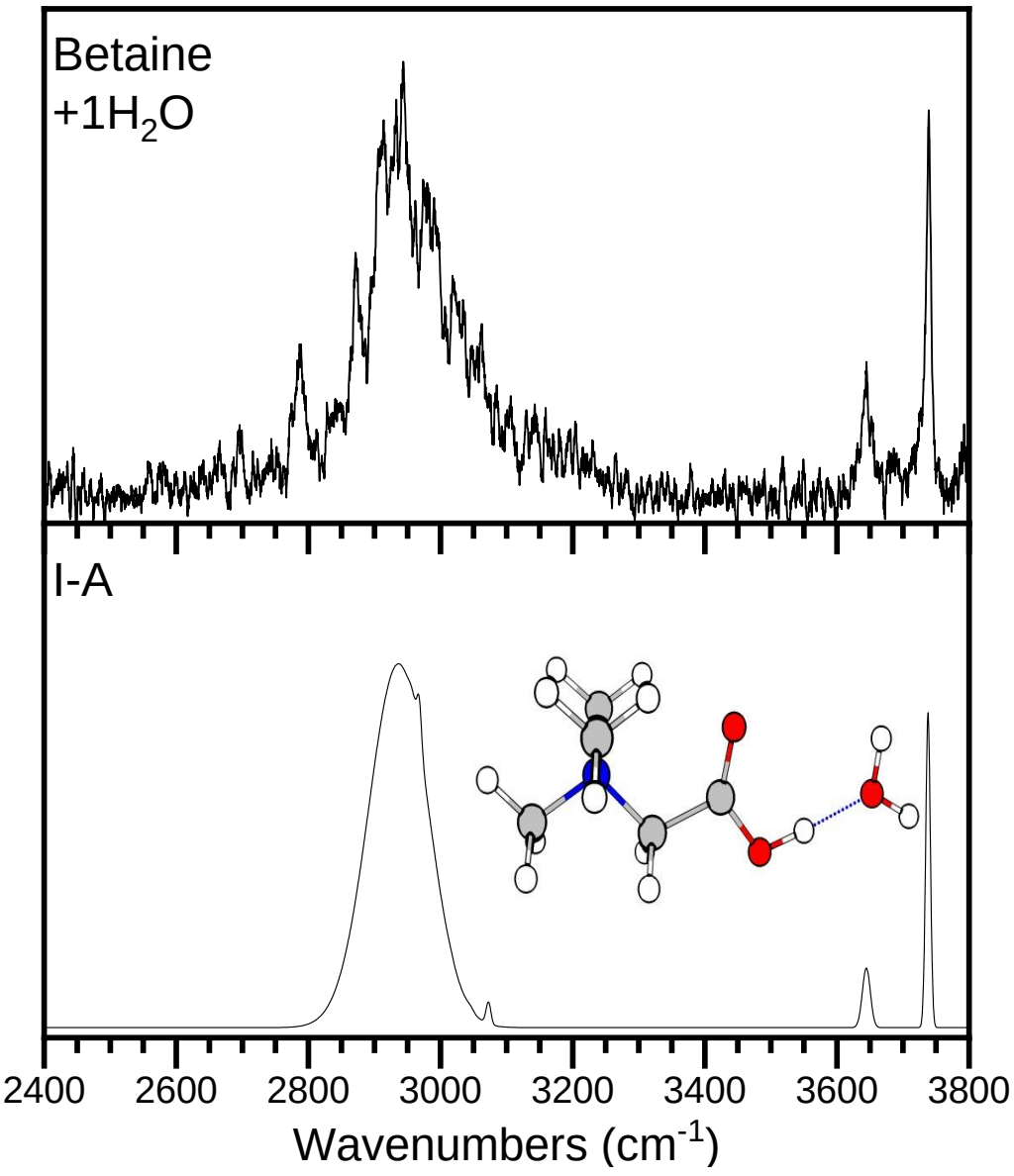


Figure 6.5. IRPD spectra of Betaine(H_2O) compared to lowest energy harmonic spectrum calculated at the cam-B3LYP/def2-TZVP/GD3BJ level.



the first water and carboxylic acid C=O bond to form a four-member ring, or it may weakly bind to the shielded quaternary amine. The basin hopping program was used to validate this and make sure that there were not additional conformers that were lower in energy. After retaking Brett's original data utilizing H_2O as the tag (IRMPD in **Figure 6.6** and **6.7**), it is possible to see that there are some very weak features below 3200 cm^{-1}. In IRMPD, the signal strength is correlated to the $(power)^2$ of the laser as well as the binding energy of the water. Therefore, as the laser moves to lower power, which means lower energy photons, there is less signal. Also, the features found below 3200 cm^{-1} usually correspond to hydrogen bonded features which can be perturbed by photons. Therefore, a two-color approach was also utilized to help visualize the weak hydrogen bonded features that may be missing in the IRMPD approach. This approach will be discussed in more detail when discussing Betaine$(H_2O)_3$ since it was the first system that the two-color approach was used for and was optimized on.

The cold IRMPD spectrum, warm IRMPD spectrum, and two-color spectrum for Betaine$(H_2O)_2$ are shown in **Figure 6.6**. Briefly, the main trap pump laser was set to 3403 cm^{-1} which is the largest feature in the cold IRMPD spectrum. The power of the pump laser was lowered to reduce depletion of population and to encourage warming of the clusters, and the probe TOF laser was scanned. Looking at the resulting warm spectrum in **Figure 6.6** (middle panel), the features above 3200 cm^{-1} are nearly identical to the cold spectrum which is not surprising because the binding energy of the water is less than the wavelength of the laser in those regions (binding energy is estimated to be ~3300 cm^{-1}), hence it only requires one photon to detach the water regardless of temperature. However, there is one small feature present at 3032 cm^{-1} which is also weakly present in the cold spectrum, though not in the same shape (it is extremely blobby in the cold spectrum). This peak was then utilized by setting the TOF probe laser to 3032 cm^{-1} and

Figure 6.6. (Top panel) Cold IRPD spectrum of Betaine$(H_2O)_2$. (Middle panel) Warm IRPD spectrum of Betaine$(H_2O)_2$ obtained by fixing the main trap pump laser to 3403 cm^{-1}. (Bottom panel) Resulting linear two-color spectrum obtained by fixing the TOF probe laser to 3032 cm^{-1} and scanning the main trap pump laser.

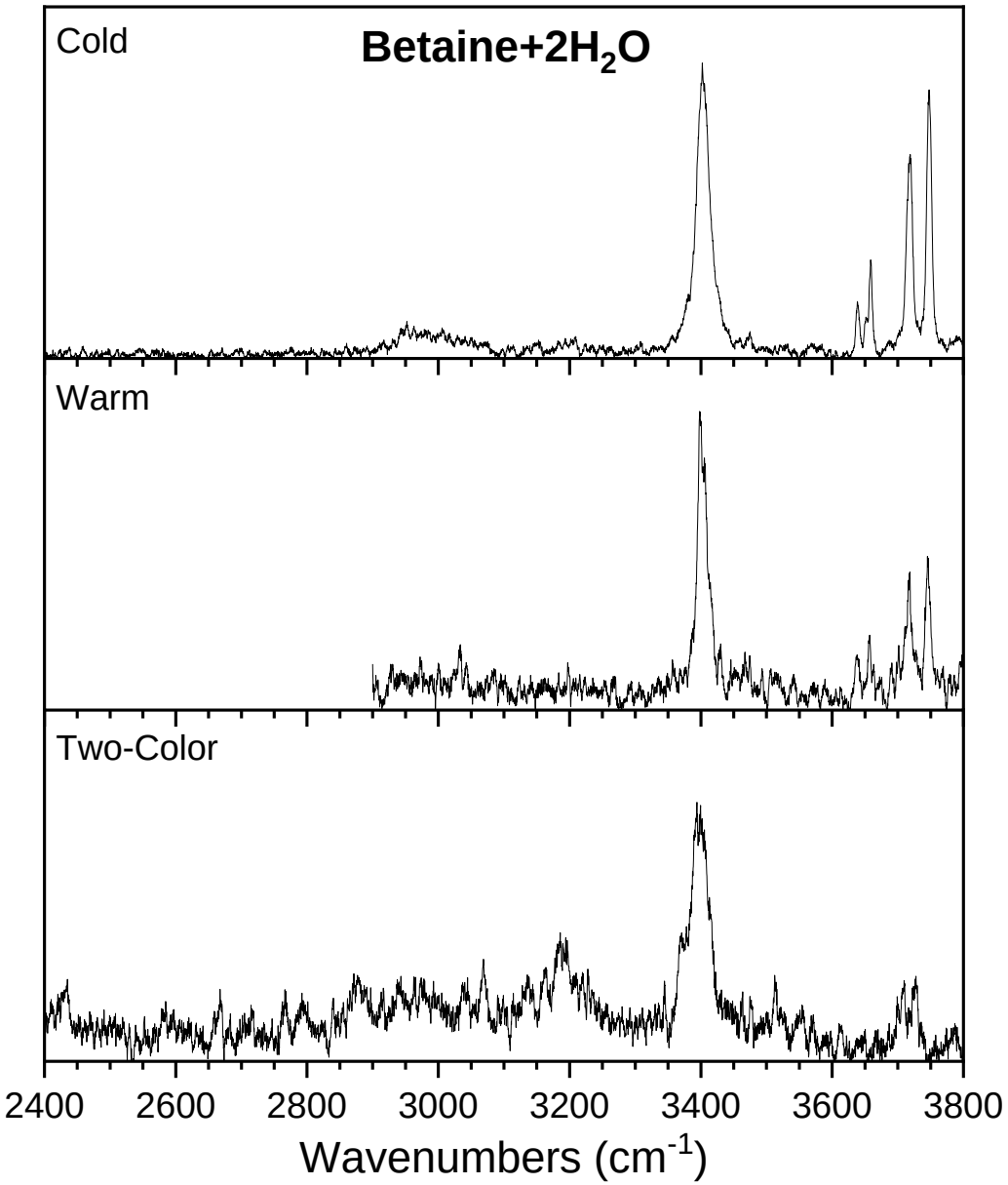
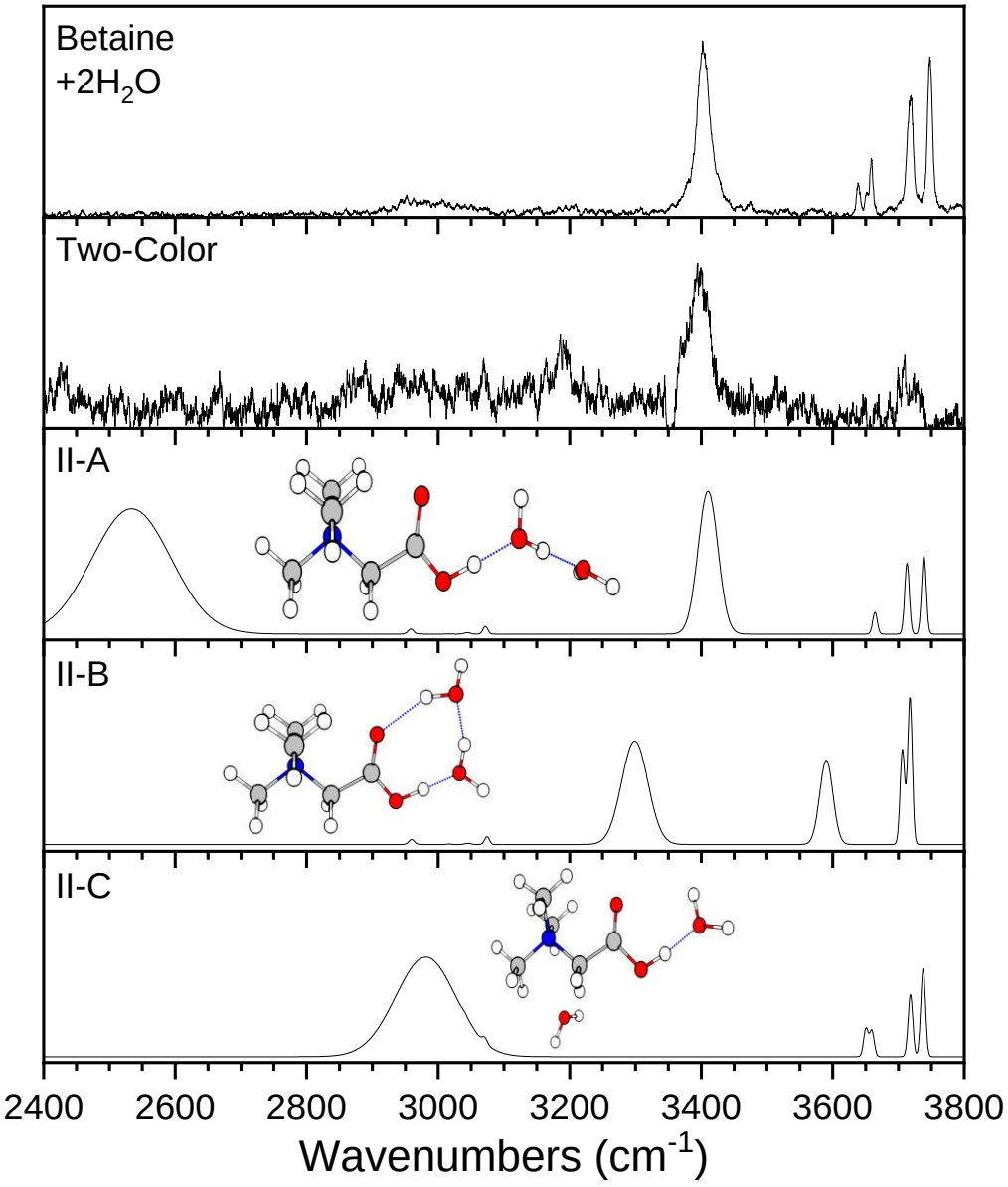


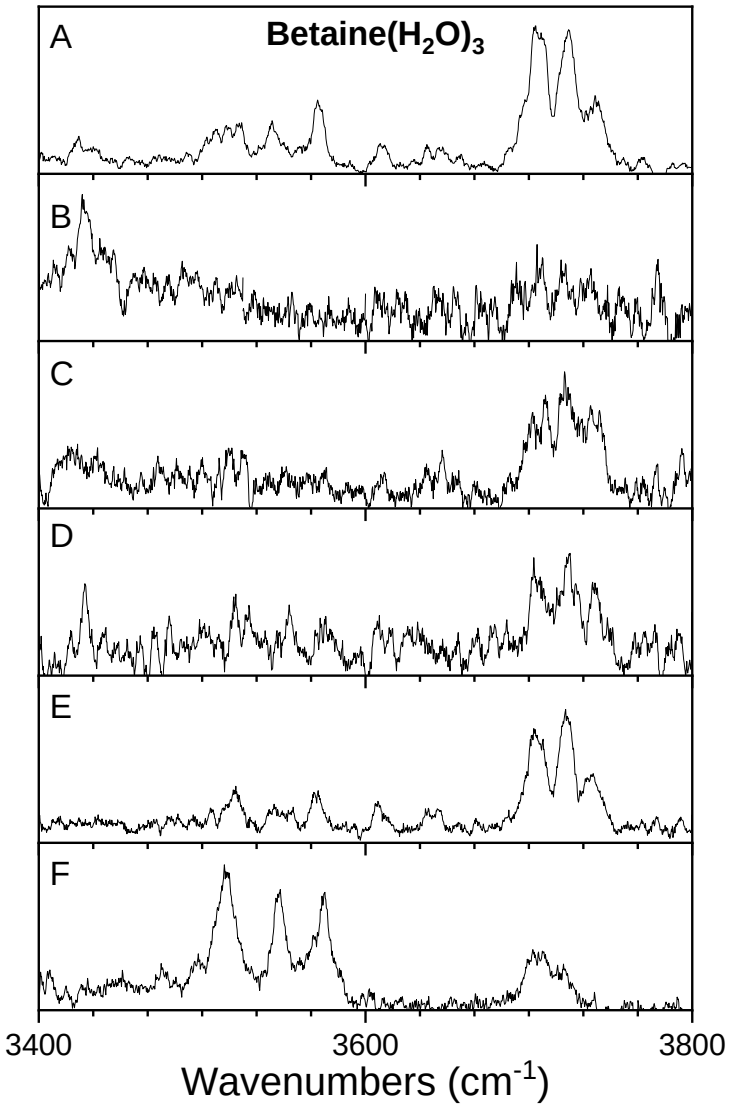
Figure 6.7. IRMPD spectrum and linear two-color spectrum of Betaine$(H_2O)_2$ compared to the harmonically calculated spectra of the three lowest energy conformers at the cam-B3LYP/def2-TZVP/GD3BJ level.



scanning the main trap laser. The resulting spectrum is the two-color linear spectrum shown in **Figure 6.6** and **6.7**. It is linear since the peak intensity is not correlated to the square power of the laser and the spectrum should be free of all artifacts from the messenger tag and photon interaction. The lowest energy conformers are also shown in **Figure 6.7.** The four features in the cold IRMPD spectrum found at 3640, 3660, 3716 and 3750 cm^{-1} have red shifted slightly in the two-color spectrum and decreased in intensity and features below 3200 cm^{-1} have gained intensity. The intensity change is not surprising since the signal being obtained is not related to the $(power)^2$ of the laser. From the cold spectrum alone, the features seem to agree well with conformer II-A, where the second water hydrogen is bound directly to the first water with both its O-H bonds free, and II-C, where the second water is bound to the shielded quaternary amine. However, with details from the two-color spectrum (specifically with the feature at 3149 cm^{-1}), it seems that conformer II-B, where the second water is hydrogen bonded directly to the first water and carboxylic acid C=O bond to form a four-member ring, cannot be discredited and may account for the feature found at 3149 cm^{-1} which traditionally can be described by a hydrogen bound water O-H which would be more susceptible to perturbation with IRMPD.

