Infrared Spectroscopic and Theoretical Characterization of a Selection of Biologically Relevant Gaseous Ionic Complexes

by

Jonathan K. Martens

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

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

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

Abstract

The structures of gaseous sodiated poly(alanines), $Ala_{8-12}Na^+$, were studied using infrared multiple photon dissociation (IRMPD) spectroscopy complimentarily to molecular dynamics and quantum chemical methods. Experimental results were obtained in both the 3000-3600 cm^{-1} region and the 1000-1800 cm^{-1} fingerprint region. Replica-exchange molecular dynamics (REMD) simulations using the multipolar-based polarizable AMOEBA force field were used to explore the conformational potential energy surface. A selection of resulting structures was extensively treated with a wide range of sequentially higher-order quantum chemical methods (from the RI-B3LYP-D/def2-SVP level to the RI-CC2/def2-TZVPP//RI-MP2/def2-SVP level). Calculated results suggest an α-helical conformation to be the lowest-energy for all $Ala_{8-12}Na^+$ peptides, and that only for Ala_8Na^+ does an accessible low energy structure exist that is also likely to contribute to the room temperature experiments. Experimental results confirmed these suggestions. All $Ala_{8-12}Na^+$ peptides have a free C-OH band in the 3575 cm^{-1} region resulting from the free C-terminal C-OH of the α-helical conformation, a necessary observation for α-helical poly(alanine) peptides. As well, bound N-H bands were found to create somewhat of an α-helical signature in the 3315-3435 cm^{-1} region. For a given peptide, the order of vibrational modes follows the order of N-H$\cdots$O=C hydrogen bond lengths (longer bond length gives higher frequency N-H vibration). Additionally for α-helical conformations, calculated N-H$\cdots$O=C hydrogen bond lengths consistently decrease from the N-terminus ($\approx$2.15 Å) to the C-terminus ($\approx$1.90 Å). This results in a consistent and similar N-H stretching band between 3315-3425 cm^{-1} of ordered N-H modes.

Proton-bound and sodium-bound-dimer ions $(Ala_6)_2H^+$, $(Ala_6)_2Na^+$, $(Ala_{12}Ala_6)H^+$, $(Ala_{12}Ala_6)Na^+$, $(Ala_{12})_2Na^+$ were considered under similar experimental and computational methodologies; however, IRMPD results were obtained only for $(Ala_6)_2H^+$. Computations indicate that in these sodium-bound and proton-bound dimers, α-helical type

conformations are dominant for all but the smallest dimers ($(Ala_6)_2H^+$, $(Ala_6)_2Na^+$), suggesting that the interaction of Ala_6H^+ or Ala_6Na^+ with neutral Ala_6 is not sufficiently stabilizing to the helical conformation. In the globular conformations, intermolecular and intramolecular hydrogen bonds and coordination of the Na^+ are favoured at the expense of a more extensive intramolecular helical-type hydrogen bond network. The lowest-energy globular conformation of Ala_6Na^+ has both C-terminal COOH groups, one N-terminal NH_2 group and an additional two C=O groups interacting with Na^+. Such a coordination of the Na^+ would not be possible if either or both peptides were helical, where instead, a low-coordination with three or four C=O groups would result. The relative stability of the α-helical conformation was found to be largely related to the stabilization of the macrodipole of the helix. Interaction of two peptides was, in some cases, found to favour formation of α-helical structure even where the isolated peptides would be globular, for example $Ala_{12}H^+$. Interaction of the protonated NH_3^+ group with the C-terminal C=O of the neutral peptide places the positive charge somewhat on the exterior of the structure, apparently limiting the unfavourable electrostatic interaction of the positive charge with the dipole (positive) of the helix. $(Ala_{12})_2Na^+$ is a somewhat obvious case, where both monomer units of the dimer are calculated to be independently helical. Although when isolated, $Ala_{12}Na^+$ is calculated to be completely helical through all 12 residues, more complete coordination of the sodium by C=O groups (disrupting, to a small extent, the helix) is calculated to occur. This is likely a result of the interaction of the macrodipole of the helix and the sodium cation being weaker as a result of the interaction between the N-terminus of $Ala_{12}Na^+$ and the C-terminus of the neutral Ala_{12}.

Additionally, a series of gaseous ionic clusters formed from guanidinium chloride, $[Guan_xCl_{(x-1)}]^+$ and $[Guan_xCl_{(x+1)}]^-$, were studied under a similar computational methodology. IRMPD spectra were obtained in both the 3000-3600 cm^{-1} region and the 1000-1800 cm^{-1} fingerprint region. REMD simulations were found to be necessary for locating conformational minima in clusters containing more than four guanidinium ions ($x \geq 4$).

Experimental spectra were found to be well modelled at the B3LYP-D/def2-SVP level, in most cases. M06/6-311+G(d,p) vibrational analysis was found to be an improvement over B3LYP-D/def2-SVP analysis, at only a moderate increase in computational cost, while the significantly more costly MP2/6-31G(d) results did not offer improvement to the M06/6-311+G(d,p) results over a limited test set. Although only practically possible for the smallest clusters, consideration of anharmonic vibrational frequencies at the B3LYP/6-311+G(d,p) level offered significant improvement over harmonically calculated modes. Generally, smaller clusters were noted to largely adopt high-symmetry conformations, while larger clusters formed cage-like structures with a small number of guanidinium ions contained in the interior.

Acknowledgements

I am very grateful to my supervisor, Prof. Terry McMahon, for the exceptional scientific supervision, personal guidance, and for the many opportunities made available to me throughout my degree. Put simply, Terry's supervision allowed me a PhD experience much beyond expectation. Not only was it a successful and life-shaping experience for me, but it was, as well, very enjoyable. I will keep many great memories from research trips, group dinners and conferences which were directly a result of Terry's generosity.

Additionally, I was very fortunate to have excellent supervision from Dr. Gilles Ohanessian and Dr. Carine Clavaguéra during my time at École Polytechnique. I gained a tremendous amount of experience and knowledge from their guidance and direction (which, generously, continued well after my time at Polytechnique was already finished).

I was very fortunate to work with a number of very talented people in Canada and in France. Firstly and most importantly, I want to acknowledge Dr. Rick Marta who was both an ideal co-worker and friend. Rick taught me a great amount about chemistry, both through his impressive knowledge and his excellent teaching and communication abilities. Working together in the lab with my sister, Sabrina, was an opportunity that I thoroughly appreciated and one which not many people are lucky enough to have. With Sabrina and Rick, working together in the lab and relaxing together outside of the lab was always enjoyable, interesting and enriching. I also am thankful for the experience gained while working with Dr. Isabelle Compagnon and Edith Nicol during experiments at CLIO. As well, I would like to thank everyone else at DCMR for a great experience and for the friendly and kind atmosphere in the lab.

In no particular order I also want to acknowledge Dr. Richard Smith, Dr. Carey Bissonnette, Dr. Jake Fischer, Prof. Bill Power, Prof. Marcel Nooijen, Prof. Bob Le Roy and all the other people who gave me superb academic and laboratory instruction while at

the University of Waterloo. I am also grateful to Vincent Steinmetz (CLIO), Dr. Philippe Maître (CLIO), Dr. Joël Lemaire (CLIO) and the staff and operators at CLIO for the great support and assistance with experiments in Orsay. Dr. Ashwani Sharma (Polytechnique), Dr. Rob Nieckarz, Dr. Ronghu Wu, Kristina Lekin and Michael Makahnouk were great co-workers and I am thankful for the many helpful discussions. I also would like to thank Cathy van Esch for her nearly continual assistance during my degree (and the associated patience and kindness).

My parents were, in many different ways, crucial to all my academic successes and I am very appreciative of them. My interest to chemistry was directly a result of the influences of my grandfather, Dr. Dimitrios Favis, and my uncle, Prof. Basil Favis. I am additionally thankful to Basil for the very valuable guidance and advice in many important decisions which shaped the direction of my university experience.

Finally, I would like to acknowledge the University of Waterloo, the government of Ontario, NSERC and the vice-presidency for external relations at École Polytechnique for financial support.

Table of Contents

List of Figures

List of Tables

Nomenclature

B3LYP	A three-parameter Becke-Lee-Yang-Parr functional
CID	Collision Induced Dissociation
CLIO	Centre Laser Infrarouge d'Orsay
CRM	Charge Residue Model
DFT	Density Functional Theory
ECD	Electron Capture Dissociation
ESI	Electrospray Ionization
FEL	Free Electron Laser
FELIX	Free Electron Laser for Infrared eXperiments
FT-ICR	Fourier Transform Ion Cyclotron Resonance
GGA	Generalized Gradient Approximation
GPCRs	G-Protein-Coupled-Receptors
HeNe	HeliumNeon laser
IEM	Ion Evaporation Model

IR	Infrared
IRMPD	Infrared Multiple Photon Dissociation
IVR	Intramolecular Vibrational Redistribution
KTP	Potassium Titanyl Arsenate
KTP	Potassium Titanyl Phosphate
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
M	Molarity, moles L^{-1}
m/z	Mass/Charge
MP2	Møller-Plesset perturbation theory, second order
MS	Mass Spectrometry/Mass Spectrum
Nd:YAG	Neodymium-doped Yttrium Aluminium Garnet
OPA	Optical Parametric Amplifier
OPO	Optical Parametric Oscillator
PES	Potential Energy Surface
QIT	Quadrupole Ion Trap
RF	Radio Frequency
RI	Resolution of the Identity
RMSD	Root Mean Square Deviation

SCF	Self-Consistent Field
SP	Single Point (electronic energy calculation)
TFA	TriFluoroacetic Acid
TFE	TriFluoroEthanol

Chapter 1

Introduction

The functionality and properties of proteins and other biological molecules are largely determined by their structure; therefore, an understanding of the role of amino acid sequence in determining the larger-scale conformation of peptides and proteins is crucial in the prediction of physical properties and the understanding of biochemical mechanisms. The structure of a folded protein is controlled by a complex combination of factors, all of which begin with the sequence of amino acids. Sequence determines non-covalent interactions, such as hydrogen bonding and hydrophobic interactions, which are heavily influential in determining all levels of structure. These intramolecular non-covalent interactions result in the formation of secondary structure (local structure) such as helices, sheets, and turns. Metal-ion complexation also plays a large role in many biochemical systems, and between one-quarter and one-half of proteins incorporate metal-ions (metalloproteins).[1] Many studies have focused on the influence of metal-ions on the stability of gas-phase structures and conformations of amino acids (zwitterionic conformations particularly) and secondary structure of peptides and proteins.[2,3]

A recently published structure by Warne et al. of a β_1-adrenergic G-protein-coupled-receptor (GPRC) serves as an illustrative example.[4] GPCRs are integral membrane proteins

whose function involves signal transduction from the outside to the inside of cells. Agonists (such as hormones or other activating molecules) bind to the protein from the exterior of the cell resulting in large conformational changes and activation of the protein, finally resulting in changes in the levels of intracellular messengers. The β_1-adrenergic receptor was found to have an α-helical extracellular loop defining the entrance to the ligand binding site. At the C-terminal end of the α-helical loop, a sodium cation was found, interacting with the α-helical dipole and thus stabilizing the helical conformation of the loop. It has been suggested that this sodium-bound α-helical loop, regulating the entrance to the ligand- binding site, may be a common feature of G-protein-coupled-receptors that bind in a rapid and reversible manner.[4] Drugs inhibiting β-andrenergic receptors have been used to modulate heart function (β-blockers), and are currently being pursued as potential drugs for the treatment of obesity through inhibition of adrenaline-based metabolic stimulation. Thus, understanding the relationship between amino acid sequence, secondary structure and the complexation of metal-ions is clearly of the utmost importance to understanding biochemical structure and function.

Due to the size and complexity of most proteins and peptides found in biological systems, it can be difficult to understand the influence of individual driving forces in the formation of the structural features in these molecules. Minimization of interactions between hydrophobic regions of proteins and the solvent, usually water, plays a large role. To simplify the situation, a combination of mass spectrometry and gas phase spectroscopy can be used to examine the individual interactions that are present in a typical peptide structure.

Ion spectroscopy of small peptides and other model systems has received much attention in recent years.[2,5,6] Spectroscopy of isolated gas-phase ions provides an ideal experiment for the structural study of biological molecules as well as benchmarking of theoretical models. Experimental advances now allow for the long-term trapping of gas-phase ions,

from even very large and fragile biomolecular complexes, in high-vacuum systems. This isolated environment allows the examination of structural features in the absence of intermolecular and solvent interactions. These experiments are also ideally suited to be modelled computationally, since they deal with the isolated environments naturally found in theoretical models. The combination of computational and experimental methodology opens a vast source of information, and will be discussed in more detail in the following chapters. The spectroscopy described here will remain confined to the infrared vibrational region, as a large amount of information can be obtained about biomolecular structure from bond vibrations in the 1000-2000 cm^{-1} fingerprint region, as well as in the higher-energy 3000-3600 cm^{-1} region of N-H and O-H stretches.

Chapter 2

Experimental and Computational Methods

2.1 Experimental Methods

2.1.1 Infrared Multiple Photon Dissociation Spectroscopy

Infrared multiple photon dissociation (IRMPD) spectroscopy provides a vibrational spectrum of gaseous ions through the observation of wavelength-dependent fragmentation in a collection of irradiated ions. Monitoring the extent of fragmentation induced by a tunable light source, allows the use of a mass spectrometer as a highly sensitive detector for vibrational spectroscopy measurements. IRMPD spectroscopy is a form of consequence (or action) spectroscopy where a result of the event is monitored, in this case fragmentation resulting from resonant photon absorption. The necessity for using this indirect measurement technique rather than a direct absorption experiment derives from the low sensitivity of direct absorption methods which require number densities of approximately 10^{10} molecules per cm^3. Coulombic repulsion in a group of ions of the same charge limits

maximum practicably attainable number densities to levels quite significantly below this ($\lessapprox 10^8$ molecules per cm^3).[7,8]

IRMPD was first demonstrated in the early 1970's.[9] At that time, tunable infrared light sources having sufficient intensities to induce fragmentation in most covalently bound molecules did not exist and the technique was, therefore, quite limited. A static-frequency, resonant laser used for IRMPD was a new fragmentation tool (and mechanism) in mass spectrometry/mass spectrometry (MS/MS, MS^n) experiments, similar to collision induced dissociation (CID). Starting in the mid 1970s, the groups of Dunbar, Beauchamp, Lee and Eyler were the first to use a tunable infrared laser to induce wavelength dependent fragmentation and observe a vibrational spectrum of gaseous ions.[10–15] In many of these early experiments, two lasers were used; a tunable, resonant laser to vibrationally excite ions and a second, non-resonant laser was used to dissociate the vibrationally excited ions (more discussion of this technique will follow later in the chapter). Following these first experiments up to the present time, one can clearly see that advances in experimental techniques, such as the free electron laser (FEL), the optical parametric oscillator (OPO) and electrospray ionization (ESI) sources, have greatly advanced IRMPD studies. A detailed discussion of the mechanism and instrumentation will follow in this section.

The simplest treatment of molecular vibrations involves their treatment as quantum harmonic oscillators.[16] Starting from the classical harmonic oscillator, the frequency, ν, and potential energy, U, of a diatomic vibration can be expressed in the following way, where k is the force constant, $\mu = (m_A m_B)/(m_A + m_B)$ is the reduced mass and x, the displacement from the equilibrium position:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{2.1}$$

$$U = \frac{1}{2}kx^2 \tag{2.2}$$

The energy levels of a quantum harmonic oscillator can be obtained by solving the time-independent Schrödinger equation with the potential energy expressed as Equation 2.2 and have the following form, where v=0,1,2,3... and $h\nu = \hbar\omega$:

$$E_n = \hbar\omega(v + \frac{1}{2}) \tag{2.3}$$

In this description, energy levels in a given mode are evenly separated by integer values of $\hbar\omega$ and the zero-point energy, $E_0 = 1/2\hbar\omega$, is obtained when v=0. A parabolic potential, centred on the equilibrium bond distance, describes the potential energy.

Unfortunately, this description presents a couple of inconsistencies with real molecular bonds. Description of the potential energy by a quadratic function proposes an infinitely increasing restoring force as the bond length moves farther from the equilibrium distance (in other words, it does not incorporate the concept of dissociation or a dissociation energy). This assumption is generally adequate in cases where the displacement about the equilibrium distance is small, however, for large displacements (and especially at energies approaching the dissociation limit) this is clearly not the case. At the dissociation limit of a bond, the potential energy will reach a plateau and the restoring force will be zero. Full treatment of these deviations from harmonic behaviour is accomplished through anharmonic treatment of the vibration and is quite difficult (practically, only accomplished for relatively small molecules), however, its conceptual consideration is necessary to properly understand the IRMPD process. The most obvious problem associated with harmonic treatment is associated with the use of a quadratic potential energy function. The use of a Morse function allows the incorporation of a dissociation limit, where D_e is the dissociation limit, x is the deviation from the equilibrium distance, and $\alpha = \hbar\omega\sqrt{\mu/(2D_e)}$.

$$U = D_e(1 - e^{\alpha x})^2 \tag{2.4}$$

In the simplest form, anharmonic energy levels can be described by the following formula, where $\chi = \hbar\omega/4D_e$:

$$E_n = \hbar\omega(v + \frac{1}{2}) \left[1 - \chi(v + \frac{1}{2})\right] \tag{2.5}$$

As χ is a small positive value, the anharmonic expression requires that there will exist decreasing level spacings as the ladder of vibrational levels is climbed and the dissociation limit is approached.

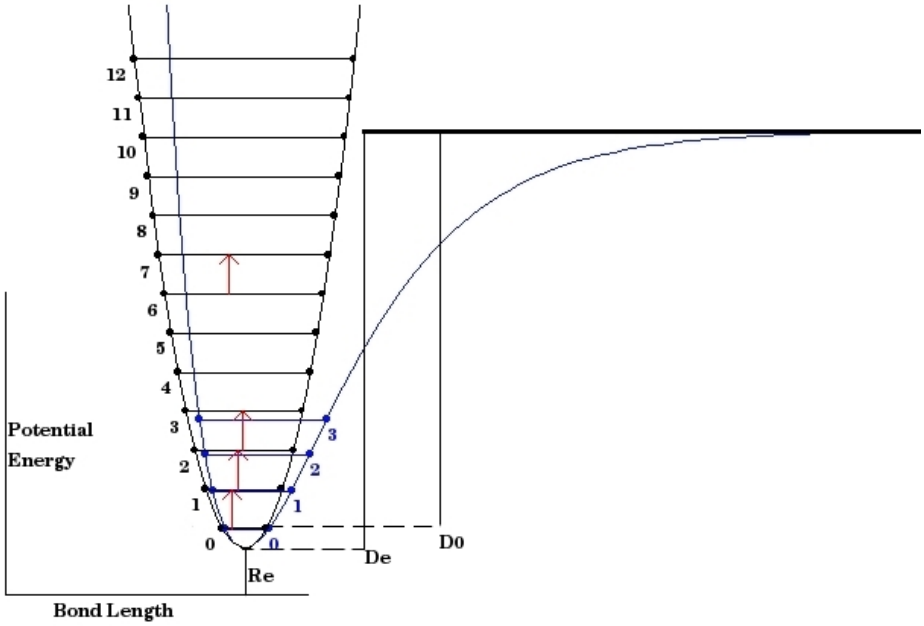

Figure 2.1 – Harmonic potential (**Black**) and anharmonic potential (**Blue**), where D_e is the well depth, D_0 the dissociation energy from zero-point energy and R_e the equilibrium bond distance.

With these considerations in mind, it is now possible to describe the IRMPD process itself. The IRMPD mechanism begins with the absorption of a resonant photon. Intramolecular vibrational redistribution (IVR) of the photon energy from the resonant or

"bright" state into many anharmonically coupled, background or "dark" vibrational states follows and allows for the system to again absorb photons resonant with the fundamental ($v = 0 \rightarrow v = 1$) transition of the bright state. Although a quantitative and explicit treatment of the IVR mechanism and IVR rates is rather complicated, a qualitative description and consideration of the consequences resulting from this process provides valuable insight into the IRMPD process. IRMPD is often referred to as non-coherent multiple photon absorption, in contrast to coherent multiple photon absorption which would represent a "ladder-climbing" mechanism ($v = 0 \rightarrow v = 1 \rightarrow v = 2$) and would be inhibited by the decreasing level spacings associated with going to higher vibration levels in anharmonic modes (referred to as the anharmonic bottleneck).[17] If one considers that tens to hundreds of photons must be absorbed to provide the minimum amount of energy required to fragment most bonds (each mole of photons at 500 cm^{-1} has 6 kJ and at 2000 cm^{-1}, 24 kJ), the process of coherent absorption-induced fragmentation seems quite unlikely. After the absorption of a few photons, the decreased level spacings would mean that the previously resonant photons would now be off-resonant to the given mode. There have been cases in which chemical barriers have been used to isolate parts of a molecular system from participating in IVR,[18] but in general it can be assumed that re-distribution quickly (on the picosecond timescale) occurs statistically and throughout the entire system.[19] Fragmentation is then observed through the lowest-energy mechanism or, in some cases, multiple mechanisms.

As the ion absorbs many photons, the internal energy distributed throughout the many background states increases, resulting in an increased density of populated states and width of the absorption band. At sufficiently high internal energies (depending on the size of the system) it can be said that the ion is in a quasi-continuum of states and the absorption band is sufficiently wide that it is difficult to consider the absorption as a resonant process, in reference to the original frequency. The size and internal energy dependence of the density of states, $\rho(E)$, can be seen in the following quasi-classical

approximation of the density of states, where E represents the vibrational energy, s is the number of vibrational degrees of freedom and ω_i the frequency of the ith vibrational mode.[7,8]

$$\rho(E) = \frac{1}{(s-1)!} \frac{E^{s-1}}{\prod_{i=1}^{s} \hbar\omega_i} \tag{2.6}$$

Although a more complicated expression is necessary for an accurate description, the correlation of increasing density of states with increasing internal energy and size of the system is illustrated by Equation 2.6. This increase in density of states at high internal energies allows for the absorption of a wider range of photon energies following initial resonant absorption. Further discussion can be found in a 2006 review[7] as well as other discussions related to IVR and IRMPD.[20–27]

As mentioned above, tens to hundreds of photons are required to induce fragmentation under typical circumstances, demanding the use of high-intensity light sources. The first IRMPD experiments used a two-laser probe-pump setup, since at that time high-power tunable infrared lasers did not exist. However, development of the FEL and OPO have significantly widened the scope of IRMPD studies. The FEL user facilities at the Centre Laser Infrarouge d'Orsay (CLIO)[28,29] of the Université Paris-Sud (Orsay, France) and the Free Electron Laser for Infrared eXperiments (FELIX)[30,31] (Utrecht, Netherlands) have been very successful in their application to IRMPD experiments. Table-top OPO setups are used in many labs as they are relatively inexpensive and simple to operate and maintain. Both will be discussed in more technical detail below. A vast number of IRMPD spectra can now be found in the literature, with a large number originating from the FEL user facilities. In a notable illustration of the strength of the FEL, spectra of different charge states of cyctocrome c, a 104 amino acid protein, recorded the main stretching bands (amide I and amide II modes) as well as determined from the position of the amide I

band that the protein was mostly α-helical.[32] All spectra presented in this thesis were obtained using the CLIO FEL for the 1000-2000 cm^{-1} fingerprint region and an OPO/OPA, also at CLIO, for the 3000-3600 cm^{-1} spectral region.

Finally, a convenient way to present the degree of fragmentation of the parent ion(s) to fragment ions is the IRMPD efficiency.

$$\texttt{IRMPD}_{efficiency} = -log\left(\frac{I_{parent}}{I_{parent} + \sum I_{fragments}}\right) \tag{2.7}$$

Expressing the extent of fragmentation in this way eliminates total ion current (TIC) signal fluctuations from appearing in the final spectrum (for example, instabilities in the spray current from an electrospray source). In an ideal case, IRMPD efficiencies in the range 0.05-0.2 are sought, however it is not always possible to maintain fragmentation in this range, as differences in absorption intensities between modes can require largely different laser powers. For this reason, some spectra in the following chapters will be presented not as a single scan over the full spectral range, but will be composed of a few scans over different ranges finally pieced together. For example, when studying peptides, it is often advantageous to divide the scan over the fingerprint region into three pieces, with the amide I band, amide II band and the 1000-1400 cm^{-1} regions being obtained in separate scans. By changing the laser intensity in each part of the scan (in practice, the irradiation time is changed and the laser intensity is not modified) relatively consistent levels of fragmentation can be obtained over the full spectral range. This is in contrast to a single scan where either the amide I band (high intensity) would be completely saturated at high laser power, or the 1000-1400 cm^{-1} region (low-intensity) would show no fragmentation at weak laser power.

Mass Spectrometry

IRMPD spectroscopy requires a population of trapped ions, as well as an ability to quantitate the extent of fragmentation resulting from the absorption of resonant infrared light. Ion trap mass spectrometry provides the solution to both of these requirements. Both Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry and quadrupole ion trap (QIT) mass spectrometry[33] (as well as a few other examples with various other types of mass spectrometers[12,34,35]) are currently and successfully being used for IRMPD spectroscopy.[36,37] The work presented in this thesis mainly used FT-ICR mass spectrometry for the collection of IRMPD spectra.

Fourier Transform Ion Cyclotron Resonance (FT-ICR) Mass Spectrometry

Since the first publication on the topic of FT-ICR mass spectrometry in 1974 by Comisarow and Marshall,[38] the technique has seen enormous growth and application. Initially demonstrated by the collection of a spectrum, with 8:1 signal to noise ratio, of CH_4^+ at a resolution of 0.005 amu, the FT-ICR technique has evolved into what is possibly the most powerful mass spectrometric technique currently available.[39] Protein identification from sub-attomole samples, trapping and detection of Megadalton ions, non-destructive detection and ultrahigh-mass accuracy (in notable examples up to 10^{-9} Da[40]) clearly illustrate the strengths of this technique.[41]

The following is a brief description of ion behavior in the cell of the FT-ICR.[41] Ions travelling in a spatially uniform magnetic field, $\vec{B}$, are subject to a force given by the Lorentz force equation

$$\texttt{Force} = \texttt{mass} \cdot \texttt{acceleration} = m\frac{dv}{dt} = q\vec{v} \times \vec{B} \tag{2.8}$$

where m is the mass of the ion, q is the charge of the ion and $\vec{v}$ is the velocity. The cross product of $\vec{v}$ and $\vec{B}$ defines the the direction of the Lorentz force to be perpendicular to the plane defined by $\vec{v}$ and $\vec{B}$ and results in a trajectory of the ion corresponding to radius, r, shown in Figure 2.2. Defining the velocity in the xy plane to be v_{xy}, writing Equation 2.8 in terms of the angular acceleration ($|dv/dt| = v_{xy}^2/r$)

$$\frac{mv_{xy}^2}{r} = qv_{xy}B \tag{2.9}$$

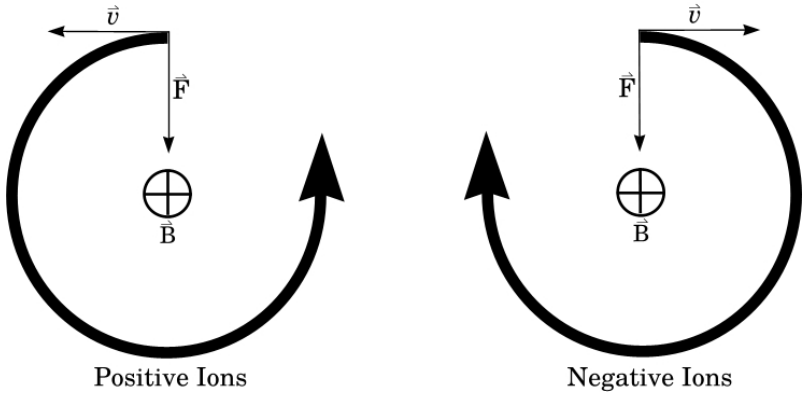

Figure 2.2 – The Lorentz force experienced by an ion in a uniform magnetic field.

and substituting in the angular velocity about the z-axis, $\omega = v_{xy/}r$, produces the following expressions:

$$m\omega^2 r = qB\omega r \tag{2.10}$$

$$\omega = \frac{qB}{m} \tag{2.11}$$

Writing Equation 2.11 in terms of frequency, ν , instead of velocity, one obtains the ion cyclotron frequency, ν_c, in Hz, B in T, m in amu, and q in terms of the total number, z, of elementary charges, e, of the ion ($q = ze$):

$$\nu_c = \frac{\omega}{2\pi} = \frac{1.536 \times 10^7 B}{m/z} \tag{2.12}$$

A few important observations should be made about Equation 2.12. Firstly, ions of the same m/z ratio will have the same ICR frequency regardless of their velocity. From a practical standpoint, this is important because not all ions of the same m/z are required to enter the cell with the same velocity. Secondly, the radii of the cyclotron motion of the ions will be defined as the following (where the average xy translational energy is: $kT \approx m\langle v_{xy}^2 \rangle/2$):

$$r = \frac{mv_{xy}}{qB} = \frac{1}{qB}\sqrt{2mkT} \tag{2.13}$$

From this expression it can easily be calculated that in a 7.0 T magnetic field, a singly charged ion at room temperature and of mass 10.0 amu will have a radius of cyclotron motion of approximately 0.01 mm, one of mass 1000 amu a radius of 0.1 mm, and one of mass 100,000 amu a radius of about 1 mm. From Equation 2.12 it can be shown that for ions with m/z in the range of approximately 20 to 10,000, the frequency of cyclotron motion will be in the range of kHz to MHz. From this we can see that in a 7.0 T magnetic field an ion of 1000 m/z, with its radius of motion of 0.1 mm, has a velocity of ≈70 m/s (note that if the same ion was excited to a radius of 1 cm it would have a velocity of ≈7 km/s).

The magnetic field only traps ion motion in two dimensions and a small potential is applied to two trapping plates to confine them in the third dimension as seen in Figure 2.3. While trapped in cyclotron motion, the ions can be excited by the excitation plates to bring them closer to the two detection plates where they induce an image current. A Fourier transformation of the time-domain image current produces a frequency domain signal which can be related to the mass of the ions by Equation 2.12. A radio frequency

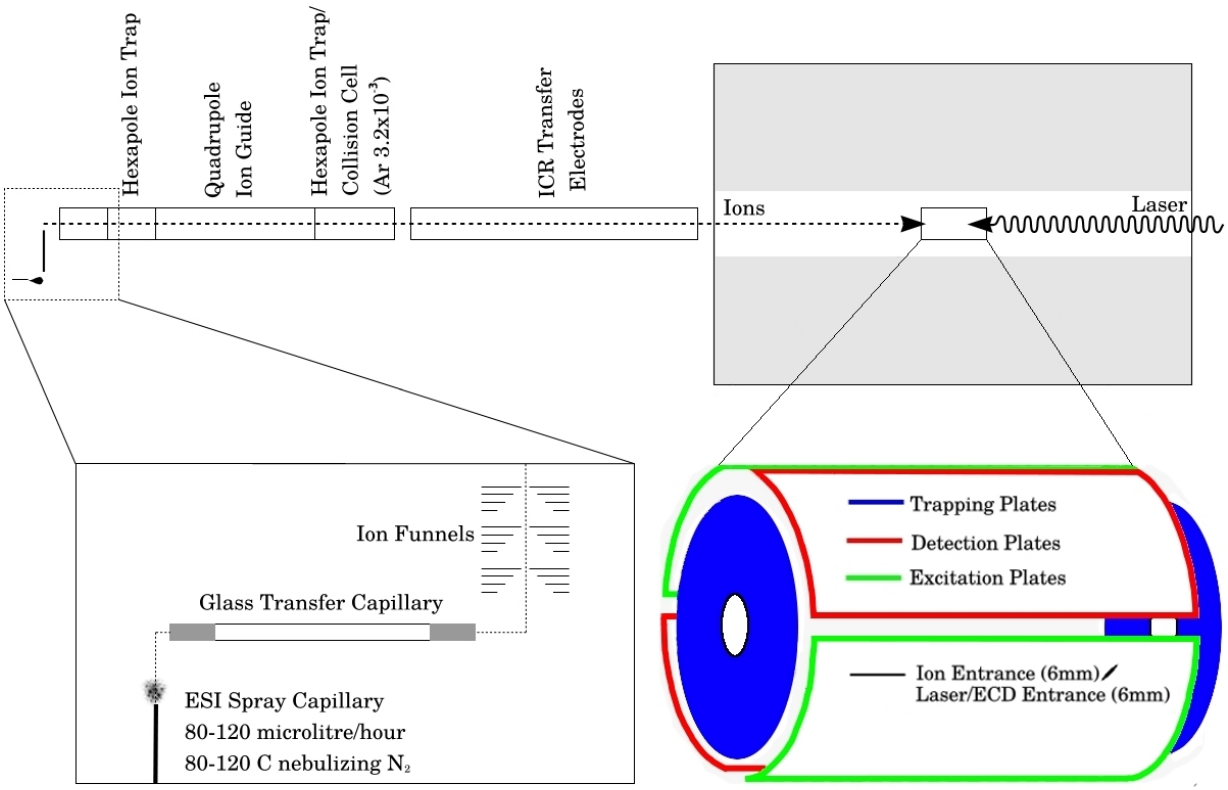

Figure 2.3 – Example schematic of an ESI source, mass spectrometer and ICR cell, approximately based on the ESI/FT-ICR instrument at CLIO.[42]

(RF) pulse from the excitation plates, resonant with the cyclotron motion of an ion of particular m/z, excites those ions to a larger radius of motion. This excitation process can be used for multiple purposes. As the radius of motion of the ions increases, their velocity (and kinetic energy) also increases. If a collision gas is present in the cell, this excitation can be used to induce collisional dissociation. As well, specific ions can be "ejected" if the RF signal is set to increase the cyclotron radius beyond the physical dimensions of the cell. If all ions are to be removed from the cell, the trapping plates are most often used (an event called quenching) by applying a small potential across them. In general, high-vacuum conditions are maintained in the ICR cell to minimize collisions, however, when desired it is possible to introduce gas through a molecular leak valve for the purpose of collision induced dissociation or observation of ion-molecule reactions.

In application to IRMPD experiments, the FT-ICR has many advantages. The main

strengths of the FT-ICR as a mass spectrometer discussed up to this point have related to high-mass resolution and high-mass range. High-mass range is not particularly important, as most ions studied by IRMPD are of relatively low m/z (>1000 m/z). Moderately high mass resolution is important for resolving and trapping/ejecting nominally isobaric ions. More importantly, the ability to efficiently trap ions for long periods of time (and as a result, irradiate them for long periods of time) in an essentially collision-free environment proves very useful, and provides a clear advantage over other trapping techniques. The ability to eject ions of specific m/z , in addition to isolating only ions of a specific m/z is also very useful. A final advantage is the capability to adjust the trapping parameters, allowing the ion packet in the ICR cell to be, to some extent manipulated in order to optimize the overlap with the laser beam.

Electrospray Ionization

To this point, the discussion has omitted a very important aspect of the mass spectrometer, that is, the ion source. While many ion sources exist and cover a wide range of applications, perhaps the most widely used and versatile is the electrospray ion source. Electrospray ionization (ESI) is a remarkable ionization technique, both in its simplicity and in its capabilities, producing ions from peptides, nucleic acids, carbohydrates, inorganic salts, polar organic compounds, as well as from a wide range of polymers. The main practical limitation in ESI is that the analyte must be sufficiently polar to be ionized in solution. All experimental results that will be presented in this thesis used ESI to produce ions.

ESI can be thought of as a soft ionization technique that takes ions from solution and directly transfers them to the gas-phase. In most cases this statement seems to hold true, however in some cases it is apparently somewhat more complicated and it is useful to consider how the transfer from solution to gas-phase occurs. Figure 2.3 shows the main components of the ESI source: the spray capillary, the counter electrode, the transfer

capillary (not present in all source designs) and the entrance to the mass spectrometer. The ionization process occurs in the atmospheric pressure region of the instrument before isolated ions enter the first stages of the mass spectrometer. The use of a transfer capillary serves two purposes, first as a small orifice to the mass spectrometer and secondly as a region to desolvate ions. The transfer capillary is often aligned orthogonally to the spray capillary and the entrance to the mass spectrometer in order to further desolvate and separate ions from evaporated solvent as they traverse the increasing vacuum gradient on their path to the mass spectrometer. Solution is provided by a microsyringe and syringe pump at flow rates usually on the order of 80-200 μl/hour. Typically, polar solvents such as mixtures of methanol and water or acetonitrile and water, containing analyte concentrations on the order of 10^{-6}-10^{-3} mol L^{-1} (M) , are used. A potential difference of 3-5 kV is applied between the spray capillary and the counter electrode.

The electric field at the tip of the spray capillary can be approximated using Equation 2.14 (assuming a large, planar counter electrode) where V_c is the applied potential difference, r_c is the radius of the capillary and d is the distance separating the capillary and the counter electrode.[43]

$$E_c = \frac{2V_c}{r_c \ln(4d/r_c)} \tag{2.14}$$

Under typical conditions (for example, a potential of 3 kV, spray capillary radius of 0.25 mm and a distance between spray capillary and counter electrode of 1.5 cm) the electric field at the tip of the spray capillary is $\approx$4x10^6 V/m. Such high electric fields cause polarization of the solution and, assuming polarities used for positive ion experiments (an analogous description applies for negative ions), an increased population of positive ions at the surface. This leads to a distortion of the meniscus and results in the formation of a Taylor cone.[43] If the surface tension of the liquid is overcome, a jet and subsequently a spray, of charged droplets will form as the positive charges in the jet repel each other. As

these charged droplets are accelerated towards the counter electrode, the solvent evaporates and the diameter of the droplets becomes smaller. Because the mass of the droplets becomes smaller, while they do not lose charge, the process which previously occurred at the capillary tip (polarization, Taylor cone formation and then jet and droplet (spray) formation) begins to happen again due to repulsion between the charges. This results in coulombic explosions to smaller dropplets and the process continues to repeat itself. From this point, there are two theories about the formation of gas-phase ions from the small charged droplets, the ion evaporation model (IEM) and the charge residue model (CRM). The charge residue model is essentially a continuation of the series of events described above leading to smaller and smaller droplets. As the droplets get smaller and smaller, desolvation will eventually result in a bare ion that contains all of the charges that remained in the droplet formed after the last coulombic explosion. This model has been used to explain the occurrence of electrosprayed ions of large molecules (proteins, DNA, polymers) having multple charges.[44] In the IEM model, the very small droplets resulting from the repeated series of coulombic explosions described above begin to actually eject ions themselves, rather than continuing to split into smaller droplets.[45,46] It has also been shown that in cases where a high salt concentration is present in addition to a large analyte molecule, both mechanisms occur, IEM followed finally by CRM.[47]

Free Electron Laser

The FEL uses a beam of accelerated, relativistic electrons in a periodic magnetic field to produce light with a wide range of energies depending on the energy of the electron beam and the strength of the magnetic field. Considering different variations of FELs currently in existence, radiation from microwave to X-ray frequencies can be generated.[48] The electrons are accelerated to relativistic velocities before entering an undulator where they experience a periodic magnetic field inside a laser cavity. This series of magnets causes the electrons

to "wiggle" in a sinusoidal motion and at each turn in direction they emit photons. The energy of the emitted light can be modified by altering the kinetic energy of the electrons or by changing the magnetic field. The FEL at CLIO is meant to produce an easily and rapidly tunable beam, accomplished by having the undulator magnets on motors which can be quickly adjusted to change the magnetic field. In practice, this works over an *approx*1200 cm^{-1} energy window, outside of which the power of the beam quickly decays. The energy of the electrons is chosen to adjust the position of this energy window, for example 45 MeV electrons will produce a beam which can be scanned from approximately 1000 cm^{-1} to 2000 cm^{-1} with adjustment of the undulator magnets. A detailed description of the operating principles and details of the FEL at CLIO can be found in literature.[49–51]

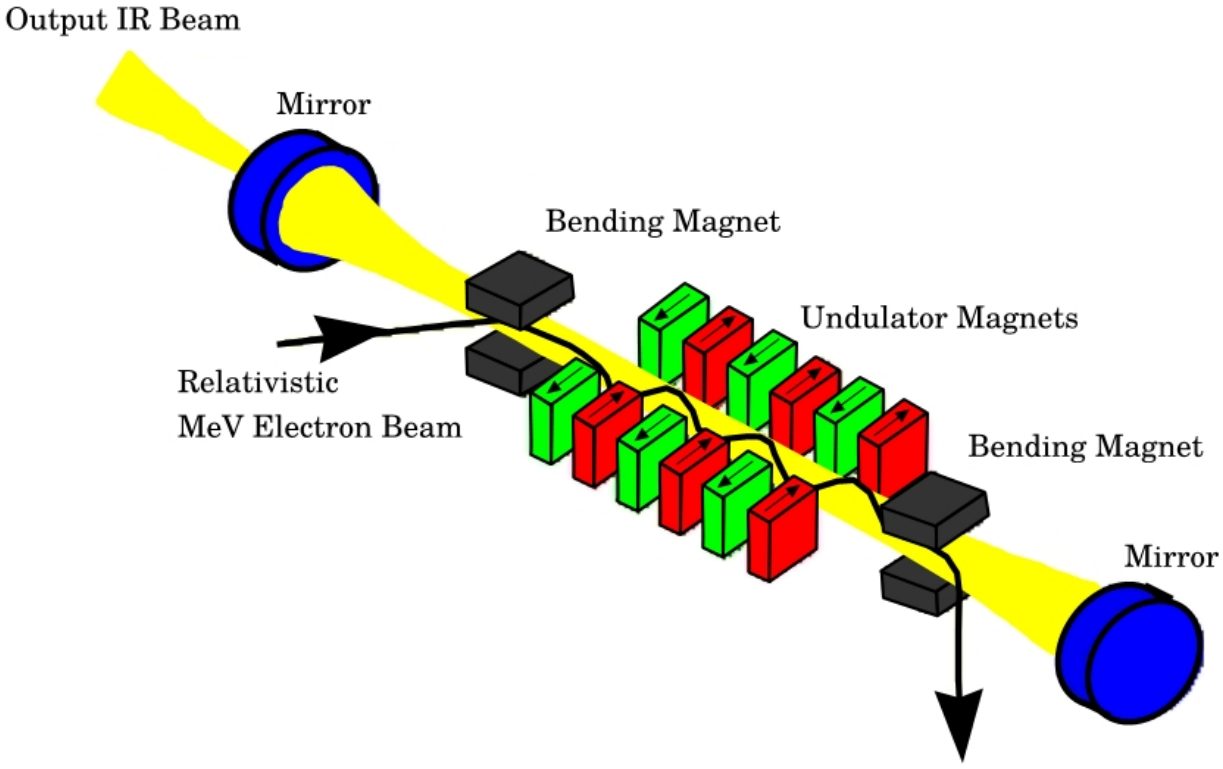

Figure 2.4 – Example schematic of a free electron laser. The black line represents the trajectory of the electrons and the yellow shaded area the emitted photons, which exit the cavity through a small hole in one mirror of the cavity.[52]

Optical Parametric Oscillator

The optical parametric oscillator (OPO) was developed in the 1960's as a device for use as a tunable mid-infrared light source.[53] Emission from an OPO is the result of the division of a pump beam into two output beams called the signal and idler by a nonlinear crystal. The sum of the frequencies of the output beams is equal to that of the input beam.

$$\omega_p = \omega_s + \omega_i \tag{2.15}$$

To control the frequencies of the two output beams, the alignment of the optical axis of the crystal, relative to the pump beam, can be varied. In an OPO apparatus the optical crystal will be contained within an optical resonator or cavity. Escape of some of the amplified beam through one of the resonator mirrors produces the output beams of the OPO called the signal and idler (the signal is the higher-energy beam and the idler is the lower energy beam). Often, after the OPO stage, the desired beam is passed through an optical parametic amplification (OPA) stage. This consists of passing the signal and/or idler beam through another nonlinear optical crystal(s) inline with a beam having the same energy as the pump beam, allowing the production of more intense beams with the same energy as the original signal and idler.