With the addition of another water, the IRMPD cold spectrum of Betaine$(H_2O)_3$ is shown in **Figure 6.8A** and **6.9**. Initially, the three-water system was used to determine if the two-color method was viable on the current instrumentation. Following a similar methodology as Yang *et. al.,*[19] a warm spectrum of the Betaine$(H_2O)_3$ was taken at a temperature of 150 K in the main trap and is shown in **Figure 6.8B**. As expected, many of the features that were once strong in the cold spectrum had lost intensity upon warming. However, there is a feature at 3428 cm^{-1} that became intense in the warm spectrum, which is just above the noise in the cold spectrum. Seeing as this peak is intense in the warm spectrum and weak in the cold spectrum, it was an ideal choice for the

Figure 6.8. (A) Cold IRMPD spectrum of Betaine$(H_2O)_3$. (B) Warm IRMPD of Betaine$(H_2O)_3$ obtained by increasing main trap temperature to 150 K. (C) Semi-warm IRMPD spectrum of Betaine$(H_2O)_3$ obtained by fixing the main trap pump laser to 3720 cm^{-1} and having the main trap laser fire 470 μs before ion extraction. (D) Warm IRMPD spectrum of Betaine$(H_2O)_3$ obtained by fixing the main trap pump laser to 3720 cm^{-1} and having the main trap laser fire at the same time as ion extraction. (E) Cold IRMPD of Betaine$(H_2O)_3$ obtained by fixing the main trap pump laser to 3720 cm^{-1} and having the main trap laser fire 100 μs after ion extraction. (F) Resulting two-color linear spectrum obtained by fixing the TOF probe laser to 3426 cm^{-1} and scanning the main trap pump laser under the timing conditions of spectrum (D).

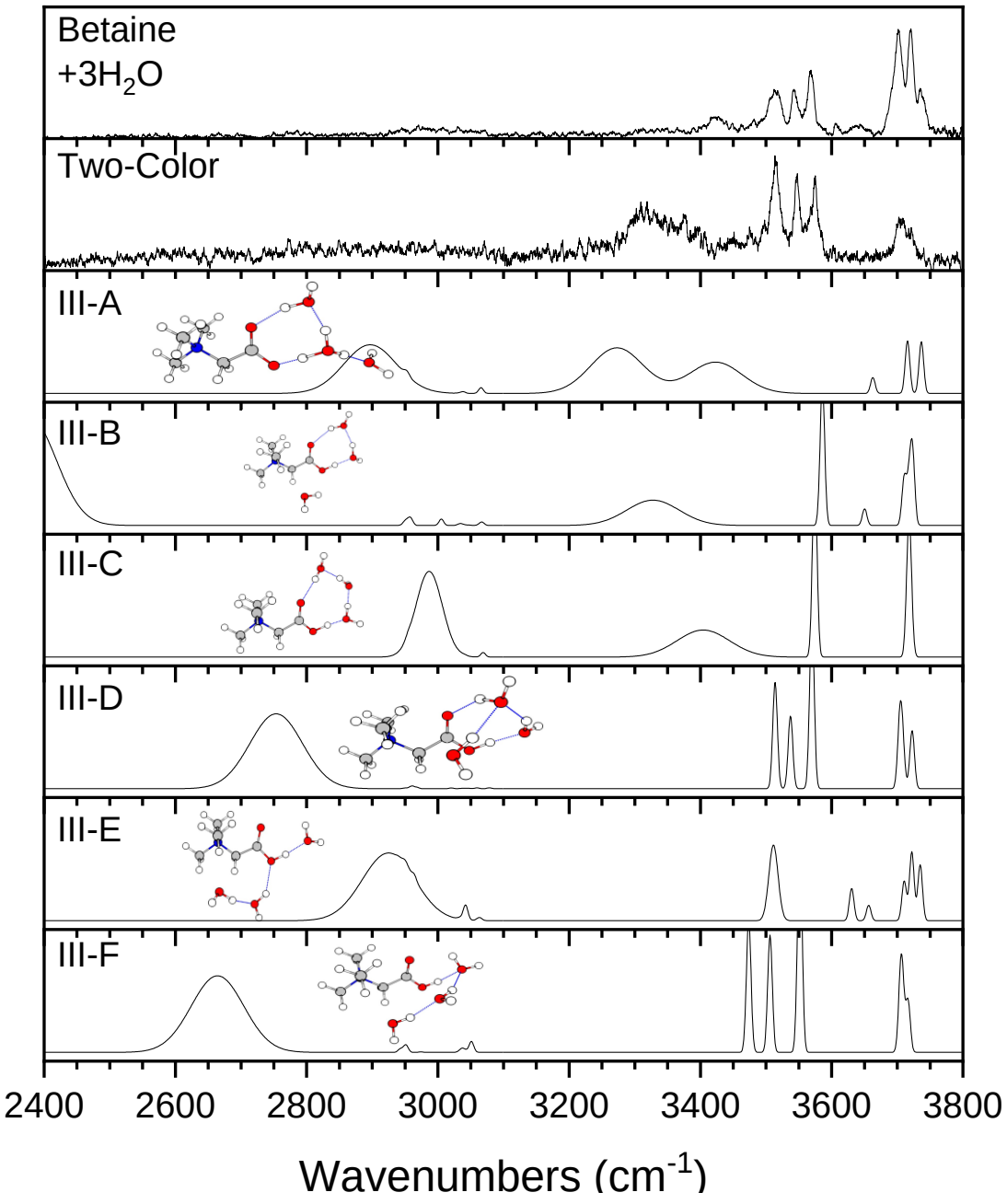


two-color process. Therefore, the main trap pump laser was fixed to 3720 cm^{-1} and the laser power was turned down to minimize dissociation of the water in the main trap and encourage warming of the ions, and the probe laser was scanned. The timings of the main trap had to be optimized to make sure that the warm ions did not re-cool before extraction. At the original timings (see Chapter 2), the laser would fire ~470 μs before ion extraction out of the trap. The resulting spectrum with this timing scheme is shown in **Figure 6.8C**. The features above ~3700 cm^{-1} have a different shape and intensity than those in **Figure 6.8A** in which the ions are thermalized to 10 K. However, a clear definitive feature at 3428 cm^{-1} is also not seen. This implies that the laser might be warming up the ions yet the 470 μs until extraction is allowing these ions to re-cool. The timing between laser firing and ion extraction was then minimized to 0 μs and the resulting spectrum is shown in **Figure 6.8D**. All intense features in the cold spectrum (such as those between 3700 and 3760 cm^{-1}) had lost much intensity when warming the complex which was expected, yet the feature at 3428 cm^{-1} had increased upon warming. Since the spectrum in **Figure 6.8D** is similar to that of **Figure 6.8B** which is the complex at 150 K, we know that this laser timing scheme is warming the ions up similarly. To validate that this timing was optimized, the timing was changed to have the main trap pump laser fire 100 μs after extraction and this spectrum is shown in **Figure 6.8E**. Since not all of the ions would be evacuated from the trap instantaneously, an optimal timing to minimize the re-cooling of the complexes would be slightly after trap extraction. Comparing this to the original cold spectrum (A), it is noted that all peaks are identical and therefore the optimized timing for the two-color experiments are when the laser firing and ion extraction occur at the same time. Though the spectra taken from the 150 K main trap and the main trap laser set on resonance with a 0 μs delay between laser firing and trap extraction are similar, it was necessary to be able to utilize the main trap laser for this so that 1) the laser set up was optimized for the two-color

experiments and 2) there might be additional "warm peak" bands that are intense at 150 K however are not as intense with the laser.

Utilizing the new timing and the peak position of the warm feature, the TOF probe laser was fixed to 3428 cm^{-1} and the main trap laser was scanned. The resulting spectrum is shown in **Figure 6.8F** and **6.9**. The calculated lowest energy conformers are shown in **Figure 6.9**. From the one laser cold spectrum, it is possible to see that there will be at least two conformers present. As there are three waters in this system, 7 spectral features are expected: 6 water O-H bands and one carboxylic acid O-H band. In the cold one laser spectrum there are at least 9 spectral features in the 3400 to 3800 cm^{-1} region. Upon initial investigation, conformer III-D and III-F seem to match the spectral features of the major conformer. However, since III-D is lower in energy and also seems to match more closely with the experimental spectrum due to having more accurate spacing of its features in the 3500 to 3600 cm^{-1} region as well as lining up with the experimental spectrum more accurately, it can be tentatively assigned as the major conformer. Yet this conformer does not account for the very small peaks found at 3423, 3608, and 3640 cm^{-1} or the other large features at 3700 and 3735 cm^{-1}. The lowest energy DFT conformer, Conformer III-A, seemed to contain some of the missing peaks at 3423, 3640, 3700 and 3735 cm^{-1} and therefore cannot be ruled out. III-A also contains extra hydrogen bonded features which may have been very weak or perturbed with the IRMPD method and therefore the two-color experiment is necessary to determine whether or not this conformer was indeed there. Conformers III-B, III-C, and III-E also contain some of the peaks that were not accounted for with just III-D. With the additional information obtained by the two-color spectrum, it is possible to see that there is a large and broad feature at 3312 cm^{-1}. This feature implies that conformers III-A, III-B, and III-C cannot be definitely excluded. Though conformer III-E does not have this broad feature in its spectrum, it also cannot be disregarded since

Figure 6.9. IRMPD and linear two-color spectrum of Betaine$(H_2O)_3$ compared to the harmonically calculated spectra of the six lowest energy conformers at the cam-B3LYP/def2-TZVP/GD3BJ level.



its extra feature at 3512 cm^{-1} overlaps with the experimental spectrum as well as the calculated spectrum for III-D. IR-IR double resonance is necessary to definitively assign this spectrum but cannot be done until a second reaction trap that is mass selective is installed.

The IRMPD cold spectrum of Betaine$(H_2O)_4$ is shown in **Figure 6.10** (top panel) and **6.11**, with its lowest energy conformers is shown in **Figure 6.11**. Similarly to the previous two systems, the warm spectrum of this cluster was obtained (**Figure 6.10**, middle panel) and its warm feature at 3076 cm^{-1} was used to obtain the two-color spectrum shown in **Figure 6.10** (bottom panel) and **6.11**. From the one laser cold spectrum, the assignment for one of the conformers seems straightforward, yet leaves the assignment for the second conformer ambiguous. The four features and their respective spacing found between 3400 and 3520 cm^{-1} led to a direct assignment of conformer IV-B, in which two waters form a form member ring with the carboxylic acid and are connected to the two other waters which are closer to the shielded protonation site. However, there is a shoulder off of these four main peaks found between 3520 and 3580 cm^{-1} in the cold spectrum. Without data from the two-color spectrum, conformers IV-A, IV-C, and IV-D cannot be discredited. Conformer IV-A has the lowest DFT energy and has the second lowest MP2 energy, 29 cm^{-1} higher than the lowest energy conformer, see **Table 6.1**. This leads to the original hypothesis that the second conformer might be IV-A. In this conformer, the waters hydrogen bond such that they make two four-member rings including the carboxylic acid. Due to the hydrogen bonding network, the proton detaches from the betaine, turning the amino acid into a neutral compound, creating a protonated water network. Because of the nature of this sensitive bonding structure, IRMPD can perturb the system such that it is not possible to see the peaks between 3000 and 3400 cm^{-1} since each incident photon absorbed can change the H-bonding framework. Therefore, the two-color spectrum may assist spectral assignment.

Figure 6.10. (Top panel) Cold IRMPD spectrum of Betaine$(H_2O)_4$. (Middle panel) Warm IRMPD spectrum of Betaine$(H_2O)_4$ obtained by fixing the main trap pump laser to 3450 cm^{-1}. (Bottom panel) Resulting linear two-color spectrum obtained by fixing the TOF probe laser to 3076 cm^{-1} and scanning the main trap pump laser.

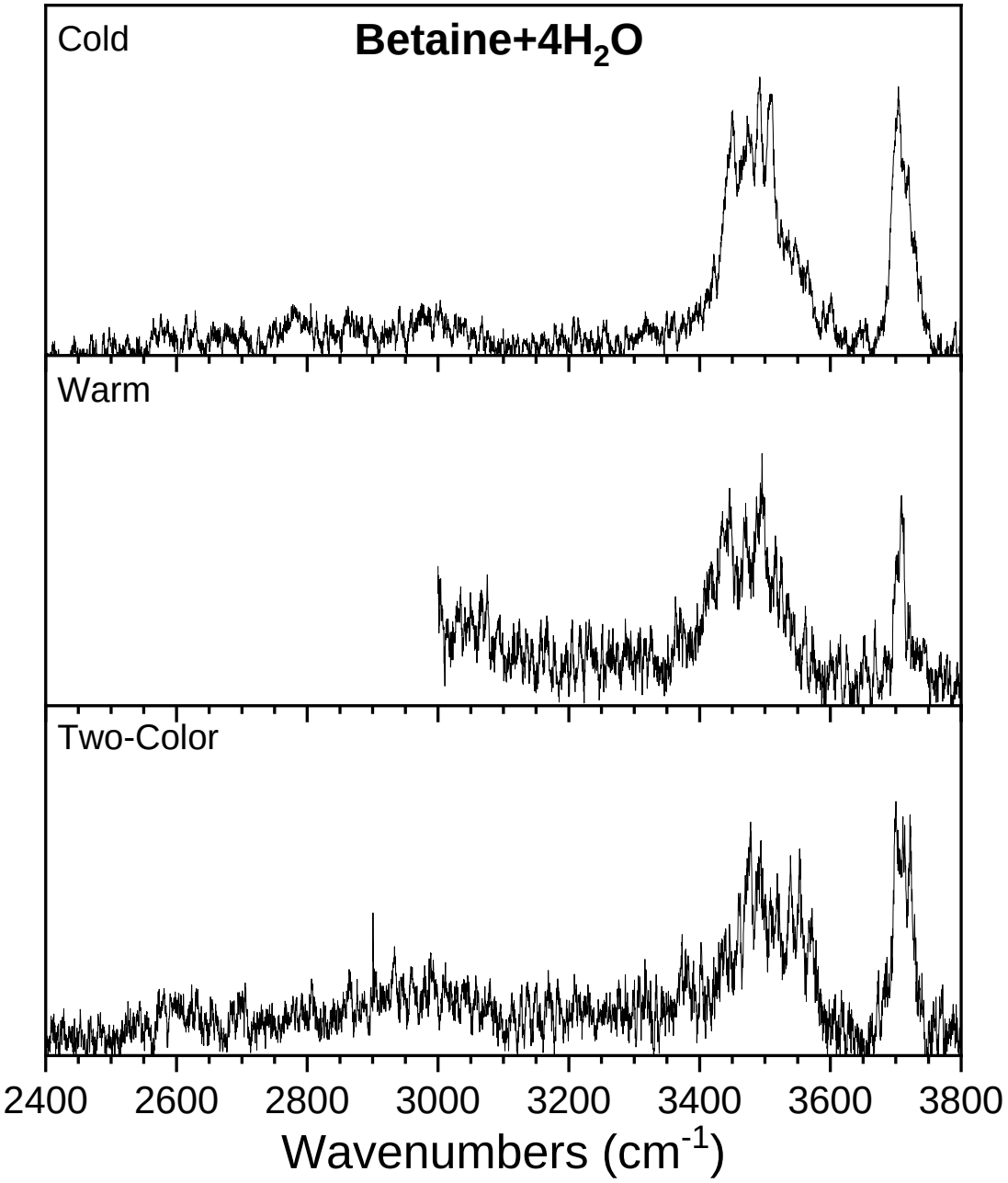
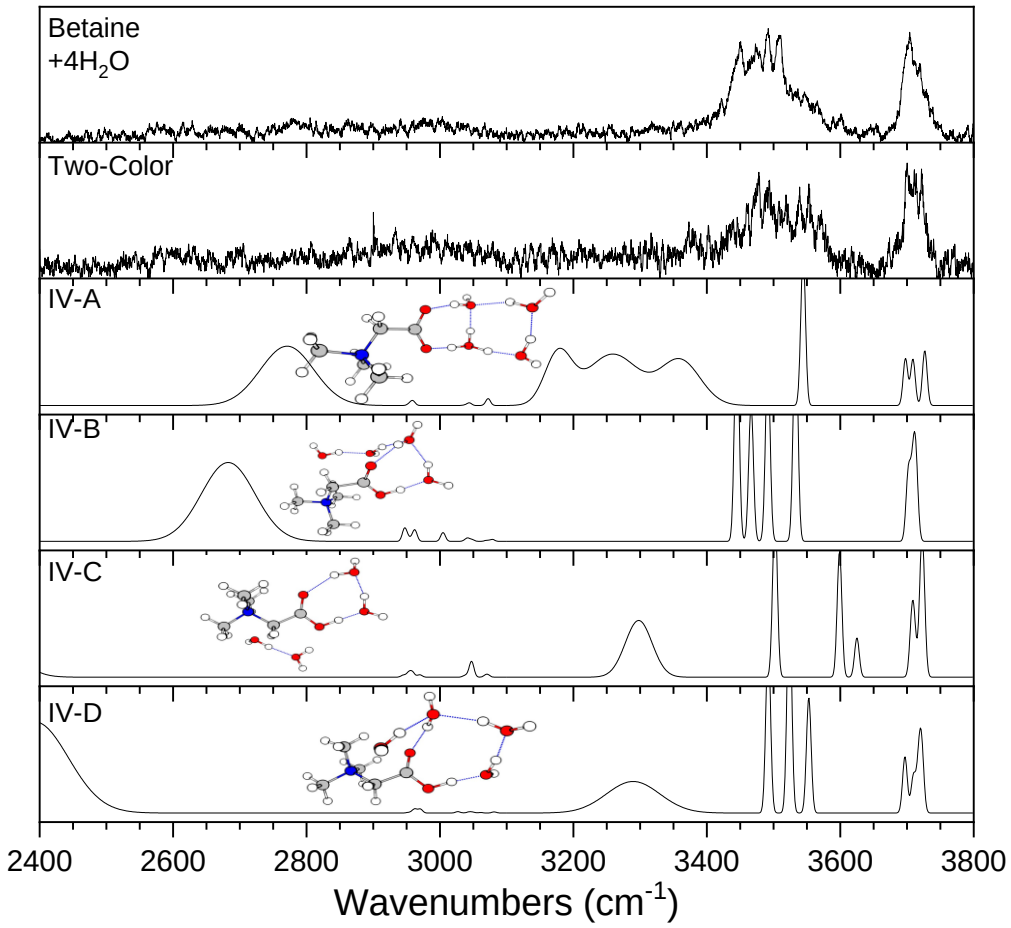