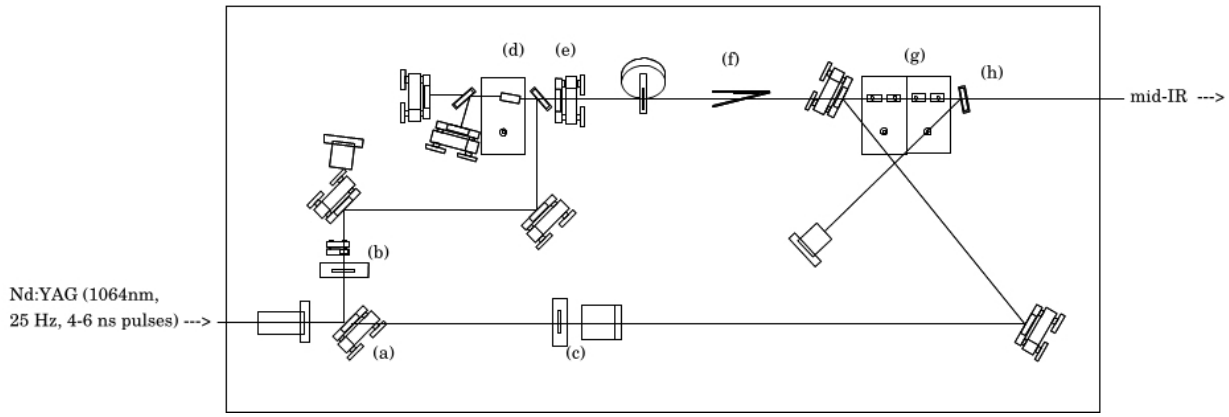

Figure 2.5 – Example schematic of OPO/OPA apparatus.[54]

Figure 2.5 Legend:

(a) Beam splitter

(b) KTP doubling crystal/variable wave plate/532 nm reflector

(c) Variable wave plate

(d) OPO crystal

(e) Exit of OPO stage/some of the 532 nm beam is back-reflected to the source

(f) Filter

(g) Four KTA crystals of the OPA stage

(h) Filter for removing residual 1064 nm

Specifically the OPO/OPA (LaserVision, USA) at CLIO provides an output beam in the 2800-3600 cm^{-1} region for IRMPD experiments (at 3600 cm^{-1} pulses of 12-13 mJ and at 2500 cm^{-1} pulses of 3.5 mJ, with ≈3-4 cm^{-1} bandwidth) from an Nd:YAG (Innolas Spitlight 600, München, DE) pump source.[55–57] This 25 Hz pump laser provides the OPO apparatus 4-6 ns, 550 mJ pulses at 1064 nm with a bandwidth of about 1 cm^{-1}.[54] After entering the apparatus, the pump beam is split in two directions, one going to the OPO and one going to a delay before the OPA. The beam going to the OPO is doubled to 532 nm in a potassium titanyl phosphate (KTP) crystal. The idler and/or signal from the OPO are combined with the 1064 nm pump beam and difference-frequency mixed in the OPA (composed of four counter-rotated potassium titanyl arsenate (KTA) crystals). After the OPA, the remaining 1064 nm pump beam is removed with a dichroic filter and a polarizer is used to isolate either the idler or the signal, which are vertically and horizontally aligned, respectively. Crystal positions in the OPO and OPA are controlled remotely by computer and in this way the OPA output intensity can be optimized and the energy of the output

beam varied (the OPO cavity must be manually optimized for signal intensity). To obtain maximum output intensity, the crystal positions of the OPA were usually re-optimized for each 200 cm^{-1} spectral region during experiments.

Enhanced IRMPD Using a Non-resonant Pump Laser

As previously mentioned, initial IRMPD experiments used fixed wavelength infrared sources, such as CO_2 or CO lasers, to observe fragmentation through the absorption of multiple photons. At the time, intense tunable light sources, such as the FEL, did not exist, and the tunable light sources that did exist lacked sufficient power. For this reason, Lee and Eyler conducted some of the earliest IRMPD experiments, with the intention of obtaining spectral information, using a two-laser, probe-pump setup.[12,13] They cited the absorption of one or a few photons as being sufficient to bring a molecule, of about 10 or more atoms, into the quasi-continuum region of vibrational states. Here the density of vibrational-rotational energy levels is sufficiently high that a broad range of photon energies in the IR region will connect two states in the quasi-continuum (illustrated in Figure 2.6). Therefore, after absorption of a few resonant photons from a low-intensity tunable source, a non-resonant pump source of any IR frequency can be used to induce fragmentation, while retaining the overall wavelength dependence of the process.[58,59]

In a detailed analysis of the effectiveness of a two-laser setup, Eyler et al. qualitatively analyzed vibrational relaxation between the probe and pump pulses. Using protonated bis(methoxy diethyl) ether ($C_6H_{15}O_3^+$) as an example, fragmentation as a function of the delay between the probe and pump lasers was examined while holding all other parameters constant. In Figure 2.7 below, one can see that at very short delay times ($\leq$50ms) there is a strong enhancement to IRMPD, while already at about 300 ms there is no enhancement to IRMPD and the extent of fragmentation is approximately equal to the extent when only the pump laser was applied.[13]

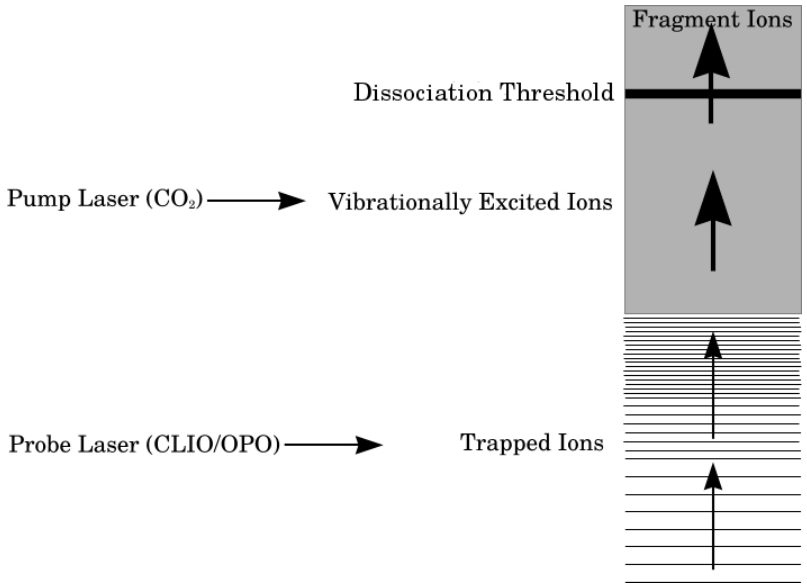

Figure 2.6 – The sequence of events in a two-laser IRMPD experiment.[13]

In addition, the effect of the probe laser power was examined and it was determined that the use of a probe-pump combination allows the use of probe lasers with only 30% of the power that would be required if a pump laser were not used. Finally, reversing the order of the resonant probe and non-resonant pump lasers in the experiment was never found to result in an increase in fragmentation above the sum of fragmentation resulting from use of both lasers alone.

More recently, the use of a CO_2 laser as a pump source for an OPO/OPA IRMPD experiment was demonstrated for various $[Mn(ClO_4)]^+$ and $[Mn_2(ClO_4)_3]^+$ clusters with water (using the same FT-ICR setup at CLIO as was used for the experiments that will be discussed in this thesis).[57,60] These observations were similar to those discussed previously.[13] Essentially no fragmentation was observed in either of the single-laser experiments, but a significant enhancement was seen in the two-laser experiment where only the probe source was at a resonant frequency.

Figure 2.8 shows example mass spectra of $Ala_{11}Na^+$ obtained under different irradiation conditions (probe/pump configurations) with experimental parameters similar to those used for experiments that will be discussed later in this thesis. Other studies dis-

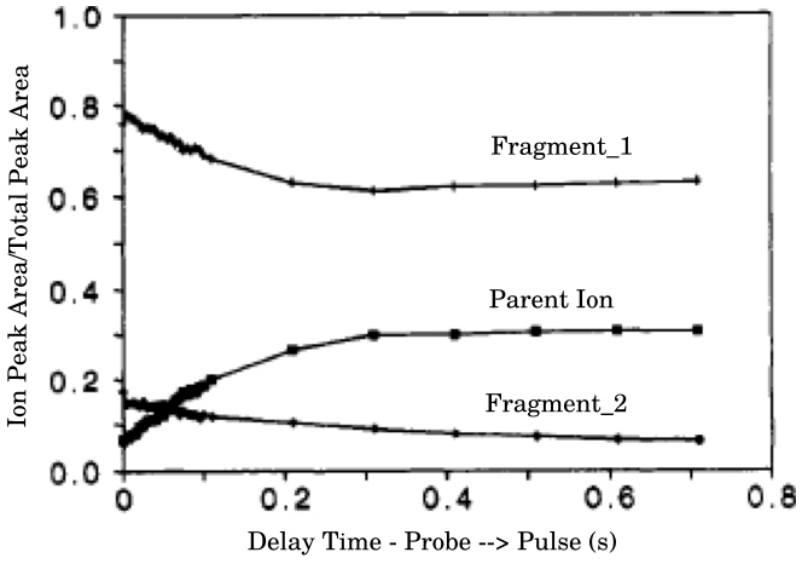

Figure 2.7 – IRMPD as a function of delay between probe and pump events.[13]

cussed here[57,60] reported similar spectral comparisons with nearly equivalent enhancement to fragmentation as is seen in Figure 2.8. It is important to note from Figure 2.8 not only the benefit of using the probe-pump setup, but also the simple necessity! For $Ala_{11}Na^+$ (in Figure 2.8) and other sodiated poly(alanine) peptides that will be discussed extensively later, it would be nearly impossible to obtain an IRMPD spectrum using the OPO/OPA (with a reasonable signal to noise ratio), without the use of both lasers. Finally, it is important to note the extremely low fragmentation levels when only the CO_2 laser is used, providing a very low level of background (non-resonant) fragmentation.

Details of the FT-ICR/OPO/FEL/CO_2 Experiment

To this point, discussion of the instrumentation has been mostly general and not specific to the instruments used in the experiments to be discussed in this thesis. In this section the FT-ICR/OPO/FEL/CO_2 at CLIO used for the experiments in this thesis will be discussed in detail.

The FT-ICR mass spectrometer at CLIO is a 7.0 T Bruker Apex Qe instrument fitted with an electrospray ionization source (as was seen schematically in Figure 2.3).[42,61]

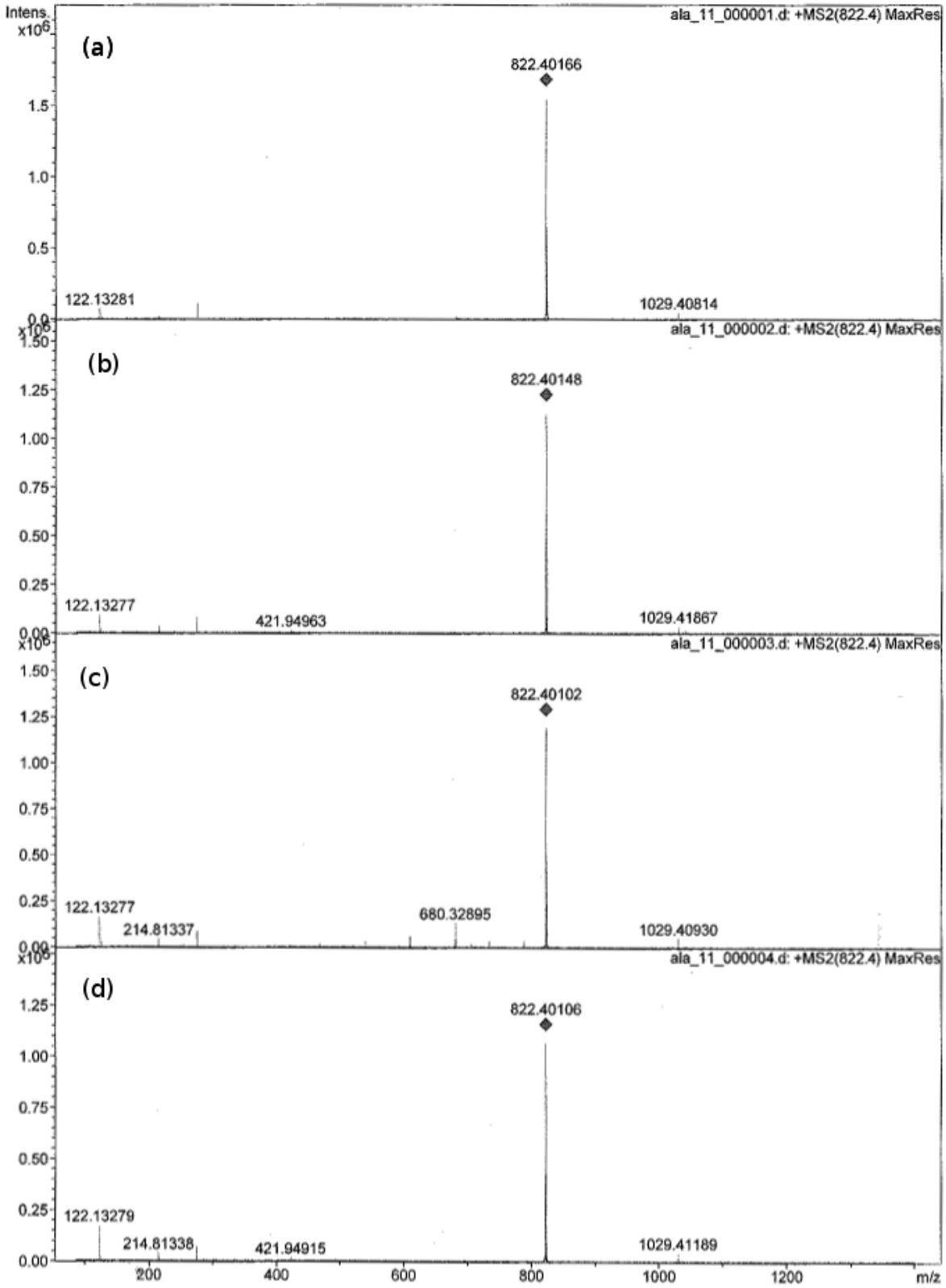

Figure 2.8 – IRMPD of $Ala_{11}Na^+$ under different fragmentation conditions (irradiation time 5.0 seconds): (a) CO_2 laser (20ms/pulse) only (b) OPO/OPA on resonance at 3365 cm^{-1} only (c) CO_2 laser (20 ms/pulse) and OPO/OPA on resonance at 3365 cm^{-1} (d) no laser.

The electrosprayed ions first reach a transfer capillary, followed by a series of ion funnels and a hexapole ion trap.[62] Here it is possible to do source accumulation and typically the ions are accumulated for 10-50 ms before they are injected into a quadrupole (where mass selection is possible in the case of an MS/MS experiment)[i]. After the quadrupole, ions are accumulated in a hexapole collision cell (Argon, ≈3.0x-10^3 mbar) for 100-2000 ms. After the collision cell are a series of transfer electrodes which the ions traverse for a variable amount of time, usually on the order of 0.5-2.5 ms before entering the ICR cell. The laser beam used to irradiate trapped ions enters the cell through a Brewster window from the opposite side of the magnet. The laser beam should be fairly well focused (but not too well because the ion packet is reasonably large), as it must pass through a 3.45 mm aperture in an electron capture dissociation (ECD) cathode before reaching the cell. The cell has a 6 mm hole in each trapping plate to allow the entrance of ions and the laser beam (from opposite sides). The instrument has two ring-electrodes which decelerate ions immediately before they enter the cell. The second of the two electrodes is actually composed of two half-ring electrodes and is able to apply a "sidekick" potential. The sidekick pushes the ions slightly off axis to provide better trapping by reducing their linear motion, however it was also noted that the adjustment of this potential has a significant influence on both the ion intensity and the degree of fragmentation. In a few cases, adjustment of this parameter away from the commonly used range was necessary to observe fragmentation.

Figure 2.9 is a slightly simplified representation of the configuration of the mirrors, lenses, and shutters used to control the lasers on their path to the ICR cell. The OPO beam travels from the OPO apparatus through a shutter and a 2 m focusing mirror before being directed to the cell. The FEL enters vertically and a small portion is split with a Zn:Se window and directed to a spectrophotometer. The CO_2 laser passes through a shutter and

[i]Usually, in the case of IRMPD experiments, mass selection is performed in this quadrupole because it has been observed that this produces higher-levels of fragmentation than when mass selection is done in the ICR cell. This is likely because there is a difference in the kinetic energy distribution of ions if they are trapped in-cell or in the quadrupole, resulting in different trajectories. The trajectories of ions isolated in the quadrupole presumably then have better beam overlap than of those isolated in-cell.

a gold mirror before passing through a hole in the centre of the last mirror used to direct the FEL/OPO beams into the ICR cell. By using a holed mirror and with the FEL/OPO beam being reflected as close to the centre of the cell as possible, close spatial overlap between the FEL/OPO and the CO_2 beams can be attained. For alignment assistance, a HeNe laser can enter from a window directly in front of the ion funnels and traverse the entire length of the instrument, exiting the Brewster window on the opposite end of the magnet. This provides an important assistance when the alignment of the FEL/OPO/CO_2 with the ion packet is lost.

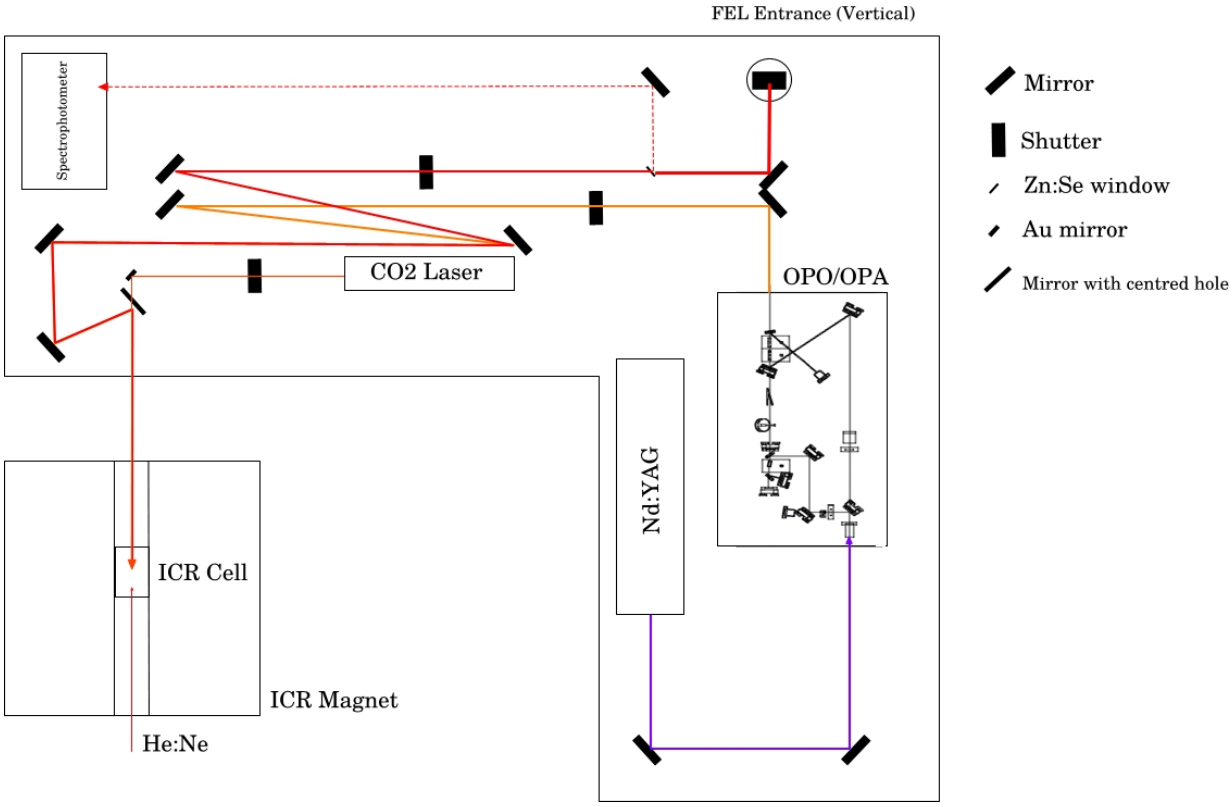

Figure 2.9 – Example schematic of laser paths from source to ICR cell.

It is useful to consider the sequence of events involved in a single data acquisition step during the experiment. Figure 2.10 below qualitatively shows the ordering and relative durations of events in this process. The cell is first quenched to remove any ions that were previously trapped. If an additional ionization method is to be applied (or gas is to be ionized (in the case of an electron ionization experiment) it is done immediately

after quenching. At this point, mass selection or ejection can be performed followed by a variable amount of time for IRMPD. The entire IRMPD time is shown in Figure 2.10 as the dashed line, however what happens during this time is slightly more complicated and is expanded upon in Figure 2.11. The IRMPD time can be set from 0-10 s and up to an undefined limit using a manual modification to the event sequence script. In the figure it appears that only one sequence of laser pulses is applied to the ions, however in reality, this sequence of probe/pump pulses is applied at 25 Hz. After irradiation the sequence is concluded with excitation of the ions and finally detection. In practice, this procedure is repeated 4-16 times per spectral point to accumulate a higher signal intensity.

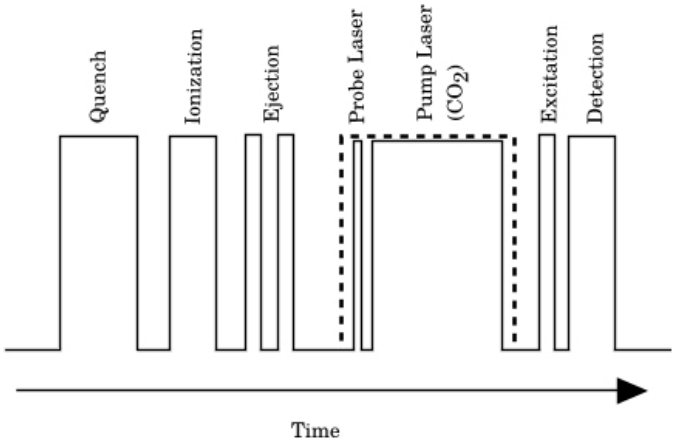

Figure 2.10 – Sequence of events ions undergo after entering the ICR cell. Approximate event times: Quench - 10 ms, Ionization - 7.0 ms, Ejection - each 0.6 ms.

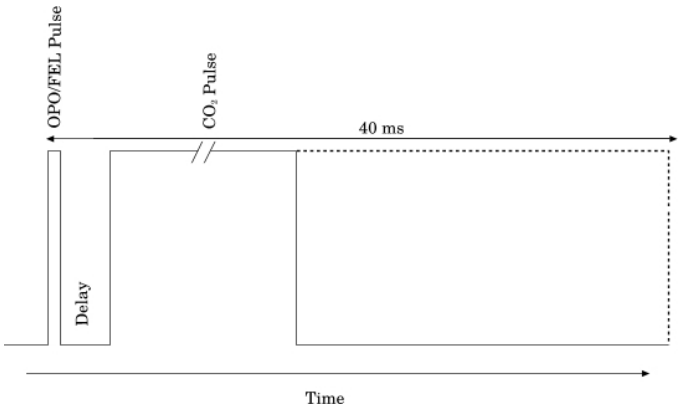

Figure 2.11 – Time sequence of laser pulses in the two-laser IRMPD experiment.

Figure 2.11 provides more detail of the sequence of laser pulses in the case of the two-laser experiment. The overall sequence is triggered by the 25 Hz signal of the probe laser (Nd:YAG of the OPO/OPA or FEL). In the case of the OPO/OPA, the 25 Hz pulse is 4-6 ns and in the case of the FEL it is an 8 μs macro pulse composed of $\approx$500 approximately picosecond pulses at 62.5 MHz. This pulse is followed by a delay of 1.0 μs and finally the CO_2 pulse. The CO_2 pulse is variable up to $\approx$40 ms when the next trigger of the 25 Hz signal arrives (in practice it is usually not applied for more than 25 ms). Two shutters, one in line with the OPO/FEL beam and one in line with the CO_2 beam are triggered by the FT-ICR at the start of the IRMPD window.

2.2 Computational Methods

The experiments discussed in the previous sections provide results which contain an enormous amount of information relating to the structural and energetic properties of gas-phase ions. However, the IRMPD spectra produced from these experiments are not always susceptible to trivial interpretation, especially with increasing size of the system. Chemical intuition coupled with general vibrational spectroscopic considerations can provide a qualitative interpretation, however a vast amount of additional information is available. To access this information it becomes necessary to use computational methods to predict the energetic, structural and vibrational properties of the ions in advance of the interpretation of experimental spectra.

After the development of quantum mechanics in the beginning of the 20th century, it quickly became apparent that essentially all properties of chemical systems could, in principle, be calculated using first principles. However, it was not until the advent of modern computational resources that this actually became practical. Beginning with the work of Pople in the 1970's,[63] for which he was awarded the 1998 Nobel prize in chemistry, the use

of theoretical chemistry as a predictive and explanatory tool has expanded enormously, and is today one of the primary tools of both experimental and theoretical chemists. Currently a wide range of computational tools exists for calculating chemical properties based on a balance between time, computational resources and desired accuracy.[64] The specific methods that were used for the calculation of results that will be presented in this thesis will be discussed in the following sections.

It is interesting to note that gas phase ionic systems are ideally suited for theoretical modelling. Since ions trapped in an ICR cell are under high-vacuum (10^{-9}-10^{-10} torr) and experience very few collisions (tens of seconds between collisions) they can accurately be assumed to be in an isolated environment. This means that it is unnecessary to include solvent interactions in the theoretical models while considering gas phase ions (perhaps, excluding solvent effects surviving the electrospray process).

2.2.1 Molecular Mechanics and Force Fields

Molecular mechanics is conceptually the simplest of computational chemistry methods, as it is based on the principles of classical mechanics. The advantages of such calculations are the speed and low computational cost, allowing for application to very large systems such as proteins and other large biomolecular systems. Well known packages which use molecular mechanics implementations, such as AMBER,[65] CHARMM[66] and GROMOS,[67] have seen wide usage for calculating properties of biomolecular systems. However, these implementations are not always sufficiently accurate in the prediction of energies or potential energy surfaces to be used directly in the preparation and selection of structures for more costly quantum mechanical methods.

In the molecular mechanics approach, the energy of a system is determined through the evaluation of the force field energy function. The parametrization of this function

is based on experimental values or values determined from high-level *ab initio* methods. Examples used in the following discussion will focus on the AMOEBA[68] force field, as all results involving molecular mechanics discussed in this thesis used this force field.[69]

All force fields contain a set of terms for bonded interactions and a set of terms for non-bonded interactions.

$$E = E_{bonded} + E_{non-bonded} \tag{2.16}$$

This expression will be expanded to varying degrees depending on the desired accuracy, demonstrated below for an AMBER like force field in Equation 2.17 and secondly for the AMOEBA force field in Equation 2.18.

$$E = [E_{bond} + E_{angle} + E_{torsion}]_{bonded} + [E_{vdW} + E_{ele}]_{non-bonded} \tag{2.17}$$

$$E = [E_{bond} + E_{angle} + E_{b\theta} + E_{oop} + E_{torsion}]_{bonded} + [E_{vdW} + E_{ele}^{perm} + E_{ele}^{ind}]_{non-bonded} \tag{2.18}$$

In Equation 2.18, the first terms in the AMOEBA expression are the bonded terms and correspond to the bond stretching, angle bending, bond-angle cross term, out-of-plane bending, and torsional rotation, while the non-bonded terms are the van der Waals, permanent electrostatic and electronic polarization interactions.[68] In the AMOEBA expression, the bond energy of atoms directly connected or separated by one atom is determined by a more sophisticated empirical expression including both anharmonic and coupling terms (the exact expressions for the bond lengths and angles are as defined in the MM3 force field).[68,70]

For AMOEBA, two specific features are included in addition to the standard functions for bonded interactions. Out of plane bending is accounted for using a Wilson-Decius-Cross function to separate angle bending into in-plane and out-of-plane. Torsions are treated by a Fourier expanded torsional function, where a special form is used in the case of torsions involving two adjacent trigonal centres (such as in the peptide backbone). This is an important feature in the modelling of peptides due to the increased rotational barrier arising from the partial double bonded character of the peptide bond. The van der Waals and the electrostatic terms treat intermolecular and separated intramolecular interactions. Interactions, such as hydrogen bonding, are very sensitive to these parameters. The torsional term and the non-bonded terms are responsible for treating the internal degrees of freedom in the system and are therefore critical to calculations seeking conformational information or energy comparisons between conformations. In AMOEBA, a buffered 14-7 function, shown to provide the best fit to *ab initio* and experimental results, is used.

$$U_{vdw}(ij) = \varepsilon_{ij}(\frac{1.07}{\rho_{ij} + 0.07})^7(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2) \tag{2.19}$$

Here $\rho = R_{ij}/R_{ij}^0$, where R_{ij}^0 is the minimum energy distance and R_{ij} is the actual distance. ε_{ij} is the depth of the potential well as defined below:

$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\varepsilon_{ii}^{1/2} + \varepsilon_{jj}^{1/2})^2} \quad \texttt{and} \quad R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2} \tag{2.20}$$

In this formulation, every atom is a van der Waals site where the site is centred on the atomic nucleus for non-hydrogen atoms. For an H-atom bound to another atom, the van der Waals site is centred somewhere along the bond at a percentage of the total bond length determined by a reduction factor. This is again similar to the formulation used in the MM3 force field, and has been shown to match both experimental and *ab initio* results more closely than centring on the nucleus, for the H-atom.

In polar systems the electrostatic term is usually the largest contribution to the energy of the system.[68] Commonly, this term is modelled using a Coulombic potential and fixed partial atomic charges. An obvious approximation in the use of fixed charges is the inability to include effects of polarization by inter- or intramolecular interactions.[71,72] One of the main advantages of the AMOEBA force field is the inclusion of both induced and permanent atomic multipoles. The permanent atomic multipoles, M_i, are included for each atom, i, where q is the point charge, μ the dipoles and Q the quadrupoles.

$$M_i = [q_i, \mu_{ix}, \mu_{iy}, \mu_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixz}, ..., Q_{izz}]^t \tag{2.21}$$

Fixed partial charges as well as dipoles and quadrupoles are taken from high-level quantum chemical calculations of small molecules. Especially for clusters and hydrogen bound complexes, electronic polarizabilities plays a large role in the description of energies. In the AMOEBA procedure, point dipole moments are induced for each atom consistently with the experienced electric field. Molecular polarization is then accomplished with a mutual induction scheme with distributed atomic polarizabilities modelled on Thole's damped interaction method.[73,74] This means that each induced dipole on any atom, i, will polarize all other atoms until convergence of all sites is reached. Local damping is included in this model through the use of a smearing function to smear one of the atomic multipoles of each interacting pair, where $u = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the effective distance as a function of linear separation and atomic polarizabilities (α) of each atom i and j, and a is dimensionless parameter controlling the strength of damping.

$$\rho = \frac{3a}{4\pi}e^{-au^3} \tag{2.22}$$

The use of this multipolar expansion to describe the electrostatic charge in combination with full and explicit treatment of polarization effects is crucial to the description of hy-

drogen bonding and for obtaining accurate energetics. It has been shown that, when used in conformational searches, traditional force fields incorporating only fixed partial charges will only identify about 50% of the conformations, while the polarizable AMOEBA force field usually identifies ≈80%.[71,75–77]

2.2.2 Molecular Dynamics and the Replica Exchange Method

When considering large and complicated molecular systems, it can be necessary to consider their structure and properties dynamically.[78] Bond vibrations within a molecule are very rapid, however the formation of secondary structure and motion of large or rigid groups can be of a significantly longer time scale. Only a short summary of this treatment will be discussed here and a more detailed description can be found in the references.[79,80]

Table 2.1 – Timescales of peptide and protein dynamics.[79]

Bond Vibration	10^{-14} - 10^{-15} s
Helix motions	10^{-9} - 10^{-6} s
Helix - coil transitions	$> 10^{-7}$ s

A system described by the previously developed molecular mechanics framework is composed of a set of particles treated as point masses. This system will experience a set of movements described by the classical equations of motion related to the interactions described in the force field equations above. The force, F_i, exerted on the particle of mass, m_i, and coordinate vector, r_i, as a function of the potential energy of the system (the force field energy), is defined as the negative change of the potential energy, U, as a function of the position, r_i.

$$F_i = -\frac{dU}{dr_i} \tag{2.23}$$

Newton's second law states that the rate of change of the momentum of the particle is equal to the force, F_i. The velocity of the particle is defined as $dr_i/dt = p_i/m_i$, where p_i, is the momentum, a_i the acceleration and t the time.

$$\frac{dp_i}{dt} = F_i = m_i \frac{dv_i}{dt} = m_i a_i \tag{2.24}$$

Restricting the scenario to one dimension, x, the position of the particle, $x(t)$, after propagation of Δt is obtained from the Taylor series

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{d^2x(t)}{dt^2}\frac{\Delta t^2}{2} + (\texttt{higher-order terms}) \tag{2.25}$$

where the first three terms (the position, $x(t)$, velocity, $dx(t)/dt$, and acceleration, $d^2x(t)/dt^2$) are evaluated numerically and the higher-order terms are approximated or neglected. Returning to Equation 2.24, the acceleration can be defined in terms of the force and force-field potential energy.

$$a = \frac{d^2x(t)}{dt^2} = \frac{F_x}{m} = -\frac{1}{m}\frac{dU}{dr_i} \tag{2.26}$$

Many algorithms for the integration of the equations of motion exist, however one of the more popular is the Beeman algorithm[81] which is used in the TINKER dynamics program. Although all atoms are in constant motion, which defines the force, F, to be a constantly changing property, the use of a sufficiently small Δt allows one to assume a constant F and acceleration, a_0. Integration of the acceleration yields velocities and positions in the following way:[82]

$$\frac{d^2x}{dt^2} = a_0 \tag{2.27}$$

$$\frac{dx}{dt} = v_{(t)} = v_0 + a_0 t \tag{2.28}$$

$$x(t) = x_0 + v_0 t + \frac{1}{2} a_0 t^2 \tag{2.29}$$

where v_0 and x_0 are the velocity and position at $t = 0$. Working within the above mentioned small Δt assumption, one can write Equation 2.29 in terms of the $n + 1$ position of time to obtain the following expressions:

$$v_{n+1} = v_n + a_n \Delta t \tag{2.30}$$

$$x_{n+1} = x_n + v_n \Delta t + \frac{1}{2} a_n (\Delta t)^2 \tag{2.31}$$

Considering

$$v_{n-1/2} = \frac{(x_n - x_{n-1})}{\Delta t} \tag{2.32}$$

one obtains the velocity at time n

$$v_n = v_{n-1/2} + a_n \frac{\Delta t}{2} = \frac{x_n - x_{n-1}}{\Delta t} + a_n \frac{\Delta t}{2} \tag{2.33}$$

and through substitution of v_n in 2.31, the following expression for the x_{n+1} position:

$$x_{n+1} = 2x_n - x_{n-1} + a_n(\Delta t)^2 \tag{2.34}$$

Here, the coordinates after $n + 1$ time steps are given in terms of only the previous two coordinates and the acceleration at time step n. It is then possible to calculate the velocity at step n after the coordinates at step $n + 1$ have already been obtained if one assumes constant velocity over two sequential time steps.

$$v_n = \frac{x_{n+1} - x_{n-1}}{2\Delta t} \tag{2.35}$$

Equations 2.34 and 2.35 constitute one of the original formulations of integration of the equations of motions in molecular dynamics simulations. This formulation, the Verlet method, can be improved by abandoning the assumption that the velocities are constant over two sequential time steps and constitutes the Beeman method. With the acceleration defined as

$$a(t) = a_n + \frac{t(a_n - a_{n-1})}{\Delta t} \tag{2.36}$$

the following expressions (derivation found in the references[81,82]) give the Beeman formula:

$$x_{n+1} = x_n + v_n\Delta t + \frac{1}{6}(4a_n - a_{n-1})(\Delta t)^2 \tag{2.37}$$

$$v_{n+1} = v_n + \frac{1}{6}(2a_{n+1} + 5a_n - a_{n-1})\Delta t \tag{2.38}$$

In addition to the more accurate velocity calculation, a feature of the Beeman algorithm is the velocity dependence of x_{n+1}. This allows a technique, referred to as a Berendsen

thermostat[83] to be used, where the system is coupled to an external heat bath and velocities are scaled to maintain constant temperature (in contrast to constant energy) throughout the simulation.

When studying increasingly large and complicated molecular systems an increasing number of minima on the potential energy surface will exist. In the case of an extensive and complicated potential energy surface it can happen that a simulation becomes locked in a minimum, or region of minima, and does not explore the entire surface (in other words, the rate of isomerization can be slow enough that it is unlikely to be observed during the simulation time). The replica-exchange method (REM) has emerged as a popular tool in the exploration of potential energy surfaces of large and complex molecular systems. The simultaneous calculation of numerous molecular dynamics trajectories at different temperatures allows the exploration of a wide range of conformations at high temperatures, where the barriers are not as significant, and the sampling of these conformations by exchange to the lower temperatures.

In a replica-exchange molecular dynamics (REMD) simulation, a series of M identical replicas are simultaneously and independently calculated in canonical ensembles at M temperatures. That is, there are an equal number of replicas and unique temperatures. Throughout the course of the simulation, at repetitive time intervals, exchanges of either the replicas' coordinate and momentum vectors or their temperatures are attempted. The probability of exchange between replicas of indices i and j is determined by the Metropolis criterion below, where X represents the replica, E the potential energy of the respective state, $P_{exchange}$ the probability of the exchange occurring and T the temperature of the replica:[84]

$$P_{exchange} = \texttt{min}[1, e^{((\frac{1}{k_B T_i} - \frac{1}{k_B T_j})(E(_i) - E(_j)))}] \tag{2.39}$$

Only exchanges between neighbouring replicas are attempted, because for increasingly different temperature, the probability of successful exchange decreases exponentially.