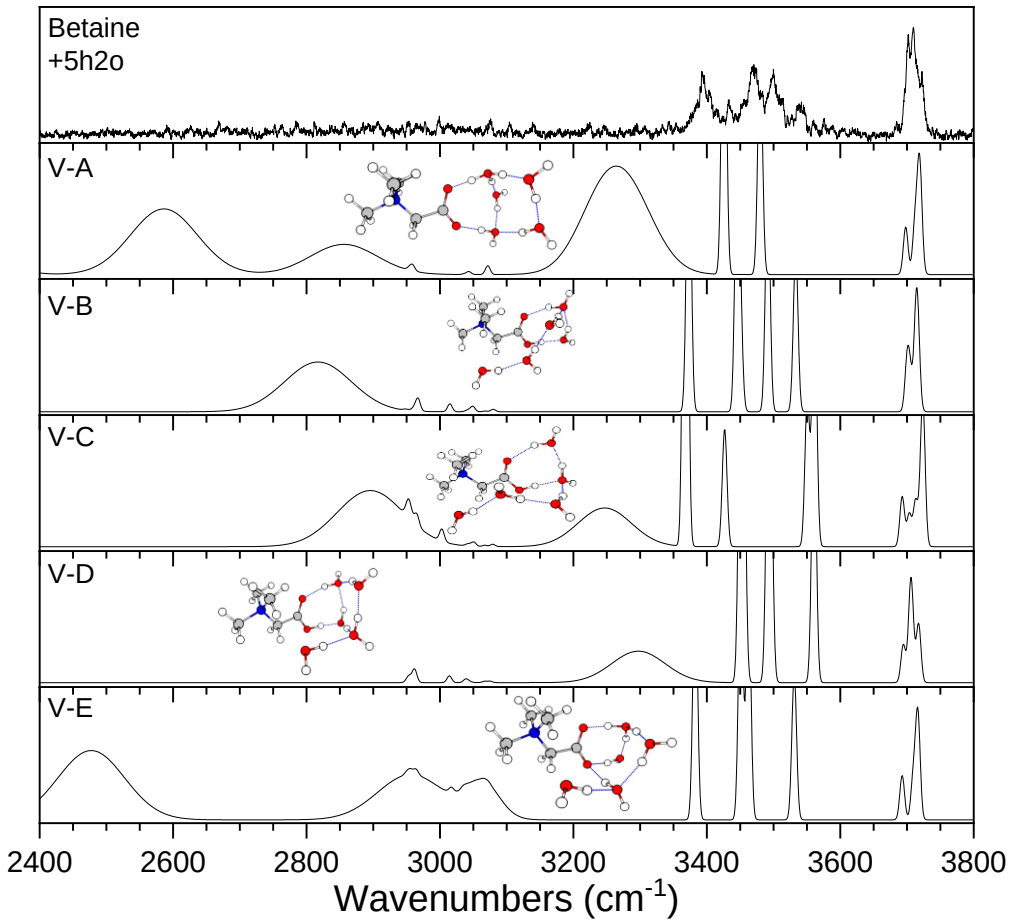
Figure 6.11. IRMPD spectrum and linear two-color spectrum of Betaine$(H_2O)_4$ compared to the harmonically calculated spectra of the four lowest energy conformers the cam-B3LYP/def2-TZVP/GD3BJ level.



Unlike in the previous two systems, there are no clear features below 3200 cm^{-1}. However, the two-color spectrum makes the features between 3522 and 3582 cm^{-1} much more defined and pronounced. There does not seem to be just one feature in this region, pointing to either more than one conformer or that IV-A is not the second conformer. Yet, since there are no obvious features between 3000 and 3400 cm^{-1}, this lends itself to the fact that IV-A is probably not the second conformer but without IR-IR double resonance experiments, it cannot confidently be ruled out completely. This leaves conformers IV-C and IV-D. For IV-C, the peaks at 3600 and 3625 cm^{-1} do not line up with the experimental spectrum perfectly and correspond to symmetric stretches of the water, which generally in the past experiments within the Garand Group[1-2, 15, 26] has been consistent with experimental spectra. Therefore conformer IV-C can be ruled out, leaving conformer IV-D. The spectrum of conformer IV-D consists of three peaks between 3492 and 3552 cm^{-1} which are consistent yet slightly red shifted from the three peaks revealed through the two-color experiment. This conformer also has a peak at 3290 cm^{-1} which may agree yet be redshifted with respect to the experimental feature at 3347 cm^{-1}. Without IR-IR double resonance experiments, IV-D can only be tentatively assigned as the second conformer, yet IR-IR double resonance experiments would be necessary for confidence and to make sure IV-A is not within the spectrum.

The IRMPD cold spectrum of Betaine$(H_2O)_5$ along with the predicted harmonic spectra of the lowest energy conformers is shown in **Figure 6.12**. The two-color spectrum could not be acquired due to complications with the instrument, see **Section 6.5**. Regardless of conformer assignment, it is important to note that all low energy conformers, minus V-E, contain a four-member ring between the carboxylic acid and two water molecules. The difference between conformers, V-B, V-C, V-D, and V-E are the placement of the other three water molecules. As

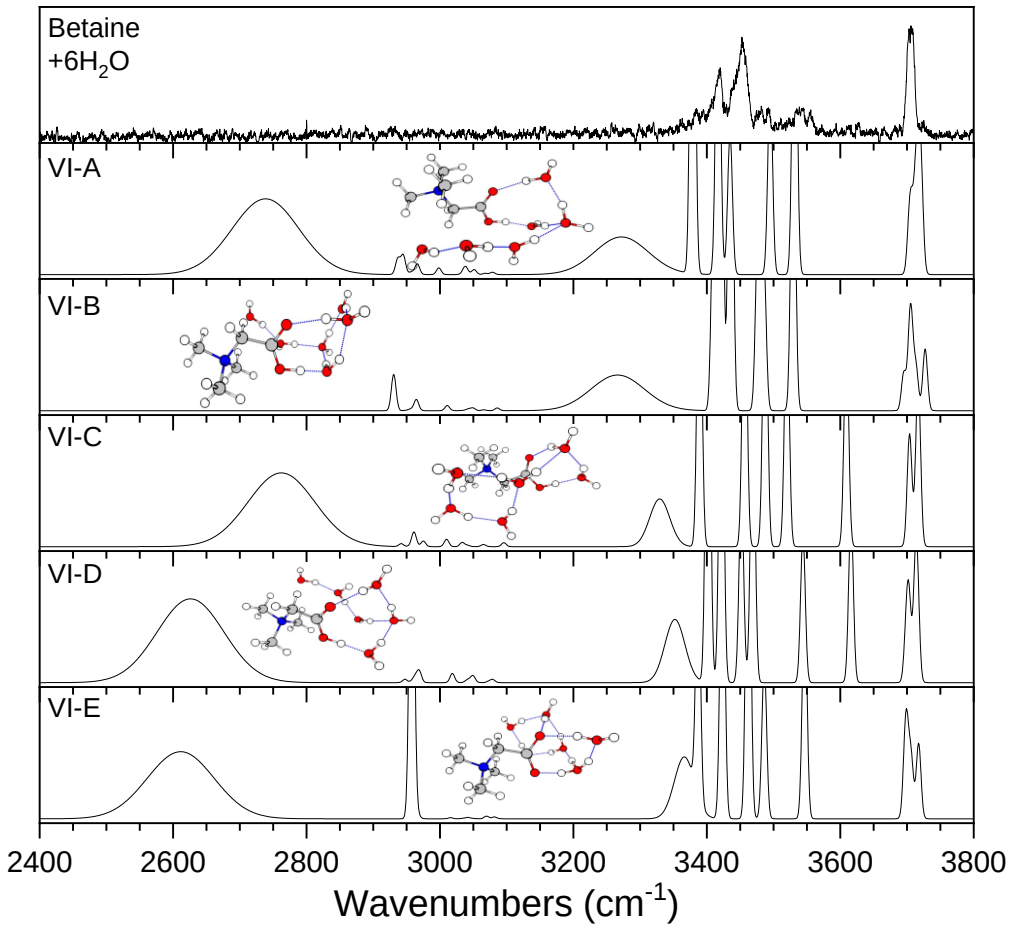
Figure 6.12. IRMPD spectrum of Betaine$(H_2O)_5$ compared to the harmonically calculated spectra of the five lowest energy conformers at the cam-B3LYP/def2-TZVP/GD3BJ level.



with Betaine$(H_2O)_4$, the one major conformer can be easily assigned to the second lowest energy conformer, V-B, due to the placement and spacing of the peaks between 3365 and 3558 cm^{-1}. There is one peak in this region that is not accounted for by V-B at 3434 cm^{-1} which means there must be a minor conformer. Upon first inspection, V-A, V-C and V-E have a peak at 3434 cm^{-1}, however V-C also seems to have a peak at 3563 cm^{-1}. It is ambiguous whether the second peak at 3563 cm^{-1} is in the experimental spectrum. V-E might be the more likely option since all other peaks computationally line up with conformer V-B's peaks leaving its only unique peak to be that at 3434 cm^{-1}. Though without IR-IR double resonance measurements, a definitive determination cannot be made. V-A cannot be discounted because all peaks above 3400 cm^{-1} either correlate with the one missing peak or overlap with an already determined peak. Further, the two-color experiment is needed to see if there are features below 3360 cm^{-1}. Looking at the final two low energy conformers, V-D and V-E, both conformers do not have any unique peaks from V-B which would make it hard to disentangle their spectral features. V-D has a similar peak compared to V-C at 3563 cm^{-1} which makes this conformer hard to confirm or deny within the spectrum, while V-E has no unique peaks. With IR-IR double resonance spectroscopy, it would be possible to definitively distinguish between V-B and V-C/V-E, yet V-E would be more difficult to disentangle meaning that it cannot be completely ruled out.

Finally, the IRMPD cold spectrum of Betaine$(H_2O)_6$ along with the lowest energy conformer computations is shown in **Figure 6.13**. Similarly to Betaine$(H_2O)_5$, the two-color spectrum could not be acquired due to complications with the instrument, see **Section 6.5**. Due to the sharpness of the experimental peak at 3706 cm^{-1}, it seems conceivable that there is only one conformer present; there are four apparent peaks between 3360 and 3580 cm^{-1}, with a couple of those peaks being a molding of two computational peaks due to their experimental broadness. With

Figure 6.13. IRMPD spectrum of Betaine$(H_2O)_6$ compared to the harmonically calculated spectra of the give lowest energy conformers at the cam-B3LYP/def2-TZVP/GD3BJ level.



the IRMPD cold spectrum alone, it does not seem to be possible to eliminate any of the proposed conformers since it was shown especially with Betaine$(H_2O)_4$ that the two-color experiments can elucidate peaks in the 3470-3640 cm^{-1} region as well as lower frequency regions. Just from the cold spectrum it seems possible to be able to eliminate conformers VI-C and VI-D due to their calculated peak at 3610 cm^{-1} which is not in the experimental spectrum, however this would be tentative pending investigation with the two-color technique.

6.5 Future Work

Further experimental data must be gathered to confidently assign and exclude conformers. Most importantly, the two-color spectrum of Betaine$(H_2O)_{5-6}$ as well as IR-IR double resonance spectrum of Betaine$(H_2O)_{2-6}$ must be acquired.

As stated in **Section 6.4**, when trying to obtain the two-color spectrum for Betaine$(H_2O)_{5-6}$, the instrument encountered some technical difficulties one day to the other. Initially it was thought to have been that the binding energy of the water for these clusters was too low and therefore these experiments would not have been possible. Computations were run to be able to affirm this suspicion, yet the calculations confirmed that the binding energy was similar to the previous cluster sizes and therefore the experiment should have work. However, it was discovered that there was no power coming out of the main trap pump laser due to one crystal being misaligned and therefore the laser required a full recalibration. In the future, once recalibration has occurred, obtaining the two-color spectrum for these two cluster sizes would be necessary.

For IR-IR double resonance experiments, an instrument modification is required. To be able to obtain IR-IR double resonance experiments on these water clusters utilizing the water as the tag, a second mass-selecting reaction is needed on the instrument. When solvated, all of the

peaks observed in the Betaine-water clusters spectra above 3300 cm^{-1} are due to the stretches of the water in the system. As solvation increases, the peaks become more undefined, "blobby", and blue shifted. Thus, a water stretching peak that is in Betaine$(H_2O)_4$ may also be in Betaine$(H_2O)_{5-6}$. So if the main trap laser is set to "burn" out a conformer, it would be burning the conformer in the four-water cluster, but also in all other conformers in different cluster-sizes which contains a water stretch with the same frequency. Therefore, the larger clusters may be losing one or two waters, artificially making more Betaine$(H_2O)_4$. When the TOF laser is then scanned, the resulting spectrum would be a collection of different species, cold Betaine$(H_2O)_4$ and warmer Betaine$(H_2O)_{5-6}$ which may have lost a water tag in the main trap stage. The same issue occurs when trying to obtain IR-IR ion-dip spectrum since some of the features might not be "unique" to just that one cluster size. Another reaction trap which is mass-selective (currently being developed in the Garand group), must be placed before the current reaction trap to isolate just the cluster size of interest. From there, these IR-IR double resonance experiments can be performed.

6.6 Conclusions

The IRPD/IRMPD spectra of Betaine$(H_2O)_{0-6}$ were retaken and these systems were used to determine whether two-color experiments can work on the current instrument set up. As this experiment is adapted from Mark Johnson's technique on a similar instrument,[19] they are able to do the two-color experiments with both lasers in the TOF (achieving separation and lack ability to re-cool), while on this instrument one laser is in the main trap and the other is in the TOF. Due to the lack of species separation in the trap as well as the possibility for re-cooling, it wasn't sure if this experiment would be able to occur. The two-color spectra for Betaine$(H_2O)_{2-4}$ were taken and show that these experiments are viable on the current instrumentation. It also opens the possibility

for novel experimentation for this instrument on systems that have very sensitive hydrogen bonding networks. Computations were run for Betaine$(H_2O)_{0-6}$ and tentative assignments were made for all species. In general, preferential solvation is on the carboxylic acid free O-H and water-water hydrogen bonding is favored over weakly binding to the shielded quaternary amine. In the future, IR-IR double resonance experiments would need to be performed to confirm the assignments of Betaine$(H_2O)_{3-6}$. In particular, these double resonance experiments would need to be occur under the two-color framework in order to determine whether the lowest energy structure with DFT for Betaine$(H_2O)_{3-4}$, exists in the spectrum. Validation of these conformers would help verify that solvation of Betaine may lead to a neutral amino acid and a protonated water network.

6.7 References

(1) Fischer, K. C.; Sherman, S. L.; Garand, E. Competition between Solvation and Intramolecular Hydrogen-Bonding in Microsolvated Protonated Glycine and beta-Alanine. *J Phys Chem A* **2020,** *124* (8), 1593-1602.

(2) Fischer, K. C.; Sherman, S. L.; Voss, J. M.; Zhou, J.; Garand, E. Microsolvation Structures of Protonated Glycine and l-Alanine. *J Phys Chem A* **2019,** *123* (15), 3355-3366.

(3) Alfthan, G.; Tapani, K.; Nissinen, K.; Saarela, J.; Aro, A. The effect of low doses of betaine on plasma homocysteine in healthy volunteers. *Br J Nutr* **2004,** *92* (4), 665-9.

(4) Atkinson, W.; Elmslie, J.; Lever, M.; Chambers, S. T.; George, P. M. Dietary and supplementary betaine: acute effects on plasma betaine and homocysteine concentrations under standard and postmethionine load conditions in healthy male subjects. *Am J Clin Nutr* **2008,** *87* (3), 577-85.

(5) Imbard, A.; Benoist, J. F.; Esse, R.; Gupta, S.; Lebon, S.; de Vriese, A. S.; de Baulny, H. O.; Kruger, W.; Schiff, M.; Blom, H. J. High homocysteine induces betaine depletion. *Biosci Rep* **2015,** *35* (4).

(6) Olthof, M. R.; van Vliet, T.; Boelsma, E.; Verhoef, P. Low dose betaine supplementation leads to immediate and long term lowering of plasma homocysteine in healthy men and women. *J Nutr* **2003,** *133* (12), 4135-8.

(7) Schwab, U.; Torronen, A.; Toppinen, L.; Alfthan, G.; Saarinen, M.; Aro, A.; Uusitupa, M. Betaine supplementation decreases plasma homocysteine concentrations but does not affect body weight, body composition, or resting energy expenditure in human subjects. *Am J Clin Nutr* **2002,** *76* (5), 961-7.