The work presented in this thesis used 12 replicas over a temperature range of 200-600 K with an exponential distribution of temperatures (200, 221, 244, 270, 298, 330, 364, 402, 445, 491, 543, 600 K). The use of an exponential distribution increases the difference between temperatures as the temperature increases. As the temperature increases so does the internal energy and therefore the exchange rate. Since it is convenient to have relatively constant exchange rates between replicas, the exponential temperature distribution is useful. 499 exchanges were attempted (total number of steps is 500) with 10,000 1.0 femtosecond dynamics steps between each attempt, for a total simulation time of 5.0 ns. Sequential REMD simulations were run using the lowest-energy structure from the previous simulation, until the lowest-energy structure produced between two subsequent runs was reproduced (usually accomplished in 3-4 simulations). Figure 2.12 shows trajectories of the lowest, middle and highest temperatures in a sample simulation, clearly illustrating that over the course of the 5 ns simulation time, each trajectory is at some point found in every replica.

A previously modified version of the Tirex program developed by Penev et al.[85,86] was used for the computations.[87] This software performs the REMD simulation using the TINKER molecular dynamics program.[69] Replicas are run in parallel, one on each processor, and temperature information is swapped at each successful exchange. The program uses the REMD procedure detailed above based on the work of Okamoto et al.[84] Hundreds of minimum energy structures are produced from each REMD simulation where, under certain conditions, a slight dependence on the starting structure of the simulation was observed. This will be discussed in more detail using specific examples in the following sections of the text. A number of methodologies can be applied in the selection of structures from the results. In general, criteria such as the root-mean-square deviation (RMSD)

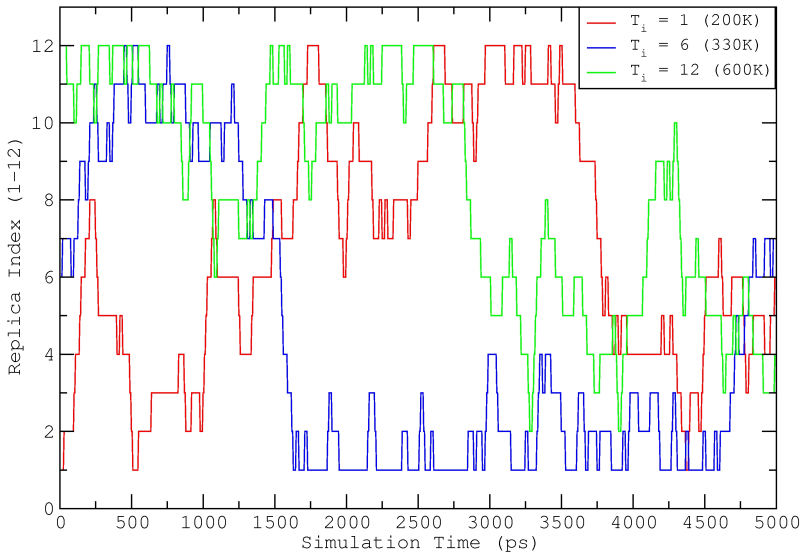

Figure 2.12 – Distribution of trajectories over replica indices as a function of simulation time

.

and the gyration radius are used to divide structures into classes. The gyration radius is useful in identifying large structural variations, for example, distinguishing α-helices from globules, while the RMSD calculation is useful for identifying and removing identical structures or closely related groups of structures.

The root-mean-square deviation (RMSD) is the difference in coordinates of two equivalent atoms in two superimposed conformations of the same molecule. This is often a primary tool used during the selection of REMD structures to consider at higher-levels of computation. For the RMSD values presented in the following sections, a superposition of the two structures is performed, minimizing the RMSD between all atoms in the structure. The RMSD is defined in Equation 2.40 below.

$$\texttt{RMSD} = \sqrt{\frac{1}{n}\sum_{n=1}^{n} r_n^2} \tag{2.40}$$

Here, r_n is the difference in coordinates between corresponding atoms n in two structures, each composed of N atoms. RMSD values reported here refer to the superposition and

comparison of all atoms.

Additionally, the gyration radii were calculated for all REMD structures, which provides some general insight into the size and compactness of structures (for example, automatically distinguishing globular and α-helical structures). The gyration radius is defined as the mass-weighted root-mean-square-distance between the atoms of the molecule and the centre of mass of the molecule.

$$\mathtt{R}_{gyr} = \left(\frac{\sum_{n=1}^{N} m_n r_{cm,n}^2}{\sum_{n=1}^{N} m_n} \right)^{1/2} \tag{2.41}$$

In Equation 2.41, m_n is the mass of atom n and $r_{cm,n}$ is the scalar distance between the atom n and the centre of mass of the molecule, for all N atoms of the molecule.

2.2.3 Quantum Chemical Methods

Computational quantum chemistry involves solving the Schrödinger equation to obtain properties originating from the electronic structure of molecules. While for the smallest of species exact accuracy is attainable, the vast majority of chemical systems require approximate treatment. The following discussion intends to provide some qualitative background about the computational procedure, which centres around the practically necessary approximations. Only a summary of the illustrative results within the procedure is provided here, and more complete derivation and background can be found in the references.[16,80,88–90]

The time-independent Schrödinger equation is an eigenequation with its eigenfunction as the molecular wavefunction, Ψ, and eigenvalues as the various energy levels, E_n, of the system.

$$H\Psi = E\Psi \tag{2.42}$$

The Hamiltonian operator, H, is a function of i electrons and a nuclei where m is the mass of the particle, Z is the charge, r the distance of separation, and $\hbar$ is Planck's constant over 2π.

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_a \frac{\hbar^2}{2m_a}\nabla_a^2 - \sum_i \sum_a \frac{e^2 Z_a}{r_{ia}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{a<b} \frac{e^2 Z_a Z_b}{r_{ab}} \tag{2.43}$$

Here the first two terms on the right hand side represent the kinetic energy of the electrons and nuclei, the third term represents the electronic-nuclear attraction and the fourth and fifth terms represent the electronic and nuclear repulsion, respectively. Ψ, as the wave-function of the molecule, is a function of the positions of all particles. The kinetic energy terms are eigenvalues of the respective kinetic energy operators, where the Laplacian has the following form:

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \tag{2.44}$$

Although the Hamiltonian expression above appears rather simple, at least relative to how powerful the results are, one can see in the expression there exist multiple instances of correlation between particles. This results from the fact that each particle is somewhat dependent on every other particle in the system and introduces a large degree of complexity. The Born-Oppenheimer approximation neglects coupling between the nuclei and electronic motion, based on the enormous mass difference ($m_{electron} = 9.1 \times 10^{-31}$ kg and $m_{proton} = 1.7 \times 10^{-27}$ kg) between the nuclei and electrons. Within this approximation, the electronic Schrödinger equation is solved for a fixed set of nuclear positions:

$$(H_{el} + V_{Nuc})\Psi_{el}(r; R_{fixed}) = E_{el}\Psi_{el}(r; R_{fixed}) \tag{2.45}$$

Here the kinetic energy of the nuclei is assumed to be independent of the electrons, defining that (in the negative nuclear-electronic term) correlation between electrons and nuclei is neglected. The positive nuclear-nuclear repulsive potential energy term, V_{Nuc}, can be evaluated separately and as a constant at fixed nuclear coordinates. A couple of initial notes must be made about the electronic wavefunction, which will, for now, be referred to as Φ. The wavefunctions must be orthonormal,

$$\iiint \Phi_i^* \Phi_i dx dy dz = 1 \tag{2.46}$$

$$\iiint \Phi_i^* \Phi_j dx dy dz = 0 \quad i \neq j \tag{2.47}$$

antisymmetric (change sign upon interconversion of the coordinates of any two electrons, due to the Pauli principle) and must include spin. This is accomplished by constructing the wavefunction using a Slater determinant (or a set of Slater determinants), as demonstrated below for n electrons, where $1/\sqrt{n!}$ is a normalization factor and χ_N represents a spin-orbital (the product of spatial orbital and a spin function, usually labelled as α or β):

$$\Phi_{SD} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} \tag{2.48}$$

Interchange of coordinates of two electrons can be accomplished by interchanging two rows, which by definition, changes the sign of the determinant.

The variational principle states that the energy, E_g obtained from the initial (guess) wavefunction will be equal to or higher in energy than the true value, E_0. Iterative minimization can then be used to obtain continually lower (and more accurate) approximations

to the electronic energy.

$$E_0 \leq E_g = \frac{\int \Phi H \Phi dr}{\int \Phi^2 dr} \tag{2.49}$$

Under the Born-Oppenheimer approximation, the Hamiltonian for variable electronic coordinates r_i and fixed nuclear coordinates R_a becomes the following (using atomic units):

$$H_{elec} = KE_e + V_{ne} + V_{ee} + V_{nn} \tag{2.50}$$

where

$$KE_e = -\sum_i \frac{1}{2}\nabla_i^2 \tag{2.51}$$

$$V_{ne} = -\sum_a \sum_i \frac{Z_a}{|R_a - r_i|} \tag{2.52}$$

$$V_{ee} = \sum_i \sum_{j>i} \frac{1}{|r_i - r_j|} \tag{2.53}$$

$$V_{nn} = \sum_a \sum_{b>a} \frac{Z_a Z_b}{|R_a - R_b|} \tag{2.54}$$

Again, V_{nn}, the nuclear-nuclear repulsion is a constant for a given set of nuclear positions, and is treated separately. The expectation value for the energy of the system can then be expressed as Equation 2.55.

$$E = \sum_{i=1} h_i + \frac{1}{2} \sum_{i=1} \sum_{j=1} (J_{ij} - K_{ij}) + V_{nn} \tag{2.55}$$

Here, h_i includes the kinetic energy and electrostatic potential energy of the nuclei and electrons. J_{ij} is the Coulomb integral, accounting for the classical repulsion between electrons and K_{ij}, the exchange integral, is a quantum mechanical term with no classical equivalent, reducing the Coulomb term in the case of electrons with the same spin. The factor of 1/2 is necessary because the repulsion of each electron to all other electrons is included and results in double counting.

$$h_i = \langle \phi_i(1) | - \frac{1}{2} \nabla_i^2 - \frac{Z_a}{|R_a - r_i|} | \phi_i(1) \rangle \tag{2.56}$$

$$J_{ij} = \langle \phi_i(1) \phi_j(2) | \frac{1}{|r_i - r_j|} | \phi_i(1) \phi_j(2) \rangle \tag{2.57}$$

$$K_{ij} = \langle \phi_i(1) \phi_j(2) | \frac{1}{|r_i - r_j|} | \phi_j(1) \phi_i(2) \rangle \tag{2.58}$$

It is possible to define F_i, the Fock operator, describing for a single electron, the attractive potential to all nuclei, the repulsive potential to all other electrons and the kinetic energy:

$$F_i = h_i + \sum_j (J_j - K_j) \tag{2.59}$$

Using the set of Fock operators, F_i, and the set of MOs, ϕ_i, a set of eigenvalue problems can be constructed of the form:

$$F_i\phi_i = \varepsilon_i\phi_i \tag{2.60}$$

Conceptually, the set of HF equations is solved using an iterative self-consistent field (SCF) approach where the total energy of the system is described by Equation 2.61.

$$E = \sum_i \varepsilon_i - \frac{1}{2}\sum_{ij}(J_{ij} - K_{ij}) + V_{nn} \tag{2.61}$$

$$\varepsilon_i = \langle\phi_i|F_i|\phi_i\rangle = h_i + \sum_j (J_{ij} - K_{ij}) \tag{2.62}$$

It is important to note from this result that the total energy is not equal to the sum of all orbital energies. Additionally, within this approximation, electron correlation is entirely unaccounted for. This is a result of each electron only "feeling" a static and average potential of all other electrons. This approximation is valid under some circumstances. Unfortunately, in the case of most real chemical situations this is not the case, making it necessary to introduce a post-Hartree-Fock method to account for electron-electron correlation effects.

Density Functional Theory

The Hohenberg and Kohn theorems form the basis of density functional theory (DFT).[91] First, they show that a given ground-state electron density uniquely determines the external potential (the nuclei) and thus the hamiltonian and wavefunction, in what is referred to now as the Hohenberg-Kohn existence theorem. Their second theorem was the formulation of a variational principle for DFT, where the density is the property to be varied. The main advantage of pursuing such a direction to obtain the energy of a system was not realized

until Kohn and Sham demonstrated the use of a Hamiltonian for a fictitious system of non-interacting electrons. Here the Hamiltonian is separable and can be constructed as a sum of one-electron operators. These are slater determinants of the one-electron eigenfunctions and have eigenvalues which are sums of the one-electron eigenvalues. One then chooses the fictitious non-interacting system of electrons to have the same density as that of some real system where the external potential (positions and charges of nuclei) is by definition the same in both systems. The energy functional can then be written in the following way:

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)] \tag{2.63}$$

$T_{ni}[\rho(r)]$ represents the kinetic energy of the non-interacting electrons and $V_{ne}[\rho(r)]$ and $V_{ee}[\rho(r)]$ represent the potential energy of the classically interacting nuclei-electrons and electrons-electrons respectively. The following two Δ terms are corrections to this expression, namely the kinetic energy difference of the interacting electrons and all the various quantum aspects of the electron-electron interactions. For comparison to the previously discussed method, Equation 2.63 can be written explicitly as Equation 2.64 where $\rho = \sum_{i=1}^{N} \langle \chi_i | \chi_i \rangle$

$$\begin{aligned} E[\rho(r)] = \sum_{i}^{N} \left(\left\langle \chi_i | -\frac{1}{2}\nabla_i^2 | \chi_i \right\rangle - \left\langle \chi | \sum_{k}^{nuclei} \frac{Z_k}{|r_i - r_k|} | \chi_i \right\rangle \right) \\ + \sum_{i}^{N} \left\langle \chi_i | \frac{1}{2} \int \frac{\rho(r')}{|r_i - r'|} dr' | \chi_i \right\rangle + E_{xc}[\rho(r)] \end{aligned} \tag{2.64}$$

Here the corrective terms have been grouped together in a variable referred to as the exchange-correlation energy, the functional form of which will be discussed next.

In consideration of the exchange-correlation functional, it is worth noting that there

are two main general ways of treatment, the local density approximation (LDA)[ii] and the generalized gradient approximation (GGA).[92] In LDA, the exchange correlation functional is determined solely from the density at that location, while in GGA the gradient of the density is also considered in addition to the local density.

Currently, one of the most widely applied density functionals (as well as one of two functionals employed in the work contained in this thesis) is the B3LYP functional composed of the LYP correlation functional and the B3 form of Becke's exchange functional.[93–95] The LYP functional is a GGA correlation functional containing three empirical parameters.[96] The exchange functional is a hybrid version of Becke's GGA "B" functional where the exact exchange of the non-interacting system is calculated in an HF-type step. This accounts for a fairly large, but not explicitly known, portion of the exchange related energy. The exchange-correlation treatment in the B3LYP functional can then be summarized as below, where $a = 0.20$, $b = 0.72$, $c = 0.81$:

$$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{HF} + b\Delta E_x^B + (1-c)E_c^{LSDA} + cE_c^{LYP} \tag{2.65}$$

The B3LYP functional proves surprisingly versatile and accurate over a very wide range of chemical systems. In the work presented in this thesis, all DFT results are obtained with either the B3LYP functional or the M06 functional.[97,98] The M06 functional, is a recently introduced functional with improved treatment of dispersive and non-covalent interactions. Dispersion is implicitly accounted for through parametrization and accounts for most of the energy associated with dispersive interactions without an explicit empirical correction.[99] M06 appears to outperform the B3LYP functional for calculation of biomolecules, especially in the calculation of vibrational modes (more attention and detailed examples will be discussed on this topic later in the text).

[ii]The local spin density approximation (LSDA) refers to the LDA method with inclusion of spin.

The inability to properly describe dispersive interactions is a notable failure of many density functional methods.[100,101] One method to correct for this is to use an empirical correction of the following form:

$$E_{DFT-D} = E_{ks-DFT} + E_{disp} \tag{2.66}$$

where

$$E_{disp} = -s_6 \sum_{i=1}^{N_{at-1}} \sum_{j=i+1}^{N_{at}} \frac{\sqrt{C_6^i C_6^j}}{R_{ij}^6} f_{damp}(R_{ij}) \tag{2.67}$$

$$f_{damp}(R_{ij}) = \frac{1}{1 + e^{-d(R_{ij}/R_r - 1)}} \tag{2.68}$$

Here, R_r is the sum of van der Waals radii and d is a parameter controlling the steepness of the damping function. While this empirical dispersion correction has been shown to improve results,[87,102] it often over corrects for dispersion and therefore must be used with some degree of caution.

Basis Sets

Computationally, the electronic wavefunction, ϕ, is a mathematical construction of linear combinations of atomic orbitals (LCAO) , φ_i.

$$\phi = \sum_{i=1}^{N} a_i \varphi_i \tag{2.69}$$

Here, the set of N functions, φ_i, and coefficients, a_i, makes up the basis set. Using a larger basis set (larger N) increases the completeness of the mathematical description of the

wavefunction, however this comes at an increased computational cost. In nearly all cases, Gaussian-type functions are used to model the atomic orbital-type functions in LCAO wavefunctions.[103] These Gaussian-type atomic orbitals, when normalized and in cartesian coordinates, have the following form:

$$\phi(x,y,z;\alpha,i,j,k) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left[\frac{(8\alpha)^{i+j+k} i!j!k!}{(2i)!(2j)!(2k)!}\right]^{1/2} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)} \tag{2.70}$$

In this function of coordinates (x, y, z), α controls the width of the function, while parameters i, j, k are integer parameters which are used to determine the orbital "type" For example, a spherical s-type orbital results when $i, j, k = 0$ and when one of i, j, or k equals one then a p-type orbital exists. In such a case, there are three possible orbitals resulting from $i = 1, j = 1$ or $k = 1$, each lying along one of the x, y, z axes and labelled as the P_x, P_y, P_z type orbitals. If one continues this scheme, in the case that two of i, j, k are equal to one, the d-type functions will result. In this case, there are six possible d-functions, where the $x^i y^j z^k$ term in Equation 2.70 have possible values of $x^2, y^2, z^2, xy, xz, yz$. These orbitals are referred to as the Cartesian functions, while the more commonly encountered $xy, xz, yz, x^2 - y^2, z^2$ functions are referred to as the canonical functions and can be obtained from linear combinations of the cartesian functions. Similar combinations can be constructed from the cartesian functions of higher $i + j + k$ indices.

In order to be adequate for the description of atomic orbitals, a few problematic features of gaussian functions require that some modifications be made. Firstly, if one were to directly use a Gaussian orbital to describe an atomic orbital, it would quickly become apparent that the amplitude of the function decays too rapidly as one moves away from the nucleus. In addition to this, Equation 2.70 does not provide any way of including the radial nodes that exist in atomic orbitals. Both of these problems are dealt with by

the use of linear combinations of M Gaussian-type orbitals in the following way:

$$\varphi(x, y, z; \{\alpha\}, i, j, k) = \sum_{a=1}^{M} c_a \phi(x, y, z; \alpha_\alpha, i, j, k) \tag{2.71}$$

c is a coefficient that can be used to optimize the contribution of individual Gaussian functions in the sum and for normalization. In most modern basis sets, one or many sums of Gaussian functions are used to model each orbital. The number of sums used to describe each orbital is given the label ζ, and a basis set describing each orbital with, for example, three sums of Gaussians can be referred to as a triple-ζ basis set. One way of increasing the mathematical flexibility while maintaining computational efficiency, is to use fewer sums for orbitals that require less mathematical flexibility; in other words, have less variation between molecules. In many modern basis sets, core orbitals are typically described by only a single set of Gaussian orbitals, as they are less dependent on changes to the bonding environment, while valence orbitals are described by many sets. These types of basis sets are referred to as "split-valence" basis sets. For example, the common Pople basis set 3-21G uses a sum of three Gaussian functions to describe the core orbitals and both a set of two Gaussian functions and a single Gaussian function to model the valence orbitals. The 6-31G basis set is constructed in a similar way, where the core orbitals are represented by a set of six Gaussian functions, while the valence are represented by a set of three and a single function. One can continue this scheme of basis set expansion and obtains increasingly mathematically flexible functions that better model the atomic orbitals. The increasing computational cost of such a methodology presently limits commonly used basis sets to the triple-ζ or quadruple-ζ level.[104]

In reality, molecular orbitals are more complicated than simple combinations of the occupied atomic orbitals, requiring inclusion of additional mathematical flexibility. Polarization functions give basis functions additional spatial flexibility through the inclusion of atomic functions of higher angular momentum than the occupied orbitals. For example,

the the 6-31G(d) basis set includes d-type functions for all non-hydrogen atoms and the 6-31G(d,p) basis set includes d-type functions on all non-hydrogen atoms as well as p-type functions on all hydrogen atoms. Finally, diffuse functions can be added to a basis set when additional flexibility to the orbital size is thought to be necessary, such as is the case with anions.

In general, optimized geometries and vibrational frequencies are obtained using a moderate-size split valence basis set, while higher-level corrections to the electronic energy are obtained with larger basis sets and possibly a higher accuracy method. A typical scheme of a calculation for a moderately sized organic molecule (10-50) atoms is shown below:

```
MP2/6-311+G(d,p)//B3LYP/6-31G(d)
```
(2.72)

Here the method and basis set combination appearing after "//" is the level of theory used for the geometry optimization and vibrational analysis steps of the calculation, while the method and basis set combination before "//" is the level of theory used to obtain the final electronic energy of the structure optimized at the first level of theory. In this example, a split-valence, double-zeta basis set with d-type polarization functions centred on all non-hydrogen atoms is used to obtain geometries and frequencies, while a larger split-valence, triple-zeta basis set with diffuse functions and d-type polarization functions placed on all non-hydrogen atoms and p-type polarization functions placed on hydrogen atoms. This approach is justified because structural parameters tend to be sufficiently accurate at lower levels of theory, while electronic energies benefit from higher-level corrections (as well, they are significantly less computationally demanding). Results contained in this thesis were obtained using similar methodology and levels of theory as discussed here.

Resolution of the Identity Approximation

The resolution of the identity (RI) approximation is an alternative approach for computing the electronic Coulomb interaction by expansion in a set of auxiliary functions. This approximation allows the four-centre-two-electron repulsion integrals to be replaced by three-centre integrals, and is capable of rendering their calculation ≈5-10 times more computationally efficient (with the improvement increasing with basis set size).[105–107] Very small errors are associated with this approximation (especially with the use of specialized basis sets) and in most cases are less than the error associated with basis set incompleteness.[108,109]

The RI approximation was used for all quantum chemical calculations in the present work with the exception of those using the M06 functional. The def2-SVP and def2-TZVPP basis sets belong a series which have been specifically optimized as auxiliary basis sets for calculations employing the resolution of the identity (RI) approximation and were exclusively used for all RI jobs.[110] These two basis sets are approximately comparable to the more common double and triple zeta 6-31G(d)/cc-pVDZ and 6-311G(d,p)/cc-pVTZ sets, respectively. Any calculation using this approximation (the majority of calculations discussed in the following sections) will have a prefix attached to the method label following the example, RI-B3LYP-D.

Post Hartee-Fock Methods

As mentioned previously in the description of HF theory, electron correlation effects are completely neglected in an HF calculation where each electron only experiences the static field of all other electrons. Møller-Plesset perturbation theory yields an estimate of the electronic correlation effects, and a brief description of it is presented below.[111–114] The procedure begins by separating the Hamilitonian, H, into two parts

$$H = H_0 + \lambda V \tag{2.73}$$

Here H_0 is a solvable reference Hamiltonian operator and V is some perturbation operator with associated coefficient λ. H_0 is chosen to be the non-interacting Hamiltonian (sum of one-electron operators over all electrons) from HF theory, where there are n basis functions.

$$H_0 = \sum_{i=1}^{n} F_i \tag{2.74}$$

The perturbed HF wavefunction is expanded in a power series where truncation at each, n, level of the expansion produces the various MPn methodologies. From the MP0 level, the energy, E_0, is equal to the sum of the eigenvalues of the one-electron operators, which is simply the sum of the orbital energies in HF theory. This is not the same as the final electronic energy obtained from HF theory, as the electronic repulsion is counted twice (repulsion of every electron to all other electrons is included in the sum). At the MP1 level, E_1 corrects for this double-counting and the sum of corrections from MP1 and MP0 result in the proper HF electronic energy.

$$< \Psi^{(0)}|H_0|\Psi^{(0)} >= E^{(0)} \tag{2.75}$$

$$< \Psi^{(0)}|V|\Psi^{(0)} >= E^{(1)} \tag{2.76}$$

$$E^{(0)} + E^{(1)} =< \Psi^{(0)}|H_0|\Psi^{(0)} > + < \Psi^{(0)}|V|\Psi^{(0)} >=< \Psi^{(0)}|H|\Psi^{(0)} >= E_{HF} \tag{2.77}$$

MP2 is, therefore, the first level of perturbation theory which improves upon the energy obtained from HF theory, E_{HF}.

The second-order correction, $E^{(2)}$, is a sum over determinants involving only double excitations, while higher-levels, such as MP4 include triply and quadruply excited determinants. Although quite significantly more computationally costly than an HF calculation (MP2 scaling by N^5 and MP4 scaling by N^7, where N is the number of basis functions, N), the resulting energies account for ≈80-90% of the correlation energy at the MP2 level and ≈95% at the MP4 level. The MP2 correction will in all cases be a negative correction to the HF energy, however, it is possible to be an overestimation of the correlation energy since it is not a variational method.

Approximate Coupled Cluster Theory

Another common post-HF methodology is coupled cluster (CC) theory, which is one of the most accurate methods for incorporation of electron correlation. Although only mentioned briefly here, a more detailed description can be found in the provided references.[80,88,115] While the higher accuracy implementations of this theory are exceedingly computationally expensive (CCSDT, for example, scales as N^8 where N is the number of basis functions), approximate implementations of this theory (which is, of course, still itself an approximation) can be much more computationally affordable (CC2 scales as N^5, the same as MP2).[116] CC2 energies are reported to have a quality similar to MP2 energies with slight improvements, however, they are obtained not through a perturbative process, but rather through an approximation to coupled cluster singles and doubles (CCSD). Implementation of the CC2 method using the RI approximation for the two electron integrals extends the applicability to larger systems (up to ≈1500 basis function).[117]

2.2.4 Additional Considerations

Vibrational Frequencies

Harmonic vibrational frequencies can be calculated from the nuclear Schrödinger equation, where N is the number of atoms, m_i are the atomic masses, $V(\mathbf{q})$ is the potential energy as a function of the $3N$ nuclear coordinates, $\mathbf{q}$, and $\Xi(\mathbf{q})$ is the nuclear wavefunction.[80,88]

$$\left[-\sum_{i}^{N}\frac{1}{2m_i}\Delta_i^2 + V(\mathbf{q})\right]\Xi(\mathbf{q}) = E\Xi(\mathbf{q}) \tag{2.78}$$

The potential energy can be approximated by a second-order Taylor expansion around the stationary geometry, $\mathbf{x_0}$, on the potential energy surface (PES) .

$$V(\mathbf{x}) \approx V(\mathbf{x_0}) + \left(\frac{dV}{d\mathbf{x}}\right)^t(\mathbf{x}-\mathbf{x}_0) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^t\left(\frac{d^2V}{d\mathbf{x}^2}\right)(\mathbf{x}-\mathbf{x}_0) \tag{2.79}$$

The energy of $V(\mathbf{x_0})$ can be defined to be zero, and because $\mathbf{x_0}$ is a stationary point, the first derivative is also zero. The Hessian matrix, $\mathbf{H}$, is a $3N\times3N$ matrix composed of the second derivatives of the energy with respect to the nuclear coordinates.

$$V(\Delta\mathbf{x}) = \frac{1}{2}\Delta\mathbf{x}^t\mathbf{H}\Delta\mathbf{x} \tag{2.80}$$

$$\left\{-\sum_{i=1}^{3N}\left(\frac{1}{2m_i}\frac{\partial^2}{\partial x_i^2}\right) + \frac{1}{2}\Delta\mathbf{x}^t\mathbf{H}\Delta\mathbf{x}\right\}\Psi_{nuc} = E\Psi_{nuc} \tag{2.81}$$

The Hessian, $\mathbf{H}$, is then transformed to mass-weighted coordinates and diagonalized. The resulting eigenvectors are the mass-weighted vibrational normal coordinates and the eigen-

values, ϵ_i, are the mass-weighted force constants related to the vibrational frequencies in the following expression:

$$\nu_i = \frac{1}{2\pi}\sqrt{\epsilon_i} \tag{2.82}$$

Six of the $3N$ eigenvalues will be zero (five for linear molecules), related to translations and rotations. If all other vibrational frequencies are found to be non-imaginary, the geometry is a minimum on the potential energy surface, and if one imaginary frequency is found the geometry represents a transition state structure. The harmonic vibrational frequencies calculated in this way tend to be quite significantly in error from the real values. This is a result of the harmonic approximation itself (assuming quadratic behaviour of the potential about the equilibrium geometry) as well as errors associated with the method and basis set. Fortunately, this error tends to be rather systematic and is easily improved by the use of a scaling factor, where a reduction of the vibrational frequency by 1-10%, in most cases, accounts for a large portion of the error. Generalized scaling factors for the most common computational methodologies (method/basis set combination) are largely available in the literature, and are widely employed and verified.[118–121] In many cases, application of unique scaling factors to different regions of the vibrational spectrum has been shown to be more appropriate than a single generally applied scaling factor.[60,122]

Thermodynamic properties

After obtaining the electronic energy, optimized coordinates and vibrational analysis for a given molecular system, the necessary information for computing all thermodynamic quantities is available. Many electronic structure programs (including TURBOMOLE 6.1 and Gaussian 09) have incorporated programs to calculate many thermodynamic properties (U_T, S_T^o, etc.). A brief generalization of the statistical thermodynamics involved begins with noting the partition function, its separability within the ideal gas approximation

(particles assumed to be non-interacting), and finally the factorization of the molecular partition function.[80,88]

$$Q(N,V,T) = \frac{[q(V,T)]^N}{N!} = \frac{[q_{trans}(V,T)q_{rot}(T)q_{vib}(T)q_{elec}(T)]^N}{N!} \tag{2.83}$$

Here, the following forms of the partition functions are used:

$$q_{trans}(T,V) = \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} \frac{k_B T}{P} \tag{2.84}$$

m is the mass, k_B is the Boltzmann constant, T the temperature, h is Planck's constant, and P is the pressure. There are no parameters in the translational partition function which need to be taken from electronic structure calculations.

$$q_{rot}(T) = \frac{\pi^{1/2}}{\sigma_r}\left(\frac{T^{3/2}}{(\Theta_{r,x}\Theta_{r,y}\Theta_{r,z})^{1/2}}\right) \tag{2.85}$$

Here, $\Theta_r = h^2/8\pi^2 I k_B$ where I is the moment of inertia and σ is the symmetry number.

$$q_{vib}(T) = \prod_{i=1}^{3n-6(5)} \frac{1}{1-e^{-\Theta_v/T}} \tag{2.86}$$

Here, $\Theta_v = h\nu_i/k_B$ where ν_i represents the $3n$-6 ($3n$-5 for linear molecules) vibrational modes obtained from the vibrational analysis, where n is the number of atoms comprising the molecule.

$$q_{elec}(T) = \omega_0 e^{-\epsilon_0/k_B T} \approxeq \omega_0 \tag{2.87}$$

Here, ω_0 is the degeneracy of the state with electronic energy, ϵ_0. All excited states are usually assumed to be inaccessible and with the ground electronic state defined to be zero, the partition function equals ω_0, in this case the electronic spin multiplicity.

The molar internal energy and the entropy can be determined from the partition function using the following expressions:

$$U_T = N_A k_B T^2 \left(\frac{\partial \ln q}{\partial T} \right)_V \tag{2.88}$$

$$S_T^o = N_A k_B \left[\frac{\partial T \ln q}{\partial T} + \ln N_A + 1 \right] \tag{2.89}$$

from which the enthalpy and free energy follow ($P^0 V = nRT, N_a k_B = R$).

$$H_T^o = U_T + P^o V = U_T + RT \tag{2.90}$$

$$G_T^o = H_T^o - T S_T^o \tag{2.91}$$

Once free energies are calculated, the distribution of relative abundances within a set of isomers can be obtained using Equation 2.92 below.

$$K = \frac{N_B}{N_A} = \frac{P_B}{P_A} = e^{\frac{\Delta G_T^o}{RT}} \tag{2.92}$$

Within a set of ions having a thermal distribution of internal energies, two isomers in equilibrium and related by a unimolecular reaction will have relative abundances exponentially related to the difference in free energy between them, ΔG_T^o. For example, at 289 K if

isomer **A** is 5 kJ mol^{-1} higher in free energy than isomer **B** they will exist in 7.5:1 ratio while at a 20 kJ mol^{-1} difference, 3200:1. All discussions of thermodynamic values that follow refer to values calculated for 298.1 K.

Practical Considerations

REMD simulations were performed with the modified Tirex program and the atomic multipole-based, polarizable AMOEBA potential as previously discussed. B3LYP-D results were calculated using the Turbomole 6.1 package,[123] while M06 and B3LYP results were calculated from the Gaussian 09 package.[124] Turbomole calculations used the computational resources of the Laboratoire des Mécanismes Réactionnels (DCMR) at École Polytechnique, Paris and Gaussian 09 jobs used the resources of the Shared Hierarchical Academic Research Computing Network (SHARCNET) at the University of Waterloo, the Centre Informatique National de l'Enseignement Supérieur (CINES), and DCMR. Visualization of computational results was accomplished using VMD[125] and Molden.[126]

Chapter 3

Sodiated Poly(Alanine)

3.1 Introduction

The most common secondary structural feature of proteins is the α-helix (comprising $\approx$30% of structure of the average protein).[127] Of all amino acids, alanine is known to have the highest propensity to form α-helical structure.[127] Previous experimental results in the literature indicate that sodium cationized poly(alanine) peptides twelve residues and larger form α-helical structures in the gas phase at room temperature. This is in contrast to protonated poly(alanine) peptides, where α-helical conformations are not seen even for significantly larger sizes ($\geq$ 20 residues). Sodiation occurs at the C-terminus where the sodium cation interacts favourably with the macrodipole of the helix. This factor, as well as the balance between coordination of the sodium (greater in globular structures) and formation of strong hydrogen bonds in the α-helical structure, are consistent with the existence of a size-dependent transition from globular to α-helical structure. Ion mobility spectrometry results are shown in Figure 3.1 and indicate that this transition occurs somewhere between 8-12 residues, however the exact transition is indistinguishable within the resolution of these measurements.[128] Relative cross section refers to the following quantity,

where n is the number of residues and 14.50 Å^2 is the calculated increase per residue for the α-helical conformation:

$$\Omega_{rel} = \Omega_{measure} - 14.50n \tag{3.1}$$

Using this relation, one would expect that for a series of α-helical peptides of increasing size, the relative cross section would remain constant, while for a series of globular peptides, the relative cross section would decrease with increasing peptide size. This behaviour is consistently observed for the protonated (globular) and sodiated (α-helical) peptides in Figure 3.1. For smaller peptides, the cross sections for helical and globular peptides are largely similar, making distinguishment on this basis alone difficult;[128] however, it is likely that sodiated peptides down to $\approx Ala_{10}Na^+$ "are largely helical".[128]

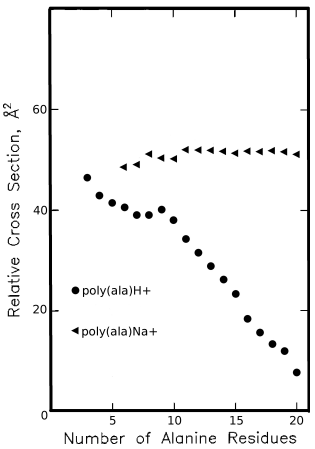

Figure 3.1 – Peptide cross section as a function of size from IMS results.[128]

REMD results also indicate that protonated $Ala_{12}H^+$ is a globular peptide in the gas phase (shown in Figure 3.2). Neutral Ala_{12} however, is calculated (both in the current work and in previous studies[129]) to have α-helical structure in the gas phase (also in Figure 3.2).

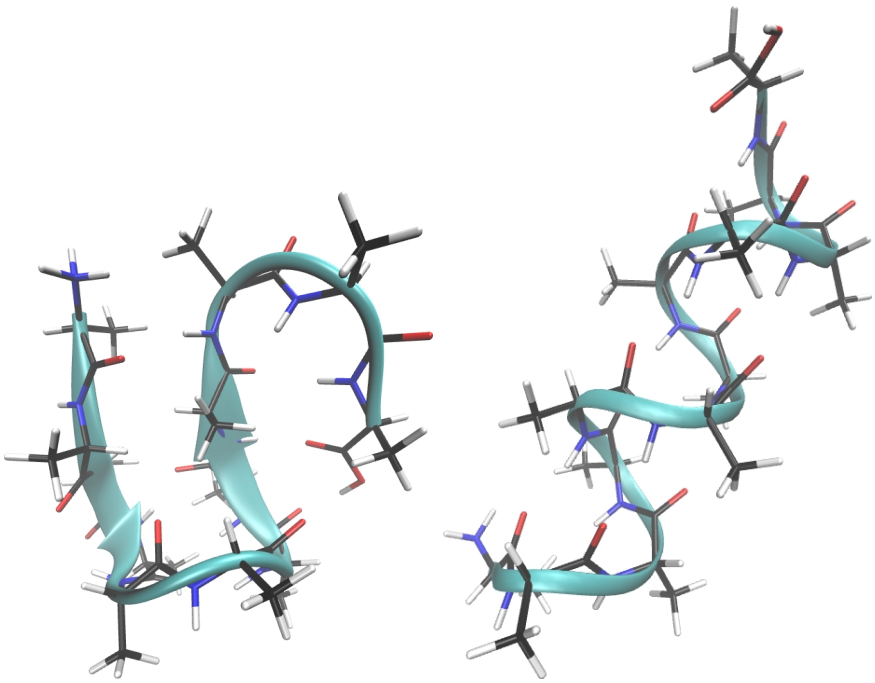

Figure 3.2 – Lowest-energy **AMOEBA** structure for $Ala_{12}H^+$ from REMD results (left) and lowest-energy **AMOEBA** structure for Ala_{12} from REMD (right).

A key influence on the structure of the α-helix is the formation of a macrodipole along the axis of the helix created by the combination of individual dipoles of each residue and follows the direction of the amidic carbonyl groups (negative at C-terminus).[130] This dipole can be utilized to stabilize helical structure by the addition of an amino acid with a negatively charged sidechain at the N-terminus or a positively charged sidechain at the C-terminus.[131] In place of substituting an amino acid in the peptide, a positive ion bound to the C-terminus or a negative ion bound to the N-terminus can also be used to stabilize the helical conformation through a charge-dipole interaction. A protonated peptide therefore, (assuming protonation to occur at the N-terminus) will have a lower tendency to form α-helical structure as the additional positive charge at the N-terminus further destabilizes the dipole moment. This is clearly illustrated by the REMD nomenclatureREMDReplica Exchange Molecular Dynamics results shown in Figure 3.2 above. Conversely, sodiation instead of protonation occurs at the C-terminus and offers a stabilizing effect to the helical structure, as will be seen in the following sections.

3.2 Computational Results

The goal of the computations discussed in this section was to obtain calculated infrared vibration spectra of a small selection of the lowest-energy conformations of each sodiated poly(alanine) under consideration. For these sizes of systems, it is very difficult to explore the entire potential energy surface with chemical intuition alone and the assistance of REMD simulations was fully taken advantage of.

The following abbreviations will be used to refer to selected levels of calculation in the remainder of the text.

(AMOEBA) REMD structures optimized using the AMOEBA force field

(B3LYP-D) RI-B3LYP-D/def2-TZVPP//RI-B3LYP-D/def2-SVP[i]

(M06) M06/6-311G+(d,p)//M06/6-31G(d,p)

(MP2) RI-MP2/def2-TZVPP//M06/6-31G(d,p)

(CC2) RI-CC2/def2-TZVPP//M06/6-31G(d,p)

REMD simulations required approximately 2 days each over 12 processors and 8 gb of memory. Geometry optimizations at the RI-B3LYP-D/def2-SVP level were calculated using one processor and 1000 mb of memory, each requiring approximately one week to complete, while B3LYP-D/def2-SVP frequencies were calculated using four processors, 4000 mb of memory and required approximately one week to finish. Each M06/6-31G(d,p) geometry optimization and frequency calculation was computed using 8-12 processors and 4000-30,000 mb of memory with job times ranging from five to 15 days. RI-MP2 jobs were completed in parallel over 8 CPUs, where RI-MP2/def2-SVP geometry optimizations

[i]Frequencies were obtained without the use of the RI approximation.

and frequency calculations each required approximately 15-20 days, while RI-MP2/def2-TZVPP single point (SP) electronic energy calculations required approximately 2-3 days (on the RI-MP2/def2-SVP geometries). **MP2** calculations are fully correlated, while **CC2** results are fully correlated only for Ala_8Na^+ but use the frozen-core approximation for Ala_9Na^+ and $Ala_{10}Na^+$. In calculated structures the following colours are associated with elements: **Carbon**, **Oxygen**, **Nitrogen** and **Sodium**. Calculated structures are named with the abbreviated size of peptide followed by either a four number index from the REMD output labels or "helix" for the α-helical structure (for example, A8_helix and A8_4740). Individual residues within a given peptide are numerically indexed, starting from the N-terminus.

3.2.1 REMD Results

Many candidate structures resulted from REMD simulations. In most cases, four or more sequential REMD simulations (generations) were completed until the lowest-energy results of subsequent simulations were consistent. For each size of peptide, simulations were initiated from a selection of user-generated starting structures to ensure that the same lowest-energy structures resulted and that the full potential energy surface was properly explored (in total an average of 8-10 generations for each size). The table below shows the number of structures resulting from one REMD simulation as a function of peptide size, as well as the number of REMD candidate structures that were further considered using DFT methods (labels defined above).

REMD structures were analyzed using energetic and structural considerations. Each REMD structure was compared with all others by RMSD (all atoms) allowing for automated elimination of overly similar structures. After this structural sorting, about 20-30 unique structures were selected from the ≈50 lowest-energy structures spanning approximately 20-25 kJ mol^{-1} in the **AMOEBA** force field from the lowest-energy REMD re-

Table 3.1 – Summary of structures considered at each level of computation for each $Ala_{8-12}Na^+$ peptide. **AMOEBA** results are from a single REMD simulation.

	AMOEBA	**B3LYP-D**	**M06**	**MP2**	**CC2**
Ala_8Na^+	1975	27	5	4	4
Ala_9Na^+	958	19	4	4	4
$Ala_{10}Na^+$	1324	21	4	4	4
$Ala_{11}Na^+$	744	31	4	2	-
$Ala_{12}Na^+$	1604	24	4	2	-

sult. These structures were then further examined at higher levels of theory, as will be described below. REMD simulations were initiated from both helical and globular starting structures. The lowest-energy structures resulting from only a single simulation were not identical from both starting structures, however similar "types" of structures were produced. For Ala_8Na^+, similar globular structures resulted from both the simulation starting from a helical structure and the simulation starting from a globular structure. As well, for $Ala_{11}Na^+$ partially helical structures resulted from both the globular and fully helical simulations. The process of selecting structures to consider at higher levels of theory included the lowest-energy results from both starting geometries.

Table 3.2 below compares the energies of the four lowest-energy (**B3LYP-D** level) structures for each peptide. While it is clear that the **AMOEBA** force field energies do not directly match the DFT calculated free energies, a couple of interesting points should be noted. The magnitude of relative energies is quite similar between the force field energies and the calculated free energies. For $Ala_{9-12}Na^+$ the helix is either correctly calculated to be the lowest-energy structure or very close to it ($\leq$ 2.2 kJ mol^{-1}), while for Ala_8Na^+ it is significantly higher in energy (35.2 kJ mol^{-1}). While it will be seen later in the chapter that this is not strictly accurate (or consistent with the quantum chemical results), the identification of a structural transition at Ala_8Na^+ will be shown to be consistent with other results later in the chapter. Finally, energetics are the most consistent between **AMOEBA** and **B3LYP-D** for $Ala_{11}Na^+$ and $Ala_{12}Na^+$ where all eight of the structures

are of largely helical nature, suggesting that AMOEBA is well-suited to treat peptides containing helical structural features.

Table 3.2 – Comparison of **AMOEBA** energies to **B3LYP-D** calculated free energies (298 K, kJ mol^{-1}). The ordering of calculated structures in the tables below follows the energetic ordering that will be used later in the chapter.

Ala_8Na^+	**AMOEBA**	**B3LYP-D**
A8_helix	35.2	9.6
A8_4740	9.6	0.0
A8_4217	5.7	2.1
A8_7740	0.0	12.3
Ala_9Na^+	**AMOEBA**	**B3LYP-D**
A9_helix	1.1	0.0
A9_3809	0.0	27.2
A9_6182	3.2	21.7
A9_3017	4.5	35.6
$Ala_{10}Na^+$	**AMOEBA**	**B3LYP-D**
A10_helix	2.2	0.0
A10_2555	15.8	15.5
A10_1433	3.7	16.4
A10_0366	0.0	10.2
$Ala_{11}Na^+$	**AMOEBA**	**B3LYP-D**
A11_helix	0.0	0.0
A11_3329	15.0	8.4
A11_9390	16.7	6.7
A11_6849	21.9	6.7
$Ala_{12}Na^+$	**AMOEBA**	**B3LYP-D**
A12_helix	0.0	0.0
A12_3913	13.0	6.8
A12_3365	9.1	4.9
A12_3347	13.3	0.7

3.2.2 Comparison of Energies by Level of Theory

RI-B3LYP-D/def2-SVP geometries and B3LYP-D/def2-SVP frequencies as well as RI-B3LYP-D/def2-TZVPP electronic energies (**B3LYP-D**) were obtained for the selected

AMOEBA structures. The most extensive calculations were performed on $Ala_{10}Na^+$ and provide a benchmark for the other peptides in the series. Energies of all calculations for $Ala_{10}Na^+$ are contained in Table 3.3 below. To verify the validity of these results with respect to the known failures of many DFT methods, RI-MP2/def2-TZVPP electronic energies for the RI-B3LYP-D/def2-SVP geometries were obtained. Consideration of the RI-MP2/def2-TZVPP electronic energy correction did not significantly change the results from the **B3LYP-D** level, with negligible average differences except for A10_1433 (15.2 kJ mol^{-1}). Analysis of two $Ala_{10}Na^+$ structures (A10_helix, A10_2555) at the RI-MP2/def2-TZVPP//RI-MP2/def2-SVP level further confirmed the general validity of the **B3LYP-D** level for calculation of relative energetics within this series of peptides, where the difference in relative energy of A10_helix and A10_2555 between levels is a negligible 0.4 kJ mol^{-1}. Additionally, optimized geometries and frequencies at the M06/6-311G+(d,p)//M06/6-31G(d,p) (**M06**) level were obtained. The **M06** results are qualitatively consistent with the **B3LYP-D** and RI-MP2/def2-TZVPP//RI-MP2/def2-SVP results, although tend to decrease the relative energy of all helical conformations as shown in the table below (a more extensive comparison will be seen in Table 3.22). Results at the M06/cc-pVTZ//M06/cc-pVDZ level for A10_2555 and A10_helix showed only a 0.3 kJ mol^{-1} difference from the M06/6-311G+(d,p)//M06/6-31G(d,p) level in both electronic and free energies. Finally, the relative energy difference between A10_helix and A10_2555 at the **MP2** level and **CC2** level are only slightly lower than results at the **M06** level and very consistent with electronic energies calculated at the RI-MP2/def2-TZVPP//RI-MP2/def2-SVP level. At each level of calculation, except for A10_2555, significant improvements in relative energies result when the single point electronic energy term is included.

Continuing to use $Ala_{10}Na^+$ as an example, the **M06** level apparently overestimates the helical stabilization (in reference to the RI-MP2/def2-TZVPP//RI-MP2/def2-SVP result) by approximately 8.7 kJ mol^{-1} in electronic energy, while the **B3LYP-D** result only differs from the RI-MP2 electronic energy by an insignificant 0.4 kJ mol^{-1}. Addition-

Table 3.3 – Relative electronic (E) and free energies (G, 298 K) for the four lowest-energy calculated structures of $Ala_{10}Na^+$ (kJ mol^{-1}).

	A10_helix		**A10_2555**		**A10_1433**		**A10_0366**	
	E	G	E	G	E	G	E	G
RI-B3LYP-D/def2-SVP	0.0	0.0	34.8	19.1	-3.1	-0.2	0.2	-4.9
B3LYP-D	0.0	0.0	35.9	20.1	18.1	21.0	20.0	14.9
M06/6-31G(d,p)	0.0	0.0	44.3	47.9	29.8	47.0	26.3	49.2
M06	0.0	0.0	40.5	44.2	37.3	54.5	34.7	57.5
M06/cc-pVTZ//M06/cc-pVDZ	0.0	0.0	40.8	44.5	-	-	-	-
RI-MP2/def2-TZVPP//RI-B3LYP/def2-SVP	0.0	0.0	35.7	20.0	33.3	36.2	31.9	26.8
RI-MP2/def2-SVP	0.0	-	23.4	-	-	-	-	-
RI-MP2/def2-TZVPP//RI-MP2/def2-SVP	0.0	-	35.5	-	-	-	-	-
MP2	0.0	0.0	34.9	38.6	36.0	53.2	32.5	55.3
CC2	0.0	0.0	36.7	40.4	30.6	47.8	26.0	48.8

ally, in terms of free energies, the stabilization of the helical structure is calculated to be 24.1 kJ mol^{-1} higher by **M06** than by **B3LYP-D** (more than half of the total value). This suggests that the **M06** calculations could in general have a tendency to somewhat overestimate the stability of the α-helix relative to globular structures (in reference to RI-MP2/def2-TZVPP//RI-MP2/def2-SVP results).

3.2.3 Structural Comparisons Between Levels of Theory

Figures 3.3, 3.4 and 3.5 show structural superpositions at the **AMOEBA**, **B3LYP-D**, **M06**, and RI-MP2/def2-SVP levels. These figures show firstly, that the main structural features and interactions remain consistent between levels, and that the largest variations between structures are found between the **AMOEBA** and the quantum chemical methods. In the two helical peptides, the variations between structures of each level appear to be consistent throughout the structure (for example, the **AMOEBA** structure is consistently shifted from the others throughout the entire structure). The overlay of the lowest-energy globular $Ala_{10}Na^+$ conformation, A10_2555, does not show such a consistent shift between the **AMOEBA** structure and the others, but a rather random set of structural variations.

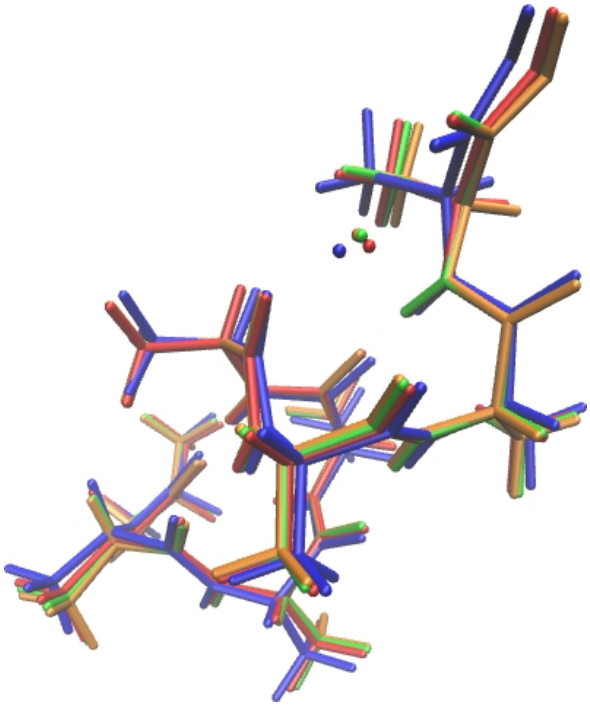

Figure 3.3 – Overlay of computed structures for helical Ala_8Na^+ - A8_helix.

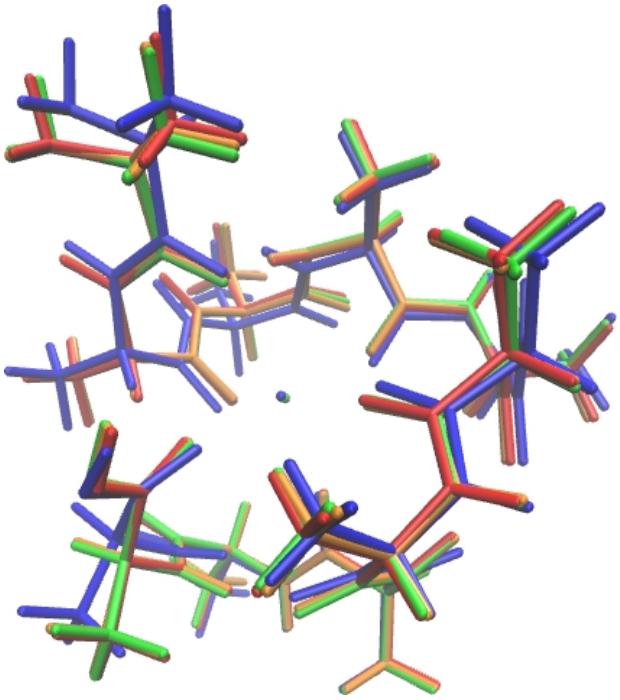

Figure 3.5 – Overlay of computed structures for globular $Ala_{10}Na^+$ - A10_2555.

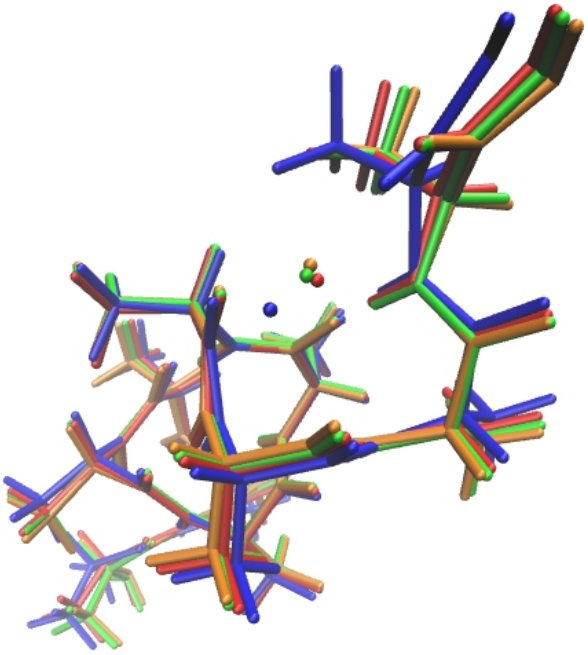

Figure 3.4 – Overlay of computed structures for helical $Ala_{10}Na^+$ - A10_helix.

Colours in the above figures represent the following levels of theory: **AMOEBA**, **RI-B3LYP-D**, **M06** and **RI-MP2/def2-SVP**.

Table 3.4 – Average RMSD between calculated structures by type and at each level of theory. Averages are over all sizes and structures and relative to **M06** geometries (Å).

	Helices	Globules
RI-MP2/def2-SVP	0.106	0.129
B3LYP-D	0.131	0.249
AMOEBA	0.422	0.948

RMSD comparisons between structures calculated at different levels of theory are shown in Table 3.4 for Ala_8Na^+ and $Ala_{10}Na^+$ relative to the **M06** structures. **AMOEBA** structures consistently show larger variations from the **M06** results, while **B3LYP-D** and RI-MP2/def2-SVP deviations from the **M06** structures are significantly smaller.

Table 3.5 contains the calculated gyration radii of the three structures shown in the overlays at each level of theory. Although the variations are small, the trends are very consistent. **AMOEBA** in all cases calculates the least compact structures, **B3LYP-D** calculates the most compact (except for A10_2555) while the **M06** and RI-MP2/def2-SVP results are very similar. The observation that **B3LYP-D** predicts more compact structures, relative to the RI-MP2/def2-SVP results, for the helical conformations while not for the globular conformations is consistent with observations in the literature that overly compact structures result in situations of strong hydrogen bonding for **B3LYP-D** calculations.[87]

Table 3.5 – Comparison of gyration radii for selected structures by level of theory (Å).

	RI-MP2/def2-SVP	**M06**	**B3LYP-D**	**AMOEBA**
A8_4740	4.33	4.33	4.30	4.41
A8_helix	4.75	4.74	4.73	4.76
A10_2555	4.83	4.83	4.84	4.91
A10_helix	5.43	5.44	5.41	5.47

RMSD comparisons provide an indication of overall structural similarity, however comparison of individual parameters is necessary for a better insight into the relationship between structural features and calculated vibrational properties. Table 3.6 and 3.7 lists lengths of all hydrogen bonds and sodium interactions by the level of theory with average deviations and percent differences from RI-MP2/def2-SVP values at the bottom. **AMOEBA** bond lengths are significantly larger than those calculated by RI-MP2/def2-SVP and by both DFT methods. For A10_2555, both **B3LYP-D** and **M06** structures have slightly larger average bond lengths than RI-MP2/def2-SVPstructures, while for A10_helix

B3LYP-D average bond lengths are shorter than RI-MP2/def2-SVP and **M06** are slightly longer. In both cases, **M06** average bond lengths are approximately 1.5% larger than RI-MP2/def2-SVP values. The **B3LYP-D** bond lengths for A10_helix are more substantially contracted for hydrogen bonds and only very slightly contracted for sodium coordinating interactions.

Table 3.6 – Comparison of structural features at different levels of theory for A10_2555 (Å).

	AMOEBA	**B3LYP-D**	**M06**	RI-MP2/def2-SVP
NH_2-NH(2)	2.33	2.12	2.12	2.09
C=O(1)-NH(4)	2.52	2.03	2.17	2.11
C=O(5)-NH(8)	2.23	1.93	2.02	1.94
C=O(7)-NH(10)	2.10	1.91	1.96	1.92
C=O(1)-Na^+	2.44	2.44	2.39	2.41
C=O(2)-Na^+	2.52	2.50	2.57	2.47
C=O(4)-Na^+	2.36	2.28	2.32	2.28
C=O(6)-Na^+	2.48	2.44	2.37	2.37
C=O(8)-Na^+	2.41	2.23	2.23	2.23
C=O(10)-Na^+	2.3	2.41	2.36	2.36
Ave. ΔRI-MP2/def2-SVP	*0.15*	*0.01*	*0.03*	-
Ave. % ΔRI-MP2/def2-SVP	*6.81*	*0.50*	*1.49*	-

Table 3.7 – Comparison of structural features at different levels of theory for A10_helix (Å).

	AMOEBA	**B3LYP-D**	**M06**	RI-MP2/def2-SVP
NH_2-NH(2)	2.31	2.11	2.10	2.08
C=O(1)-NH(4)	2.35	2.15	2.14	2.10
C=O(1)-NH(5)	2.31	2.01	2.11	2.11
C=O(2)-NH(6)	2.05	1.88	1.96	1.91
C=O(3)-NH(7)	2.14	1.93	2.03	1.97
C=O(4)-NH(8)	2.03	1.86	1.92	1.89
C=O(5)-NH(9)	2.08	1.87	1.93	1.90
C=O(6)-NH(10)	1.92	1.83	1.92	1.86
C=O(10)-Na^+	2.20	2.22	2.27	2.25
C=O(8)-Na^+	2.27	2.25	2.27	2.27
C=O(7)-Na^+	2.31	2.25	2.30	2.26
Ave. ΔRI-MP2/def2-SVP	*0.12*	*-0.02*	*0.03*	-
Ave. % ΔRI-MP2/def2-SVP	*6.06*	*-1.06*	*1.55*	-

3.2.4 General Structural Description

Normally, protonated amino acids and peptides are not found in zwitterionic form in the gas phase. Complexation of amino acids and peptides with ions or solvent molecules (such as water or ammonia) has been shown to stabilize the zwitterionic form.[6,132] In most cases, however, complexation with a singly charged metal cation, such as Na^+ is not sufficient to stabilize the zwitterionic isomer. Preliminary calculations on Ala_8Na^+ and $Ala_{10}Na^+$ confirmed the instability of the zwitterionic structures relative to the non-zwitterionic conformations. The three most stable structures (at the RI-B3LYP-D/def2-SVP level) of each size were modified by proton transfer from the C- to N-terminus (the rest of the structure remained unchanged) followed by re-optimization of the geometries. The zwitterionic helical conformations of Ala_8Na^+ and $Ala_{10}Na^+$ were found to be especially unfavoured and with significant structural deviations from the α-helical conformation resulting after geometry optimization. Zwitterionic helices would be especially unfavoured as both ends of the helix would be destabilized by matching of the negative charge at the C-terminus and positive charge at the N-terminus with the same charges of the macrodipole in both cases. Both calculations for globular zwitterionic structures of $Ala_{10}Na^+$ simply transferred the proton back to the C-terminus, returning the original non-zwitterionic structure and therefore are not listed in Table 3.8. The globular zwitterions calculated for Ala_8Na^+ were also not found to be minima on the potential energy surface (proton was transferred back to the C-terminus), however unique higher energy structures resulted (in reference to the original non-zwitterionic structure). Electronic energies relative to the corresponding original (non-zwitterionic) structure at the RI-B3LYP-D/def2-SVP level are shown below. Additionally, as will be seen in the following section, experimental spectra for $Ala_{8-12}Na^+$ have a band that is easily attributable to a free C-OH in the 3000-3600 cm^{-1} region. This band can only result from a protonated C-terminus, which would not be present in a zwitterionic conformation of a sodiated poly(alanine) peptide. From these results it was concluded that

more extensive consideration of the zwitterionic structures was not necessary. All further discussion and results pertain to non-zwitterionic structures.

Table 3.8 – Relative electronic energies for structures resulting from optimization of zwitterionic Ala_8Na^+ and $Ala_{10}Na^+$. Energies are relative to the non-zwitterionic structure before proton transfer (RI-B3LYP-D/def2-SVP, kJ mol^{-1}).

Ala_8Na^+		$Ala_{10}Na^+$	
A8_helix_ZW	338.9	A10_helix_ZW	432.3
A8_glob_ZW	26.0	A10_glob1_ZW	-
A8_glob2_ZW	36.3	A10_glob2_ZW	-

In an α-helical structure, the C=O of a given amino acid interacts with the N-H of the amino acid four residues later with a turn of approximately 100 degress (in a right-handed manner) at each amino acid.[133] Alpha-helical structures for $Ala_{8-12}Na^+$ have very nearly identical structural features at both termini and, aside from the number of residues, are qualitatively the same. At the C-terminus, the sodium cation is coordinated with the C=O of the C-terminus in addition to C=O groups two and three residues earlier (the C=O group of the residue immediately preceding the C-terminal residue is free and pointing somewhat outwards and away from the Na^+). At the N-terminus, the NH_2 group is free, apart from a weak interaction between the nitrogen of the NH_2 with the N-H of the second residue. Additionally, the C=O of the first residue (N-terminal residue) forms hydrogen bonds with the N-H of both the fourth and fifth residues, while the N-H of the third residue is left free. All other residues of the peptide are involved in typical α-helical hydrogen bonds of the C=O(res)$\cdots$N-H(res$-$4) type.

Fewer structural generalization can be seen relating to the globular structures of different sizes of peptides. In most cases, the Na^+ interacts with the C-terminal C-OH in some form, although in some cases the C-OH of the C-terminus is free and in some cases it is not, interacting instead with the N-terminus or with C=Os of other residues in the peptide. The main characteristic difference between the globular structures and the α-helical structures in all cases is the lack of ordering and repetitive structural features in

the globular conformations. The relative stability of globular and α-helical conformations depends on a variety of factors. The higher degree of coordination of the Na^+ in the globular structures comes at the expense of strong and repeating hydrogen bonds in the helix as well as the favourable interaction between the Na^+ and the dipole of the helix. More detailed attention will be given to this topic at the end of the following section of structural descriptions. All structures and frequencies discussed in the following sections were obtained at the M06/6-31G(d,p) level (corresponding to the applied **M06** methodology for $Ala_{11,12}Na^+$ or **CC2** for $Ala_{8-10}Na^+$) unless stated otherwise.

3.2.5 Computational Results - Ala_8Na^+

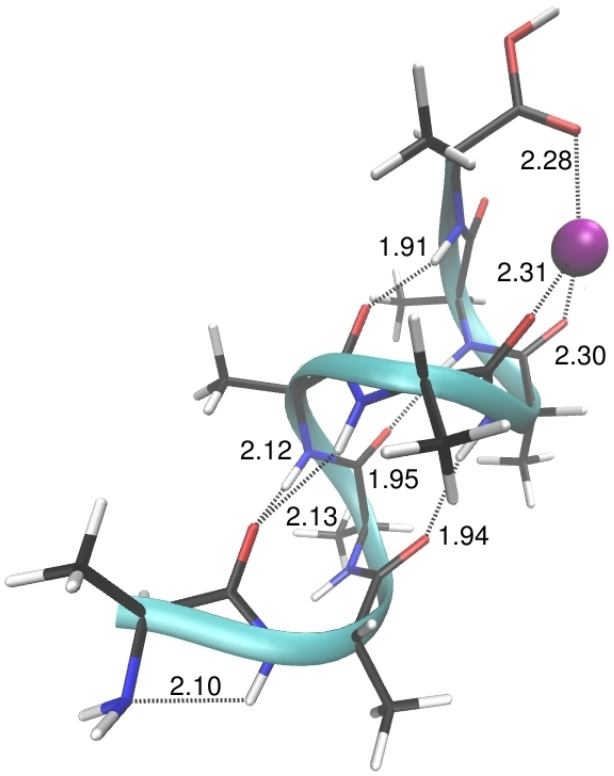

Figure 3.6 – A8_helix (**CC2**)

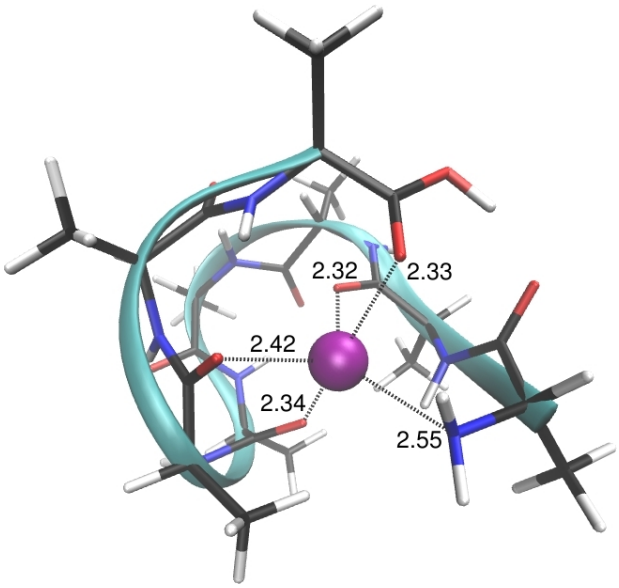

Figure 3.8 – A8_4217 (**CC2**)

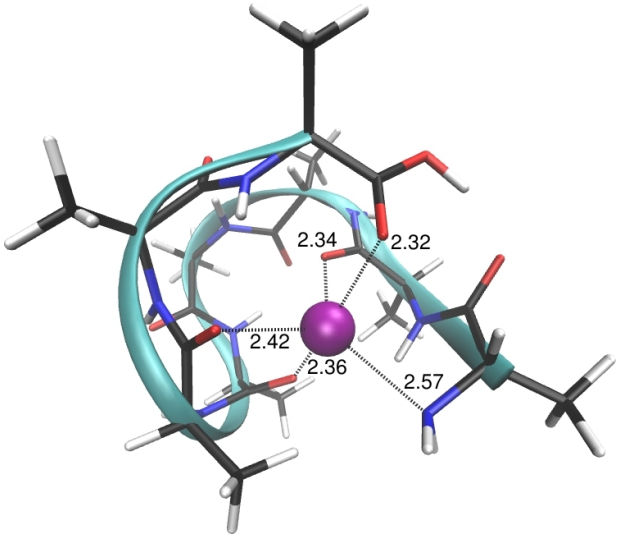

Figure 3.7 – A8_4740 (**CC2**)

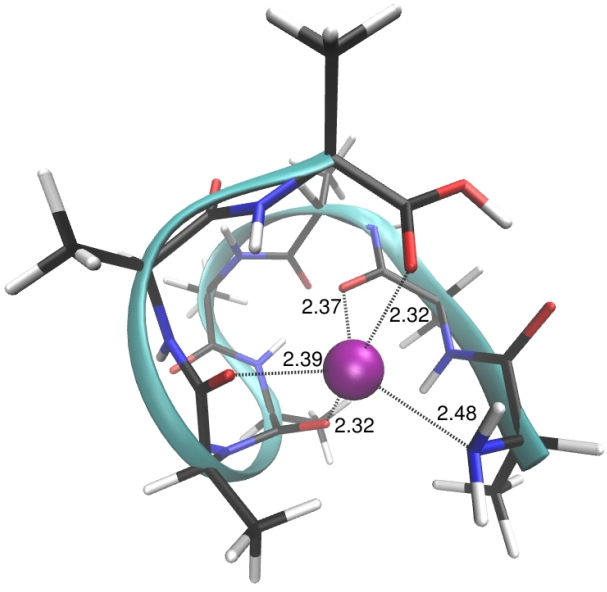

Figure 3.9 – A8_7740 (**CC2**)

Table 3.9 – Relative electronic (E) and free energies (G, 298 K) for Ala_8Na^+ (kJ mol^{-1}).

	A8_helix		**A8_4740**		**A8_4217**		**A8_7740**	
	E	G	E	G	E	G	E	G
RI-B3LYP-D/def2-SVP	39.9	25.5	0.0	0.0	-0.1	1.7	-2.1	7.5
B3LYP-D	23.0	9.6	0.0	0.0	0.4	2.1	2.8	12.3
M06/6-31G(d,p)	0.0	0.0	9.9	24.2	20.7	29.3	11.3	35.3
M06	0.0	0.0	18.3	32.5	16.8	35.8	21.2	45.2
RI-MP2/def2-TZVPP//RI-B3LYP-D/def2-SVP	0.0	0.0	9.0	22.4	9.7	24.9	12.9	35.8
MP2	0.0	0.0	11.2	25.4	10.1	29.1	13.0	37.0
CC2	0.0	0.0	4.7	19.0	4.2	23.3	5.7	29.7
RI-MP2/def2-SVP	0.0	0.0	-6.6	2.6	-	-	-	-
RI-MP2/def2-TZVPP//RI-MP2/def2-SVP	0.0	0.0	7.8	17.0	-	-	-	-
RI-CC2/def2-TZVPP//RI-MP2/def2-SVP	0.0	0.0	2.6	11.9	-	-	-	-

After extensive analysis of **AMOEBA** structures and further consideration at the **B3LYP-D** level, the four lowest-energy structures finally considered at the higher level methods for Ala_8Na^+ are shown above. The calculated α-helical structure, A8_helix, is calculated at the RI-CC2/def2-TZVPP//RI-MP2/def2-SVP level to be 11.9 kJ mol^{-1} lower in free energy than the lowest-energy globular structure. The three globular structures, although relatively similar structurally, span a free energy range of 10.7 kJ mol^{-1}, at the **CC2** level. Stabilization of the α-helix is significantly reduced when results at the **B3LYP-D** level are considered, where the lowest-energy globular structure A8_4740, is lower in free energy by 9.6 kJ mol^{-1}. Inclusion of larger basis set electronic energies increase the relative stability of the α-helical structure in nearly all cases.

A8_helix

The α-helical structure of Ala_8Na^+ has three helical hydrogen bonds between the C=O of residues (2,3,4) and the NH of (6,7,8) with lengths of 1.94, 1.95 and 1.91 Å, respectively. Additionally, there is a hydrogen bond shared between the C=O of residue one and the NH of residues (4,5) with lengths of 2.12 and 2.13 Å, respectively. The sodium cation is coordinated by C=O (5,6,8) at distances of 2.31, 2.30 and 2.28 Å, respectively.

Finally the nitrogen of the N-terminus hydrogen bonds with NH(2) at a distance of 2.10 Å, while NH(3) and C=O(7) are free.

Ala_8Na^+ Globular structures

At the **CC2** level the three lowest-energy globular structures calculated for Ala_8Na^+ (A8_4740 (G19.0 kJ mol^{-1}), A8_4217 (G23.3 kJ mol^{-1}) and A8_7740 (G29.7 kJ mol^{-1}) have largely similar geometries, with hydrogen bonds between C=O(1)-COH (1.76, 1.75, 1.79 Å[ii]), C=O(3)-NH(5) (1.88, 1.87, 1.81 Å), C=O(4)-NH(7) (1.95, 1.93, 2.06 Å) and C=O(7)-NH(4) (1.93, 1.88, 1.91 Å) as seen in the figures above. The Na^+ is coordinated by the N-terminus (2.57, 2.55, 2.48 Å), as well as C=O(2) (2.34, 2.32, 2.37 Å), C=O(5) (2.36, 2.34, 2.32 Å), C=O(6) (2.42, 2.42, 2.39 Å), C=O(8) (2.32, 2.33, 2.32 Å). NH(3) is free.

The main structural differences between these globular structures are found around the N-terminus. In A8_4740, the N-terminal NH_2 interacts with Na^+ (2.57 Å), forms a hydrogen bond with NH(2) (2.33 Å) and does not interact with the C-OH of the C-terminus. A8_4217 differs by a small displacement at the N-terminus in which the NH_2 moves away from NH(2) (2.50 Å) bringing it closer to the C-OH of the C-terminus (2.63 Å) and the Na^+ (2.55 Å). Finally, for A8_7740, the NH_2 interacts with Na^+ (2.48 Å) and weakly with the C-OH (2.43 Å), however the interaction with NH(2) is significantly weaker in this case (3.31 Å). RMSD comparisons between these three structures (relative to the lowest-energy structure) describe, in a general sense, the magnitude of difference between structures and are provided in Table 3.10. These RMSD values are consistent with general structural similarity and indicate the magnitude of structural difference, relative to A8_4740, is greater for A8_7740, than for A8_4217.

[ii]Bond lengths are listed following specification of intramolecular interactions and ordered by increasing relative free energies at the **CC2** level.

Table 3.10 – RMSD between Ala_8Na^+ globular structures, relative to lowest-energy globular structure (**CC2**, Å).

A8	**RMSD**
A8_4740	0.000
A8_4217	0.847
A8_7740	1.221

3.2.6 Computational Results - Ala_9Na^+

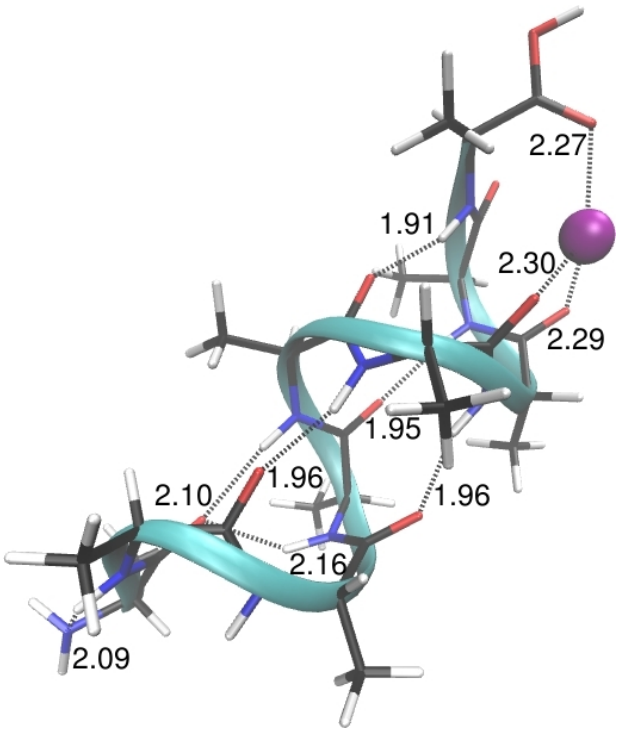

Figure 3.10 – A9_helix (**CC2**)

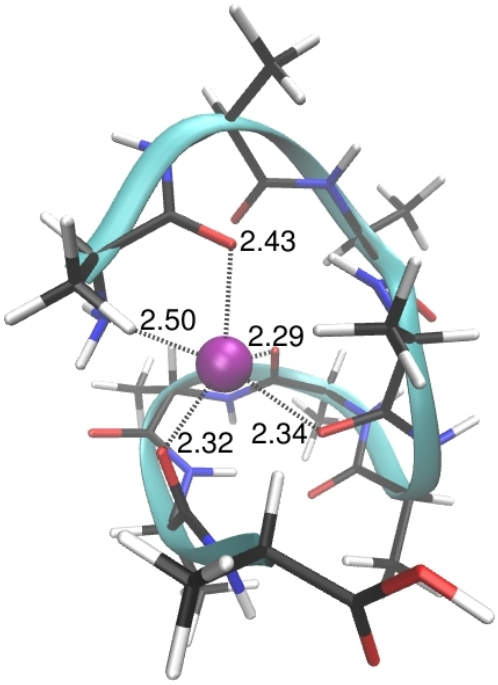

Figure 3.12 – A9_6182 (**CC2**)

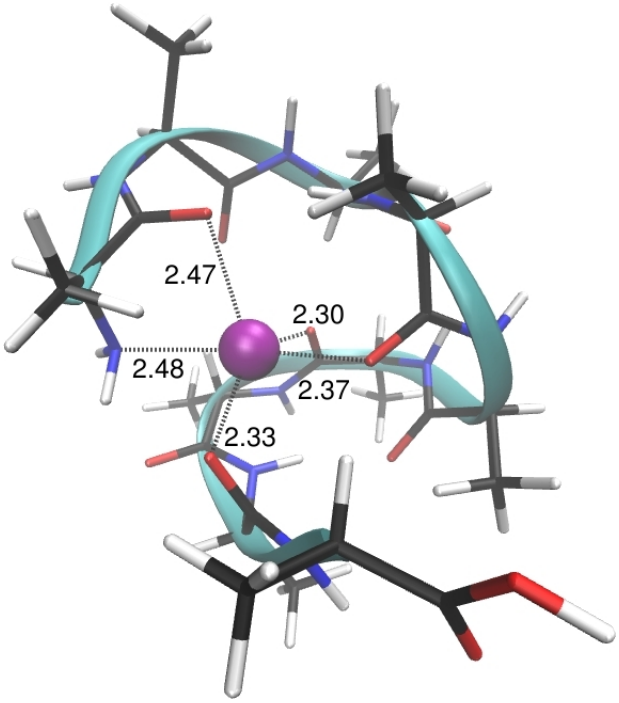

Figure 3.11 – A9_3809 (**CC2**)

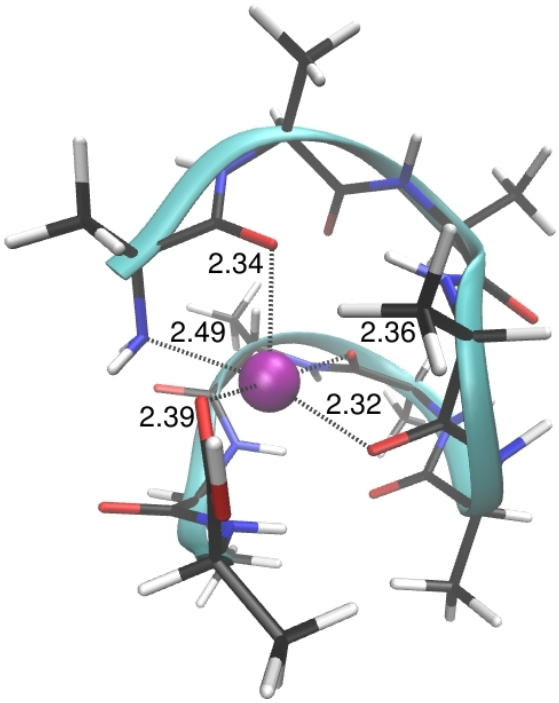

Figure 3.13 – A9_3017 (**CC2**)

Table 3.11 – Relative electronic (E) and free energies (G, 298 K) for Ala_9Na^+ (kJ mol^{-1}).

	A9_helix		**A9_3809**		**A9_6182**		**A9_3017**	
	E	G	E	G	E	G	E	G
RI-B3LYP-D/def2-SVP	0.0	0.0	7.6	13.8	5.4	7.1	-2.7	13.6
B3LYP-D	0.0	0.0	18.3	27.2	17.3	21.7	19.3	35.6
M06/6-31G(d,p)	0.0	0.0	33.4	29.7	38.6	31.9	52.5	37.8
M06	0.0	0.0	37.7	41.5	39.8	46.5	48.9	63.6
RI-MP2/def2-TZVPP//RI-B3LYP-D/def2-SVP	0.0	0.0	40.1	49.0	39.2	43.6	42.9	59.2
MP2	0.0	0.0	35.3	39.1	37.0	43.7	42.2	56.8
CC2	0.0	0.0	30.6	34.3	32.1	38.8	35.5	50.1

The four **B3LYP-D** structures finally selected for consideration at the higher levels of theory are shown above. For Ala_9Na^+, the α-helical structure was calculated to be the lowest-energy structure at all levels of theory considered in terms of free energies. Already at the **B3LYP-D** level the α-helix is stabilized by 27.2 kJ mol^{-1} in free energy over the lowest-energy globular structure, A9_3809, while at the **M06** level it is 41.5 kJ mol^{-1} more stable and at the **CC2** level 34.4 kJ mol^{-1}. Consideration of a higher level electronic energy consistently increases the stability of the α-helical structure over the globular structures, although general trends in the energetic ordering of structures are quite consistent between different levels.

A9_helix

The α-helical structure of Ala_9Na^+ has four helical hydrogen bonds between the C=O of residues (2,3,4,5) and the NH of (6,7,8,9) with lengths of 1.96, 1.96, 1.95 and 1.91 Å, respectively. Additionally, there is a hydrogen bond shared between the C=O of residue one and the NH of residues (4,5) with lengths of 2.16 and 2.10 Å, respectively. The sodium cation is coordinated by C=O (6,7,9) at distances of 2.30, 2.29 and 2.27 Å, respectively. Finally the nitrogen of the N-terminus hydrogen bonds with NH(2) at a distance of 2.09 Å, while NH(3) and C=O(8) are free.