(8) Schwahn, B. C.; Wang, X. L.; Mikael, L. G.; Wu, Q.; Cohn, J.; Jiang, H.; Maclean, K. N.; Rozen, R. Betaine supplementation improves the atherogenic risk factor profile in a transgenic mouse model of hyperhomocysteinemia. *Atherosclerosis* **2007,** *195* (2), e100-7.

(9) Stampfer, M. J.; Malinow, M. R. Can lowering homocysteine levels reduce cardiovascular risk? *N Engl J Med* **1995,** *332* (5), 328-9.

(10) Song, Z.; Zhou, Z.; Deaciuc, I.; Chen, T.; McClain, C. J. Inhibition of adiponectin production by homocysteine: a potential mechanism for alcoholic liver disease. *Hepatology* **2008,** *47* (3), 867-79.

(11) Sarkar, P. K.; Lambert, L. A. Aetiology and treatment of hyperhomocysteinaemia causing ischaemic stroke. *Int J Clin Pract* **2001,** *55* (4), 262-8.

(12) Schoeller, D. A. Changes in total body water with age. *Am J Clin Nutr* **1989,** *50* (5 Suppl), 1176-81; discussion 1231-5.

(13) Voss, J. M.; Fischer, K. C.; Garand, E. Accessing the Vibrational Signatures of Amino Acid Ions Embedded in Water Clusters. *J Phys Chem Lett* **2018,** *9* (9), 2246-2250.

(14) Voss, J. M.; Kregel, S. J.; Fischer, K. C.; Garand, E. IR-IR Conformation Specific Spectroscopy of Na(+)(Glucose) Adducts. *J Am Soc Mass Spectrom* **2018,** *29* (1), 42-50.

(15) Fischer, K. C.; Voss, J. M.; Zhou, J.; Garand, E. Probing Solvation-Induced Structural Changes in Conformationally Flexible Peptides: IR Spectroscopy of Gly3H(+).(H2O). *J Phys Chem A* **2018,** *122* (41), 8213-8221.

(16) Voss, J. M.; Fischer, K. C.; Garand, E. Revealing the structure of isolated peptides: IR-IR predissociation spectroscopy of protonated triglycine isomers. *Journal of Molecular Spectroscopy* **2018,** *347*, 28-34.

(17) Marsh, B. M.; Voss, J. M.; Garand, E. A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *J Chem Phys* **2015,** *143* (20), 204201.

(18) Yeh, L. I.; Okumura, M.; Myers, J. D.; Price, J. M.; Lee, Y. T. Vibrational spectroscopy of the hydrated hydronium cluster ions H3O+·(H2O)n (n=1, 2, 3). *The Journal of Chemical Physics* **1989,** *91* (12), 7319-7330.

(19) Yang, N.; Duong, C. H.; Kelleher, P. J.; Johnson, M. A.; McCoy, A. B. Isolation of site-specific anharmonicities of individual water molecules in the I−·(H2O)2 complex using tag-free, isotopomer selective IR-IR double resonance. *Chemical Physics Letters* **2017,** *690*, 159-171.

(20) Wolk, A. B.; Leavitt, C. M.; Garand, E.; Johnson, M. A. Cryogenic ion chemistry and spectroscopy. *Acc Chem Res* **2014,** *47* (1), 202-10.

(21) Campbell, J. L.; Yang, A. M.; Melo, L. R.; Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J Am Soc Mass Spectrom* **2016,** *27* (7), 1277-84.

(22) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: lessons learned from tetraalkylammonium cations and their isomers. *J Am Soc Mass Spectrom* **2014,** *25* (9), 1583-91.

(23) Lecours, M. J.; Chow, W. C.; Hopkins, W. S. Density functional theory study of Rh(n)S(0,+/-) and Rh(n+1)(0,+/-) (n = 1-9). *J Phys Chem A* **2014,** *118* (24), 4278-87.

(24) Liu, C.; Le Blanc, J. C.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using differential mobility spectrometry to measure ion solvation: an examination of the roles of solvents and ionic structures in separating quinoline-based drugs. *Analyst* **2015,** *140* (20), 6897-903.

(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

(26) Sherman, S. L.; Nickson, K. A.; Garand, E. Comment on "Microhydration of Biomolecules: Revealing the Native Structures by Cold Ion IR Spectroscopy". *J Phys Chem Lett* **2022,** *13* (8), 2046-2050.

CHAPTER 7

Future Directions: How Many Waters is Necessary to Induce Zwitterion Formation in Glycine?

7.1 Introduction

The work in this thesis demonstrates how gas-phase spectroscopy is useful to simplify and analyze systems that may be complicated in nature. It makes trend analysis, such as step-wise solvation and amino acid substitution, possible for a deeper understanding of molecular processes. Equipped with this information, we can extrapolate these trends to larger systems. However, one question asked of gas-phase spectroscopists is "how do we know that what we are observing equates to a real solvated system?" The question is not unfounded since in the gas phase the instruments are under vacuum and usually cryogenically cooled. Many have wondered if cryogenic cooling affects the conformer population (see **Chapter 3**) and also, upon solvation, whether the cluster observed can be related to solution phase (e.g. if the cluster was an ice crystal or actually a model for solution phase). In the gas-phase, especially at cryogenic temperatures, it is possible that the conformations may be kinetically trapped higher energy species. As found in the work detailed in **Chapter 3**, Rice-Ramsperger-Kassel-Marcus (RRKM) theory[1] of unimolecular reaction rates can predict whether the system being probed, on the time scale of the instrument, is kinetically trapping. By utilizing this theory, it is possible to see if a very minor proportion of the conformer population would correlate to kinetic trapping. In general, the major conformer of the system would be the non-kinetically trapped conformer that would be observed in nature. The concern of whether the solvated cluster can be related to solution phase will be addressed in this chapter.

It is well known that amino acids exist as a zwitterion in solution phase. However, without solvent in the gas-phase, they exist in their non-ionized form. Though not much else is known about zwitterion formation, it is well understood that solvent is the driving force behind it. Many theorists and experimentalists are trying to understand how solvation drives zwitterion formation and more specifically, what is considered full solvation (e.g. at what water cluster size might

zwitterion formation be induced). Since glycine is the simplest amino acid, many studies have been done both experimentally and theoretically to determine at what cluster size zwitterion formation occurs.[2-11] All studies have agreed that one water molecule is not enough to induce zwitterion formation, but there has not been conclusive results from theoretical studies[2-7, 9]. Cluster sizes between two and twelve water molecules have all been suggested. There also have not been many gas-phase experiments discussing this problem. Xu *et al.*[10] utilized size-selected photoelectron spectroscopy on microsolvated glycine and determined that five water molecules were necessary to induce this formation. However, to do this experiment, an excess electron had to be attached to the respective solvated cluster of glycine. It is possible that this electron has stabilizing effects that induce zwitterion formation at an earlier stage. In another study, Oomens *and co-workers*[11] obtained gas-phase spectra of a microsolvated glycine analog, tryptophan, to answer the same question with a UV-IR technique called resonance-enhanced multiphoton ionization (REMPI).[12-13] REMPI is a multi-photon technique that relies on a UV pulse to excite the molecule that has a chromophore and a second pulse to remove an electron. The second pulse creates cation for detection via mass spectrometry. They determined that a minimum of five solvent molecules were necessary to induce zwitterionic formation. As with the previously mentioned study, the removal of the electron might be changing the electronic state of the molecule and therefore stabilizing the formation of the zwitterion. Despite potential stabilization, both of these experimental techniques yielded the same result of needing at minimum of five water molecules to see the formation of the zwitterion.

Due to the ability of being able to microsolvate amino acids in the gas-phase, Cryogenic Ion Vibrational Spectroscopy (CIVS) can be an asset in this determination. However, since CIVS

requires a protonated species for mass spectrometry, protonated glycine cannot be utilized for this since it has no capability of forming a zwitterion. A glycine analog must be created and used.

7.2 Experimental Details

All IRPD spectra shown here were obtained with our homebuilt dual cryogenic ion trap vibrational spectrometer described in detail previously (see **Chapter 2**).[14] Briefly, protonated ions were generated via electrospray ionization of an ~1 mM solution of the permanently protonated glycine analog in methanol. Hexapole ion guides transfer the ions through a series of differentially pumped regions and into a 2D octupole ion reaction trap held at 80 K by a liquid nitrogen cryostat. An inert buffer gas seeded with either H_2O or D_2O is pulsed into the trap at ~1 ms bursts to collisionally cool the ions and induce cluster formation. The ions are then transferred to the 3D quadrupole ion main trap held at 10 K by a closed-cycle helium cryostat. A He buffer gas seeded with 10% D_2 is pulsed into the trap and used to collisionally cool the clusters further and induce the formation of D_2-tagged adducts. The tagged adducts are then injected into the TOF mass spectrometer for separation and mass selection. Finally, ions are intersected by the output of a 10 Hz Nd:YAG pumped OPO/OPA infrared probe laser. Resonant absorption of a single photon resulted in the loss of the D_2 tag for the glycine adduct + $(H_2O)_{0\text{-}3}$ and the loss of a water molecules for cluster sizes larger than three waters.

7.3 Computational Details

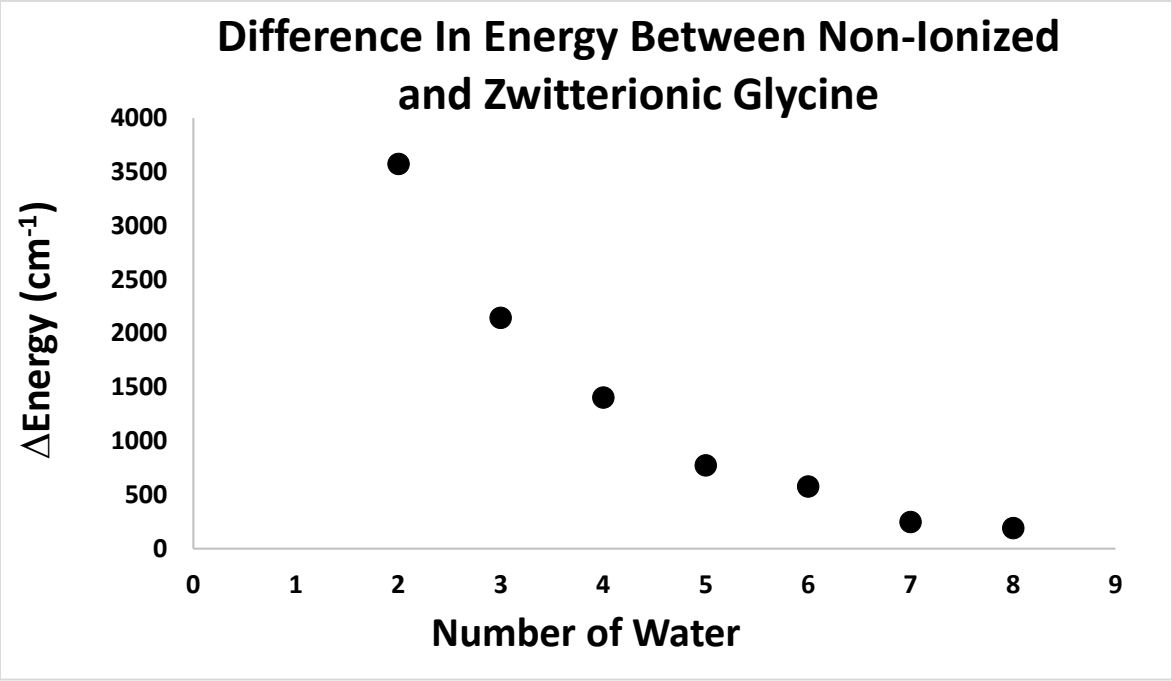
A methodical isomer structure search was done with help from the molecular mechanics basin hopping program developed by the Hopkins group at the University of Waterloo[15-18] to find the lowest energy structures. The computational search answers two critical questions: first, when

it might be possible to observe the zwitterionic structure and second, to compare the lowest energy structures to the experimental IRPD spectra. For each cluster size, the previous cluster's lowest energy isomer was used as the starting point for each optimization and there was a set of optimizations for both the non-ionized and zwitterionic form of the analog. Each water in the cluster was iteratively stepped 7,000 times by a random value between ±5° and 0.2 Å, for rotational and translational motions, respectively. The dihedrals of the glycine analog were also iteratively stepped 7,000 times by a random value between ±5°. After each iteration, the resulting structures were optimized using the AMBER force fields and duplicates were rejected by the program. All following computations were done with Gaussian 16[19] software. Each unique structure was then optimized rapidly with BLYP/3-21g and all duplicate structures and those with energies >140KJ/mol were rejected. The remaining structures were further optimized at the cam-B3LYP/6-311g level and duplicate structures and those with relative energies >50KJ/mol were rejected. The remaining structures were once again optimized at the cam-B3LYP/def2TZP level, and harmonic frequency calculations were performed on the lowest ~ 5-10 conformers each of the non-ionized and zwitterionic forms of the molecule in the cluster. These same structures were also optimized at the MP2/def2TZP level.

7.4 Results and Discussion

While determining the appropriate glycine analog, DFT and MP2 computations were run for glycine$(H_2O)_n$ systems, both non-ionized and zwitterionic, to determine at what point it would be feasible to see a zwitterionic compound in solution. The energy difference between the lowest energy conformers of the non-ionized and zwitterionic systems was plotted and is shown in **Figure 7.1**. At ~5-6 water molecules, the zwitterionic form of glycine starts becoming more favorable at

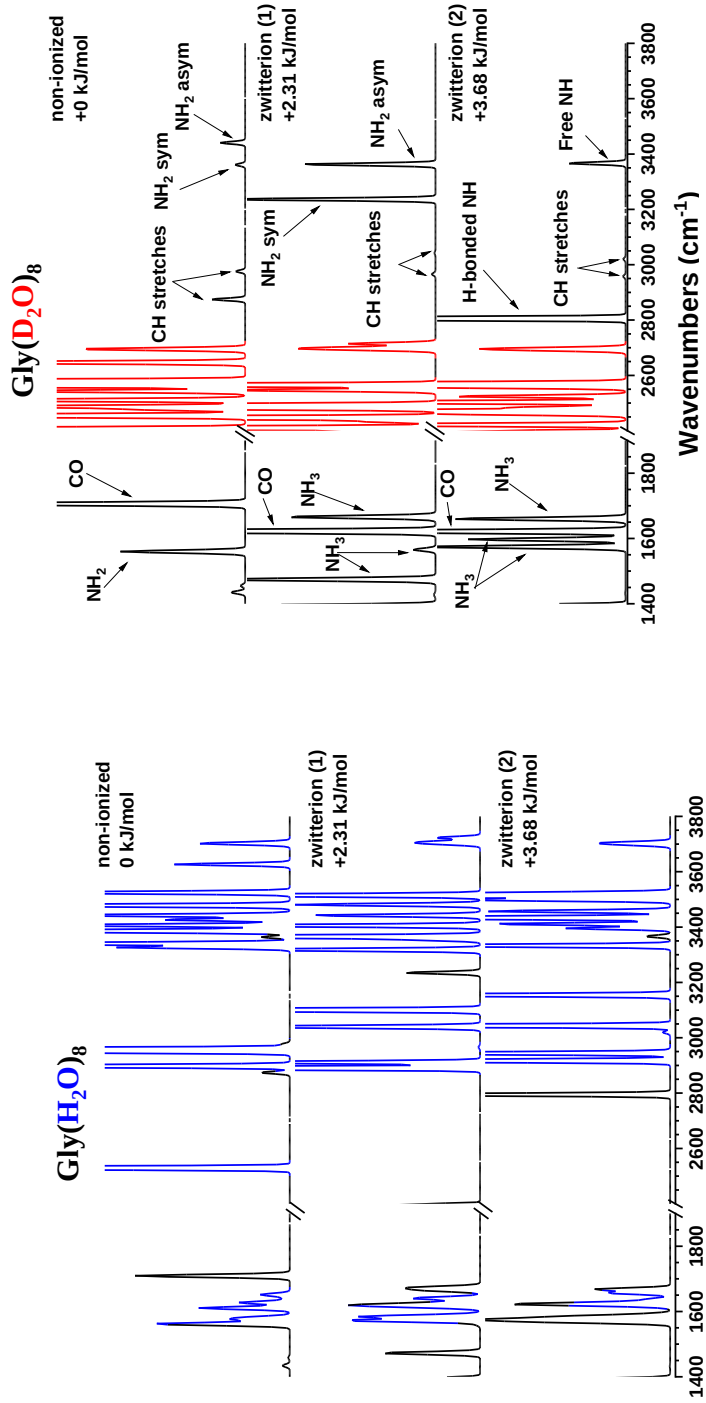
Figure 7.1. The difference in energy, in cm^{-1}, between the lowest energy conformer of the non-ionized and zwitterionic forms of $Gly(H_2O)_{0-8}$ at MP2/ def2-TZVP with ZPE corrections added (ZPE corrections found at cam-B3LYP/def2-TZVP). For n=0-1, the zwitterionic form always optimized to the non-ionized form.



~500 cm^{-1} (5.98 kJ/mol) and at ~7-8 water molecules, it is nearly iso-energetic with the non-ionized form of the molecule. Computationally, it was important to look for "signatures" of the zwitterion in solution, particularly in the amine and the C=O region. Three low energy structures, one non-ionized and two zwitterionic structures, were used in glycine$(H_2O)_8$ to figure this out. The H_2O and D_2O calculated spectra are shown in **Figure 7.2**. With H_2O alone, a lot of the water features block the stretches directly relating to the peptide. With use of D_2O, it is possible to elucidate these features due to the red shifting of the D_2O features. In general, the main difference of the zwitterionic and non-ionized form of the molecule will be in the NH_2/NH_3 region in the mid-IR and the NH_2/NH_3 and C=O region in the far-IR. The carboxyl OH group should be solvated with the addition of one water molecule and therefore cannot help with signature patterns. In general, the zwitterionic form of glycine has a redshifted amine region in the mid-IR that is much more intense. It also has a NH_3 feature before the C=O (which is also shifted from its non-ionized form) in the far-IR. Therefore, this should be the main focus of spectral analysis.