Ala_9Na^+ Globular Structures

At the **CC2** level, the three lowest-energy globular structures calculated for Ala_9Na^+ (A9_3809 (G34.3 kJ mol^{-1}), A9_6182 (G38.8 kJ mol^{-1}) and A9_3017 (G50.1 kJ mol^{-1}) have similar geometries although greater structural differences exist than were seen previously for Ala_8Na^+. For A9_3809 and A9_6182 hydrogen bonds exist between C=O(7)-NH_2 (2.45, 2.52 Å[iii]), C=O(1)-NH(4) (2.15, 2.05 Å), C=O(3)-NH(6) (2.00, 1.98 Å), C=O(5)-NH(8) (1.99, 2.01 Å) and C=O(9)-NH(9) (2.23, 2.25 Å) as seen in the figures above. The Na^+ is coordinated by the N-terminus (2.48, 2.50 Å), as well as C=O(1) (2.47, 2.43 Å), C=O(4) (2.37, 2.34 Å), C=O(6) (2.30, 2.29 Å), C=O(8) (2.33, 2.32 Å). NH(2,3,5,7) and C=O(2) are free of intermolecular interactions. In the structure A9_3017, hydrogen bonds are formed between C=O(7)-NH_2 (2.06 Å), C=O(1)-NH(4) (2.14 Å), C=O(3)-NH(5) (2.40 Å), C=O(4)-NH(9) (2.50Å) C=O(5)-NH(8) (2.16 Å) and the Na^+ is coordinated by the N-terminus (2.49 Å), C=O(1) (2.34 Å), C=O(4) (2.32 Å), C=O(6) (2.36 Å), and C=O(9) (2.39 Å), while NH(2,3,5,7) C=O(2) free.

The main structural difference between structures A9_3809 and A9_6182, and structure A9_3017 is a twist in the peptide backbone of A9_3017 at residue nine. This rotates the C-terminus and NH(9) inwards to the centre of the structure and breaks the C=O(8) interaction with Na^+, but allows the C-terminal C=O(9) to interact with the Na^+. This alteration of the structure leaves C=O(8) excluded and without any interactions, apparently significantly raising the relative energy of this structure in comparison to the lowest-energy globular structure A9_3809 (15.8 kJ mol^{-1} in free energy, **CC2**). At this size, the peptide is somewhat too large to accommodate all the Na^+ interactions without excluding the C-terminus or causing some distortion and strain to the structure. RMSD comparisons in Table 3.12 are consistent with the above description where only small differences exist between A9_3809 and A9_6182 (RMSD is less than 0.5 Å). The significant structural de-

[iii] Bond lengths are listed following specification of intramolecular interactions and ordered by increasing relative free energies at the **CC2** level.

viation at the C-terminus of A9_3017 relative to the other two structures gives the larger RMSD value of 1.519 Å.

Table 3.12 – RMSD between Ala_9Na^+ globular structures, relative to lowest-energy globular structure (**CC2**, Å).

A9	**RMSD**
A9_3809	0.000
A9_6182	0.472
A9_3017	1.519

3.2.7 Computational Results - $Ala_{10}Na^+$

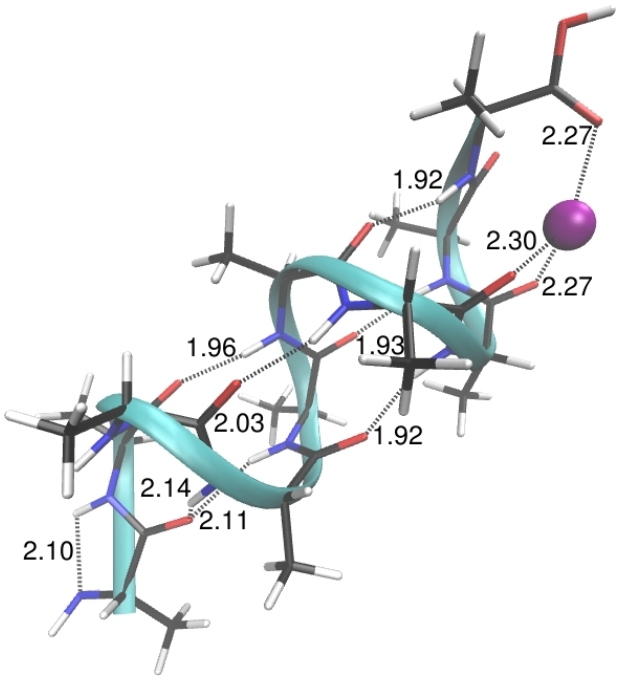

Figure 3.14 – A10_helix (**CC2**)

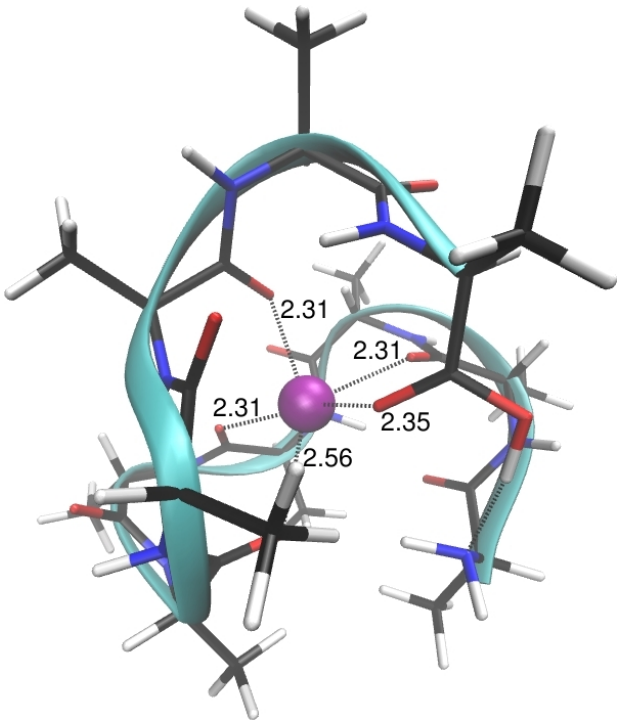

Figure 3.16 – A10_1433 (**CC2**)

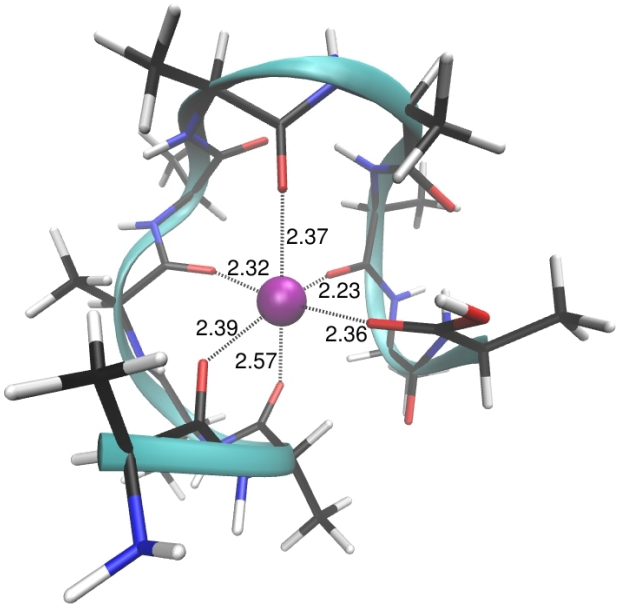

Figure 3.15 – A10_2555 (**CC2**)

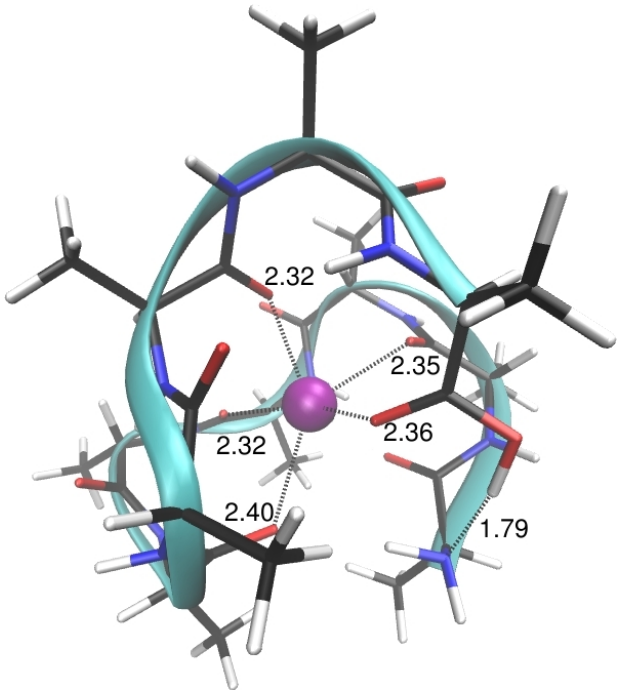

Figure 3.17 – A10_0366 (**CC2**)

Table 3.13 – Relative electronic (E) and free energies (G, 298 K) for $Ala_{10}Na^+$ (kJ mol^{-1}).

	A10_helix		**A10_2555**		**A10_1433**		**A10_0366**	
	E	G	E	G	E	G	E	G
RI-B3LYP-D/def2-SVP	0.0	0.0	34.8	19.1	-3.1	-0.2	0.2	-4.9
B3LYP-D	0.0	0.0	35.9	20.1	18.1	21.0	20.0	14.9
M06/6-31G(d,p)	0.0	0.0	44.3	47.9	29.8	47.0	26.3	49.2
M06	0.0	0.0	40.5	44.2	37.3	54.5	34.7	57.5
M06/cc-pVTZ//M06/cc-pVDZ	0.0	0.0	40.8	44.5	-	-	-	-
RI-MP2/def2-TZVPP//RI-B3LYP/def2-SVP	0.0	0.0	35.7	20.0	33.3	36.2	31.9	26.8
RI-MP2/def2-SVP	0.0	-	23.4	-	-	-	-	-
RI-MP2/def2-TZVPP//RI-MP2/def2-SVP	0.0	-	35.5	-	-	-	-	-
MP2	0.0	0.0	34.9	38.6	36.0	53.2	32.5	55.3
CC2	0.0	0.0	36.7	40.4	30.6	47.8	26.0	48.8

The four lowest-energy **B3LYP-D** identified structures selected for consideration at the higher levels of theory are shown above. For $Ala_{10}Na^+$, the α-helical structure was calculated to be the lowest-energy structure at all levels of theory considered. At the **B3LYP-D** level the α-helix is stabilized by 20.1 kJ mol^{-1} in free energy over the lowest-energy globular structure, A10_2555, while at the **M06** level it is 44.2 kJ mol^{-1} more stable and at the the **CC2** level 40.4 kJ mol^{-1}.[iv]

A10_helix

The α-helical structure of $Ala_{10}Na^+$ has five helical-type hydrogen bonds between the C=O of residues (2,3,4,5,6) and the NH of (6,7,8,9,10) with lengths of 1.96, 2.03, 1.92, 1.93, 1.92 Å, respectively. Additionally, there is a hydrogen bond shared between the C=O of residue one and the NH of residues (4,5) with lengths of 2.14 and 2.11 Å, respectively. The sodium cation is coordinated by C=O (7,8,10) at distances of 2.30, 2.27 and 2.27 Å, respectively. Finally the nitrogen of the N-terminus hydrogen bonds with NH(2) at a distance of 2.10 Å, while NH(3) and C=O(9) are free.

[iv]Comparisons between different computational methodologies for $Ala_{10}Na^+$ were discussed in detail in the introduction of this chapter.

$Ala_{10}Na^+$ Globular Structures

At the **CC2** level, globular structures A10_1433 (G47.8 kJ mol^{-1}) and A10_0366 (G48.8 kJ mol^{-1}) have very similar overall geometries, while the lowest-energy globular structure, A10_2555 (G40.4 kJ mol^{-1}), has a different set of interactions around the C- and N-terminal residues. For A10_1433 and A10_0366 hydrogen bonds exist between C=O(6)-NH_2 (2.60, 2.63 Å[v]), COOH-NH_2 (1.76, 1.79 Å), C=O(1)-NH(4) (1.96, 1.99 Å), C=O(5)-NH(8) (2.03, 1.97 Å) and C=O(7)-NH(10) (2.08, 2.17 Å) as seen in the figures above. The Na^+ is coordinated by C=O(2) (2.31, 2.35 Å), C=O(4) (2.31, 2.32 Å), C=O(6) (2.56, 2.40 Å), C=O(8) (2.31, 2.32 Å) and C=O(10) (2.35, 2.36Å). NH(2,3,5,6,7,9) and C=O(3,9) are free of intermolecular interactions. In the lowest-energy globular structure, A10_2555, hydrogen bonds are formed between C=O(1)-NH(4) (2.17 Å), C=O(5)-NH(8) (2.02 Å), C=O(7)-NH(10) (1.96 Å) and NH_2-NH(2) (2.12 Å), while the Na^+ is coordinated by C=O(1) (2.39 Å), C=O(2) (2.57 Å), C=O(4) (2.32 Å), C=O(6) (2.37 Å), C=O(8) (2.23 Å) and C=O(10) (2.36 Å), while NH(3,5,6,7,9) C=O(3) free.

Only very minor differences exist between structures A10_1433 and A10_0366, such as small variations in combinations of bond lengths, angles and torsions, and all intermolecular interactions are consistent between structures. Structure A10_2555 differs by a twist in the peptide backbone immediately before the NH_2 which rotates the N-terminus away from the centre of the structure and breaks the COOH interaction. This allows the NH_2 to hydrogen bond to NH(2), C=O(6) to more favourably interact with the Na^+ and C=O(1) to coordinate the Na^+, thereby reducing the relative free energy of A10_2555 by 7.4 kJ mol^{-1} (**CC2**). RMSD comparisons in Table 3.14 illustrate the strong similarity between structures A10_1433 and A10_0366 (RMSD of 0.001 Å). The difference in position of the N-terminus (as described above) between these two structures and the lowest-energy globular structure, A10_2555, is seen by the larger RMSD of 1.62 Å.

[v]Bond lengths are listed following specification of intramolecular interactions and ordered by increasing relative free energies at the **CC2** level.

Table 3.14 – RMSD between $Ala_{10}Na^+$ globular structures, relative to lowest-energy globular structure (**CC2**, Å).

A10	**RMSD**
A10_2555	0.000
A10_1433	1.624
A10_0366	1.623

3.2.8 Computational Results - $Ala_{11}Na^+$

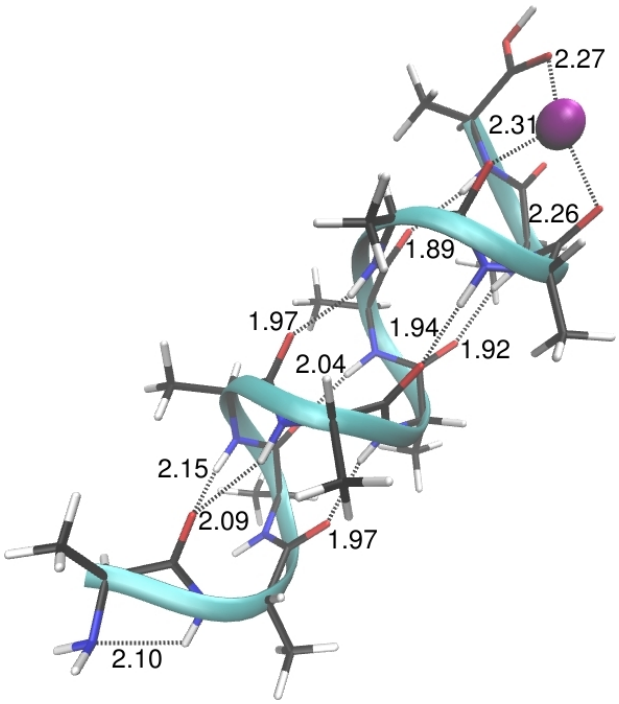

Figure 3.18 – A11_helix (**M06**)

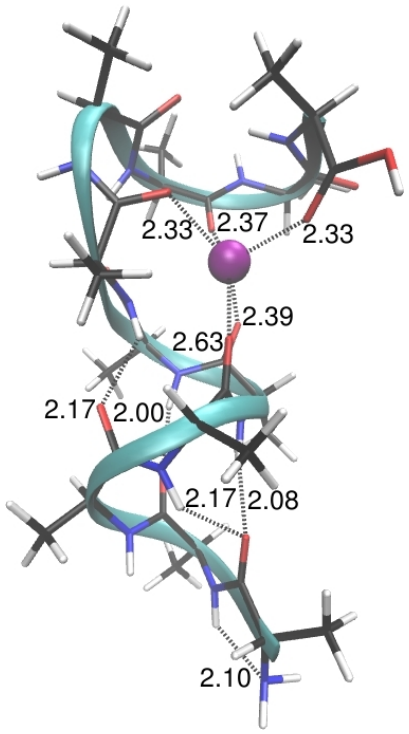

Figure 3.20 – A11_9390 (**M06**)

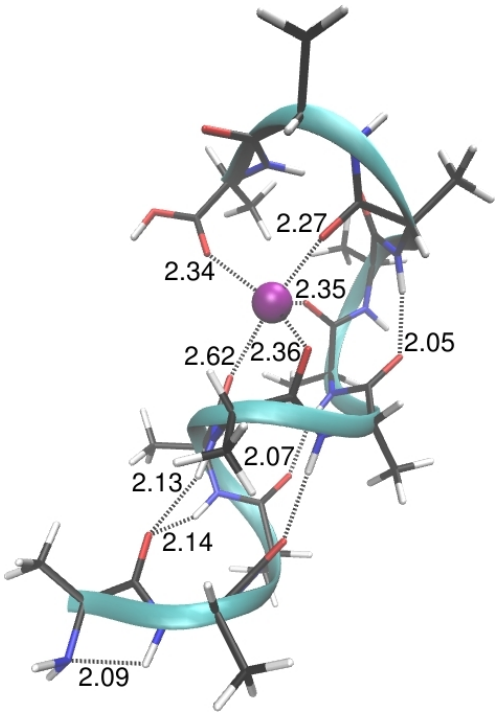

Figure 3.19 – A11_3329 (**M06**)

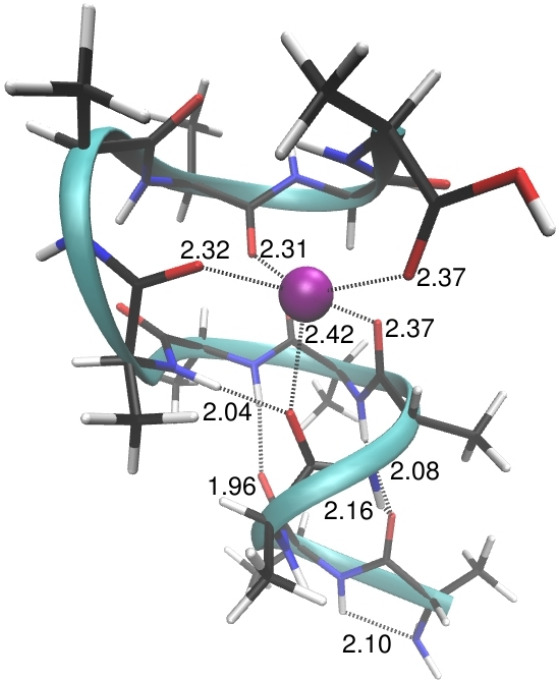

Figure 3.21 – A11_6849 (**M06**)

Table 3.15 – Relative electronic (E) and free energies (G, 298 K) for $Ala_{11}Na^+$ (kJ mol^{-1}).

	A11_helix		**A11_3329**		**A11_9390**		**A11_6849**	
	E	G	E	G	E	G	E	G
RI-B3LYP-D/def2-SVP	0.0	0.0	29.9	10.6	10.3	0.7	30.1	9.0
B3LYP-D	0.0	0.0	27.7	8.4	16.3	6.7	27.8	6.7
M06/6-31G(d,p)	0.0	0.0	31.8	21.3	25.5	35.8	32.2	35.1
M06	0.0	0.0	32.4	21.9	31.3	39.6	36.7	41.5
RI-MP2/def2-TZVPP//RI-B3LYP/def2-SVP	0.0	0.0	28.6	9.4	22.4	12.8	30.9	9.8
MP2	0.0	0.0	29.6	19.0	-	-	-	-

The four lowest-energy identified structures selected for continued consideration beyond the **B3LYP-D** level of theory are shown above. For $Ala_{11}Na^+$, the α-helical structure was calculated to be the lowest-energy structure at all levels of theory considered. At the **B3LYP-D** level the α-helix is stabilized by 8.4 kJ mol^{-1} in free energy over the lowest-energy globular structure, A11_3329, while at the **M06** level it is 21.9 kJ mol^{-1} more stable and at the **MP2** level 19.0 kJ mol^{-1}.

A11_helix

The α-helical structure of $Ala_{11}Na^+$ has six helical-type hydrogen bonds between the C=O of residues (2,3,4,5,6,7) and the NH of (6,7,8,9,10,11) with lengths of 1.97, 2.04, 1.97, 1,94, 1.92 and 1.89 Å, respectively. Additionally, there is a hydrogen bond shared between the C=O of residue 1 and the NH of residues (4,5) with lengths of 2.15 and 2.09 Å, respectively. The sodium cation is coordinated by C=O (8,9,11) at distances of 2.31, 2.26 and 2.27 Å, respectively. Finally the nitrogen of the N-terminus hydrogen bonds with NH(2) at a distance of 2.10 Å, and NH(3) and C=O(10) are free.

$Ala_{11}Na^+$ Partially Helical Structures

At the **M06** level, structures A11_3329 (G21.9 kJ mol^{-1}), A11_9390 (G39.6 kJ mol^{-1}) and A11_6849 (G41.5 kJ mol^{-1}) are helical on the N terminal half of the peptide, while the C-terminal half remains coiled around the Na^+ to maximize the degree of coordination.

For A11_3329, A11_9390 and A11_6849 hydrogen bonds exist between NH_2-NH(2) (2.09, 2.10, 2.10 Å[vi]), C=O(1)-NH(4) (2.14, 2.17, 2.16 Å), C=O(1)-NH(5) (2.13, 2.08, 2.08 Å), C=O(2)-NH(6) (1.99, 2.00, 1.96 Å), C=O(3)-NH(7) (2.07, 2.17, 2.04 Å), C=O(6)-NH(9) (2.05, 2.07, 2.00 Å) and C=O(8)-NH(11) (2.08, 2.12, 2.08 Å) as seen in the figures above. For A11_3329 and A11_9390 the Na^+ is coordinated by C=O(4) (2.62, 2.63 Å), C=O(5) (2.36, 2.39 Å), C=O(7) (2.35, 2.37 Å), C=O(9) (2.27, 2.33 Å) and C=O(11) (2.34, 2.33 Å). NH(3,8,10) and C=O(10) are free. In A11_6849, the Na^+ is coordinated by C=O(3) (2.42 Å), C=O(5) (2.37 Å), C=O(7) (2.32 Å), C=O(9) (2.31 Å) and C=O(11) (2.37 Å) while NH(3,8,10) and C=O(5,10) are free.

Only small differences exist between structures A11_3329 and A11_9390, such as small variations in combinations of bond lengths, angles and torsions. Structure A11_6849 differs by a tighter turn in the helix after residue two which brings C=O(3) to interact with Na^+ in addition to NH(7) and excludes C=O (5) from the Na^+ raising the relative free energy of A11_9390 by 19.6 kJ mol^{-1} (**M06**) in comparison to A11_3329. RMSD values in Table 3.16 are consistent with the smaller observed differences between A11_3329 and A11_9390, where the larger change in the helical structure of the backbone in A11_6849 is consistent with an RMSD of 1.755 Å.

Table 3.16 – RMSD between $Ala_{11}Na^+$ partly helical structures, relative to lowest-energy partly helical structure (**M06**, Å).

A11	**RMSD**
A11_3329	0.000
A11_9390	0.823
A11_6849	1.755

3.2.9 Computational Results - $Ala_{12}Na^+$

[vi]Bond lengths are listed following specification of intramolecular interactions and ordered by increasing relative free energies at the **M06** level.

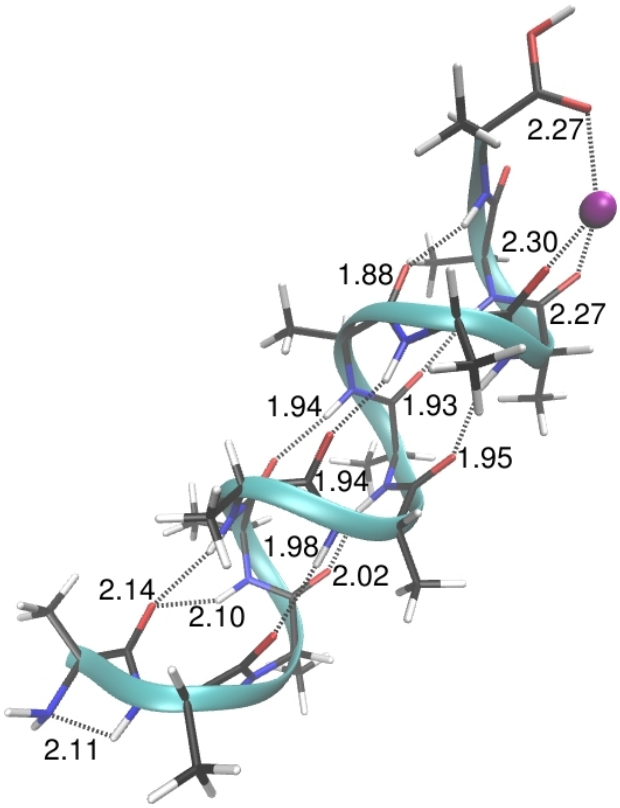

Figure 3.22 – A12_helix (**M06**)

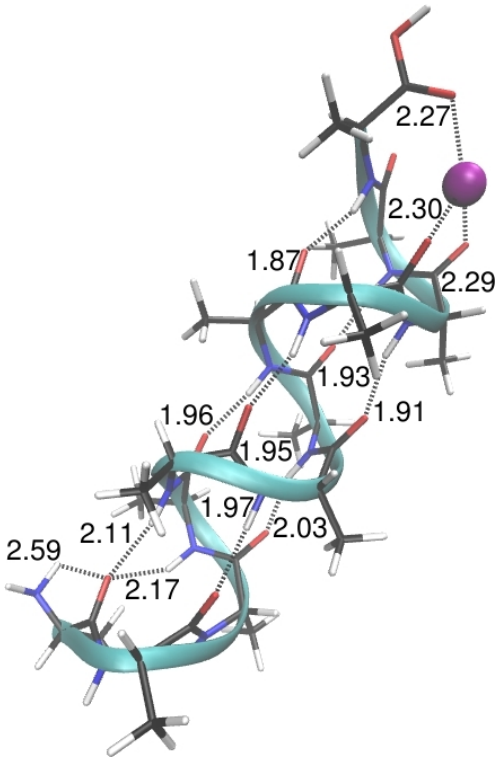

Figure 3.24 – A12_3365 (**M06**)

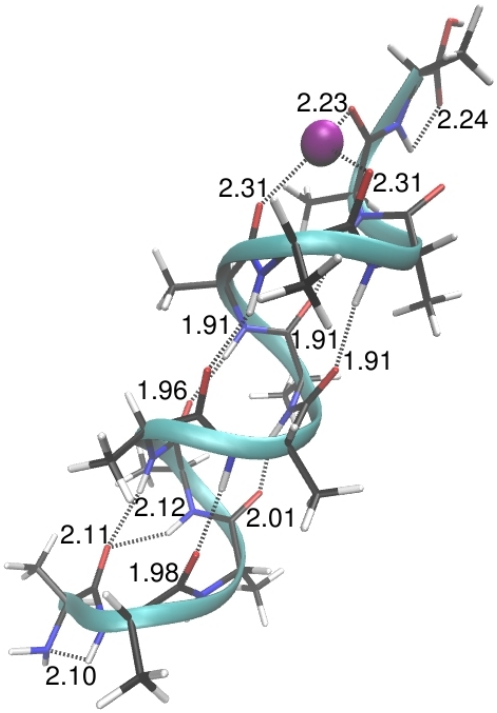

Figure 3.23 – A12_3913 (**M06**)

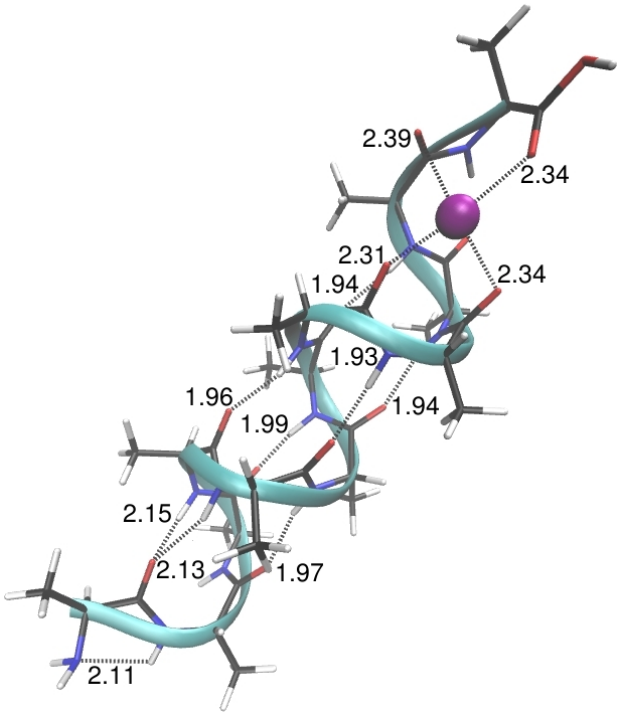

Figure 3.25 – A12_3347 (**M06**)

Table 3.17 – Relative electronic (E) and free energies (G, 298 K) for $Ala_{12}Na^+$ (kJ mol^{-1}).

	A12_helix		**A12_3347**		**A12_3365**		**A12_3913**	
	E	G	E	G	E	G	E	G
RI-B3LYP-D/def2-SVP	0.0	0.0	28.5	32.3	20.2	23.5	4.0	12.2
B3LYP-D	0.0	0.0	3.1	6.8	1.6	4.9	-7.4	0.7
M06/6-31G(d,p)	0.0	0.0	24.7	22.0	17.2	21.5	8.5	13.5
M06	0.0	0.0	20.0	17.2	13.3	17.5	14.4	19.4
RI-MP2/def2-TZVPP//RI-B3LYP/def2-SVP	0.0	0.0	5.4	9.1	2.6	5.9	-2.7	5.4
MP2	0.0	0.0	-	-	-	-	12.9	10.1

A12_helix (G0.0 kJ mol^{-1})

The four lowest-energy identified structures selected for consideration above the **B3LYP-D** level of theory are shown in Table 3.17. For $Ala_{12}Na^+$, the fully α-helical structure was calculated to be the lowest-energy structure at all levels of theory considered in terms of free energies. At the **B3LYP-D** level the α-helix is stabilized by 0.7 kJ mol^{-1} in free energy over the next lowest-energy structure, A12_3347, while at the **M06** level it is 19.4 kJ mol^{-1} more stable and at the **MP2** level 10.1 kJ mol^{-1}.

A12_helix

The fully α-helical structure of $Ala_{12}Na^+$ has seven helical hydrogen bonds between the C=O of residues (2,3,4,5,6,7,8) and the NH of (6,7,8,9,10,11,12) with lengths of 1.98, 2.02, 1.94, 1.94, 1.95, 1.93 and 1.88 Å, respectively. Additionally, there is a hydrogen bond shared between the C=O of residue 1 and the NH of residues (4,5) with lengths of 2.14 and 2.10 Å, respectively. The sodium cation is coordinated by C=O (9,10,12) at distances of 2.30, 2.27 and 2.27 Å, respectively. Finally the nitrogen of the N-terminus hydrogen bonds with NH(2) at a distance of 2.11 Å, while NH(3) and C=O(11) are free.

$Ala_{12}Na^+$ Other helical structures

At the **M06** level, structures A12_3913 (G17.2 kJ mol^{-1}), A12_3365 (G17.5 kJ mol^{-1}) and A12_3347 (G19.4 kJ mol^{-1}) are all mainly helical structures with only minor variations

of structure at either the C- or N-terminal ends of the peptide. For A12_3913, hydrogen bonds are formed between NH_2-NH(2) (2.10 Å), C=O(1)-NH(4) (2.12 Å), C=O(1)-NH(5) (2.11 Å), C=O(2)-NH(6) (1.98 Å), C=O(3)-NH(7) (2.01 Å), C=O(4)-NH(8) (1.96 Å), C=O(5)-NH(9) (1.91 Å), C=O(6)-NH(10) (1.91 Å), C=O(7)-NH(11) (1.91 Å) and C=O(12)-NH(12) (2.24 Å) while the Na^+ is coordinated by C=O(8) (2.31 Å), C=O(9) (2.31 Å) and C=O(11) (2.23 Å) and NH(3) and C=O(10) are free. For A12_3365, hydrogen bonds exist between NH_2-C=O(1) (2.59 Å), C=O(1)-NH(4) (2.17 Å), C=O(1)-NH(5) (2.11 Å), C=O(2)-NH(6) (1.97 Å), C=O(3)-NH(7) (2.03 Å), C=O(4)-NH(8) (1.96 Å), C=O(5)-NH(9) (1.95 Å), C=O(6)-NH(10) (1.91 Å), C=O(7)-NH(11) (1.93 Å) and C=O(8)-NH(12) (1.87 Å) while the Na^+ is coordinated by C=O(9) (2.30 Å), C=O(10) (2.29 Å) and C=O(12) (2.27 Å) and NH(2,3) and C=O(11) are free. For A12_3347, hydrogen bonds are formed between NH_2-NH(2) (2.11 Å), C=O(1)-NH(4) (2.15 Å), C=O(1)-NH(5) (2.13 Å), C=O(2)-NH(6) (1.97 Å), C=O(3)-NH(7) (1.99 Å), C=O(4)-NH(8) (1.96 Å), C=O(5)-NH(9) (1.93 Å), C=O(6)-NH(10) (1.94 Å) and C=O(7)-NH(11) (1.94 Å) while the Na^+ is coordinated by C=O(8) (2.31 Å), C=O(9) (2.34 Å), C=O(11) (2.39 Å) and C=O(12) (2.34 Å) and NH(3) are free.

A12_3913 differs from the A12_helix by a twist in the peptide backbone after residue 10. This essentially unravels the helix by one residue and shifts the coordination of Na^+ down the peptide chain from C=O(12,10,9) to C=O(11,9,8), leaving a C=O(12)-NH(12) interaction. A12_3365 differs from A12_helix only by movement of the NH_2 from interaction with NH(2) to interaction with C=O(1). A12_3347 differs from A12_helix by an unravelling of the helix by one residue such as for A12_3913, however in this case, the C-terminal residue continues to coordinate the Na^+ which somewhat hinders the C=O(11)-Na^+ interaction. RMSD comparisons in Table 3.18 are consistent with the observation that A12_3913 and A12_3347 have smaller differences since they both differ from A12_helix by a structural change at the C-terminus, while A12_3365 differs by a small change at the N-termius.

Table 3.18 – RMSD between $Ala_{12}Na^+$ mainly helical structures, relative to lowest-energy mainly helical structure (**M06**, Å).

A12	**RMSD**
A12_3913	0.000
A12_3365	1.888
A12_3347	1.277

Structural features are summarized in the following tables.

Table 3.19 and 3.20 Legend:

- COOH - interactions of the C-terminal C-OH
- NH_2 - interactions of the N-terminus
- Na^+ - coordination of Na^+
- C=O-N-H - hydrogen bonds between C=O and N-H groups

Table 3.19 – Summary of interactions of the C- and N-terminal groups of calculated structures (**M06**)

$\mathbf{Ala_8Na^+}$	**COOH**	$\mathbf{NH_2}$
A8_helix	free	NH(2)
A8_4740	C=O(1)	Na^+
A8_4217	C=O(1)	Na^+
A8_7740	C=O(1)	Na^+
$\mathbf{Ala_9Na^+}$	**COOH**	$\mathbf{NH_2}$
A9_helix	free	NH(2)
A9_3809	free	Na^+
A9_6182	free	Na^+
A9_3017	free	Na^+
$\mathbf{Ala_{10}Na^+}$	**COOH**	$\mathbf{NH_2}$
A10_helix	free	NH(2)
A10_2555	free	NH(2)
A10_1433	NH_2	C=O(10)
A10_0366	NH_2	C=O(10)
$\mathbf{Ala_{11}Na^+}$	**COOH**	$\mathbf{NH_2}$
A11_helix	free	NH(2)
A11_3329	free	NH(2)
A11_9390	free	NH(2)
A11_6849	free	NH(2)
$\mathbf{Ala_{12}Na^+}$	**COOH**	$\mathbf{NH_2}$
A12_helix	free	NH(2)
A12_3913	free	NH(2)
A12_3365	free	C=O(1)
A12_3347	free	NH(2)

Table 3.20 – Summary of main hydrogen bonding features in calculated structures (**M06**).

$\mathbf{Ala_8Na^+}$	$\mathbf{Na^+}$	**C=O-NH**
A8_helix	C=O (5,6,8)	NH((4,5),6-8)-C=O(1,2-4)
A8_4740	C=O(2,5,6,8)	NH(3,7,4)-C=O(5,4,7)
A8_4217	C=O(2,5,6,8)	NH(3,7,4)-C=O(5,4,7)
A8_7740	C=O(2,5,6,8)	NH(3,7,4)-C=O(5,4,7)
$\mathbf{Ala_9Na^+}$	$\mathbf{Na^+}$	**C=O-NH**
A9_helix	C=O (6,7,9)	NH((4,5),6-9)-C=O(1,2-5)
A9_3809	C=O (1,4,6,8)	NH(4,(5,6),8,9)-C=O(1,3,5,9)
A9_6182	C=O (1,4,6,8)	NH(4,(5,6),(8,9))-C=O(1,3,5,9)
A9_3017	C=O (1,4,6,9)	NH(4,(5,6),(7,8))-C=O(1,3,5)
$\mathbf{Ala_{10}Na^+}$	$\mathbf{Na^+}$	**C=O-NH**
A10_helix	C=O (7,8,10)	NH((4,5),6-10)-C=O(1,2-6)
A10_2555	C=O (1,2,4,6,8,10)	NH(4,8,10)-C=O(1,5,7)
A10_1433	C=O (2,4,6,8,10)	NH((7,8),10)-C=O(5,7)
A10_0366	C=O (1,2,4,8,10)	NH((7,8),10)-C=O(5,7)
$\mathbf{Ala_{11}Na^+}$	$\mathbf{Na^+}$	**C=O-NH**
A11_helix	C=O (8,9,11)	NH((4,5),6-11)-C=O(1,2-7)
A11_3329	C=O (4,5,7,9,11)	NH((4,5),6,7,9,11)-C=O(1,2,3,6,8)
A11_9390	C=O (4,5,7,9,11)	NH((4.5),6,7,9,11)-C=O(1,2,3,6,8)
A11_6849	C=O (3,4,7,9,11)	NH((4,5),6,7,9,11)-C=O(1,2,3,6,8)
$\mathbf{Ala_{12}Na^+}$	$\mathbf{Na^+}$	**C=O-NH**
A12_helix	C=O (9,10,12)	NH((4,5),6-12)-C=O(1,2-8)
A12_3913	C=O (8,9,11)	NH((4,5),6-11,12)-C=O(1,2-7,12)
A12_3365	C=O (9,10,12)	NH((4,5),6-12)-C=O(1,2-8)
A12_3347	C=O (8,9,11,12)	NH((4,5),6-11)-C=O(1,2-7)

3.2.10 General Comments Relating to Alpha Helical Structures

Calculations predict the lowest-energy conformations for $Ala_{8-12}Na^+$ to be the fully α-helical conformation. These helical structures are essentially the same between sizes, with the only difference being the number of residues and length of the helix. This is also demonstrated in Table 3.21 where the gyration radius increases very consistently as peptide size increases.[vii]

Table 3.21 – Radius of gyration of helical structures and the change associated with increasing peptide size for calculated structures (**M06**, Å).

	rgyr	Δ
A8helix	4.74	
A9helix	5.08	0.33
A10helix	5.44	0.36
A11helix	5.80	0.37
A12helix	6.16	0.36

In the structures A12_3913 and A12_3347, the disruption of the fully helical structure at the C-terminus effectively amounts to the breaking of the first intramolecular interactions that, if continued, would result in unwinding of the helical structure to form partial helical structures such as were seen for $Ala_{11}Na^+$. Comparison of these mostly helical $Ala_{12}Na^+$ structures and partial helical $Ala_{11}Na^+$ structures to the fully helical structures for both sizes, explores the balance between coordination of the Na^+ and formation of the strong helical hydrogen bonds. This is clearly illustrated through comparison of the relative energies of the completely helical structures and the next lowest-energy structures seen in the table below. For $Ala_{11}Na^+$, the second-lowest-energy conformation retains approximately half helical structure and for $Ala_{12}Na^+$ the next lowest-energy conformation remains mainly helical with only a small distortion.

[vii] General helical structure in these poly(alanine) peptides was discussed in more detail in the section of general structural descriptions earlier in the chapter.

Table 3.22 – Free energies (298 K) of α-helical conformations relative to the second-lowest-energy conformations. The second-lowest-energy conformations for $Ala_{8-10}Na^+$ are globular, while for $Ala_{11-12}Na^+$ the second-lowest-energy structures remain partly and mainly helical, respectively (kJ mol^{-1}).

	Ala_8Na^+	Ala_9Na^+	$Ala_{10}Na^+$	$Ala_{11}Na^+$	$Ala_{12}Na^+$
RI-B3LYP-D/def2-SVP	25.3	-13.8	-19.1	-10.6	-12.2
B3LYP-D	9.6	-27.2	-20.1	-8.4	-0.7
M06/6-31G(d,p)	-24.2	-33.4	-47.9	-21.3	-13.5
M06	-32.5	-41.5	-44.2	-21.9	-19.4
RI-MP2/def2-TZVPP//RI-B3LYP-D/def2-SVP	-22.4	-49.0	-20.0	-9.4	-5.4
MP2	-25.4	-37.7	-38.6	-19.0	-10.1
RI-MP2/def2-SVP	-2.6	-	-	-	-
RI-MP2/def2-TZVPP//RI-MP2/def2-SVP	-17.0	-	-	-	-
CC2	-19.0	-34.3	-40.4	-	-
RI-CC2/def2-TZVPP//RI-MP2/def2-SVP	-11.9	-	-	-	-

The **MP2** values for the relative free energy difference between the lowest-energy fully helical structure and the second-lowest-energy calculated structure in Table 3.22 show a couple of interesting trends. Firstly, the relative energy of the helix increases as the size is reduced from $Ala_{10}Na^+$ to Ala_8Na^+. In principle, there is a limit to how many C=O or other electron donating groups can coordinate the Na^+, such as was seen in the globular structures of Ala_9Na^+ and $Ala_{10}Na^+$ when some part of the peptide was excluded from the main structure (usually the C- or N-terminal residue). Referring to Table 3.20, one can see that coordination around the Na^+ reaches a maximum at 5-6. This however causes a significant reduction in the ordering of the rest of the structure and decreases the number of strong hydrogen bonds in the structure. For smaller-sized peptides this reduction in strong hydrogen bonds is less significant, while for larger cases where more hydrogen bonds can be formed in the helical structure (number of hydrogen bonds in helix: Ala_8Na^+ - 5, $Ala_{12}Na^+$ - 9) the stabilization involved with forming the helical structure is greater.[viii] For the helical

[viii] This trend is consistent between Ala_8Na^+ and $Ala_{10}Na^+$, however no fully globular structures were calculated for $Ala_{11}Na^+$ and $Ala_{12}Na^+$ and so the comparison cannot be made. It would be expected however, that the trend would continue and the relative energy difference between the helix and lowest-energy globular structure for $Ala_{11}Na^+$ and $Ala_{12}Na^+$ would continue to increase.

structures, the coordination of the sodium cation by three carbonyl groups is the same regardless of the size of the peptide. The stabilization of the helical structure in Table 3.22 for $Ala_{11}Na^+$ is a comparison of a fully helical $Ala_{11}Na^+$ structure to an approximately half helical $Ala_{11}Na^+$ structure. For $Ala_{11}Na^+$, the value of -19.0 kJ mol^{-1} is approximately half of the -38.6 kJ mol^{-1} for $Ala_{10}Na^+$, where the comparison is between fully helical and fully globular peptides (a similar difference in relative stabilization is observed at the **M06** and **B3LYP-D** levels).

Using $Ala_{12}Na^+$ as an example, it is clear that each modification or distortion of this structure raises the relative energy in a systematic way, and this trend would be expected to remain the same between sizes, making further exploration of other helical structures unnecessary. REMD generated conformations, such as those obtained for $Ala_{11}Na^+$ with partial helical character, were also produced for $Ala_{12}Na^+$, but not for any of the smaller peptides. Preliminary calculations at the **B3LYP-D** level of similar partial helical conformations for $Ala_{12}Na^+$ indicated that they were significantly higher in energy than the more helical conformations and continued calculations at the **M06** level (or higher) were not completed. For peptides smaller than $Ala_{11}Na^+$ REMD simulations did not produce any partially or fully helical structures. For $Ala_{12}Na^+$, the lowest-energy structures were found to be those that were closest to the complete α-helix without distortions or structural deviations from the ideal helix. Any modifications to the fully α-helical conformation led to an increase in the relative energy. This example, as well as the inability of partially helical conformations of $Ala_{11}Na^+$ to produce lower energy structures relative to the α-helix, led to the assumption that the completely α-helical conformation would be the lowest structure of a helical nature for $Ala_{8-10}Na^+$ as well.

3.3 Experimental Results

All IRMPD results were obtained at the Centre Laser Infrarouge d'Orsay (CLIO) at the Laboratoire de Chimie Physique (LCP) of the Université Paris-Sud. The CLIO free electron laser (FEL) was used to generate a photon beam in the 1000-2000 cm^{-1} fingerprint region and a benchtop OPO/OPA was used for the 3000-3600 cm^{-1} energy range. In both cases a 10 W continuous wave CO_2 (broadband, non-tunable) laser was added on top of the scanning beam (FEL or OPO) to enhance fragmentation.[57,60] This experiment was described extensively in the introduction. Both laser setups were coupled to a 7T Bruker APEX IV FT-ICR mass spectrometer where ions were generated by electrospray ionization.[26,28,29,36,42,55,56,61] Irradiation times of 0.3-1.5 s and 5.0-20.0 s were used for fingerprint (FEL/CO_2) and 3000-3600 cm^{-1} (OPO/CO_2) experiments respectively. In all experiments, solutions were prepared by dissolving a few milligrams of peptide sample (unpurified, desalted, GeneCust-Luxembourg) in a small amount of trifluoroacetic acid followed by an approximate order of magnitude dilution with trifluoroethanol. This stock solution was then further diluted with trifluoroethanol to obtain micromolar (μM) solutions. Using this method of preparation, the sodiated peptide was normally the most intense ion in the spectrum without any further or deliberate addition of sodium to the solutions. Previous studies have demonstrated the use of different solvent matrices.[134] Solutions of [90]:10 [trifluoroethanol or acetonitrile]:formic acid, 90:10 acetic acid:trifluoroacetic acid as well as pure trifluoroacetic acid and pure formic acid all yielded the same results. In the experiments presented here, as well as in the literature, it was observed that the acidic solvent matrix slowly hydrolyzes the peptides, rendering solutions useable for approximately 1-3 days.[134] Electrosprayed ions were mass selected in the quadrupole and subsequently trapped in the ICR cell where IRMPD was observed.

In general, absolute intensities deriving from an IRMPD experiment must be treated with caution. Laser power and alignment (in both the FEL and OPO experiments) can

vary widely between different days and directly affect the extent of fragmentation observed for a given set of experimental conditions. The effect of such fluctuations are especially noticeable in weakly absorbing bands, such as the 1000-1400 cm^{-1} region of sodiated poly(alanine) peptides. Using the currently discussed experimental apparatus and a typical irradiation time of ≈0.25-0.5 s, only under conditions of relatively high laser power and with a good alignment would well-defined peaks be observed.[ix] Even greater caution needs to be used when comparing experimental IRMPD efficiencies to calculated absorption intensities as these two processes are not directly related. Fragmentation by IRMPD requires efficient IVR (coupling between modes) and a suitable fragmentation mechanism, all in addition to sufficient photon absorption. This, as well as possible errors associated with theoretical models, can often cause inconsistencies between calculated and experimental intensities.

3.3.1 Fragmentation

Fragmentation was observed to occur mainly by the common peptide fragmentation mechanism with the *a*-Na^+, *b*-Na^+, and y fragments as the primarily derived ions. In both cases, the Na^+ is retained on the fragment ion. In the *y*-fragmentation mechanism the charge remains on the C-terminal containing fragment and the peptide is reduced by unit residues (for example, $Ala_{10}Na^+$ (751 m/z) produces *y*-fragment ion Ala_9Na^+ (680 m/z) (and smaller peptides) by sequentially removing amino acids from the N-terminus). Fragment ions of the *a*- and *b*-type leave the charge on the N-terminal containing fragment, in this case the result of rearrangement of the peptide and movement of the Na^+ from the C-terminus to a different location in the peptide. Fragments of the *a*-type indicate cleavage of the carbon-carbon bond in the peptide backbone, while fragments of the *b*-type indicate cleavage of the carbon-nitrogen in the peptide backbone (for singly charged ions,

[ix] With the CLIO FEL, references to high and low power pertain to the range of approximately 1.3 - 0.6 watts.

b-fragments will be observed at spacings of 28 m/z above *a*-fragments because they include an extra C=O group). Additionally, singly charged fragment ions resulting from the loss of water (-18 m/z) and the loss of ammonia (-17 m/z) (or both together (-35 m/z)) were also observed in most cases.[x] Variation in the observed fragment ions between different peptide sizes and between experiments (OPO or FEL) are likely not due to real differences in fragmentation mechanisms, but more likely due to variations in the extent of fragmentation as a result of experimental variations (for example, variations in the power or alignment of the beam). Consistent with established notation, superscripts will be added to the *a*, *b* and textity notation indicating the size of the peptide chain in the fragment ion. A list of fragments and example mass spectra (from FEL experiments since the extent of fragmentation is typically higher than in OPO experiments) during IRMPD scans are shown below. It is clear that *y*-fragmentation is the favoured pathway over *a*- or *b*-fragmentation by relative intensities of the fragment ions. Losses of -35 and -18 m/z are observed from fragment ions as well as the parent molecular ion, [M-Na^+], when the fragmentation efficiency is relatively high. In the following tables, <u>underlined</u> ions indicate the most intense fragment ion in the spectrum.

[x]This pattern of observed fragmentation is consistent with what is typically seen in low-energy collision induced dissociation (CID) experiments (*a*, *b*, *y* fragments and loss of 18 and 35).

Table 3.23 – List of IRMPD fragment ions for Ala_8Na^+ (parent ion - 609 m/z).

FEL				OPO	
m/z	type			m/z	type
591	**-18**	492	$\mathbf{a}^7$	574	**-35**
574	**-35**	467	$\mathbf{y}^6$	467	$\mathbf{y}^6$
538	$\mathbf{y}^7$	396	$\mathbf{y}^5$		
521	$\mathbf{b}^7$				

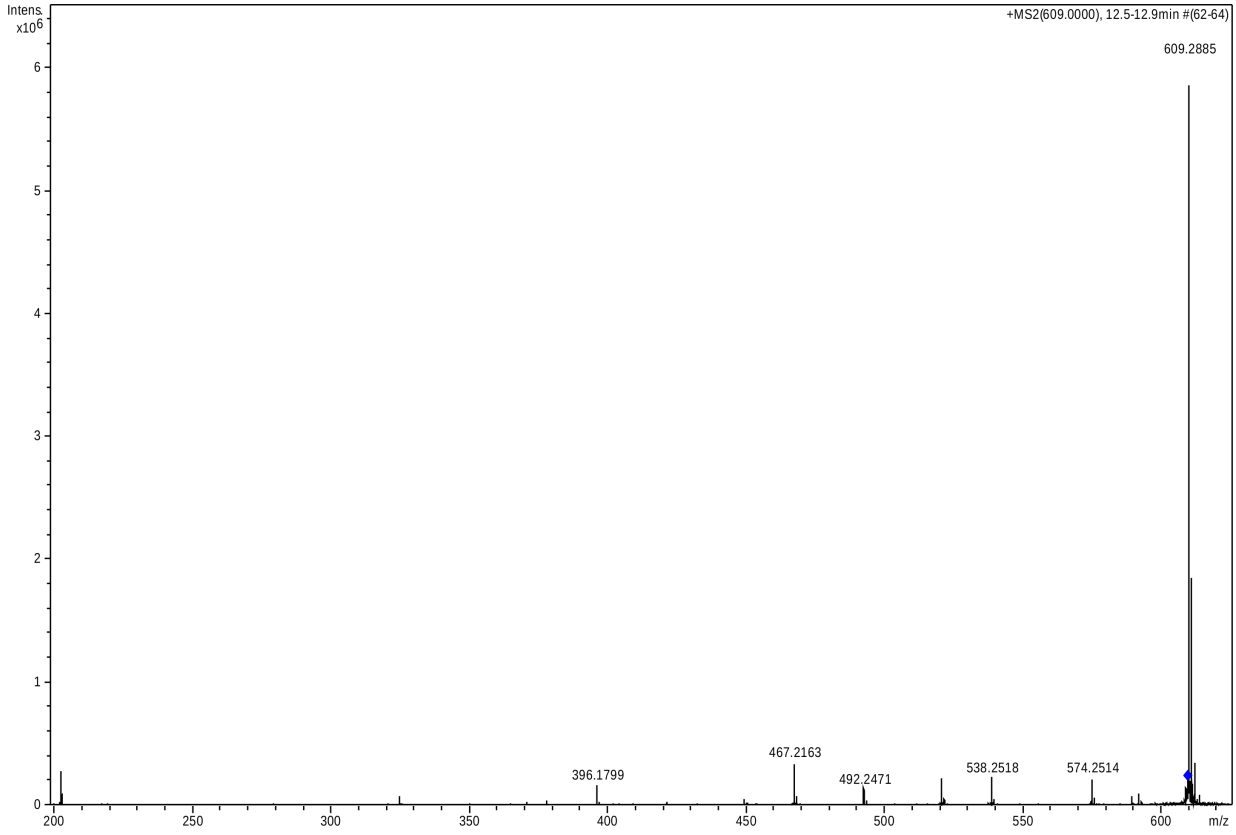

Figure 3.26 – Example IRMPD mass spectrum for Ala_8Na^+ (IRMPD observed upon FEL irradiation).

Table 3.24 – List of IRMPD fragment ions for Ala_9Na^+ (parent ion - 680 m/z).

FEL				OPO	
m/z	type			m/z	type
662	**-18**	467	$\mathbf{y}^6$	645	**-35**
645	**-35**	449	$\mathbf{b}^6$	609	$\mathbf{y}^8$
609	$\mathbf{y}^8$	421	$\mathbf{a}^6$	591	$\mathbf{b}^8$
591	$\mathbf{b}^8$	396	$\mathbf{y}^5$	563	$\mathbf{a}^8$
563	$\mathbf{a}^8$	378	$\mathbf{b}^5$	538	$\mathbf{y}^7$
538	$\mathbf{y}^7$			467	$\mathbf{y}^6$
520	$\mathbf{b}^7$			396	$\mathbf{y}^5$

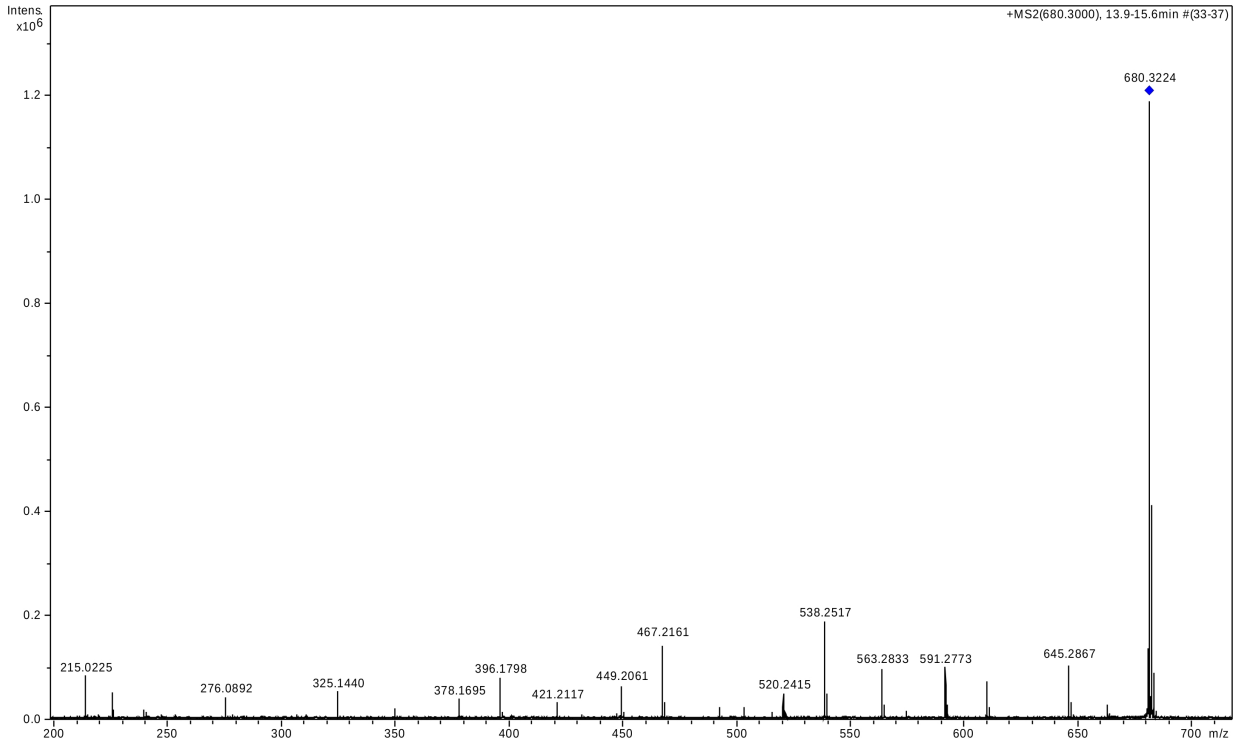

Figure 3.27 – Example IRMPD mass spectrum for Ala_9Na^+ (IRMPD observed upon FEL irradiation).

Table 3.25 – List of IRMPD fragment ions for $Ala_{10}Na^+$ (parent ion - 751 m/z).

FEL				OPO	
m/z	type			m/z	type
733	**-18**	521	$\mathbf{b}^7$	733	**-18**
716	**-35**	492	$\mathbf{a}^7$	716	**-35**
680	$\mathbf{y}^9$	467	$\mathbf{y}^6$	680	$\mathbf{y}^9$
662	**-18**	449	$\mathbf{b}^6$	662	**-18**
634	$\mathbf{a}^9$	421	$\mathbf{a}^6$	609	$\mathbf{y}^8$
609	$\mathbf{y}^8$	396	$\mathbf{y}^5$	538	$\mathbf{y}^7$
591	$\mathbf{b}^8$	378	$\mathbf{b}^5$	467	$\mathbf{y}^6$
563	$\mathbf{a}^8$	325	$\mathbf{y}^4$	396	$\mathbf{y}^5$
538	$\mathbf{y}^7$				

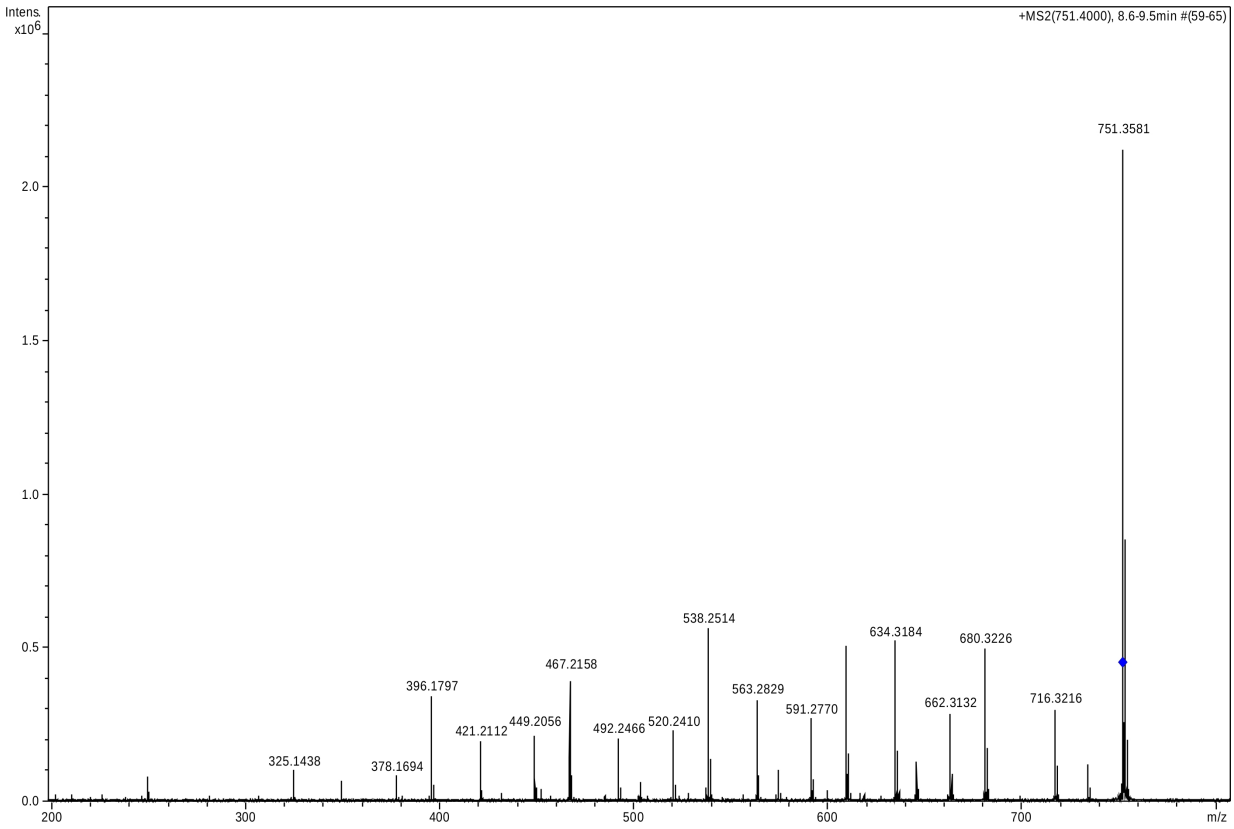

Figure 3.28 – Example IRMPD mass spectrum for $Ala_{10}Na^+$ (IRMPD observed upon FEL irradiation).

Table 3.26 – List of IRMPD fragment ions for $Ala_{11}Na^+$ (parent ion - 822 m/z).

FEL				OPO	
m/z	type			m/z	type
804	**-18**	591	$\mathbf{b}^8$	804	**-18**
787	**-35**	563	$\mathbf{a}^8$	787	**-35**
751	$\mathbf{y}^{10}$	538	$\mathbf{y}^7$	733	$\mathbf{b}^{10}$
733	$\mathbf{b}^{10}$	521	$\mathbf{b}^7$	705	$\mathbf{a}^{10}$
716	**-35**	492	$\mathbf{a}^7$	680	$\mathbf{y}^9$
705	$\mathbf{a}^{10}$	467	$\mathbf{y}^6$	609	$\mathbf{y}^8$
680	$\mathbf{y}^9$	449	$\mathbf{b}^6$	538	$\mathbf{y}^7$
662	$\mathbf{b}^9$	421	$\mathbf{a}^6$	467	$\mathbf{y}^6$
634	$\mathbf{a}^9$	396	$\mathbf{y}^5$	396	$\mathbf{y}^5$
609	$\mathbf{y}^8$				

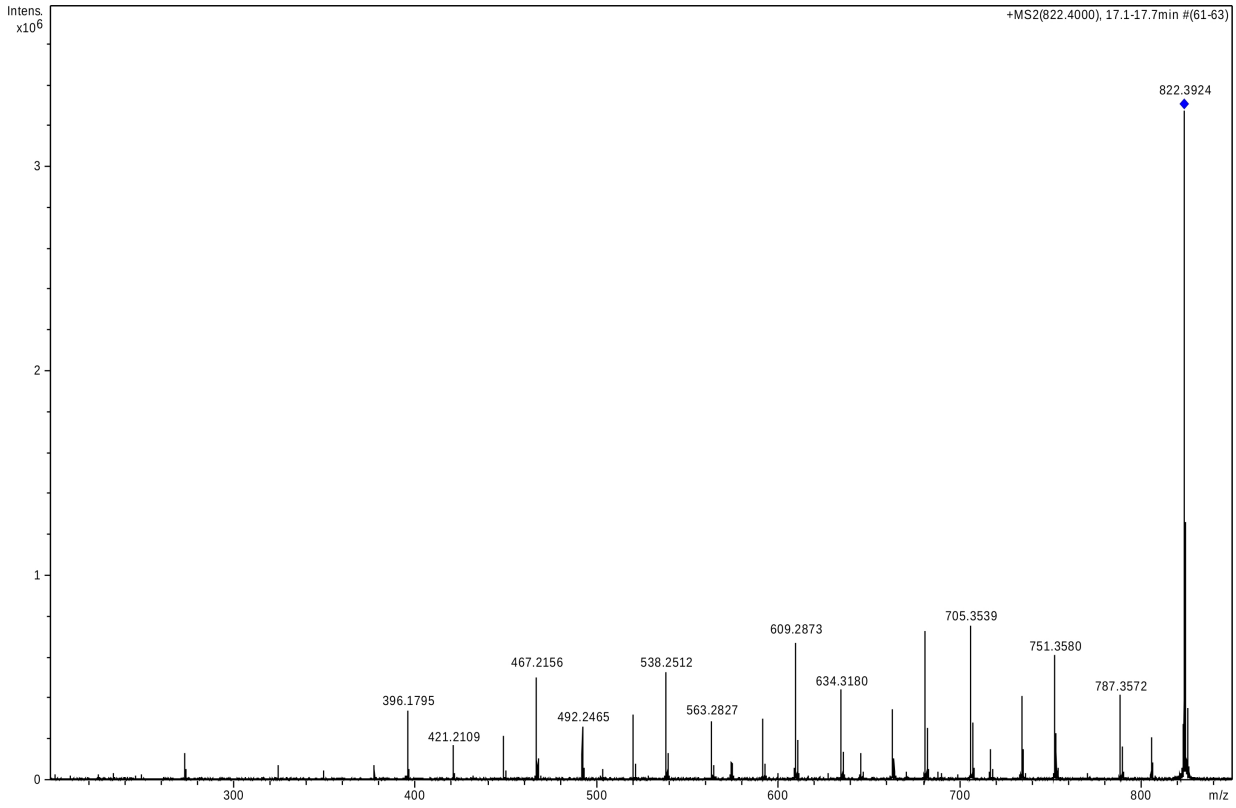

Figure 3.29 – Example IRMPD mass spectrum for $Ala_{11}Na^+$ (IRMPD observed upon FEL irradiation).

Table 3.27 – List of IRMPD fragment ions for $Ala_{12}Na^+$ (parent ion - 893 m/z).

FEL				OPO	
m/z	type			m/z	type
875	**-18**	634	$\mathbf{a}^9$	858	**-35**
858	**-35**	609	$\mathbf{y}^8$	804	$\mathbf{b}^{11}$
822	$\mathbf{y}^{11}$	591	$\mathbf{b}^8$	776	$\mathbf{a}^{11}$
804	$\mathbf{b}^{11}$	563	$\mathbf{a}^8$	751	$\mathbf{y}^{10}$
787	**-35**	538	$\mathbf{y}^7$	680	$\mathbf{y}^9$
751	$\mathbf{y}^{10}$	520	$\mathbf{b}^7$	609	$\mathbf{y}^8$
733	$\mathbf{b}^{10}$	492	$\mathbf{a}^7$	467	$\mathbf{y}^6$
716	**-35**	467	$\mathbf{y}^6$		
705	$\mathbf{a}^{10}$	449	$\mathbf{b}^6$		
680	$\mathbf{y}^9$	421	$\mathbf{a}^6$		
662	$\mathbf{b}^9$	396	$\mathbf{y}^5$		

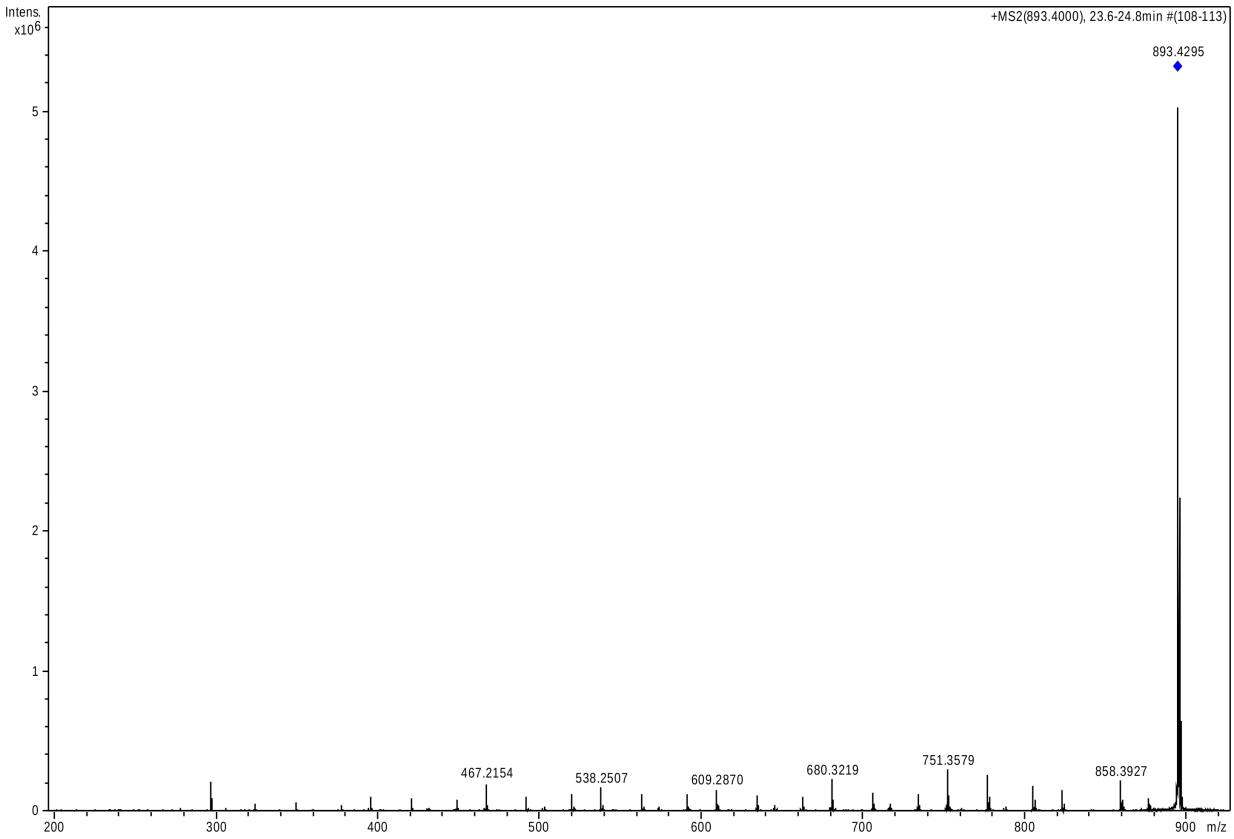

Figure 3.30 – Example IRMPD mass spectrum for $Ala_{12}Na^+$ (IRMPD observed upon FEL irradiation).

3.3.2 Fingerprint Experimental Spectra

In fingerprint spectral region experiments, trapped ions were irradiated for 0.3-1.5 seconds with the FEL beam. 42 MeV electrons were used to generate radiation in the 900-1900 cm^{-1} region, giving the peak laser intensity in the 1000-1400 cm^{-1} region where many weakly absorbing modes are found in poly(alanine)s.

Each experimental spectrum in the following section is composed of multiple scans using slightly different experimental conditions (mainly varied irradiation times). This yielded the optimal fragmentation efficiencies and confirmation of peak positions. The fingerprint region is mainly informative to the general functionality of the molecule. The $\approx$1700 cm^{-1} (amide I) region contains bands attributed to the various carbonyl vibrations (80%, with the remaining 20% character associated with C-N stretching and C-C-N deformation[135]), with the modes to the blue of 1700 cm^{-1} resulting from C-terminal C=O and the unbound C=O group one residue preceding the C-terminal residue. The 1500 cm^{-1} (amide II) region contains bands from the various amide N-H vibrations (60%, with the remaining 40% character mainly attributed to C-N stretching[135]) and finally the region below 1400 cm^{-1} (the region centred around 1300 cm^{-1} is referred to as the amide III) contains the various C-N stretches and N-H bending modes as well as C-H and CH_3 modes.

It would be expected that a peptide with extensive helical structure would have narrow amide I and II bands because of the similarities between bonding environments of each residue, aside from the C- and N-terminal residues. All calculated helical conformations were shown to have a free C-OH at the C-terminus and one free C=O group corresponding to the C=O of the residue immediately preceding the C-terminal residue. The amide II band is expected to be relatively narrow and slightly to the red of the position it would be in a non-helical peptide because of the lack of free N-H groups in the helical structure. The region below 1400 cm^{-1} is more difficult to contrast between structures and sizes in general terms due to the weakly absorbing but abundant spectral activity in this region.

Calculations predict that α-helical structures for $Ala_{8-12}Na^+$ have C=O modes to the blue of the main amide I group of C=O bands. These modes were found to be very difficult to observe experimentally owing to their relatively weak absorption intensities and their spectral location immediately beside the strongly absorbing main amide I band. This results from the practical problem that an appropriate laser power for the main band will be of insufficient intensity for observation of these low intensity bands to the blue of the main band. A separate scan with a significantly increased laser power is necessary to observe these bands. For this reason, the C-terminal and free C=O bands to the blue of the main group are not seen in the spectra in the following sections, with the exception of Ala_8Na^+ where additional experiments were made. Such experiments were not completed for other peptides because of the very limited availability of the experiment.

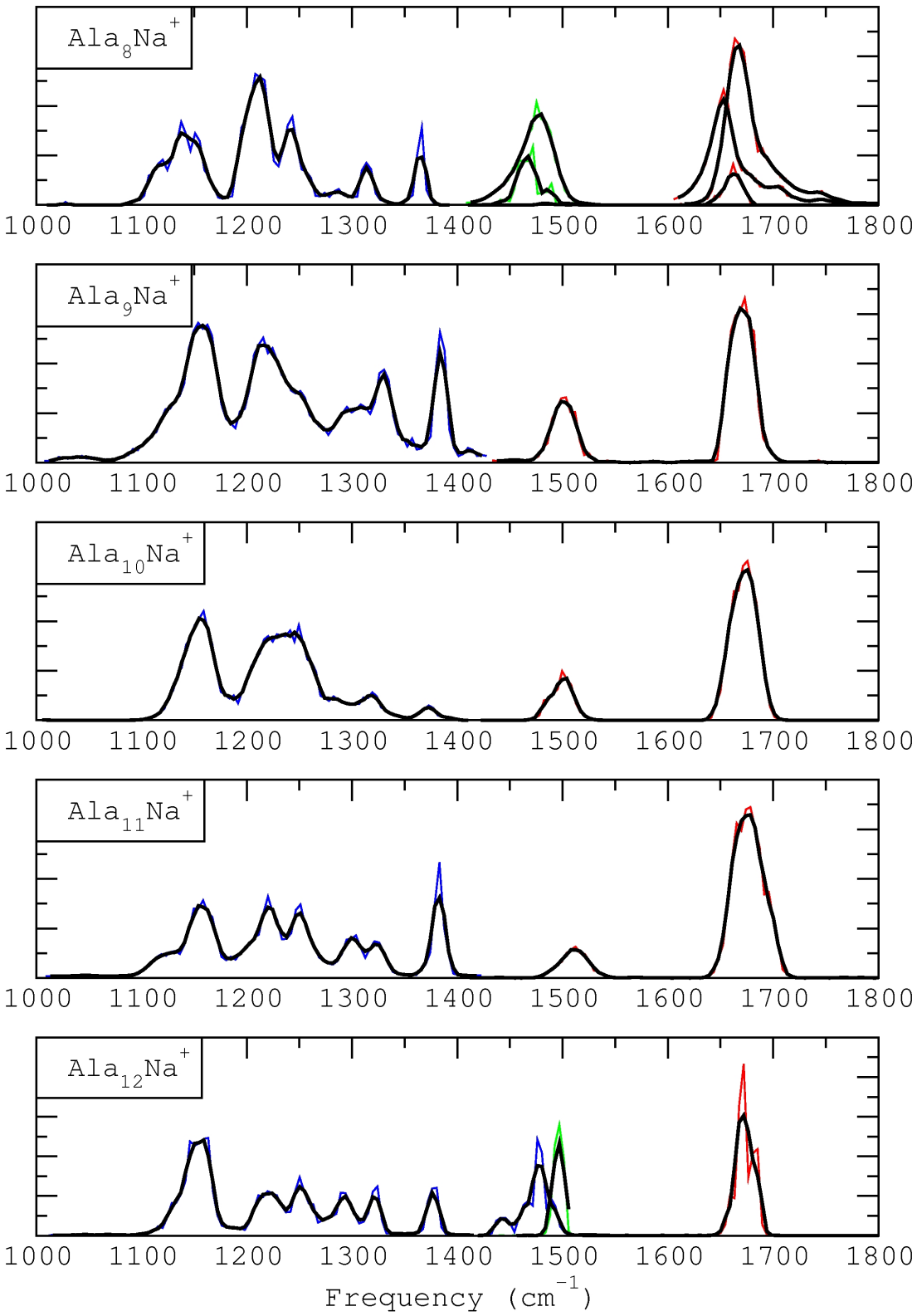

Figure 3.31 – Fingerprint IRMPD spectra for $Ala_{8-12}Na^+$. **Black** lines are a three-point moving average of the experimental data points. Fragmentation efficiencies (y-axes) are scaled arbitrarily.

Experimental spectra in the fingerprint region are shown in Figure 3.37 and a summary of spectral features is given in Table 3.28. Amide I bands in all spectra are very nearly identical. Ala_8Na^+ appears to be somewhat different, but as described previously, special attention was paid to this experiment in an effort to observe the C-terminal C=O and free C=O (7) at approximately 1730 and 1720 cm^{-1}. As a result, the two amide I scans with the highest intensity appear to be relatively more broad. The lowest intensity Ala_8Na^+ amide I scan was obtained using similar experimental conditions as were used for the other peptides and is of very similar shape and position (lower intensity is a result of scaling for presentation). Some variations between amide II bands are seen when comparing Ala_8Na^+ and $Ala_{12}Na^+$ with $Ala_{9-11}Na^+$. In these cases, additional scans were made for verification and will be discussed when comparisons to computational results are made. In the spectrum of $Ala_{10}Na^+$ a noticeable difference from the other spectra exists in the intensities of the two bands at approximately 1325 and 1375 cm^{-1}. This is most likely the result of 30-40% reduction of laser power in the region around 1300-1400 cm^{-1} relative to the 1000-1300 cm^{-1} region on that particular day.

Table 3.28 – Summary of fingerprint region spectra for $Ala_{8-12}Na^+$ (cm^{-1}).

Amide II	Maximum	Minimum	Peak Max	Width
Ala_8Na^+	1505	1420	1480	85
Ala_9Na^+	1530	1475	1500	55
$Ala_{10}Na^+$	1525	1470	1505	55
$Ala_{11}Na^+$	1540	1475	1510	65
$Ala_{12}Na^+$	1505	1430	1495	95
Amide I	Maximum	Minimum	Peak Max	Width
Ala_8Na^+	1760	1630	1665	120
Ala_9Na^+	1705	1645	1670	60
$Ala_{10}Na^+$	1705	1640	1675	65
$Ala_{11}Na^+$	1715	1640	1675	75
$Ala_{12}Na^+$	1695	1645	1675	50

3.3.3 3000-3600 cm^{-1} Region Experimental Spectra

Features in the 3000-3600 cm^{-1} region are associated with N-H and O-H bond vibrations, while C-H bond vibrations are found slightly below this region, but were not observed in the experiments discussed in this thesis. Considering poly(alanine) based peptides, most peaks in this region will result from amide N-H modes, since the only other activity in this region results from the C- and N-terminal C-OH and NH_2 groups. Since gas-phase peptides normally do not adopt zwitterionic structures, both the amine NH_2 group and the C-OH can in principle be observed. However, this is often not the case as hydrogen bonding of the C-OH group shifts the peak position significantly to the red and broadens it to the point that it is unobservable.[136,137] As well, the amine vibrations are also not always observed as they are typically weak and can be dominated by the weakly bound and free amidic NH bands usually found in a similar region. Typically, bound N-H bands are found between 3225-3425 cm^{-1}, free N-H bands between 3450-3500 cm^{-1} and NH_2 bands usually between 3430-3475 cm^{-1} for asymmetric stretching and between 3345-3375 cm^{-1} for the symmetric vibration. Probably the most distinct feature in spectra of poly(alanine) peptides in this region is the free C-OH band, which if present, is a sharp and isolated mode very consistently found at 3575 cm^{-1}. Calculated results for all peptides, $Ala_{8-12}Na^+$, indicate that the α-helical structures invariably have free C-OH and therefore this feature should be present at 3575 cm^{-1} in all spectra in this region.