At the start of this project, trimethyllysine was used as a glycine analog since it is readily available and contains a delocalized charge which is far from the glycine terminal (**Figure 7.3A**). However, when data was taken on this system, the one laser IRPD spectrum became extremely cluttered. The long chain between the glycine moiety and the protonated and shielded charge was extremely flexible thus allowing the molecule to fold in on itself. The folding of the molecule directly interacted with the glycine moiety and possibly stabilized the zwitterionic structure. Additionally, it convoluted the many of the conformers via small changes in the angles of the chain. Thus, with no water, the trimethyllysine IRPD spectrum was congested, and IR-IR double resonance could not help since all structures were extremely similar to each other. Thus, a search for a new analog continued.

Figure 7.2. Three of the lowest energy structures of Gly(H_2O/D_2O)$_8$, one in non-ionized and two in zwitterionic form found at cam-B3LYP/def2-TZVP and ZPE corrected. The features highlighted in blue correspond to H_2O stretches, the features highlighted in red correspond to D_2O stretches, and the features in black correspond to the Glycine.



The amino acid histidine looked promising as it could be modified to contain a protonated, delocalized charge in its imidazole ring and only contained a small chain between the ring and the glycine moiety. The modified histidine molecule, 1,3-Dimethylhistidine, is shown in **Figure 7.3B**. The synthesis for methylation and deprotection of BOC-His-OMe to form $His(Me)_2OH^+$ is described below and the reaction scheme is shown in **Figure 7.4**.

Starting with BOC-His-OMe, 4 equivalents of $NaHCO_3$ and 25 equivalents of MeI were added into excess acetonitrile. The mixture was refluxed mixture under N_2 at 60 degrees Celsius for 3 days. Once refluxed, the solvent was evaporated utilizing a rotovap. The formed solid was washed using chloroform and water. The water layer was extracted and evaporated. Into the remaining white powder, 1M HCl (10 mmol, 10 mL), was stirred at room temperature for 3 hours. 1M (10 mmol, 10 mL) KOH was then added to the solution and stirred at room temperature for an additional 3 hours. The neutral solution was then saturated with NaCl. The final product was extracted with chloroform and washed. The solvent was rotovaped off and the final white product was obtained. Recrystallization methods using acetonitrile were used if product looked yellow or was very impure, found by taking a mass spectrum of the product. Reaction scheme is shown in **Figure 7.4**.

The one-laser IRPD spectra of $His(Me)_2OH^+(H_2O)_{0\text{-}12}$ is shown in **Figure 7.5**. The IRPD spectrum of $His(Me)_2OH^+(H_2O)_3$ could not be taken due to low signal of tag. For the bare $His(Me)_2OH^+$ it is possible to see two strong features around 3572 and 3421 cm^{-1} corresponding to the free carboxylic acid O-H and asymmetric stretch of the amine, respectively. A weaker amine symmetric stretch feature at 3358 cm^{-1} and a couple of smaller features between 2800 and 3200 cm^{-1} corresponding to the CH_3 stretches in the imidazole ring are also present in **Figure 7.5**. Upon the addition of one water molecule, the free carboxylic acid O-H peak at 3572 cm^{-1} disappeared

Figure 7.3. (A) Trimethyllysine molecule. (B) Synthesized $His(Me)_2(OH)^+$ molecule.

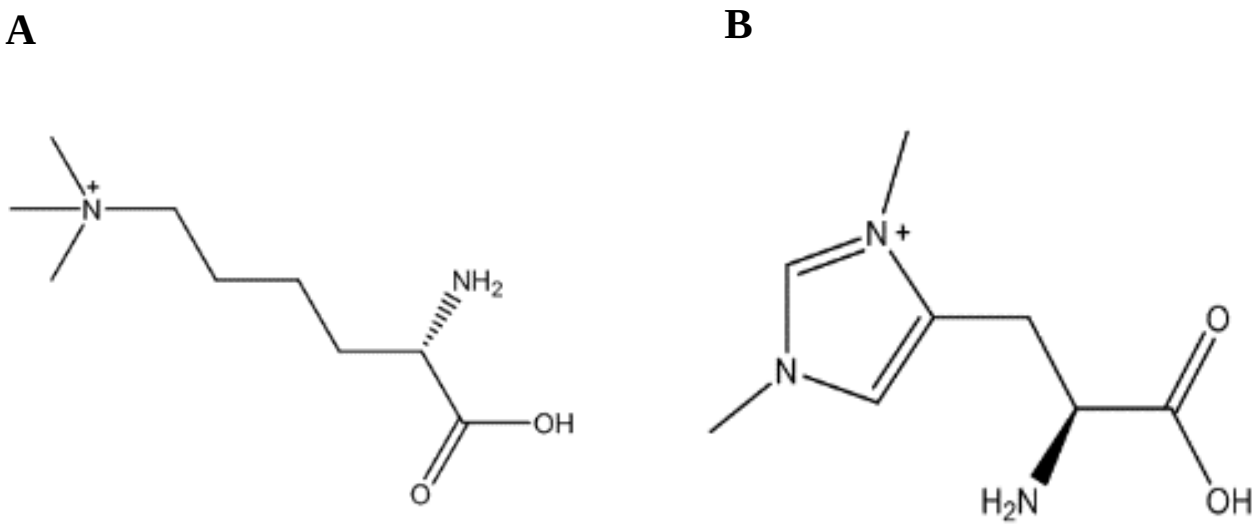


Figure 7.4. Synthesis of the methylation and deprotection of BOC-His-OMe to form $His(Me)_2OH^+$.

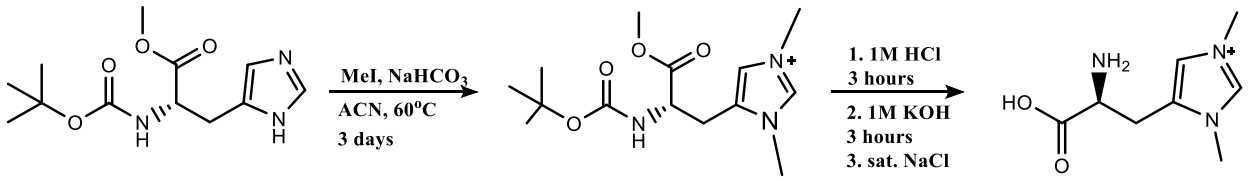
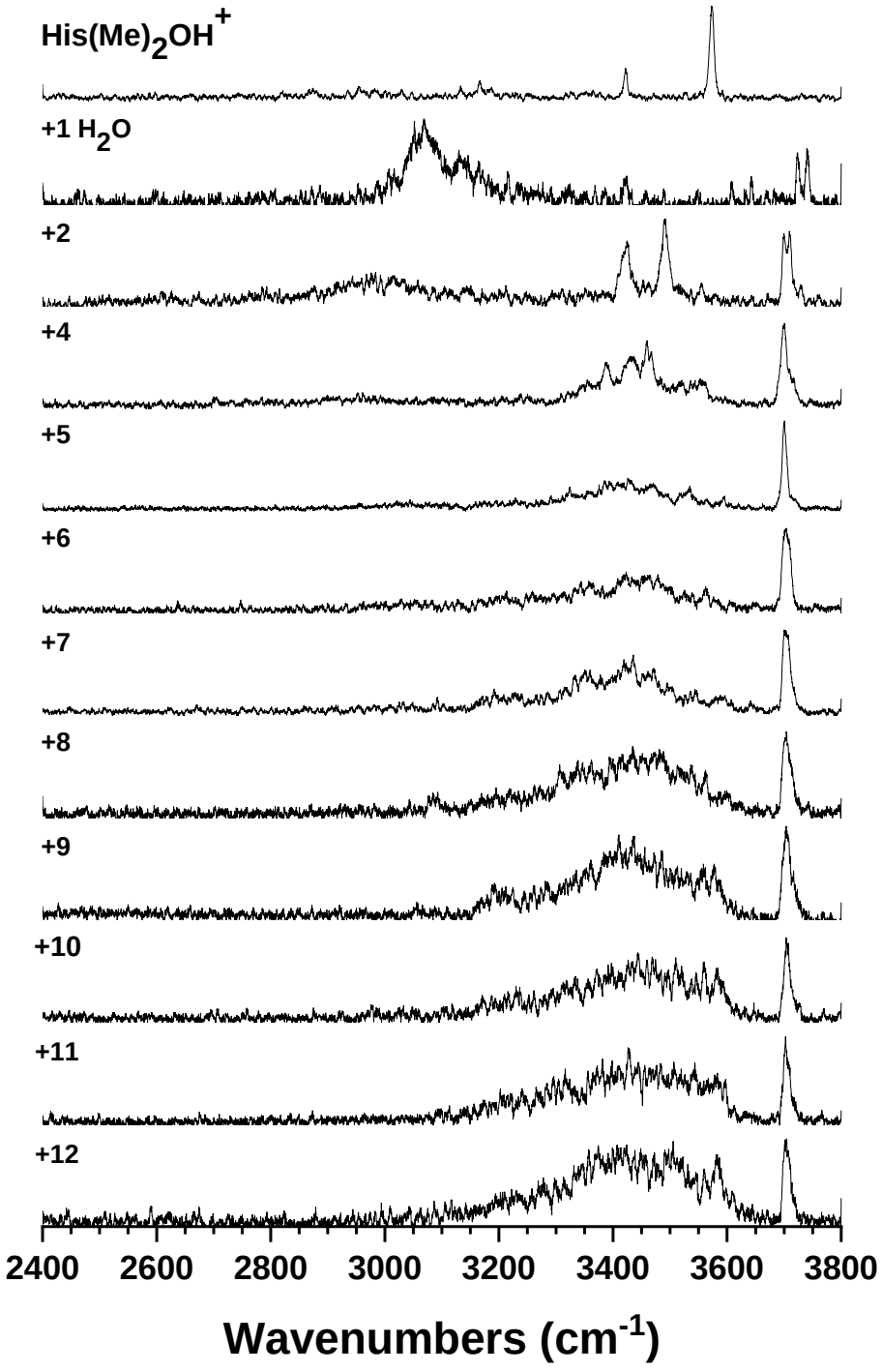


Figure 7.5. One-laser IRPD spectra of $His(Me)_2OH^+(H_2O)_{0\text{-}12}$. The IRPD spectrum of $His(Me)_2OH^+(H_2O)_3$ could not be taken due to low signal of tag.



and a large, broad feature between 3000 and 3200 cm^{-1} appeared, indicating that the water molecule preferentially binds to the carboxylic acid. This is encouraging because the same observations were made for betaine (which is similar to the modified histidine since it also has a delocalized charge)(see **Chapter 6**) In betaine and histidine alike, solvation at the carboxylic acid rather than the at the imidazole ring shows that the charge is delocalized around the ring and should not be affecting the solvation pattern around the glycine moiety. Finally, as solvation size increases, the features corresponding to the water and interactions with it become blue shifted and broaden. It is therefore necessary to then solvate with D_2O and overlay the two spectra. All features corresponding to H_2O are in blue, D_2O are in red, and the overlapping regions are related to the peptide are in purple.

Thus far, only $His(Me)_2OH^+(H_2O/D_2O)_{0\text{-}2}$ have been obtained and the results are shown in **Figure 7.6**. The IRPD spectrum for $His(Me)_2OH^+(H_2O/D_2O)_1$ and the corresponding three lowest energy calculations are shown in **Figure 7.7**. For the one water spectrum, there are four free O-H stretches that belong to a water molecule between 3600 and 3745 cm^{-1}. However, there are only two modes for a water molecule in this region, the symmetric and asymmetric modes. Since there are four peaks, there must be at least two conformers in the population. Looking at the low energy conformers found for this system, all features of all three conformers, shown in **Figure 7.7 A-C**, overlap with each other except those of the water. The calculated spectra of **A** and **B** account for all the features of the one laser IRPD spectrum. The difference between these two conformers is which oxygen is the C=O and which is the O-H.

The IRPD spectrum for $His(Me)_2OH^+(H_2O/D_2O)_2$ and corresponding low energy calculations are shown in **Figure 7.8**. For the two-water spectrum, there are four free O-H stretches that belong to the two water molecules between 3300 and 3750 cm^{-1}. These four stretches on its

Figure 7.6. IRPD spectra of $His(Me)_2OH^+(H_2O/D_2O)_{0\text{-}2}\cdot D_2$. The features in blue correspond to the H_2O spectrum, in red correspond to the D_2O spectrum, and the overlap of the two in purple correspond to the peptides.

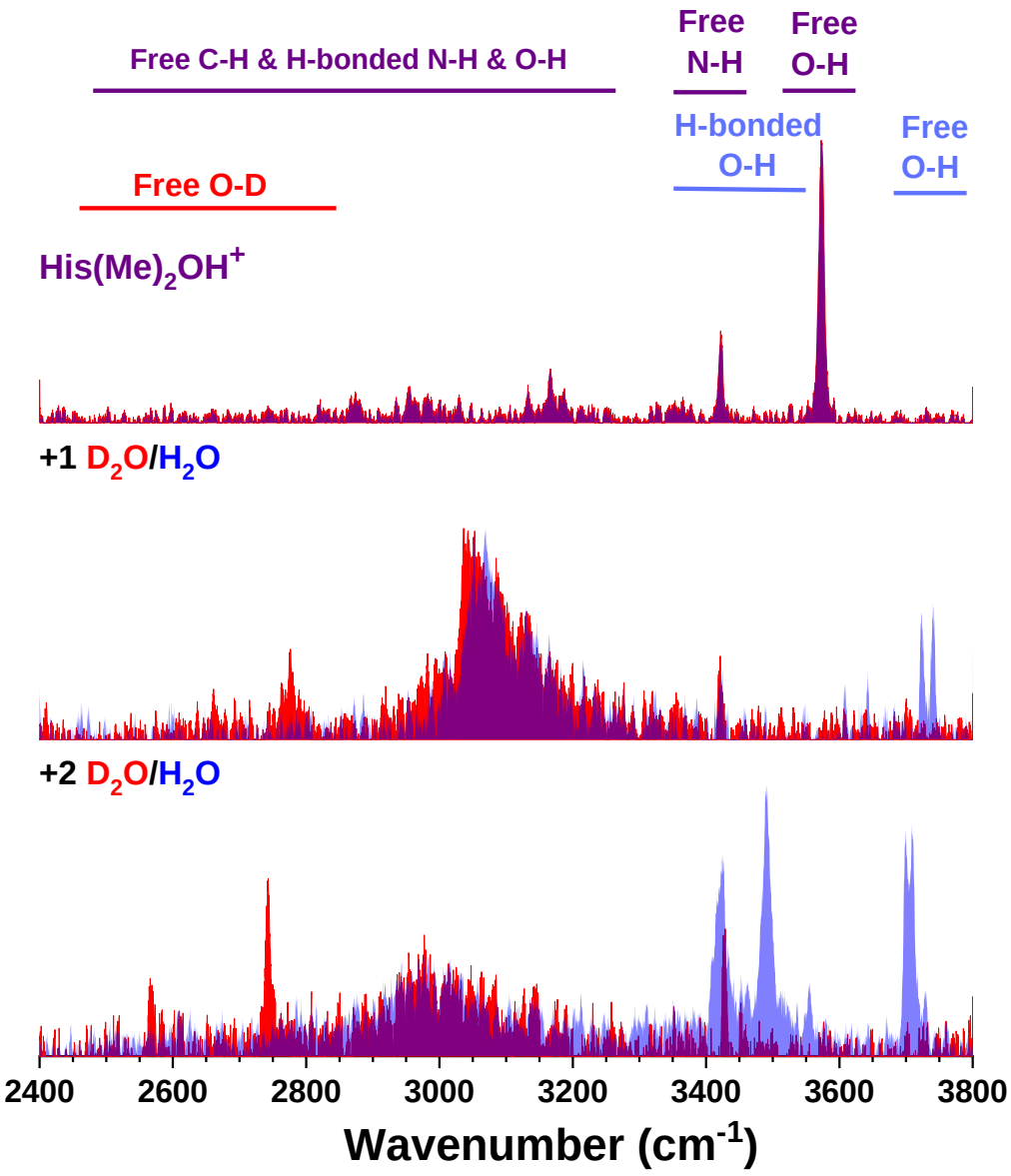


Figure 7.7. IRPD spectrum for $His(Me)_2OH^+(H_2O/D_2O)_1$ and corresponding three lowest energy conformers. Energetics and calculated spectra were found at cam-B3LYP/ def2-TZVP and ZPE corrected.