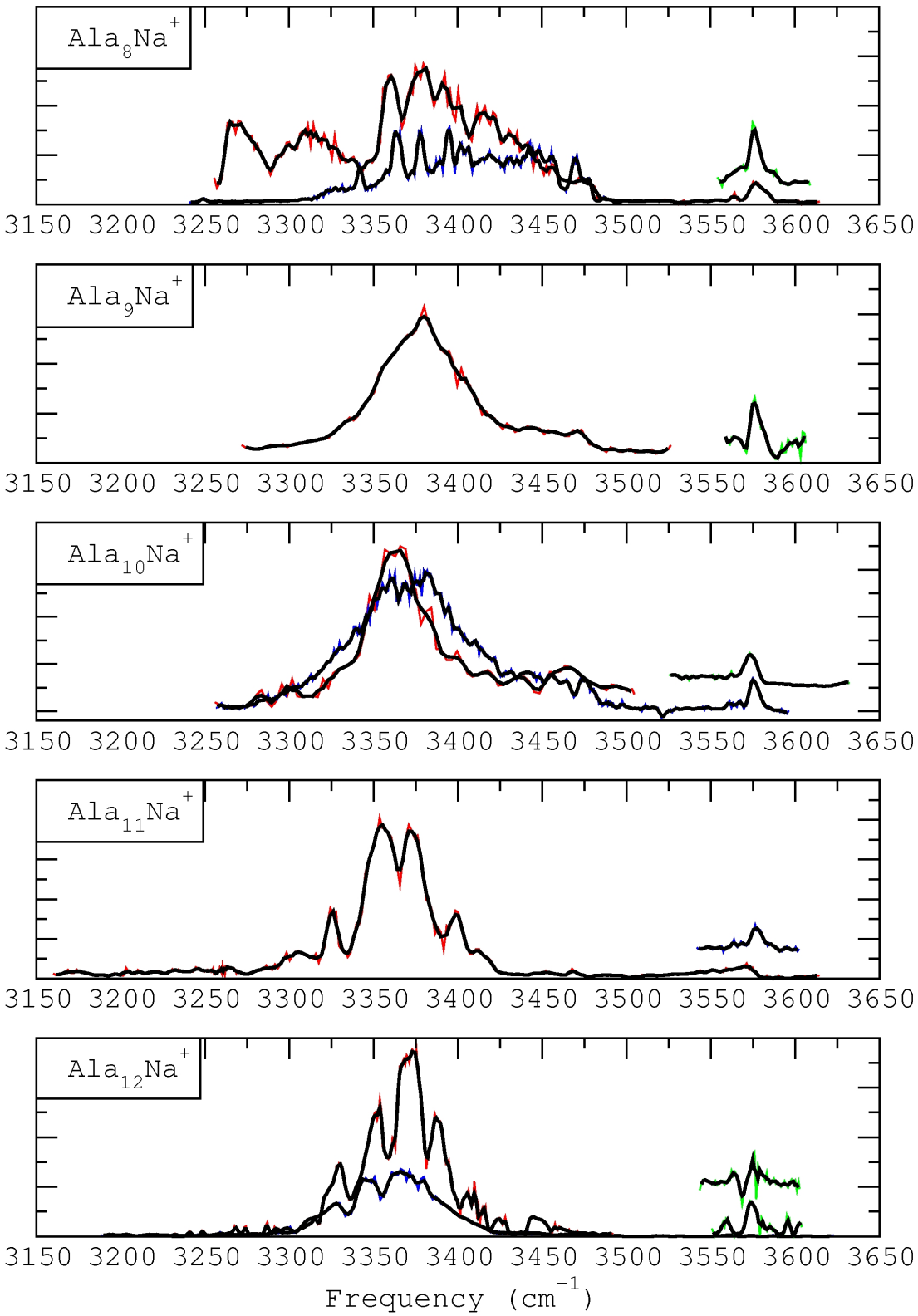

Figure 3.32 – 3000-3600 cm^{-1} region IRMPD spectra for $Ala_{8-12}Na^+$. **Black** lines are a three-point moving average of the experimental data points. Fragmentation efficiencies (y-axes) are scaled arbitrarily.

Experimental spectra in the 3000-3600 cm^{-1} region are shown in Figure 3.32 and a summary of spectral features is given in Table 3.29. The free C-OH band is observed in all spectra, where in all cases the position is twice verified to ± 2 cm^{-1} of 3575 cm^{-1}. Also common to all spectra is the large N-H stretching band mainly contained between 3350-3400 cm^{-1}. The spectrum for Ala_8Na^+ has two bands to the red of this main band which is likely contributed from a second isomer not present in the other experiments and will be discussed during comparison to calculated results. The spectrum of Ala_9Na^+ was obtained under more difficult experimental conditions which resulted in less pronounced spectral features in the N-H band. The red spectrum for $Ala_{10}Na^+$ has a lower resolution comparing to the other spectra in Figure 3.32, resulting from use of a 10.0 s irradiation time under otherwise similar experimental conditions (this reduced the number of data points collected in the experimental spectrum by half).

Table 3.29 – Summary of 3000-3600 cm^{-1} region bound N-H band for $Ala_{8-12}Na^+$ (cm^{-1}).

	Maximum	Minimum	Peak Max	Width
Ala_8Na^+	3480	3255	3375	225
Ala_9Na^+	3470	3290	3365	180
$Ala_{10}Na^+$	3485	3280	3365	195
$Ala_{11}Na^+$	3475	3290	3350	185
$Ala_{12}Na^+$	3470	3290	3375	180

3.3.4 Comparison of Frequencies Calculated at Different Levels of Theory

Ala_8Na^+ was used to benchmark the calculation of vibrational frequencies. Extensive calculations show that the B3LYP-D/def2-SVP level cannot accurately reproduce vibrational frequencies with respect to RI-MP2/def2-SVP calculations and experiment, as shown in Figure 3.33. Further comparison of Ala_8Na^+ frequencies showed that at the M06/6-31G(d,p) level, scaled vibrational frequencies were consistent with those at the RI-MP2/def2-

SVP level while still at a much lower computation cost (≈1/4 that of RI-MP2/def2-SVP).

Figure 3.33 compares the fingerprint and OPO experimental spectra for Ala_8Na^+ to calculations at the B3LYP-D/def2-SVP, RI-MP2/def2-SVP and M06/6-31G(d,p) levels. In general, M06/6-31G(d,p), RI-MP2/def2-SVP and experimental results are quite consistent, with the exception of a blue shift in the amide II band (approximately 50 cm^{-1}) in the RI-MP2/def2-SVP calculated spectrum. B3LYP-D/def2-SVP results are neither consistent in the fingerprint region nor in the OPO region, with a significant blue shift of the amide II band, as well as a significant red shift of the bound N-H band in the OPO spectrum (approximately 50 cm^{-1} in both cases). Additionally the region of zero intensity in the B3LYP-D/def2-SVP calculated spectrum between 3350 and 3400 cm^{-1} is not consistent with the experimental result.

These observations are in direct agreement with previous results from the literature. Semrouni et al.[87] showed that the B3LYP-D/def2-TZVPP//B3LYP-D/def2-SVP level accurately predicted energetic results (relative to the MP2/def2-TZVPP//MP2/def2-SVP level), however spectra calculated at this level had significant errors. This error is attributed to the calculation of overly compact structures where strong hydrogen bonds exist, exactly as was noted in the previous section for $Ala_{10}Na^+$. Additionally, Schwabe et al.[138] attribute the failure of the B3LYP-D functional in their calculations of a sample polypeptide system to medium-range correlation effects, although note that it performs significantly better than the uncorrected B3LYP functional.

3.3.5 Comparison Between Experimental and Calculated Spectra

Spectral comparisons and tables of band assignments are found in the following section with discussion and interpretation organized at the end of the section.

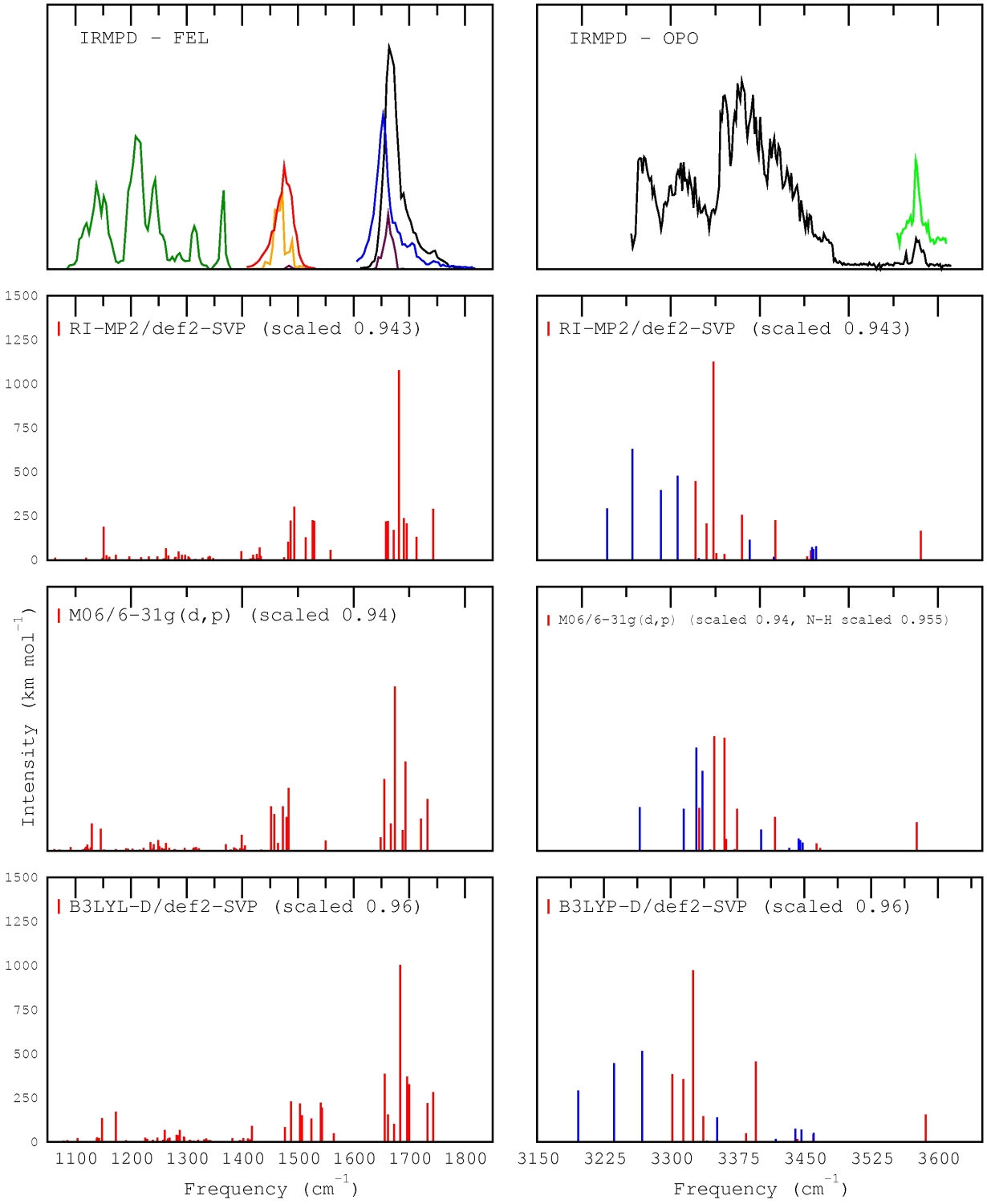

Figure 3.33 – Comparison of calculated Ala_8Na^+ spectra at different levels. For calculated 3000-3600 cm^{-1} region spectra, the modes in **blue** are for the second-lowest-energy isomer (A8_4740) and bands in **red** are for the lowest-energy (α-helical) structure (A8_helix). Evidence will be presented in the following section to support contribution from both isomers.

Fingerprint Region Spectral Results

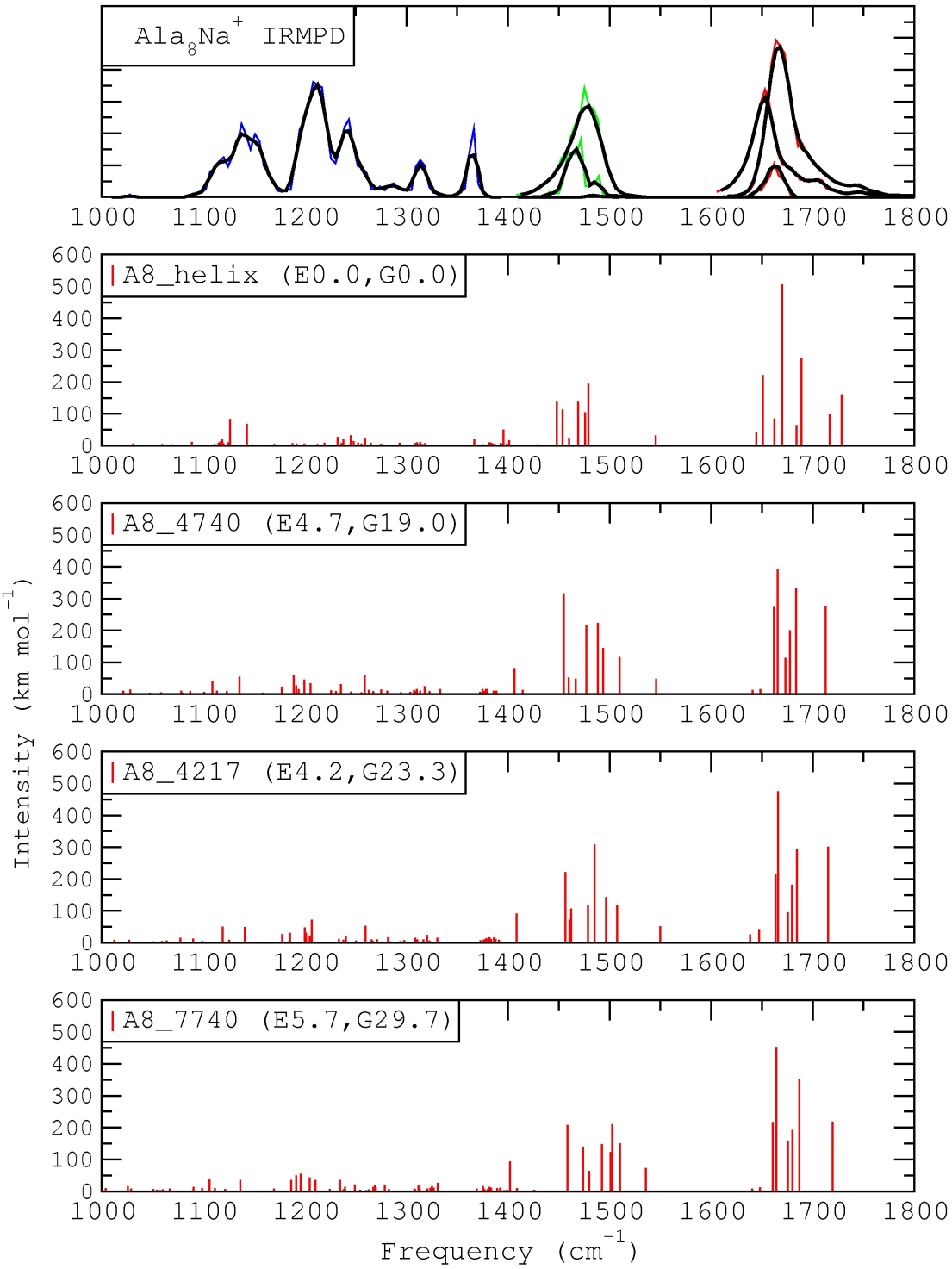

Figure 3.34 – Ala_8Na^+ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **CC2**, frequencies are scaled by 0.94)

Table 3.30 – Ala_8Na^+ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A8_helix		**A8_4740**	
1000	CH+CH_3	1000	CH+CH_3
1129	↓	1136	↓
1145	C-OH	1158	CH+NH+C-OH
1146	CH+NH+C-OH	1294	↓
1296	↓	1295	CH_3
1312	CH_3	1415	↓
1404	↓	1455	NH (2)
1434	NH (4)	1460	NH (3)
1452	NH (2)	1466	NH (6)
1457	NH (3)	1477	NH (5)
1464	NH (5)	1488	NH (7)
1473	NH (8)	1494	NH (8)
1480	NH (7)	1510	NH (4)
1483	NH (6)	1546	NH_2 Na^+
1549	NH_2 N-H (2)	1641	C=O (6)
1649	C=O (6,5)	1648	C=O (4)
1655	C=O (1)	1662	C=O (2)
1667	C=O (6)	1666	C=O (3)
1674	C=O (5,6,2)	1673	C=O (5)
1688	C=O (4,2)	1678	C=O (1)
1693	C=O (3,2)	1684	C=O (7)
1721	C=O (7)	1713	C=O (C-term)
1733	C=O (C-term)		

Table 3.31 – Ala_8Na^+ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A8_4217		**A8_7740**	
1000	CH+CH_3	1000	CH+CH_3
1141	↓	1136	↓
1178	CH+NH+C-OH	1169	CH+NH+C-OH
1297	↓	1295	↓
1298	CH_3	1305	CH_3
1409	↓	1426	↓
1456	NH (3)	1459	NH (3)
1461	NH (6)	1474	NH (2)
1462	NH (2)	1480	NH (7))
1479	NH (7)	1493	NH (8)
1485	NH (5)	1501	NH (6,5)
1497	NH (8)	1503	NH (7,6)
1508	NH (4)	1510	NH (4)
1550	NH_2 Na^+	1536	NH_2 Na^+
1638	C=O (6,4)	1641	C=O (6,4)
1647	C=O (6,4)	1648	C=O (6,4)
1664	C=O (2)	1661	C=O (2)
1666	C=O (3)	1664	C=O (7)
1676	C=O (5)	1676	C=O (3)
1679	C=O (1)	1680	C=O (1)
1685	C=O (7,2)	1687	C=O (5)
1715	C=O (C-term)	1720	C=O (C-term)

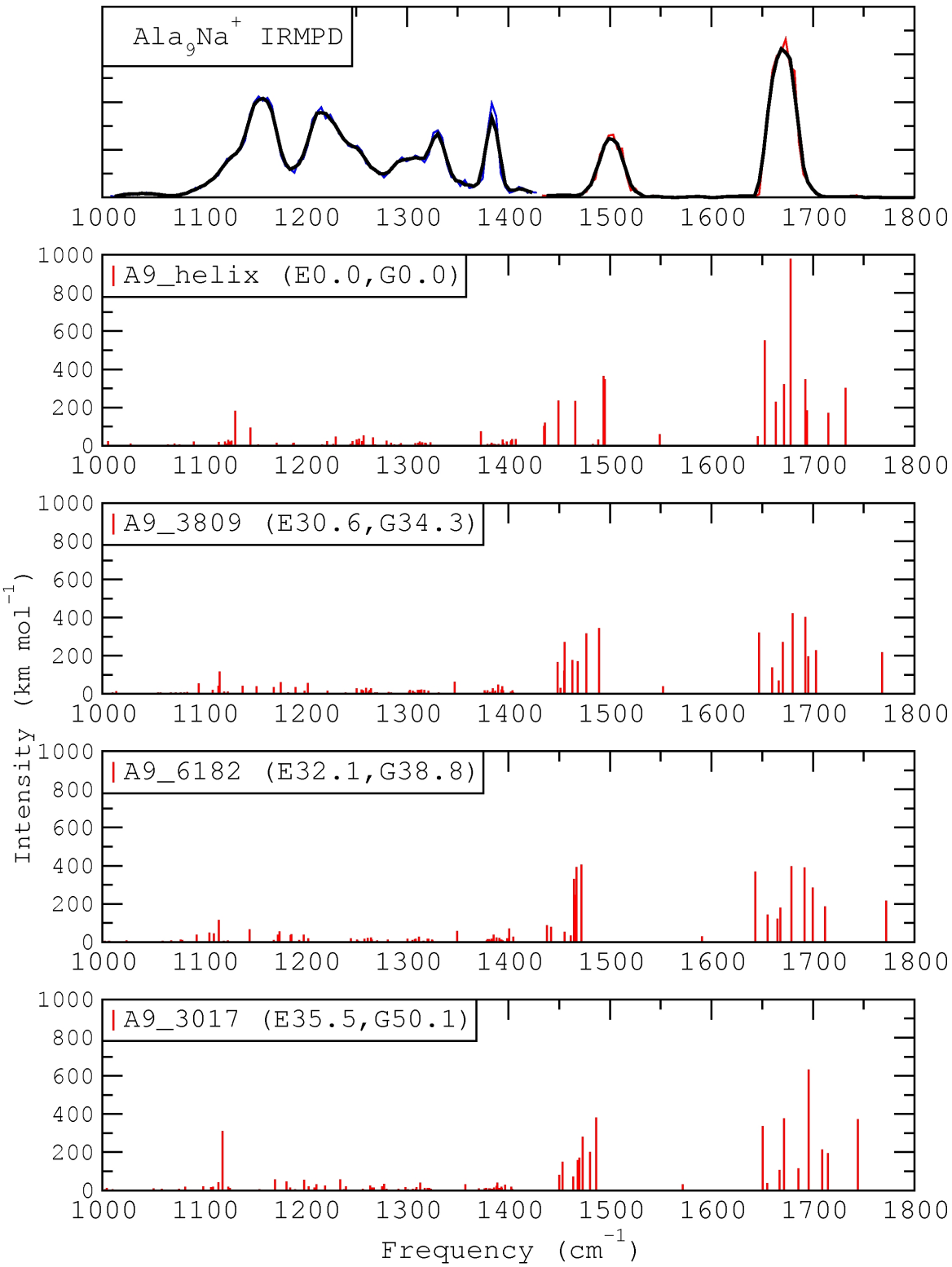

Figure 3.35 – Ala_9Na^+ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **CC2**, frequencies are scaled by 0.94)

Table 3.32 – Ala_9Na^+ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A9_helix		**A9_3809**	
1000	CH+CH_3	1000	CH+CH_3
1127	↓	1120	↓
1146	C-OH	1138	C-OH
1144	CH+NH+C-OH	1168	CH+NH+C-OH
1294	↓	1285	↓
1308	CH_3	1302	CH_3
1407	↓	1408	↓
1435	NH (9)	1448	NH (7)
1436	NH (3)	1451	NH (4,3)
1449	NH (2)	1455	NH (5)
1466	NH (4)	1455	NH (9)
1483	NH (7)	1463	NH (2)
1488	NH (8)	1468	NH (6,5,4,3)
1494	NH (5)	1477	NH (6,4)
1495	NH (6)	1489	NH (8)
1549	NH_2	1552	NH_2 Na^+
1646	C=O (6)	1647	C=O (1)
1652	C=O (1)	1660	C=O (8)
1664	C=O (7)	1666	C=O (3)
1671	C=O (4,3)	1670	C=O (6)
1678	C=O (5,2)	1680	C=O (5)
1693	C=O (4)	1693	C=O (4)
1694	C=O (3,2)	1695	C=O (7)
1715	C=O (8)	1703	C=O (2)
1732	C=O (C-term)	1768	C=O (C-term)

Table 3.33 – Ala_9Na^+ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A9_6182		**A9_3017**	
1000	$CH+CH_3$	1000	$CH+CH_3$
1120	↓	1126	↓
1145	C-OH	1155	C-OH
1169	CH+NH+C-OH	1170	CH+NH+C-OH
1281	↓	1294	↓
1301	CH_3	1299	CH_3
1419	↓	1403	↓
1438	NH (8)	1450	NH (8)
1442	NH (4,3)	1453	NH (3)
1456	NH (6)	1464	NH (5)
1462	NH (9,7)	1468	NH (2)
1465	NH (9)	1470	NH (9)
1466	NH (2)	1473	NH (7)
1467	NH (7)	1480	NH (6,4)
1472	NH (6,5)	1487	NH (6,4)
1591	NH_2 (Na^+)	1572	NH_2 (Na^+)
1643	C=O (1)	1651	C=O (1)
1655	C=O (8)	1655	C=O (3)
1665	C=O (6,5,4)	1667	C=O (4)
1668	C=O (3)	1671	C=O (6,5)
1679	C=O (6)	1686	C=O (7)
1692	C=O (5,4)	1696	C=O (5,4)
1700	C=O (7)	1709	C=O (8)
1712	C=O (2)	1715	C=O (2)
1772	C=O (C-term)	1744	C=O (C-term)

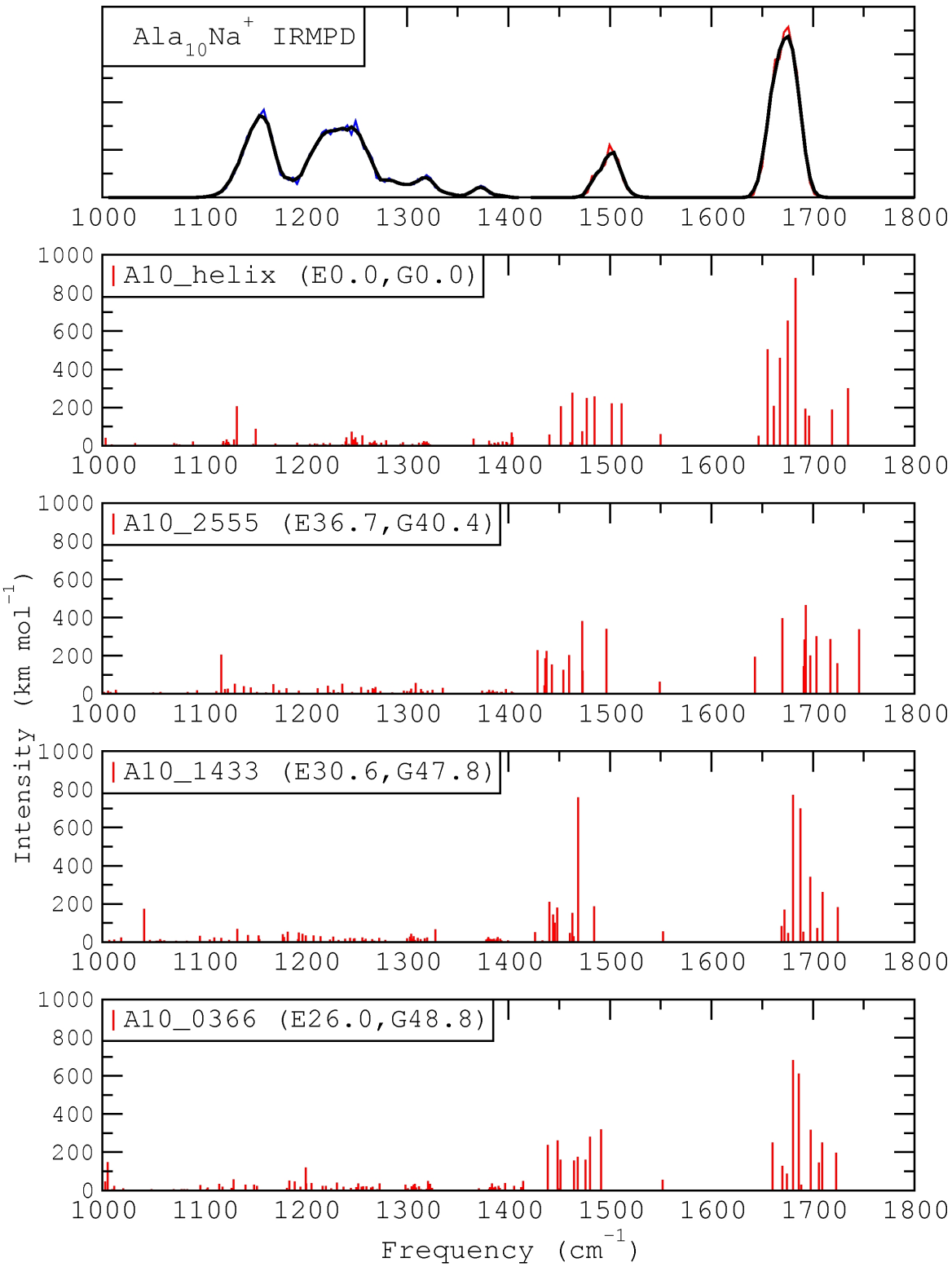

Figure 3.36 – $Ala_{10}Na^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **CC2**, frequencies are scaled by 0.94)

Table 3.34 – $Ala_{10}Na^+$ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A10_helix		**A10_2555**	
1000	CH+CH_3	1000	CH+CH_3
1130	↓	1146	↓
1133	C-OH	1117	C-OH
1203	CH+NH+C-OH	1161	CH+NH+C-OH
1296	↓	1300	↓
1305	CH_3	1303	CH_3
1404	↓	1436	↓
1441	NH (3)	1429	NH (5)
1452	NH (2)	1436	NH (3)
1461	NH (10,5,7))	1437	NH (9)
1463	NH (10)	1443	NH (6)
1473	NH (7,4)	1454	NH (7)
1477	NH (6,4)	1460	NH (2)
1485	NH (5)	1473	NH (4)
1502	NH (9)	1473	NH (8)
1512	NH (8)	1497	NH (10)
1550	NH_2 N-H (2)	1549	NH_2 N-H (2)
1646	C=O (8,7)	1643	C=O (1)
1656	C=O (1)	1670	C=O (7)
1661	C=O (6,5)	1691	C=O (5)
1668	C=O (8,4)	1692	C=O (2)
1675	C=O (6,3)	1693	C=O (9)
1683	C=O (5,2)	1697	C=O (8)
1693	C=O (4,3)	1703	C=O (6)
1696	C=O (3,2)	1717	C=O (3)
1719	C=O (9)	1724	C=O (4)
1735	C=O (C-term)	1745	C=O (C-term)

Table 3.35 – $Ala_{10}Na^+$ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A10_1433		**A10_0366**	
1000	CH+CH_3	1000	CH+CH_3
1133	↓	1141	↓
1143	CH+NH+C-OH	1152	CH+NH+C-OH
1307	↓	1300	↓
1308	CH_3	1301	CH_3
1434	↓	1412	↓
1440	NH (5)	1439	NH (5)
1444	NH (3)	1448	NH (9)
1446	NH (6)	1450	NH (4)
1448	NH (9)	1452	NH (6)
1461	NH (10)	1465	NH (2)
1463	NH (2)	1468	NH (7)
1464	NH (8)	1476	NH (3)
1469	NH (4)	1481	NH (10)
1484	NH (7)	1491	NH (8)
1552	NH_2 CO-H	1552	NH_2 CO-H
1669	C=O (6)	1660	C=O (1)
1672	C=O (1)	1670	C=O (6)
1675	C=O (8)	1674	C=O (8)
1681	C=O (7)	1680	C=O (5)
1688	C=O (5)	1686	C=O (7)
1691	C=O (4,2)	1688	C=O (2)
1697	C=O (4,2)	1698	C=O (3)
1704	C=O (3)	1706	C=O (9,4)
1709	C=O (9)	1709	C=O (9,4)
1724	C=O (C-term)	1723	C=O (C-term)

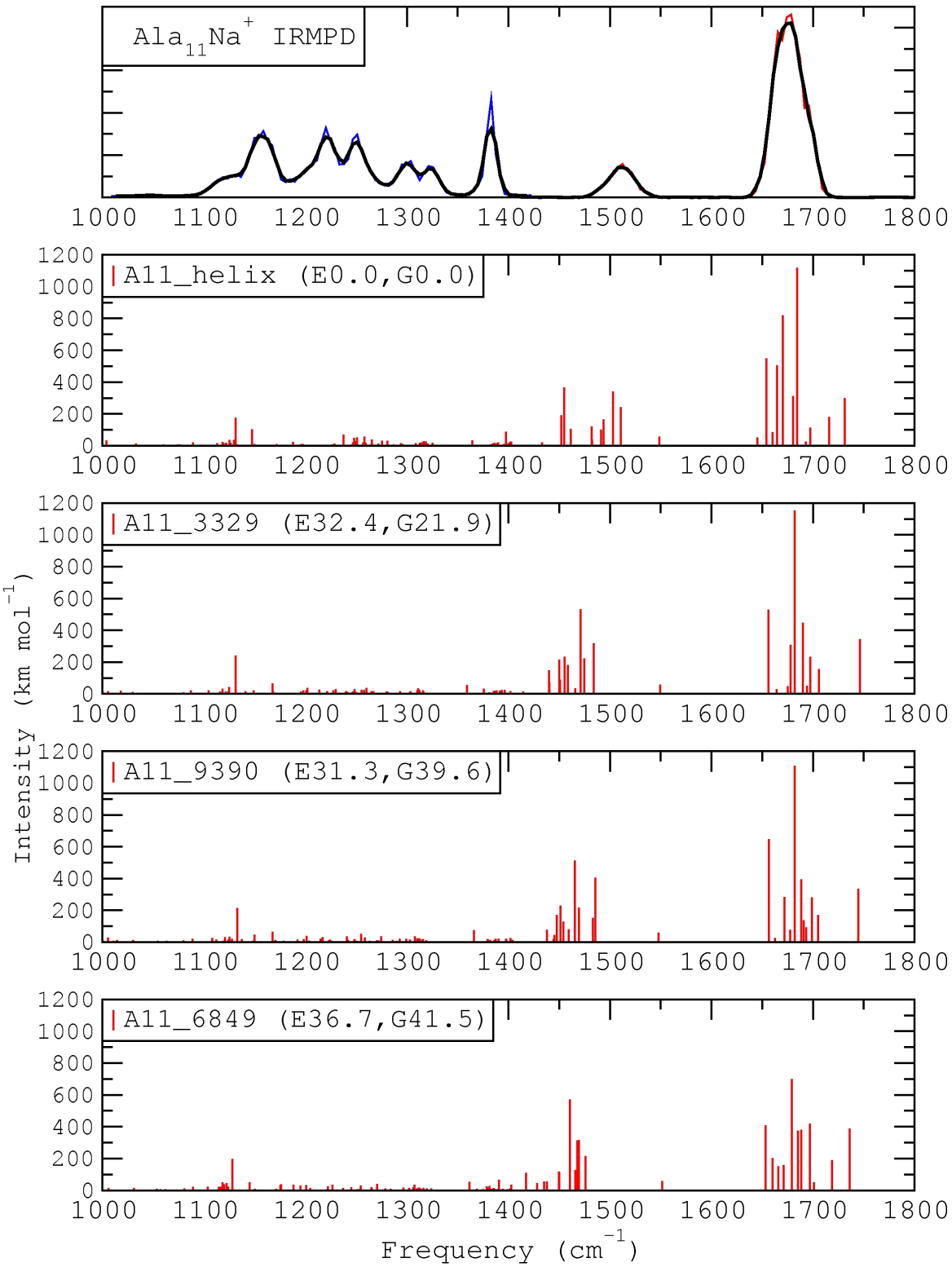

Figure 3.37 – $Ala_{11}Na^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **M06**, frequencies are scaled by 0.94)

Table 3.36 – $Ala_{11}Na^+$ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A11_helix		**A11_3329**	
1000	CH+CH_3	1000	CH+CH_3
1133	↓	1125	↓
1132	C-OH	1131	C-OH
1188	CH+NH+C-OH	1168	CH+NH+C-OH
1296	↓	1303	↓
1309	CH_3	1310	CH_3
1403	↓	1414	↓
1433	NH (4,3)	1440	NH (10,11)
1452	NH (3,2)	1441	NH (4,3,2)
1455	NH (2)	1450	NH (3,2)
1462	NH (11)	1451	NH (7)
1482	NH (7,6,5)	1455	NH (11,10)
1483	NH (10,6,5)	1459	NH (9,8)
1491	NH (6,5)	1466	NH (4)
1494	NH (7,6,5)	1471	NH (6,5,4)
1503	NH (9,8,7)	1475	NH (9,8,7,6)
1511	NH (9,8)	1484	NH (5)
1549	NH_2 N-H (2)	1549	NH_2 N-H (2)
1645	C=O (9,8)	1656	C=O (1)
1654	C=O (1)	1664	C=O (5,4)
1660	C=O (7,6)	1675	C=O (8,6)
1665	C=O (5)	1678	C=O (8,7,5)
1670	C=O (4)	1682	C=O (9,6,5,2)
1680	C=O (7)	1690	C=O (10,5,4,3,2)
1685	C=O (6)	1690	C=O (10,5,4,3,2)
1693	C=O (4,2)	1694	C=O (10,8)
1697	C=O (3,2)	1697	C=O (3,2)
1716	C=O (10)	1706	C=O (9)
1731	C=O (C-term)	1746	C=O (C-term)

Table 3.37 – $Ala_{11}Na^+$ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	**Mode**	**Freq**	**Mode**
A11_9390		**A11_6849**	
1000	CH+CH_3	1000	CH+CH_3
1137	↓	1124	↓
1133	C-OH	1128	C-OH
1167	CH+NH+C-OH	1175	CH+NH+C-OH
1303	↓	1284	↓
1308	CH_3	1302	CH_3
1444	↓	1428	↓
1438	NH (3)	1417	NH (11)
1445	NH (7)	1435	NH (7,6)
1447	NH (10)	1438	NH (6,3)
1451	NH (3,2)	1450	NH (3,2)
1454	NH (9,8)	1450	NH (9,8)
1459	NH (8,7,6)	1461	NH (7,5,4,3)
1466	NH (6,4)	1466	NH (8,5)
1469	NH (9,8)	1468	NH (10,5)
1483	NH (11)	1469	NH (10,8)
1486	NH (5)	1476	NH (6)
1548	NH_2 N-H (2)	1551	NH_2 N-H (2)
1656	C=O (1)	1653	C=O (3)
1662	C=O (4)	1660	C=O (1)
1672	C=O (6)	1666	C=O (6)
1678	C=O (7,6)	1671	C=O (7)
1682	C=O (5,2)	1679	C=O (9)
1689	C=O (10)	1685	C=O (4)
1691	C=O (4,3,2)	1688	C=O (2)
1694	C=O (8)	1697	C=O (7,8)
1699	C=O (3,2)	1701	C=O (10)
1705	C=O (9)	1719	C=O (5)
1744	C=O (C-term)	1736	C=O (C-term)

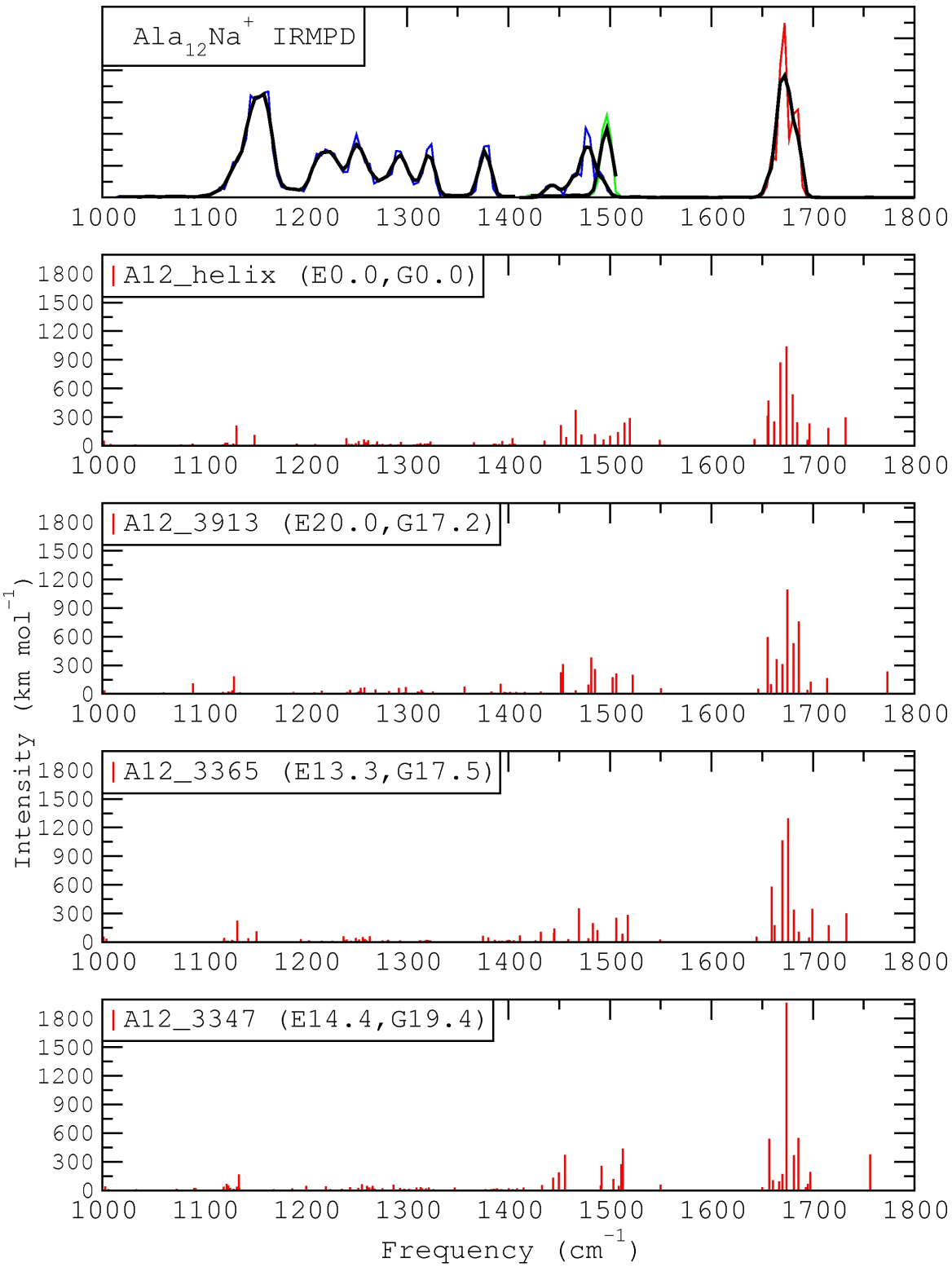

Figure 3.38 – $Ala_{12}Na^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **M06**, frequencies are scaled by 0.94)

Table 3.38 – $Ala_{12}Na^+$ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
A12_helix		**A12_3347**	
1000	$CH+CH_3$	1000	$CH+CH_3$
1134	↓	1135	↓
1132	C-OH	1134	C-OH
1171	CH+NH+C-OH	1169	CH+NH+C-OH
1301	↓	1300	↓
1307	CH_3	1302	CH_3
1406	↓	1415	↓
1436	NH (4,3)	1433	NH (4,3,2)
1452	NH (3,2)	1444	NH (11)
1457	NH (4,3,2)	1450	NH (4,3,2)
1466	NH (6)	1456	NH (4,3,2)
1472	NH (12,11)	1491	NH (9,8)
1485	NH (12,11)	1492	NH (12)
1494	NH (7,5)	1499	NH (5,4)
1500	NH (7,5)	1503	NH (9,8)
1508	NH (10,9,8,7)	1509	NH (7,6,5)
1515	NH (10,9)	1511	NH (7,6)
1520	NH (8)	1513	NH (10)
1549	NH_2 NH_2	1550	NH_2 NH_2
1642	C=O (10,9)	1650	C=O (8)
1655	C=O (8,7)	1657	C=O (1)
1656	C=O (1)	1661	C=O (5)
1662	C=O (6)	1666	C=O (9,5,4)
1668	C=O (5)	1670	C=O (11)
1674	C=O (8,4)	1674	C=O (8,9)
1680	C=O (7,6)	1681	C=O (7)
1684	C=O (8,5,2)	1685	C=O (6)
1695	C=O (4,2)	1693	C=O (4,2)
1697	C=O (3,2)	1695	C=O (10)
1715	C=O (11)	1697	C=O (3,2)
1732	C=O (C-term)	1756	C=O (C-term)

Table 3.39 – $Ala_{12}Na^+$ scaled (0.94) M06/6-31G(d,p) fingerprint region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
A12_3913		**A12_3365**	
1000	$CH+CH_3$	1000	$CH+CH_3$
1128	↓	1135	↓
1130	C-OH	1133	C-OH
1163	CH+NH+C-OH	1144	CH+NH+C-OH
1301	↓	1298	↓
1311	CH_3	1301	CH_3
1416	↓	1426	↓
1432	NH (4,3)	1432	NH (2)
1452	NH (3,2)	1445	NH (3)
1453	NH (11,4,3,2)	1445	NH (12)
1454	NH (11)	1459	NH (6,5,4,3)
1466	NH (8,7)	1470	NH (6,4)
1479	NH (9,8,7)	1479	NH (6,5,4)
1482	NH (12)	1483	NH (10,9,8,7)
1485	NH (6)	1488	NH (10,7)
1503	NH (10)	1506	NH (9,8)
1506	NH (5)	1512	NH (10,7)
1522	NH (9)	1517	NH (11)
1550	NH_2 N-H (2)	1550	NH_2 N-H (2)
1646	C=O (9,8)	1645	C=O (10,9)
1655	C=O (1)	1656	C=O (9,8,7,6)
1659	C=O (9,8,7,6)	1659	C=O (1)
1664	C=O (10,6,5)	1662	C=O (1)
1670	C=O (4)	1670	C=O (5)
1675	C=O (11,9,8)	1676	C=O (8,4)
1681	C=O (7,3)	1681	C=O (7,6)
1686	C=O (6)	1686	C=O (8,5)
1695	C=O (5,4,2)	1696	C=O (4,2)
1698	C=O (3,2)	1699	C=O (3,2)
1714	C=O (10)	1716	C=O (11)
1773	C=O (C-term)	1733	C=O (C-term)

Ala_8Na^+

In the experimental spectra shown in Figure 3.34, the main amide I band and two small bands to the blue of the main band (corresponding the C-terminal C=O(8) and the free C=O(7)) are consistent with the calculated spectrum for the helix presented below it. This is in contrast to the calculated spectra for the globular structures where there is only one calculated band above the main amide I band. In addition, the position of the amide II band is consistent with the theoretical bands of the calculations of the helical structure, while very slightly to the red of the position of the amide II band in all other calculated spectra. Due to the numerous peaks distributed throughout the entire region below the amide II band, it is not possible to make any conclusive statements, although one can say that the calculated peaks of all four structures and experimental peak positions are consistent in this region. Possibly, the A8_4740 globular structure accounts for the activity in the 1200-1250 cm^{-1} and $\approx$1310 cm^{-1} regions.