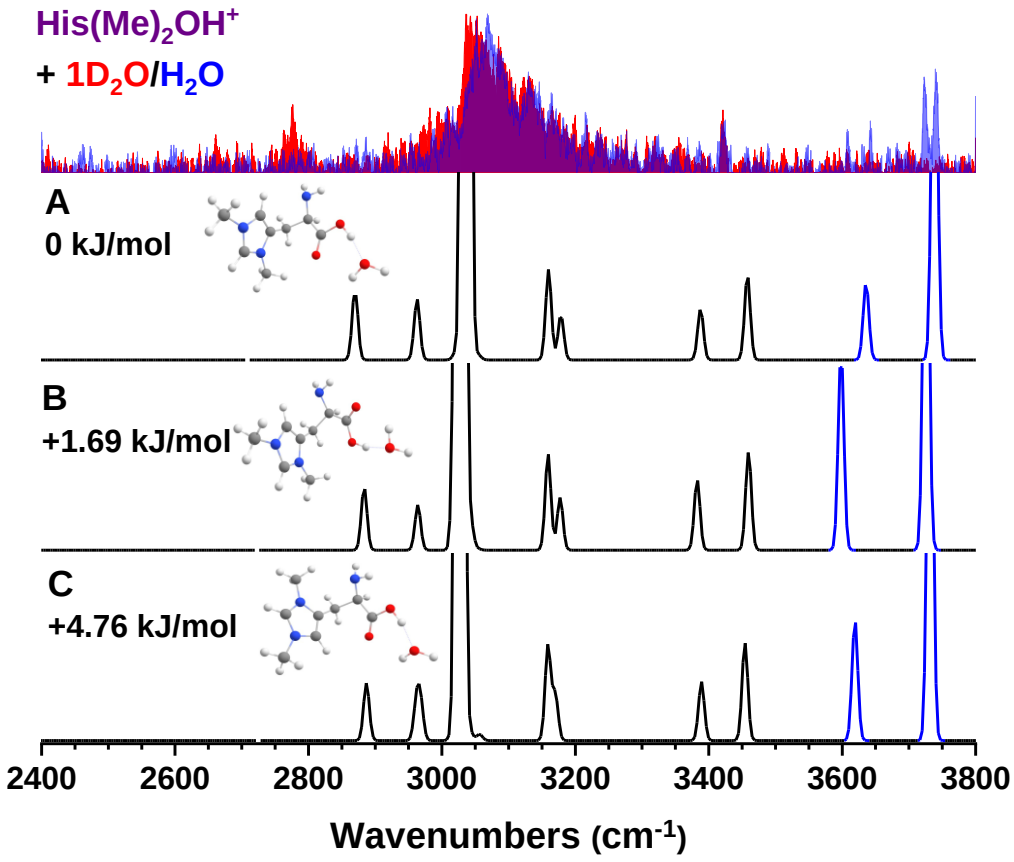
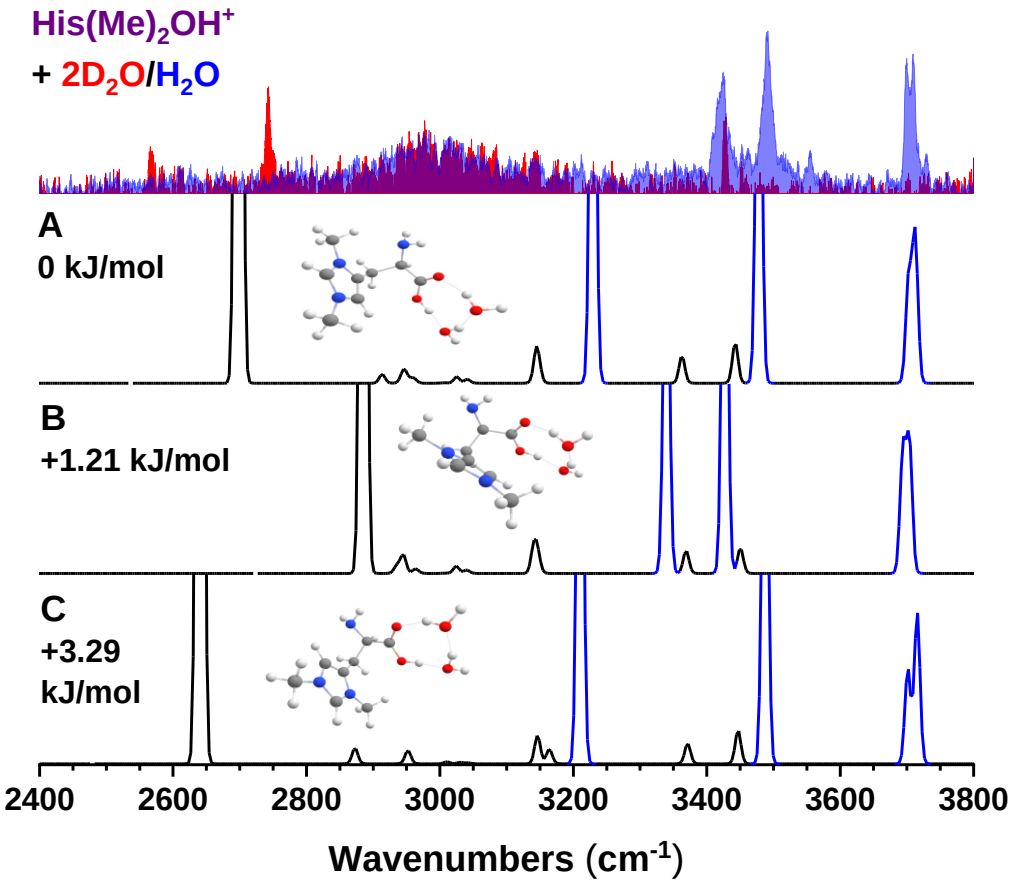


Figure 7.8. IRPD spectrum for $His(Me)_2OH^+(H_2O/D_2O)_2$ and corresponding three lowest energy conformers. Energetics and calculated spectra were found at cam-B3LYP/ def2-TZVP and ZPE corrected.



own can correspond to just one conformer since each water molecule should have a symmetric and asymmetric stretch. The region corresponding to the peptide should be able to give clues to whether it should only be one conformer. The overlay of the H_2O and D_2O spectra shows that between 3400 and 3500 cm^{-1} there are two peptide features which correspond to the symmetric and asymmetric stretching of the amine. Between 2800 and 3100 cm^{-1} there is a broad peptide feature that corresponds to an H-bonded carboxyl OH. From simply counting the features, it seems that there is only one conformer, however IR-IR double resonance spectroscopy can help fully determine this. Comparing the IRPD spectrum to the calculated structures, it seems that the low energy structure found in panel C fully describes the system. The main motif of the three lowest energy conformers is that the two waters form a four-member ring with the carboxylic acid, just as it did with betaine in **Chapter 6**. The only difference with these conformers is the position of the imidazole ring which by localized effects it affects the frequency at which water appears.

7.5 Future Work

Ideally, D_2O spectra for $His(Me)_2OH^+(H_2O/D_2O)_{3\text{-}12}$ would need to be taken and compared to the H_2O spectra. However, when trying to obtain this data, there were impurities in the solution that overlapped with the D_2O peaks. Purification techniques were attempted but did not prove fruitful in removing the impurity. Enlisting help from a synthetic chemist in the department might be beneficial. If the impurity cannot be removed, a different glycine analog must be used. Depending on which analog is selected, computations must be run to help identify conformers in the various the IRPD spectra. IR-IR double resonance experiments may occur to help determine if there is more than one conformer within the population and help assist with conformer determinations.

7.6 Conclusions

Cryogenic ion vibrational spectroscopy (CIVS) is shown to be a useful tool to identify how many water molecules are needed to induce the zwitterionic form of an amino acid. Two glycine analogs, trimethyllysine and 1,3-Dimethylhistidine were used to answer this question. Trimethyllysine was too flexible and shielded protonation site stabilized the glycine moiety thus creating a congested spectrum. The spectra of the1,3-Dimethylhistidine and its water clusters were taken up to twelve water molecules. The D_2O spectra was taken up to two D_2O molecules. Further purification techniques or a different synthesis schematic are required to mitigate impurities in the mass spectrum which hinders the acquiring of higher order D_2O spectra. Future work on other glycine analogs to use with CIVS might be necessary.

7.7 References

(1) Callear, A. B. Chapter 4 Basic RRKM Theory **1983,** *24*, 333-356.

(2) Aikens, C. M.; Gordon, M. S. Incremental solvation of nonionized and zwitterionic glycine. *J Am Chem Soc* **2006,** *128* (39), 12835-50.

(3) Fernández-Ramos, A.; Smedarchina, Z.; Siebrand, W.; Zgierski, M. Z. A direct-dynamics study of the zwitterion-to-neutral interconversion of glycine in aqueous solution. *The Journal of Chemical Physics* **2000,** *113* (21), 9714-9721.

(4) Förner, W.; Otto, P.; Bernhardt, J.; Ladik, J. J. A model study of the intermolecular interactions of amino acids in aqueous solution: The glycine-water system. *Theoretica Chimica Acta* **1981,** *60* (3), 269-281.

(5) Jensen, J. H.; Gordon, M. S. On the Number of Water Molecules Necessary To Stabilize the Glycine Zwitterion. *Journal of the American Chemical Society* **1995,** *117* (31), 8159-8170.

(6) Kassab, E.; Langlet, J.; Evleth, E.; Akacem, Y. Theoretical study of solvent effect on intramolecular proton transfer of glycine. *Journal of Molecular Structure: THEOCHEM* **2000,** *531* (1-3), 267-282.

(7) Kokpol, S. U.; Doungdee, P. B.; Hannongbua, S. V.; Rode, B. M.; Limtrakul, J. P. Ab initio study of the hydration of the glycine zwitterion. *Journal of the Chemical Society, Faraday Transactions 2* **1988,** *84* (11), 1789.

(8) Strittmatter, E. F.; Lemoff, A. S.; Williams, E. R. Structure of cationized glycine, gly.m (m = be, mg, ca, sr, ba), in the gas phase: intrinsic effect of cation size on zwitterion stability. *J Phys Chem A* **2000,** *104* (43), 9793-6.

(9) Tortonda, F. R.; Pascual-Ahuir, J. L.; Silla, E.; Tuñón, I. Why is glycine a zwitterion in aqueous solution? A theoretical study of solvent stabilising factors. *Chemical Physics Letters* **1996,** *260* (1-2), 21-26.

(10) Xu, S.; Nilles, J. M.; Bowen, K. H. Zwitterion formation in hydrated amino acid, dipole bound anions: How many water molecules are required? *The Journal of Chemical Physics* **2003,** *119* (20), 10696-10701.

(11) Blom, M. N.; Compagnon, I.; Polfer, N. C.; Helden, G.; Meijer, G.; Suhai, S.; Paizs, B.; Oomens, J. Stepwise solvation of an amino acid: the appearance of zwitterionic structures. *J Phys Chem A* **2007,** *111* (31), 7309-16.

(12) von Helden, G.; Holleman, I.; Knippels, G. M. H.; van der Meer, A. F. G.; Meijer, G. Infrared Resonance Enhanced Multiphoton Ionization of Fullerenes. *Physical Review Letters* **1997,** *79* (26), 5234-5237.

(13) von Helden, G.; van Heijnsbergen, D.; Meijer, G. Resonant Ionization Using IR Light: A New Tool To Study the Spectroscopy and Dynamics of Gas-Phase Molecules and Clusters. *The Journal of Physical Chemistry A* **2003,** *107* (11), 1671-1688.

(14) Marsh, B. M.; Voss, J. M.; Garand, E. A dual cryogenic ion trap spectrometer for the formation and characterization of solvated ionic clusters. *J Chem Phys* **2015,** *143* (20), 204201.

(15) Campbell, J. L.; Yang, A. M.; Melo, L. R.; Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J Am Soc Mass Spectrom* **2016,** *27* (7), 1277-84.

(16) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: lessons learned from tetraalkylammonium cations and their isomers. *J Am Soc Mass Spectrom* **2014,** *25* (9), 1583-91.

(17) Lecours, M. J.; Chow, W. C.; Hopkins, W. S. Density functional theory study of Rh(n)S(0,+/-) and Rh(n+1)(0,+/-) (n = 1-9). *J Phys Chem A* **2014,** *118* (24), 4278-87.

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

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

CHAPTER 8

Introduction into Fundamental Research Using Gas-Phase Spectroscopy To Probe Small Peptides

8.0 Preface

My whole life growing up my mother always stressed to me that education was important and told me that "education was the only thing that she was going to leave me behind." My mother had immigrated to this country from Argentina at 22-years old and my grandmother and aunt followed when I was 6-years old. Therefore, I grew up in a primarily Spanish speaking household and became the first person in my family to go to college in the United States and complete a double Bachelor's degree. And now, as I sit here and write my thesis, I will be the first in my family to earn a doctorate! Though we did not come from wealth, it was always important to them for me to get an education. My passion has always been with physical and analytical chemistry, and in general, physical chemistry has a lot of jargon which can be very daunting for a non-scientific audience. My family loves to be in the loop with what I am doing and does try to understand it, yet I know that even if they fully don't understand what I am studying, they are still pushing me to succeed. Therefore, I have written this chapter to explain my research to a broad, non-scientific audience and I wanted to dedicate it to my family.

Before I continue, I would also like to thank Professor Bassam Shakhashiri who created the Wisconsin Initiative for Science Literacy (WISL) at UW-Madison which is the program that gives graduate students the platform to be able to write this chapter. It is amazing that as a UW student, we are supported in our endeavors to create this chapter and it is recognized that describing our work to a non-scientific audience is actually important! I would also like to thank Professor Shakhashiri, Elizabeth Reynolds, and Cayce Osborne for dedicating their valuable time to read my chapter, provide valuable feedback, and also give encouragement in the quest to write this chapter.

8.1 Fundamental Chemistry—Why Should We Care?

When I began my graduate school experience and had to choose a group to do my PhD in, I had absolutely no knowledge of the field of gas-phase spectroscopy. There are many questions that I constantly get asked as I have gone through my PhD in this fundamental physical chemistry field: “Why does your research matter? Does this help discover something in pharmaceuticals? In the end will it help cure cancer?” The list goes on. Though fundamental in nature, my research has a larger impact when you look at the full picture; however, my particular research is far removed from practical applications.

In high school chemistry and biology classes, the teachers often tell students trends that are known to be true in chemistry without any of the facts or data behind it. For instance, one major example that teachers tell students is that the “structure of a molecule determines its function.” This is inherently true. These teachers might even tell their students particular trends, “If there is an electron withdrawing group then we expect the reaction to proceed in X way; however, there is an exception…” And as a high school student, we take this for granted, accept it as fact, and memorize the trends and exceptions. As we go on and delve further into the subject, we realize that there are a lot of reasons that these statements are true and why there might be exceptions to cases. But it isn’t until we go straight into the fundamentals, through quantum equations and derived principles that we can fully begin to understand why it is true on all levels. This is where my research steps in and helps guide scientists in many fields. We do work in trying to fundamentally understand what is occurring to then be able to explain the trends and exceptions that we see.

8.2 The Big Picture

When many people think of protein, they think of something they can ingest— food or supplements. Independent of food consumption choices, everyone knows that protein is necessary for a healthy diet in day-to-day life. Many people who do strenuous workouts supplement their diet with additional protein to help repair and grow their muscles. Besides being something that people ingest, protein is also found everywhere throughout the body and it plays many critical roles. Some of these critical roles are the proteins functioning as antibodies or serving as enzymes, which are molecules used to bring about specific chemical reactions in the body. However, there are many more roles that proteins play.[1]

There is much interest in understanding how proteins interact in the environment. However, before understanding how they interact in the environment, which can depend on many different factors, it is important to understand them in their simple state, i.e. without external influences, and then build up the complexity.

Proteins, which are very large peptides, are made up of a collection of amino acids. Each natural amino acid in general contains a carboxylic acid group connected to an amine, shown in **Figure 8.1**. A carboxylic acid group is where there is a central carbon atom (C) bonded to an oxygen atom (O) by a double bond and a hydroxyl group (-OH) by a single bond. Its last bond links the carbon atom to the rest of the protein chain. An amine group is where a nitrogen (N) is bonded to two hydrogen atoms (H). Its final bond links the nitrogen atom to the rest of the protein chain. All natural amino acids differ by their corresponding R-group, which is just an abbreviation for any group that has a carbon or hydrogen atom to attach to the “rest (R) of the molecule,” that is tacked on the base, also shown in **Figure 8.1**. Therefore, every R-group gives each of the 20

Figure 8.1. The backbone of one of the 20 natural amino acids. It consists of a carboxylic acid attached to an amino by a single carbon chain. The R-group denoted in green changes based on the amino acid.

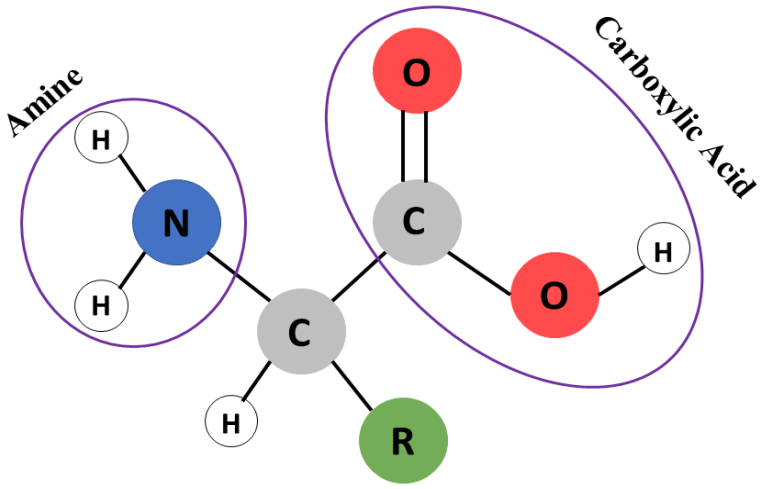
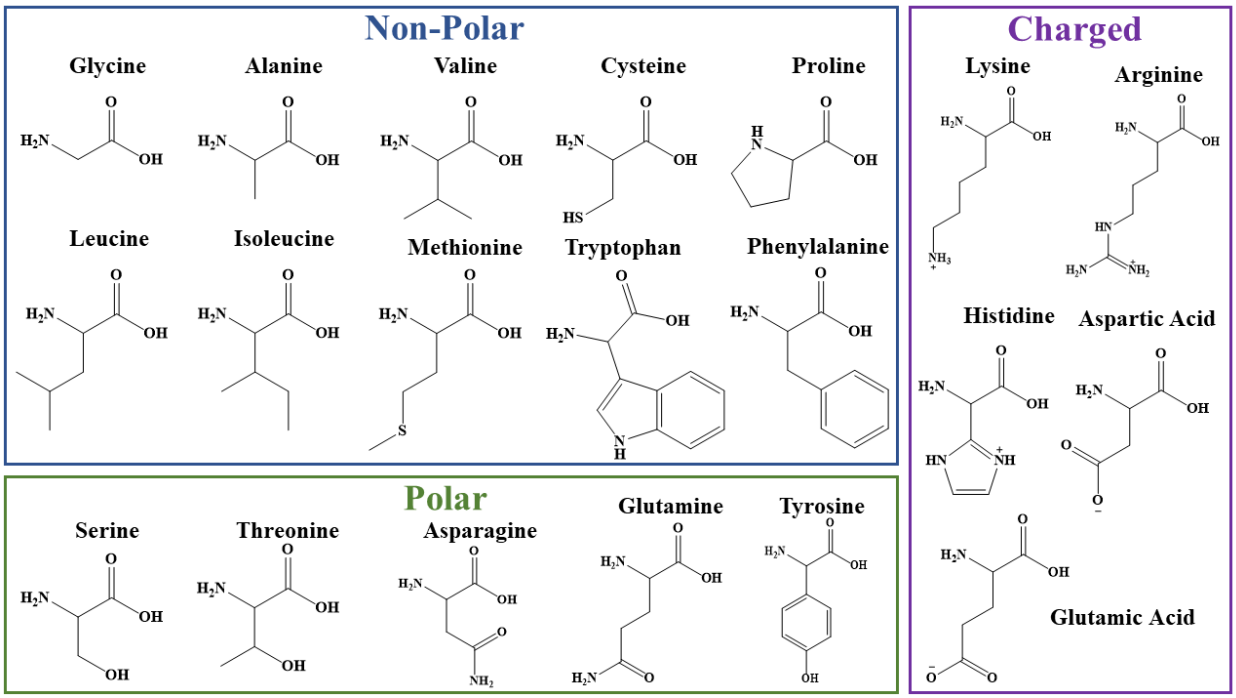


Figure 8.2. The 20 natural amino acids and their three distinct groupings.



natural amino acids their own specific properties. The natural amino acids can be divided into three distinct groups based on their side chains: non-polar (neutral molecule containing no positive or negative poles since all electrons are shared evenly), polar (neutral molecule containing positive and negative poles), and electrically charged (has a permanent positive or negative charge), shown in **Figure 8.2**. Amino acids attach together as building blocks for the peptide/protein. I think of amino acids like legos. Every different type of amino acid is a different size and might have a slightly different shape, just like the vast array of legos. Depending on how they are placed together, you might have a very different end structure (protein). And since we have heard in the past that the structure of the protein affects the function, it is important to understand why this happens. We can therefore use the amino acids like our own version of legos and build up our protein one block at a time to be able to see how every little change affects the structure and therefore the function.

8.3 Why Probe in the Gas Phase?

As I said above, we want to start very simple. There are so many techniques to be able to look at these proteins. Many techniques probe them in the solution that you would find them in, such as water. Since our bodies are ~60% water, with some regions being a higher percentage and others being a lower percentage, it makes sense to want to probe these molecules in water. However, water is a tricky molecule. In many techniques, water leads to a broadening of experimental features so much so that features turn into a blob when recording spectra. It is not then possible to actually see every little wiggle in a spectrum, and therefore interpret it. Also, in water, there are many more interactions that can occur. The water can interact with the protein

through hydrogen bonding and the water molecules can interact with each other. Therefore, a simple molecule can get very complicated extremely quickly.

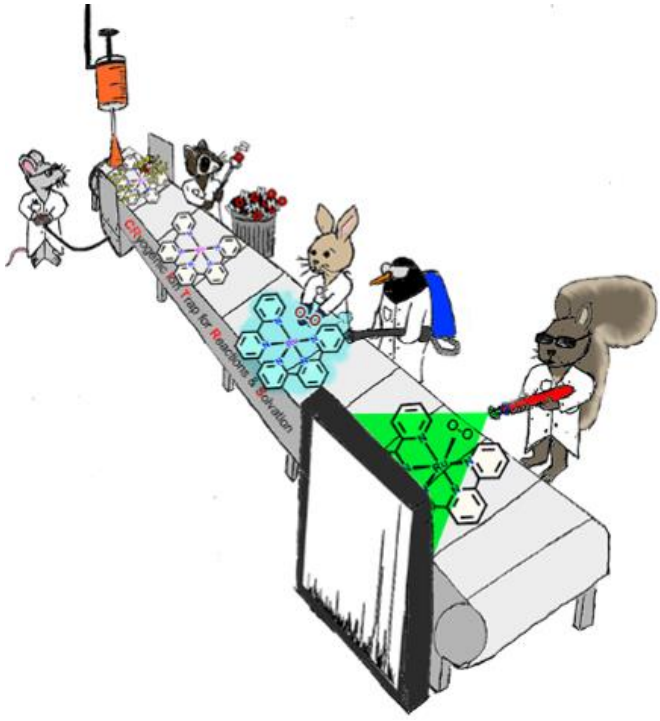
To be able to isolate the molecule outside of any solvent (a substance, which is normally a liquid, which is used to dissolve other materials in order to make a solution), we would need to put the molecule into the gas-phase. However, in the gas-phase, in order to detect your molecule, it would need to be charged in some way. Since all but five natural amino acids are neutral, which means that they do not contain a permanent positive or negative charge, it has to be manipulated for us to detect them and once manipulated, the charged species is called an "ion." Therefore, we can see the ion as it is without any solvent at all. And depending on instrumentation, it is possible to slowly solvate the molecule of interest one water molecule at a time and build up to bulk solvent. It is basically like putting salt in an empty glass and slowly adding one drop of water at a time. In the beginning, with such few drops you can see how the salt interacts with those drops, but as you keep adding more and more water, you build up to a glass of water where the salt is fully dissolved. After a certain point, it doesn't matter how many drops that you add, the extra water doesn't change the interaction between the water and the salt, this is called the bulk. This gives a better idea of the properties that the molecule might have on its own and how those properties change as a function of solvation.

8.4 The Machine

So how are we even going put our molecule into the gas-phase? We are able to probe the molecule on an instrument termed CIVS, which stands for Cryogenic Ion Vibrational Spectroscopy. If you walk into the Garand lab and see this "machine" as Etienne calls it, it looks like a daunting task, as a 3-meter instrument that has probably over 100 black cords crossing every

which way stares back at you. However, in its simplest form, it is basically an ion cooler and can be described by the cartoon that Jon Voss (PhD 2018) drew, and Erin Duffy (PhD 2017) colored in, shown in **Figure 8.3**. The figure itself is showing a ruthenium molecule going through the instrument, but the set-up is the same for a peptide. A peptide is the same as a protein since it is also a collection of amino acids, however it is usually smaller in scale.

Figure 8.3. Cartoon instrument schematic drawn by Jonathan Voss (Ph.D. 2018) and colored by Erin Duffy (Ph.D. 2017).

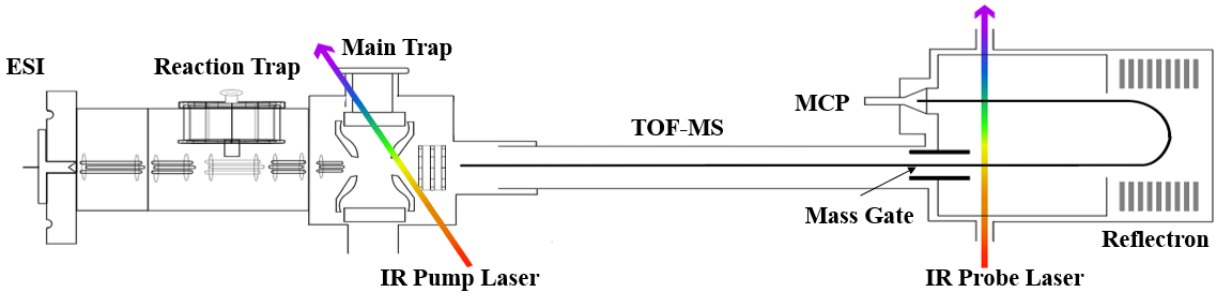


The instrument starts when the user (the lab rat) adds a liquid sample of the protonated peptide (a peptide with an additional hydrogen attached to it forming a charged species, peptide-H^+), which is our ion, being probed into a syringe that has a voltage applied to it to allow the sample to spray and go into the gas phase. We call this electrospraying the solution into the instrument. The ions then get guided through the instrument and stop at the reaction trap (the

racoon). At this stage, the user can add a solvent into the instrument through one of the gas lines. Even with one drop of water, for instance, the gas can flow over the drop and bring water molecules into the instrument to cluster around the peptide. Once clustered, the molecules get transferred from this stage into the main tagging trap (the bunny and the penguin). Here the clusters get cooled to 10 Kelvin, which is -263 degrees Celsius (very cold!), by a cryostat and a tag molecule, usually D_2, gets added to the cluster. D_2 is deuterium gas, which is an inert gas (it does not interact or chemically react with other species) and deuterium is an isotope of hydrogen; it has one additional proton than a hydrogen leading it to have a molar mass of 2 g/mol while hydrogen is only 1 g/mol. We utilize a D_2 tag for the laser stage of our instrument so we can observe a change of mass whenever something interesting is happening to our molecule through interaction with a laser. From here, the tagged cluster is taken into the last stage of the instrument, which is the laser stage (the squirrel). Whenever the laser is resonant with a vibration of the molecule, the molecule begins to move by a certain mode (bending or stretching of certain bonds). The vibrations are then redistributed by intermolecular vibration redistribution (IVR) and this motion allows the D_2 tag to pop off the cluster. As I sit here and write this next to my golden retriever, all of what I described above can be related and visualized as a human petting a dog. The dog is like our molecule and the human is like the laser. As the human is petting the dog, they move their hand from the top of the dog's belly to down near the legs, kind of like the laser scanning a certain frequency range. When you hit that "sweet spot" on a dog, the dog's lower legs start going crazy. This "sweet spot" on a dog can be related to a "resonance" of the molecule. If we imagine that there was something weakly attached to the dog's leg, as the D_2 tag is to our molecule, when the dog starts to move his legs, the weakly attached substance can be dislodged! All species are separated and hit a detector. We now then have two different masses reaching the detector when the laser is resonant (the untagged

and tagged cluster) and only one mass reaching the detector when the laser is not resonant (just the tagged cluster). By monitoring the appearance of the untagged cluster as a function of laser wavelength as we scan the laser, we can produce an IR spectrum of the cluster that we are interested in. The actual instrumental set-up is shown in **Figure 8.4**.

Figure 8.4. Real CIVS instrument schematic.



8.5 Understanding the Infrared (IR) Data

As stated above, the data that we get is basically how the peptide vibrates. Since each peptide is just a collection of amino acids, there are a couple of vibration features that we know that all the IR spectra will have. Each peptide has an O-H on the carboxylic acid (carbon double bonded to an oxygen and singly bonded to an -OH) and N-H features on the amine (nitrogen bonded to two hydrogens) and amides (nitrogen bonded to a hydrogen and to a carbon that is doubly bonded to an oxygen). The O-H and N-H features vibrate separately from each other and are dependent on the wavelength of the laser. So, what do I mean by they vibrate? If we look at **Figure 8.5**, I am showing how the O-H "vibrates." This motion is the O-H bond stretching and compressing while the rest of the molecule stays unperturbed. With an IR-laser, we typically would see a feature that represents this mode when the laser is placed around 3575 cm^{-1} (wavenumbers). Wavenumbers is defined as the number of waves per distance and is used in IR spectroscopy since

Figure 8.5. The "vibrating" stretching motion of the carboxylic O-H. The O-H bond stretches and compresses. This resonant mode is usually found around 3575 cm^{-1}.

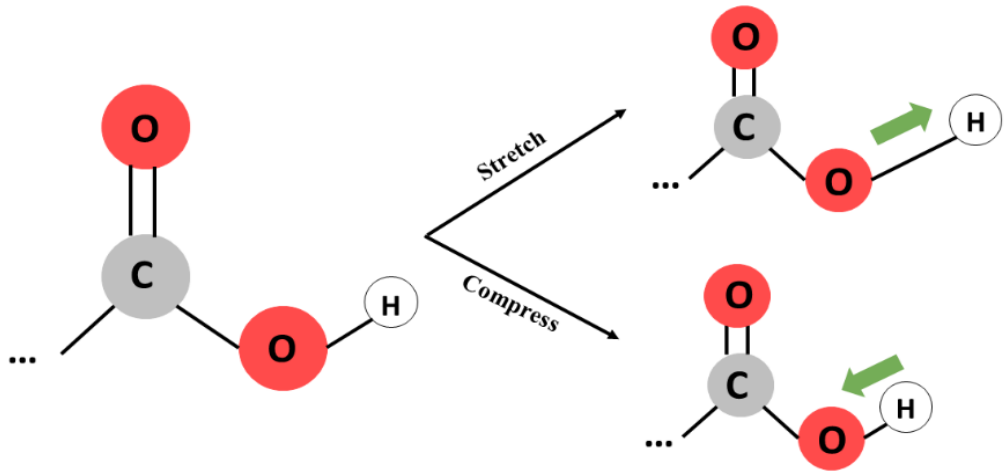
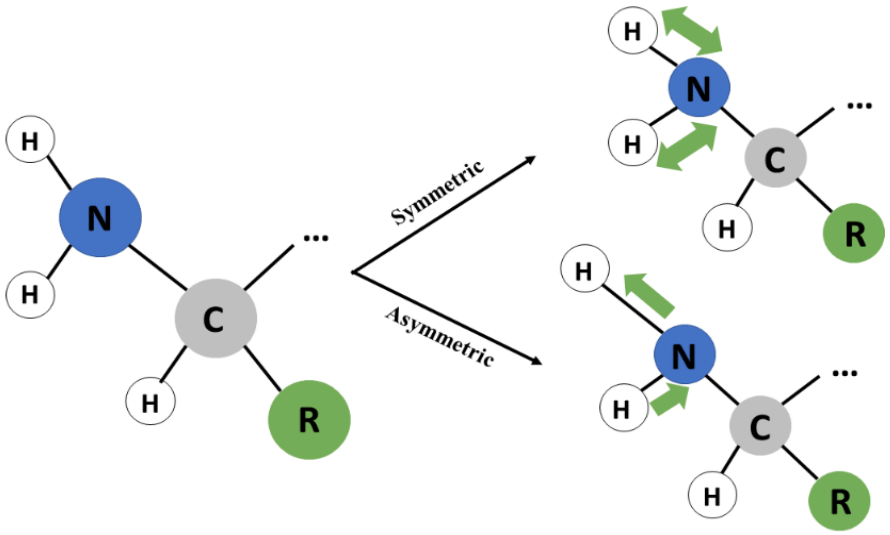


Figure 8.6. The "vibrating" stretching motion of the amine N-H. The N-H bonds stretch and compress. They can be either a symmetric mode where both N-H bonds stretch and compress in sync or they can be asymmetric mode where one N-H bond stretches while the other compresses and vice versa. These resonant modes are usually found around 3420 and 3470 cm^{-1}, respectively.



wavenumbers is proportional to the energy of the light; the higher the number, the greater the energy of light. Therefore, this mode would have a "resonant frequency" at 3575 cm^{-1}. An amine, which is at the other end of the peptide, has two different stretching motions, shown in **Figure 8.6**. It can have a "symmetric stretch" which is where both of the N-H bonds expand and compress in sync and it can have an "asymmetric stretch" where one N-H bond expands at the same time as the other compresses and then vice versa. These features usually have resonant frequencies around 3420 and 3470 cm^{-1} for the symmetric and asymmetric stretches, respectively.

When it comes to protein folding, there are interactions of the atoms within the proteins that we can probe with our IR laser. Hydrogen bonding is usually the main culprit for protein folding, both within the molecule and outside of the molecule by solvent. Hydrogen bonding is rather simple in definition. It requires there to be a hydrogen to be bonded to another atom (usually oxygen, fluorine, or nitrogen). This bond is weaker than the bonds within the molecule (the single and double bonds) but stronger than other types of interactions the molecule can have. With our IR laser, the interactions probed within (intramolecular) the folded molecule are usually by the N-H features hydrogen bonding to the C=O feature of the carboxylic acid. And once water is involved, water molecules like to hydrogen bond to the N-H and O-H features (intermolecular). When these perturbations happen from hydrogen bonding, we see a "shift" of peaks in our spectrum, usually to the smaller wavenumbers. This is termed "red shifting." By understanding where these features should be unperturbed, and then seeing the features change as a function of different R-groups (an abbreviation for the rest of the molecule) of the various amino acids, we can give information as to how the various R-groups affect the structure of the peptides.

As an example of a real IR spectrum, I am showing protonated Gly-Gly-Gly (GGG) and Ala-Ala-Ala (AAA) in **Figure 8.7**. The structures of linear GGG and AAA are shown in **Figure**

8.8. Glycine and alanine are the two simplest amino acids and only differ by the R-group: the R-group in glycine is a hydrogen and the R-group in alanine is a methyl group. As you can see, in the green region where the O-H stretch should be, there is no difference between the two peptides. However, in the blue and purple regions which are from the amine/amide and hydrogen bonded amine/amide, there are differences between the two, both in peak height and peak placement so we can assume that the methyl R-group that is alanine does something to perturb the folding structure of the backbone.

Figure 8.7. (Top panel) IR spectra of protonated Gly-Gly-Gly (GGG) and (bottom panel) Ala-Ala-Ala (AAA). The various stretching mode regions are highlighted in different colors.

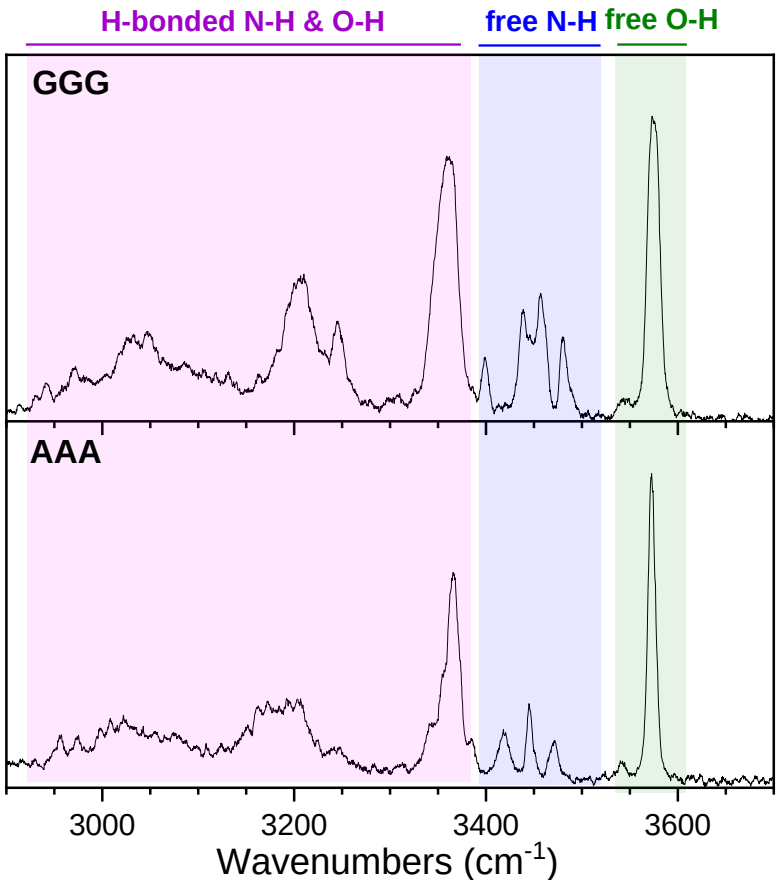
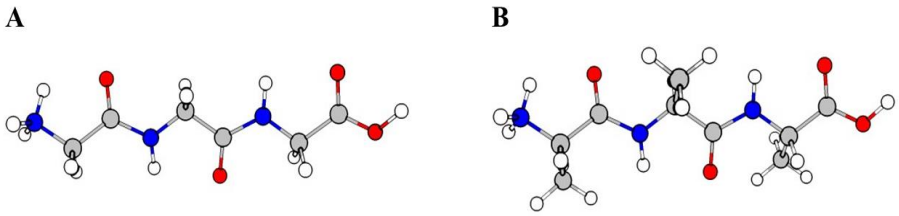


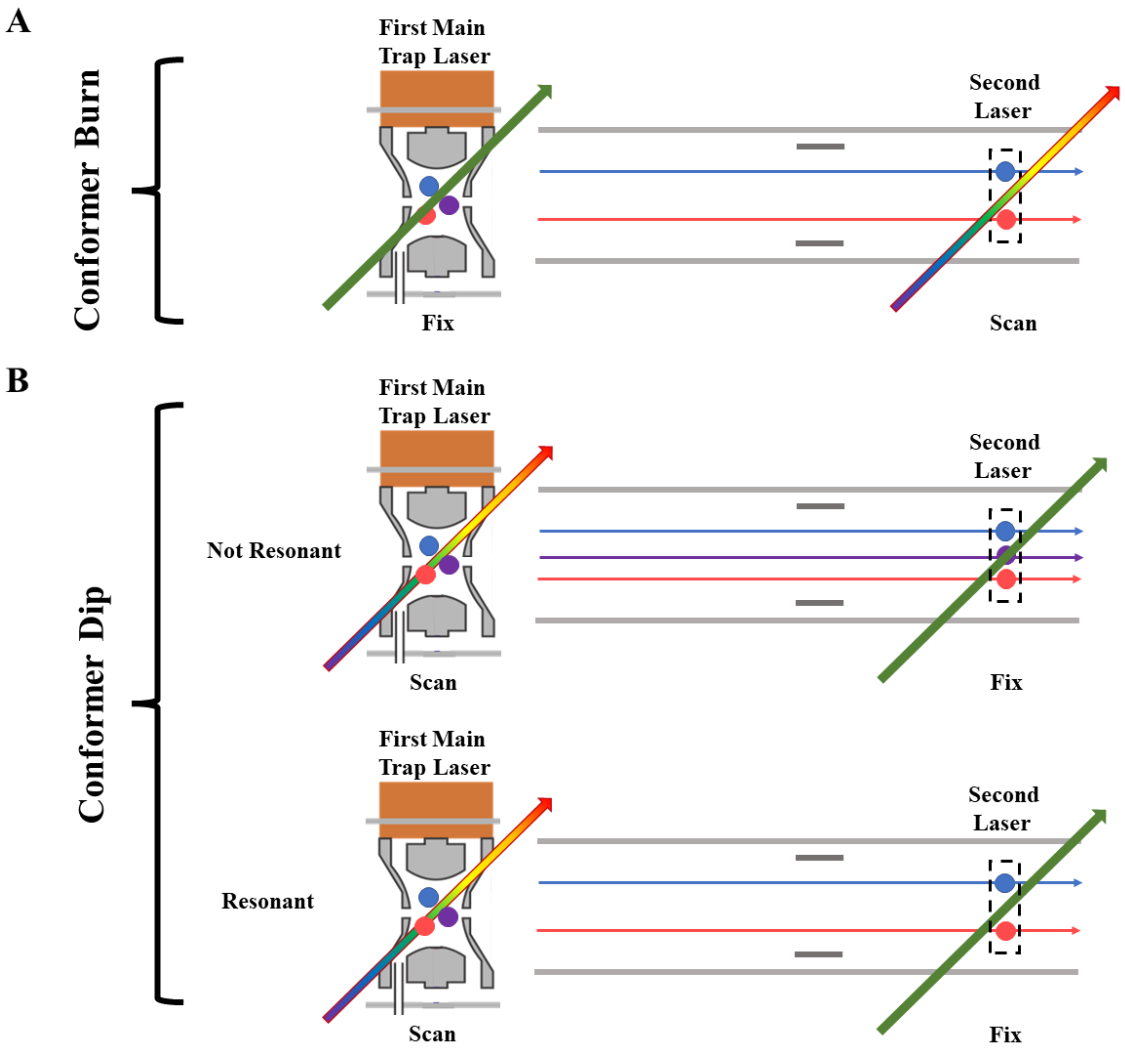
Figure 8.8. A) Protonated Gly-Gly-Gly (GGG). B) Protonated Ala-Ala-Ala (AAA). The difference between Gly and Ala is the R-group replacement of a hydrogen in Gly for a methyl group in Ala.



Since proteins and peptides are flexible, we can't expect them to be locked in one conformation, rather there are probably many conformations that exist both in solution and in the gas-phase. One way to know if there are more than one structure that you are seeing in your IR spectrum is to count the peaks. For instance, if we look at GGG we should have one O-H stretch peak, two amide N-H stretch peaks (since there are two amides), and three stretch N-H peaks that correspond to the protonated amine. This then means we should see 6 features in our IR spectrum if there was just one structure present. In just the green and blue highlighted regions there are 5 peaks and in the purple region there are at least 4 more. Since this total is larger than what we predict, then we know that there should be at minimum two different structures within this population. When a species has more than one structure that it can form into, we say that these species have multiple "conformers".

So how do we distinguish the multiple conformers? With our instrument set up, I described the case where we just had one laser at the end. However, we can add an additional laser to the main tagging trap stage and we can do IR-IR double resonance spectroscopy. Basically, this means we can do conformer selective spectroscopy. There are two techniques that we can use: fix the first laser that is in the main tagging trap and scan the second laser or we can flip the two and scan the first laser and fix the second laser. I am going to break down the two different techniques. You can also refer to **Figure 8.9** for a descriptive picture.

Figure 8.9. A) Schematic of conformer burn method. B) Schematic of conformer dip method both when the first main trap laser is not resonant and resonant.



1) Fix the first main trap laser and scan the second laser: This is what we call a conformer "burning" technique. Before describing the technique normally, I want to bring us back to the dog analogy. So, in our population, we can have more than one conformer which is like having more than one type of dog in our household. Each dog has its own quirks as well as its own "sweet spots" when petting it. If we have both dogs on their backs with something weakly attached to their foot, it is possible for me to get the spectrum of one dog without the effects of the other. If I know that one of the dogs is sensitive to scritches under his armpit and the other one isn't, I can scratch them both there at the same time. This leads one of the dogs to kick his back legs like crazy, which dislodges the weakly attached thing on their foot, while the other dog remains still. When I then look at who still has the tie on their foot, I see that the one lying still does, and then I can continue on with petting that dog until I see where their "sweet spot" is.

So now I am going to describe the technique without the analogy. Once we have the normal one laser spectrum, which is what I showed before and described in the instrumental set up, we can pick a feature that we believe might belong to one conformer (just like me scratching under the dog's armpit). From there, we fix the first laser on that wavelength. Since the laser is on, anything that is resonant with that wavelength will vibrate and pop off the D_2 tag in the main trap (as the dog did who had a sweet spot in the armpit). When the whole population leaves the main trap (me looking at which dog still has their tag on their foot), whatever does not have a tag gets kicked out of the system and what you are left with is anything that was not resonant with that wavelength, i.e. does not have that peak in the spectrum (just the dog lying still). With this technique you are getting rid of part of the population early on in the instrument and observing the spectrum of what is left through the second laser.

2) Scan the first main trap laser and fix the second laser: This is what we call a conformer "dip" technique. Before describing it normally, I want to bring back my dog analogy. So, let's say we have our two different dogs in the household, but I just want to understand where the Golden's "sweet spot" is without the effects of the other dog. I know that the golden loves getting petted on his ribcage and the other one is neutral about it. So, I am going to focus my pets there. First, I will start petting both, moving from the top of their belly to their legs, and at each step I am going to check who still has the tie around their leg. Then, I will pet the ribcage of whoever still has the tie around their leg. I am expecting my Golden to always shake the tie off when I pet his ribcage (therefore have a constant "signal"). However, if he shakes off the tie earlier when petting his belly, when I move to the second step of petting the ribcage, I won't see anyone shake their tie off because my Golden has already done it earlier! This is then a "dip" of signal in each place this happens.

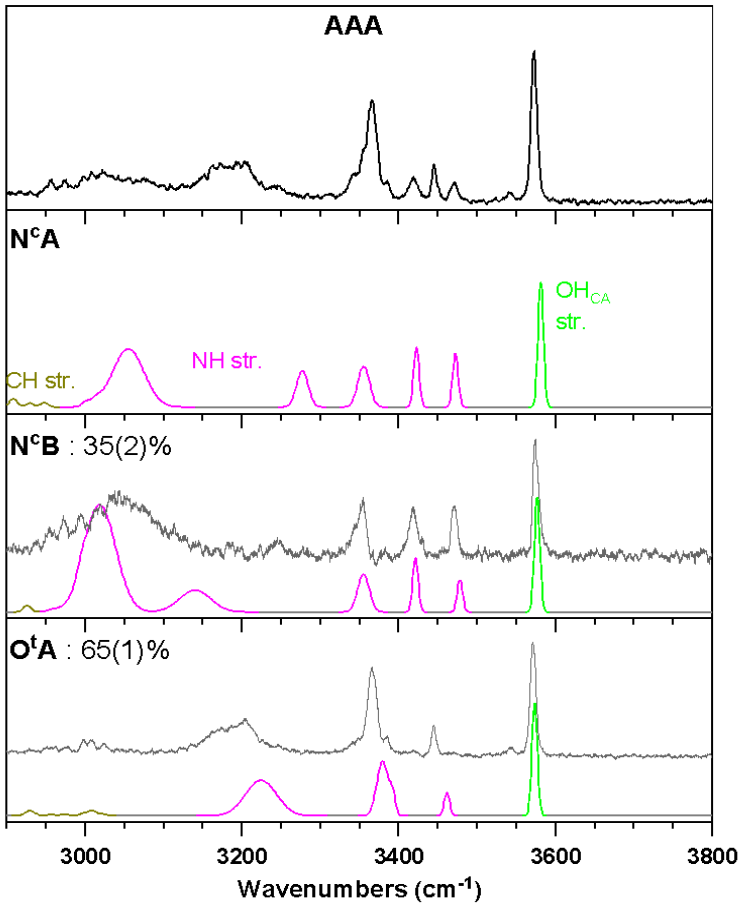
Now I will describe this technique plainly. Similar to the burn method, once we have the normal one laser spectrum, we pick a feature that we believe might belong to one conformer and we set our second laser at the end of the instrument to it (petting the ribcage of the Golden). This means that no matter what, we should see a signal from this conformer. When we scan the first main trap laser, if it is not resonant in the main trap, it passes through the rest of the instrument and to the second laser, so we see its constant signal. In the main trap, if the conformer that we are probing is resonant, the molecule vibrates and loses its D_2 tag and is kicked out of the system there (if the Golden loses its tie earlier during the pets). So, when we are probing with our second laser, we actually lose the signal that we have. The resulting spectrum is basically the dip of a signal and if we invert it, it is the spectrum of just the conformer that has this feature that we were probing.

8.6 Matching IR Data to Real Conformers

So now that we have seen some data and can physically see changes in peptide folding due to different R-groups, we want to be able to know how these peptides actually look. For this we use computations that are done on the computer. We are able to build our molecules with Gaussian 16 software[2] and utilize other programs given to us from collaborators in the Hopkins group at the University of Waterloo[3-6] to help us search ways in which the peptides can fold. From there we can optimize geometries and get vibrational frequencies of the molecules. These vibrational frequencies are calculated frequencies that predict where the molecules' various vibrations should appear. We can compare these to our spectrum and get determinations of what structures we are seeing in the gas-phase. An example of this is shown in **Figure 8.10**. Here I am showing the IR spectrum of AAA which you saw in the bottom panel of **Figure 8.7**, however, I am overlaying the conformer specific spectrum to the calculated computation. As you see, I am able to highlight what features are due to what stretches as predicted with the calculations and there is really good agreement.

Now knowing the specific conformations, we are able to make informed opinions on what we believe is occurring. We can use other sets of computations to help guide our reasoning until we get to the bottom of what is going on. And many times, it is not just one factor. There may be a combination of things occurring which leads us to the result that we obtained such as sterics of the R-group on the amino acid (which is if the group is bulky and obstructs the peptide's/protein's ability to fold), electron donating/withdrawing effects from the side chain, proton affinities, etc.

Figure 8.10. (Top panel) The IR spectrum of AAA compared to the (Bottom panels) lowest energy calculated spectra and the conformer specific ion dip spectra for each structure. Each structure's population percentage and respective errors, in parenthesis, are denoted next to their names.



8.7 Impact

This research has not just helped put another piece together of the larger puzzle that is research, but it has also taught me a lot. I did not come with this fundamental research background, and I also had to learn all these techniques from the ground up. And though the techniques might be very specialized, more importantly I learned how to think and problem solve like a scientist.

No matter what endeavors I take in the future, the skills I have learned through this research (critical thinking, working independently and in a group, perseverance, and many more skills) will be extremely beneficial and I can use them all to my advantage.

I hope as I have written this chapter you have begun to grasp some of these techniques that I use on my instrument to answer these fundamental questions. My particular research has focused on these small peptides and understanding how their R-groups and solvation affect their folding (as I introduced and discussed above). I have been able to progress the gas-phase spectroscopy community by showing and understanding that gas-phase structures do make sense and give us insight into what might be happening in solution phase. Though these questions may be fundamental in nature, they help us understand the facts that we have been told many times over without justification. They also help other scientists make informed decisions on peptides to make if they want a particular structure and therefore a particular function. This fundamental knowledge is the basis of all science.

And where do I go from here? I have learned so many valuable techniques from my work in my PhD. Not only that, I have also learned how to critically think about a problem and try to use the fundamentals to give an answer. I will be taking my knowledge and skills that I have learned and used in the Garand Group and will be going on and doing a postdoc in the Remucal Group at UW Madison. Though the science is very different as I am moving from a fundamental physical chemistry field to an environmental chemistry field, I will be able to use everything I have learned and hopefully bring new insights to this different research. In particular, I will be able to bring my expertise about how small changes in the structure of a chemical's backbone can lead to different effects in a different class of compounds, Per- and Polyfluoroalkyl Substances (PFAS), also known as the "forever chemicals."

8.8 References

(1) Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. Molecular Biology of the Cell. 4th edition. *New York: Garland Science* **2002**.

(2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.

(3) Campbell, J. L.; Yang, A. M.; Melo, L. R.; Hopkins, W. S. Studying Gas-Phase Interconversion of Tautomers Using Differential Mobility Spectrometry. *J Am Soc Mass Spectrom* **2016,** *27* (7), 1277-84.

(4) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-molecule clustering in differential mobility spectrometry: lessons learned from tetraalkylammonium cations and their isomers. *J Am Soc Mass Spectrom* **2014,** *25* (9), 1583-91.

(5) Lecours, M. J.; Chow, W. C.; Hopkins, W. S. Density functional theory study of Rh(n)S(0,+/-) and Rh(n+1)(0,+/-) (n = 1-9). *J Phys Chem A* **2014,** *118* (24), 4278-87.

(6) Liu, C.; Le Blanc, J. C.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using differential mobility spectrometry to measure ion solvation: an examination of the roles of solvents and ionic structures in separating quinoline-based drugs. *Analyst* **2015,** *140* (20), 6897-903.