Ala_9Na^+

In the experimental spectra shown in Figure 3.35 the amide I band is narrow and sharply consistent with the main amide I band of the calculated lowest-energy structure, A9_helix. The weakly absorbing carbonyl peaks to the blue of the main amide I modes are not visible in this case as additional experiments tuned specifically to observe these two modes were not obtained. Overlooking the absence of these modes, the experimental and calculated results of A9_helix are slightly more consistent than the agreement between the experiment and the other calculated results measured by the width of the peak. The experimental amide II band is consistent with the position of the most intense bands of the calculated lowest-energy structure, A9_helix, and less consistent with the amide II bands for the other calculated structures by the position of the centre of the peak (other calculated bands are somewhat red shifted). The weaker amide II bands, seen in the calculated spectrum of A9_helix to the red of the most intense modes, were not observed

under the experimental conditions of this scan. As was noted for Ala_8Na^+, the modes below the amide II band are more difficult to assign, but may be considered consistent with each of the calculated results in this spectral region.

$Ala_{10}Na^+$

In the experimental spectra shown in Figure 3.36 the amide I band is again very consistent with the main amide I band of the calculated lowest-energy structure, A10_helix. Unfortunately, it is also the case that the weakly absorbing carbonyl peaks to the blue of the main amide I modes were not observed in the experimental spectrum. However, overlooking the absence of these two small modes, one can see that the experimental amide I matches most closely the calculated result for the A10_helix structure with the other calculations predicting an amide I band either slightly too broad or slightly blue shifted. The experimental amide II band is consistent with the calculated result for A10_helix and less consistent with A10_0366. The region below the amide II band is again less than definitive, except for A10_1433 where there is clear inconsistency between calculation and experiment.

$Ala_{11}Na^+$

In the experimental spectra shown in Figure 3.37 the amide I band is consistent with the main amide I band of the calculated lowest-energy structure, A11_helix, and all other calculated spectra. The three calculated structures in addition to the fully α-helical structure (lowest-energy) are approximately 50% helical and result in very similar amide I bands in all cases. It is again the case that the weakly absorbing carbonyl peaks to the blue of the main amide I modes were not observed in the experimental spectrum. The amide II band of the experimental spectrum is in fair agreement with the calculated band of the A11_helix structure, and quite out of agreement with the other calculated results by measure of the position of the centre of the peak. The amide II modes farthest to the red in the calculated spectrum of A11_helix were not experimentally observed under the

configuration of that scan. Modes below the amide II band for all calculated results can be considered consistent with experiment.

$\mathbf{Ala_{12}Na^{+}}$

In the experimental spectra shown in Figure 3.38 the amide I,II,III bands are consistent with the bands of the calculated lowest-energy structure, A12_helix, as well as the other calculated spectra. In this case, all calculated structures are almost fully helical, with only minor differences in the sodium binding and the conformations at the terminal ends of the peptide. For this reason, the calculated spectra are too similar to be distinguished with this spectral resolution in this range.

3000-3600 cm^{-1} Region Spectral Results

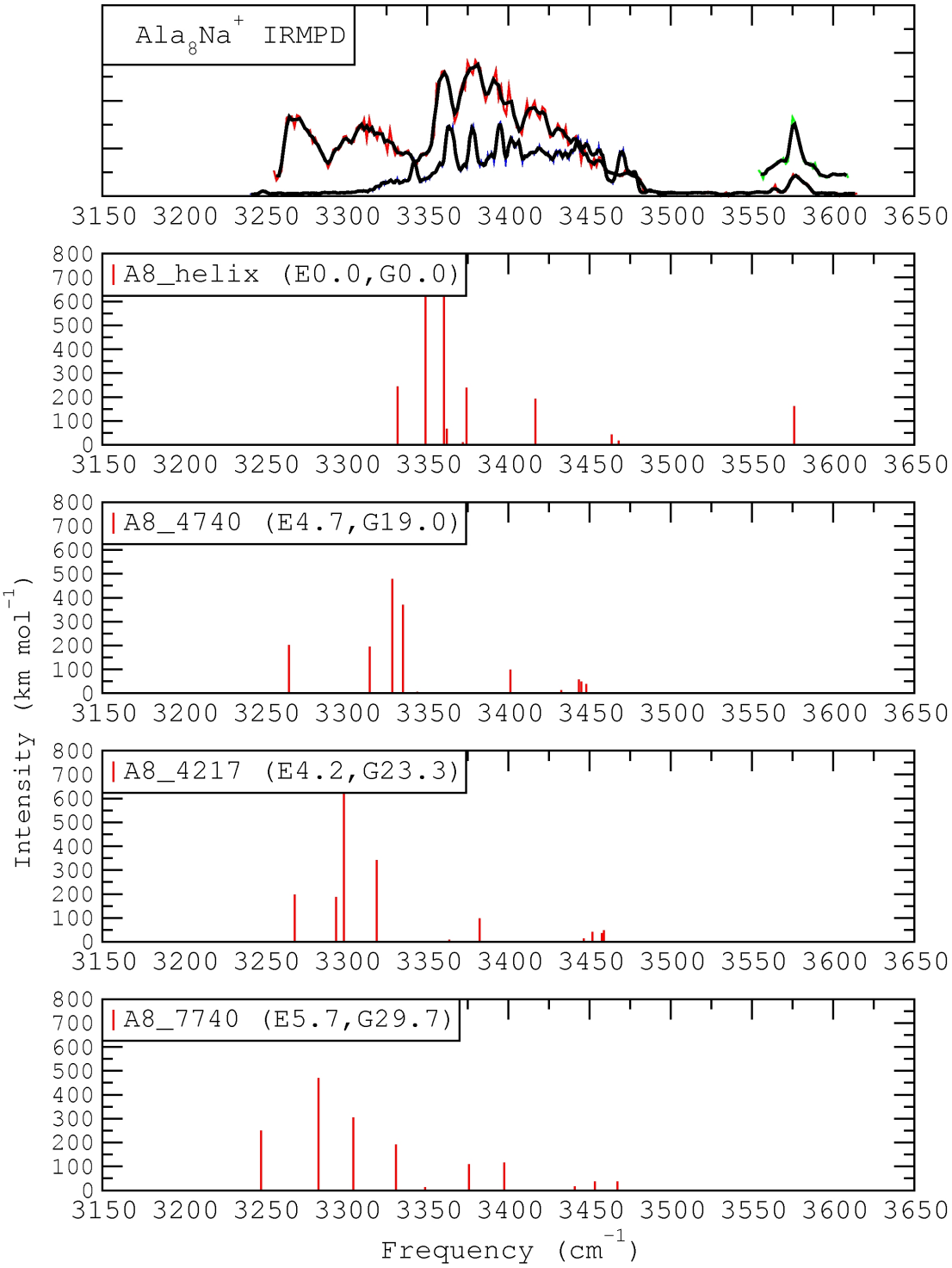

Figure 3.39 – Ala_8Na^+ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **CC2**, frequencies are scaled by 0.94 and N-H stretching modes by 0.955)

Table 3.40 – Ala_8Na^+ scaled (0.94, N-H modes scaled 0.955) M06/6-31G(d,p) 3000-3600 cm^{-1} region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
	A8_helix		**A8_4740**
3332	N-H (7) C=O (3)	3265	N-H (5) C=O (3)
3349	N-H (8) C=O (4)	3315	N-H (4) C=O (7)
3360	N-H (6,2) C=O (2),NH_2	3329	O-H C=O (1)
3362	N-H (6,2) C=O (2),NH_2	3335	N-H (7) C=O (4)
3372	NH_2 N-H (2)	3344	NH_2 Na^+
3374	N-H (5) C=O (1)	3402	N-H (8) C=O (6)
3417	N-H (4) C=O (1)	3433	NH_2 Na^+
3464	N-H (3) free	3443	N-H (2) NH_2
3468	NH_2 N-H (2)	3445	N-H (6) C=O (4)
3576	O-H free	3448	N-H (3) free
	A8_4217		**A8_7740**
3268	N-H (5) C=O (3)	3248	N-H (5) C=O (3)
3294	O-H C=O (1) + N-H (4) C=O(7)	3283	O-H C=O (1)
3299	O-H C=O (1) + N-H (4) C=O(7)	3305	N-H (4) C=O (7)
3319	N-H (7) C=O (4)	3331	N-H (7) C=O (4)
3364	NH_2 Na^+	3349	NH_2 Na^+)
3383	N-H (8) C=O (6)	3376	N-H (8) C=O (6)
3447	NH_2 Na^+	3398	N-H (6) free
3452	N-H (3) free	3441	NH_2 Na^+
3458	N-H (2) free	3453	N-H (2) free
3459	N-H (6) free	3467	N-H (3) free

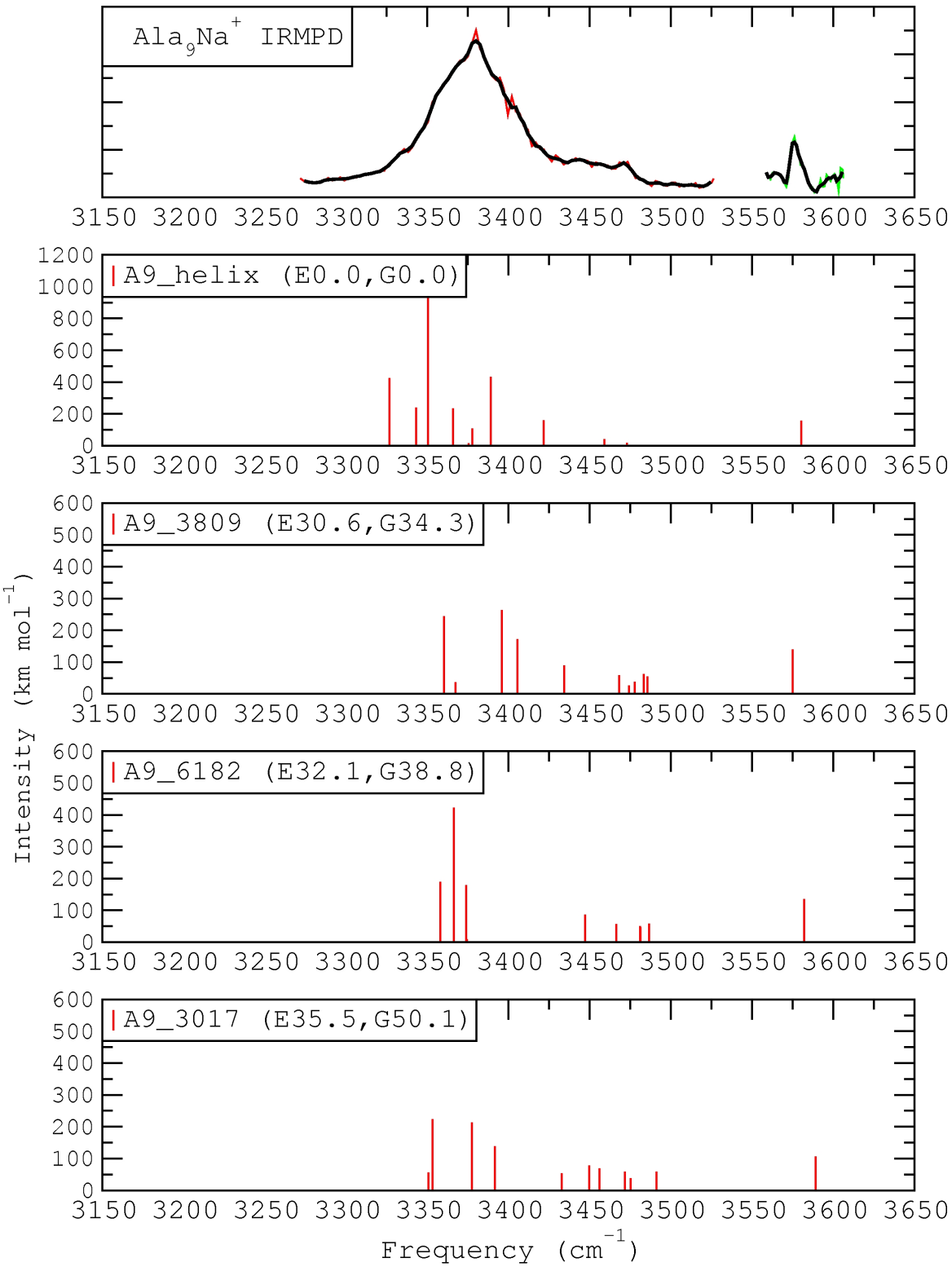

Figure 3.40 – Ala_9Na^+ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **CC2**, frequencies are scaled by 0.94 and N-H stretching modes by 0.955)

Table 3.41 – Ala_9Na^+ scaled (0.94, N-H modes scaled 0.955) M06/6-31G(d,p) 3000-3600 cm^{-1} region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
A9_helix		**A9_3089**	
3327	N-H (9) C=O (5)	3361	N-H (8) C=O (6)
3343	N-H (8) C=O (4)	3368	NH_2 Na^+
3351	N-H (7) C=O (3)	3396	N-H (6) C=O (3)
3366	N-H (2) NH_2	3406	N-H (4) C=O (1)
3376	NH_2 N-H (2)	3435	N-H (9) C=O (9)
3378	N-H (5) C=O (1)	3468	N-H (7) free
3389	N-H (6) C=O (2)	3475	NH_2 Na^+
3422	N-H (4) C=O (1)	3478	N-H (3) free
3459	N-H (3) free	3484	N-H (5) free
3473	NH_2 N-H(2)	3486	N-H (2) free
3581	O-H free	3575	O-H free
A9_6182		**A9_3017**	
3358	N-H (8) C=O (5)	3351	NH_2 Na^+
3366	N-H (4) C=O (1)	3353	N-H (4) C=O (1)
3374	N-H (6) C=O (3)	3378	N-H (8) C=O (5)
3375	NH_2 Na^+	3392	N-H (6) C=O (3)
3447	N-H (9) C=O (9)	3433	NH_2 Na^+
3467	N-H (7) free	3450	N-H (5) C=O (3)
3481	N-H (5) free	3456	N-H (9) free
3481	NH_2 Na^+	3472	N-H (7) free
3481	N-H (3) free	3476	N-H (3) free
3487	N-H (2) free	3491	N-H (2) free
3582	O-H free	3590	O-H free

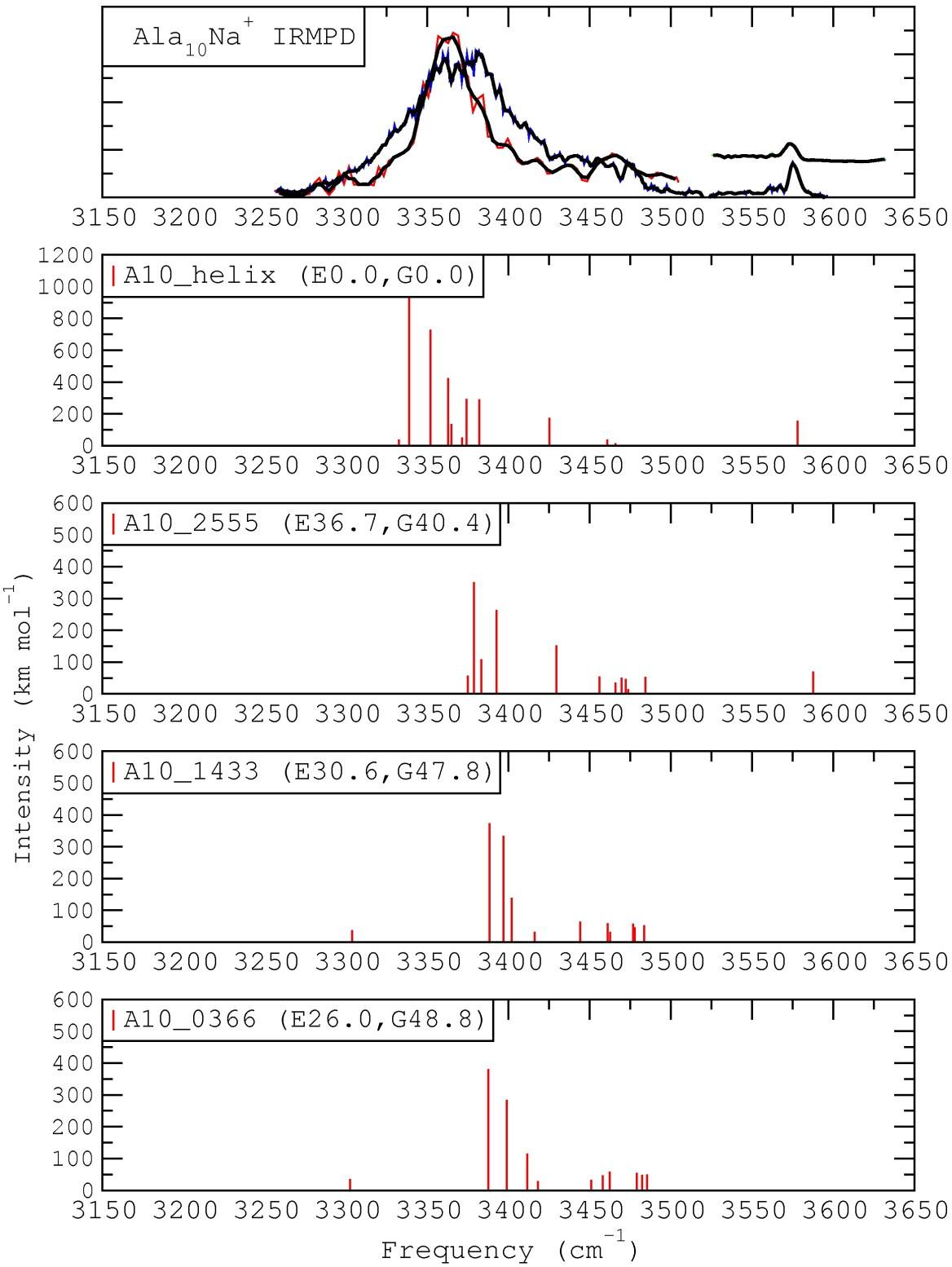

Figure 3.41 – $Ala_{10}Na^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **CC2**, frequencies are scaled by 0.94 and N-H stretching modes by 0.955)

Table 3.42 – $Ala_{10}Na^+$ scaled (0.94, N-H modes scaled 0.955) M06/6-31G(d,p) 3000-3600 cm^{-1} region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
	A10_helix		**A10_2555**
3333	N-H (10,9) C=O (6,5)	3375	NH_2 N-H (2)
3339	N-H (10,9) C=O (6,5)	3379	N-H (10) C=O (7)
3352	N-H (8) C=O (4)	3383	N-H (2) NH_2
3363	N-H (7) C=O (3)	3393	N-H (8) C=O (5)
3365	N-H (2) NH_2	3430	N-H (4) C=O (1)
3372	N-H (2) NH_2	3456	N-H (5) free
3374	N-H (6) C=O (2)	3466	N-H (3) free
3382	N-H (5) C=O (1)	3470	N-H (6) free
3425	N-H (4) C=O (1)	3472	N-H (9) free
3461	N-H (3) free	3474	NH_2 N-H (2)
3466	NH_2 N-H (2)	3484	N-H (7) free
3578	O-H free	3588	O-H free
	A10_1433		**A10_0366**
3304	NH_2 C-OH	3303	NH_2 C-OH
3389	N-H (8) C=O (5)	3388	N-H (8) C=O (5)
3397	N-H (4) C=O (1)	3399	N-H (4) C=O (1)
3402	N-H (10) C=O (7)	3412	N-H (10) C=O (7)
3416	NH_2 C-OH	3418	NH_2 C-OH
3444	N-H (7) free	3451	N-H (6) free
3461	N-H (6,5)	3458	N-H (5) free
3463	N-H (6,5)	3462	N-H (7) free
3477	N-H (3) free	3479	N-H (9) free
3478	N-H (9) free	3483	N-H (3) free
3484	N-H (2) free	3485	N-H (2)

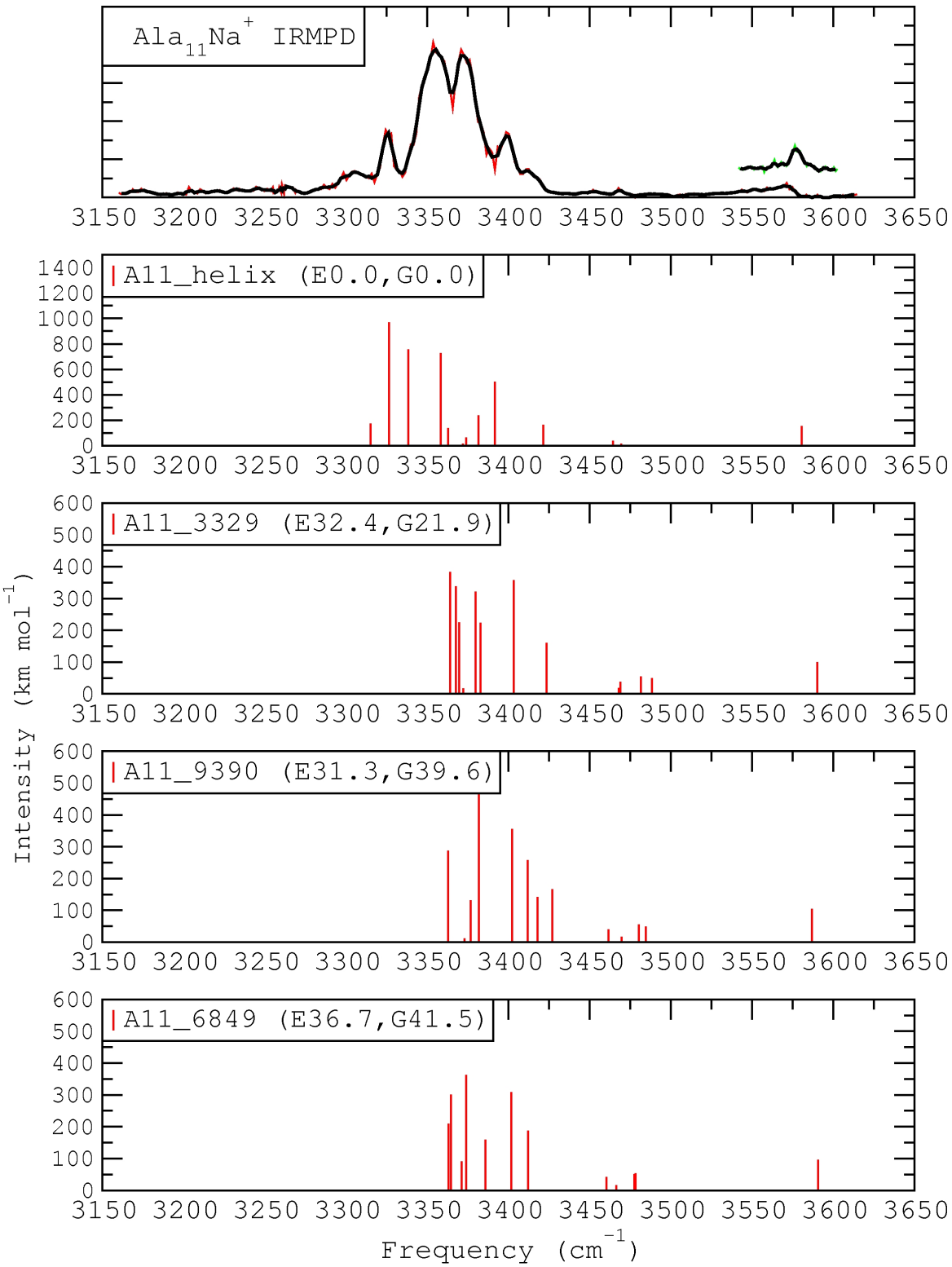

Figure 3.42 – $Ala_{11}Na^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **M06**, frequencies are scaled by 0.94 and N-H stretching modes by 0.955)

Table 3.43 – $Ala_{11}Na^+$ scaled (0.94, N-H modes scaled 0.955) M06/6-31G(d,p) 3000-3600 cm^{-1} region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
A11_helix		**A11_3329**	
3315	N-H (10) C=O (6)	3364	N-H (2) NH_2
3327	N-H (11) C=O (7)	3368	N-H (6) C=O (2)
3339	N-H (9) C=O (5)	3370	N-H (9) C=O (6)
3358	N-H (8) C=O (4)	3372	N-H (2) NH_2
3363	N-H (2) NH_2	3380	N-H (5) C=O (1)
3372	N-H (2) NH_2	3383	N-H (11) C=O (8)
3374	N-H (5) C=O (1)	3404	N-H (7) C=O (3)
3382	N-H (7) C=O (3)	3424	N-H (1) C=O (1)
3392	N-H (6) C=O (2)	3468	NH_2 N-H (2)
3422	N-H (4) C=O (1)	3469	N-H (3) free
3465	N-H (3) free	3482	N-H (10) free
3470	NH_2 N-H (2)	3489	N-H (8) free
3580	O-H free	3591	O-H free
A11_9390		**A11_6849**	
3363	N-H (2) NH_2	3363	N-H (11) C=O (8)
3373	N-H (2) NH_2	3365	N-H (2) NH_2
3377	N-H (6) C=O (2)	3371	N-H (2) NH_2
3382	N-H (5) C=O (1)	3374	N-H (5,7) C=O (2,3)
3402	N-H (7) C=O (3)	3374	N-H (5,7) C=O (2,3)
3412	N-H (11) C=O (8)	3386	N-H (9) C=O (6)
3418	N-H (9) C=O (6)	3402	N-H (6) C=O (2)
3427	N-H (4) C=O (1)	3412	N-H (4) C=O (1)
3462	N-H (3) free	3460	N-H (3) free
3470	NH_2 N-H (2)	3466	NH_2 N-H (2)
3480	N-H (10) free	3478	N-H (8) free
3485	N-H (8) free	3478	N-H (10) free
3587	O-H free	3591	O-H free

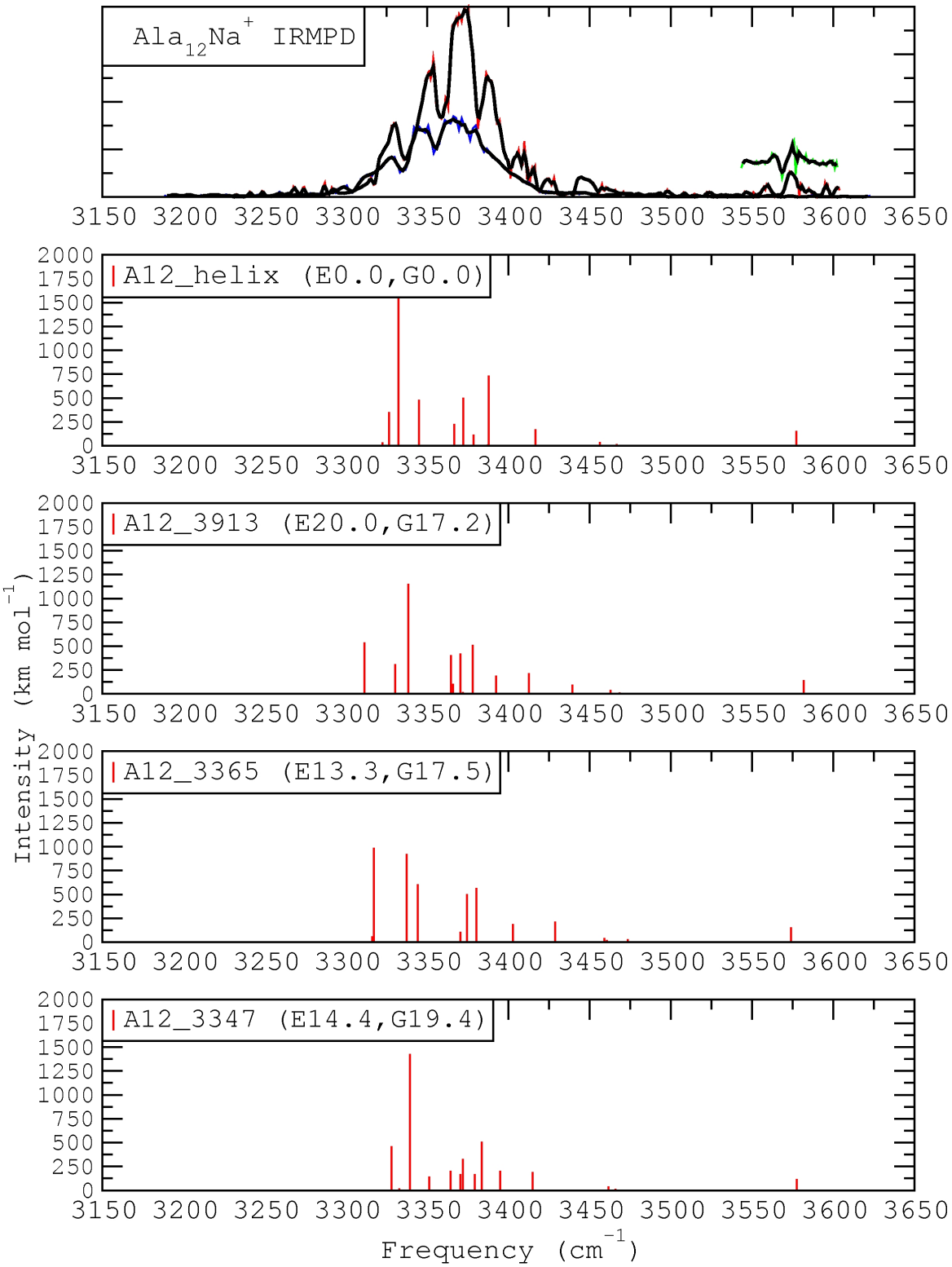

Figure 3.43 – $Ala_{12}Na^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **M06**, frequencies are scaled by 0.94 and N-H stretching modes by 0.955)

Table 3.44 – $Ala_{12}Na^+$ scaled (0.94, N-H modes scaled 0.955) M06/6-31G(d,p) 3000-3600 cm^{-1} region vibrational assignments (cm^{-1}).

Freq	Mode	Freq	Mode
	A12_helix		**A12_3347**
3323	N-H (12,11) C=O (8,7)	3328	N-H (11) C=O (7)
3327	N-H (12,11,10) C=O (8,7,6)	3333	N-H (9) C=O (5)
3332	N-H (12,11,10) C=O (8,7,6)	3340	N-H (10) C=O (6)
3345	N-H (9) C=O (5)	3352	N-H (12) C=O (10)
3367	N-H (2) NH_2	3365	N-H (2) NH_2
3372	N-H (8) C=O (4)	3371	N-H (2) (NH_2
3373	N-H (2) NH_2	3372	N-H (8) C=O (4)
3379	N-H (6) C=O (2)	3380	N-H (7) C=O (3)
3387	N-H (5) C=O (1)	3384	N-H (6) C=O (2)
3388	N-H (7,5) C=O (3,1)	3395	N-H (5) C=O (1)
3417	N-H (4) C=O (1)	3415	N-H (4) C=O (1)
3456	N-H (3) free	3462	N-H (3) free
3467	NH_2 N-H (2)	3466	NH_2 N-H (2)
3577	O-H free	3578	O-H free
	A12_3913		**A12_3365**
3312	N-H (11) C=O (7)	3316	N-H (12,10) C=O (8,6)
3331	N-H (10) C=O (6)	3317	N-H (12,10) C=O (8,6)
3338	N-H (9) C=O (5)	3337	N-H (11) C=O (7)
3365	N-H (2) NH_2	3344	N-H (9) C=O (5)
3366	N-H (6) C=O (2)	3365	NH_2
3371	N-H (2) NH_2	3371	N-H (8,7) C=O (4,3)
3372	N-H (8,7) C=O (4,3)	3375	N-H (8,7) C=O (4,3)
3378	N-H (8,7,6) C=O (4,3,2)	3380	N-H (6) C=O (2)
3393	N-H (5) C=O (1)	3403	N-H (5) C=O (1)
3413	N-H (4) C=O (1)	3429	N-H (4) C=O (1)
3440	N-H (12) C=O (12)	3459	N-H (2) free
3463	N-H (3) free	3461	NH_2
3469	NH_2 N-H (2)	3474	N-H (3) free
3582	O-H free	3574	O-H free

Ala_8Na^+

In both scans in Figure 3.39, bands between 3325-3450 cm^{-1}, corresponding to the bound N-H modes, match well to the analogous modes in the calculated A8_helix spectrum. Modes of the bound N-H groups are clearly defined in the experimental spectrum as well as modes of the single free N-H. The two modes of the NH_2 are convoluted with a bound N-H around 3365 cm^{-1} and the free N-H at 3475 cm^{-1}. The free C-OH is seen as expected at 3575 cm^{-1}. The difference between the blue and the red scans in the figure is attributed to an approximate 50% reduction of laser power in the blue scan. This is likely the cause of the differences in the 3250-3350 cm^{-1} region. The bands in the red scan of the experimental spectrum in the 3250-3350 cm^{-1} region are unaccounted for by the calculated spectrum of A8_helix, however match very well to the calculated modes of the lowest-energy globular structure, A8_4740. The α-helical conformation of Ala_8Na^+ is calculated at the RI-CC2/def2-TZVPP//RI-MP2/def2-SVP level to be 11.9 kJ mol^{-1} lower in relative free energy (298 K) (2.6 kJ mol^{-1} lower in electronic energy) than the lowest-energy globular conformation. This observation is consistent with the discussion in the introduction of this chapter that the relative stability of the α-helical structure is overestimated by the **M06** functional in comparison to results at higher levels of theory. Additionally, Ala_8Na^+ was the only size where a globular structure was calculated to be lower in free energy than the α-helical structure at any level of theory. The combined experimental and computational considerations appear to indicate that the lowest-energy globular structure, A8_4740, and A8_helix are of very similar relative energies and both contribute to the observed room temperature experimental spectrum (where the A8_helix is seemingly the dominant contribution).

Ala_9Na^+

The 3000-3600 cm^{-1} region spectrum of Ala_9Na^+ is presented in Figure 3.40. The bound N-H bands in the experimental spectrum starting at 3325 cm^{-1} correspond well to

the beginning of the bound N-H peaks in the calculated spectrum for A9_helix. As well, the experimental free N-H band, assigned at ≈3455 cm^{-1}, matches the calculated position for A9_helix (3459 cm^{-1}) very well. The calculated globular structures are less consistent with the experimental spectra, where the bound N-H modes begin approximately 25 cm^{-1} to the blue of the experimental band. Unfortunately, A9_helix is not the only structure calculated to have a free C-OH, making the free C-OH band less conclusive.

$Ala_{10}Na^+$

Both scans of Figure 3.41 show excellent agreement with the calculated spectrum of the lowest-energy structure, A10_helix. Although the resolution in this case is insufficient to distinguish individual modes, the overall shape of the bound N-H band, as well as the position, are very consistent with the calculated spectrum. Additionally, the bound N-H band at 3425 cm^{-1} and the free N-H and NH_2 bands at approximately 3460 cm^{-1} of the calculated A10_helix spectrum are also in agreement with the experimental spectra. Only the calculated spectrum of the helical structure is in reasonable agreement with the experimental bound N-H band, as all calculated globular structures are approximately 50 cm^{-1} to the blue of the experimental band. This assignment is also supported by the presence of the free C-OH band at 3575 cm^{-1}, although in the second-lowest-energy structure, A10_2555, the C-terminal C-OH is also free.

$Ala_{11}Na^+$

The experimental spectrum shown in Figure 3.42 is again in excellent qualitative agreement with the calculated spectrum of the lowest-energy structure, A11_helix. The position, width and shape of the bound N-H band seem to match only the calculated spectrum of the helical structure. The free N-H and NH_2 modes at 3470 cm^{-1} match very closely between the calculation and the experiment, as well as the bound N-H modes farthest to the blue. Due to the excellent spectral resolution of this experiment, one can clearly see that all the bound N-H modes match very well between experiment and

calculation, taking into account minor shifts in the peak positions in some cases. All calculated $Ala_{11}Na^+$ structures have a free C-terminal C-OH, preventing diagnosis on this basis.

$\mathbf{Ala_{12}Na^+}$

The experimental spectrum shown in Figure 3.43 is in excellent agreement with the calculated spectrum of the lowest-energy structure, A12_helix, and with the other calculated spectra. All calculated structures are mainly helical and have only small structural differences between them, making the differences between spectra more subtle than in the previously discussed examples. Clearly, the spectrum of the lowest-energy structure, A12_helix, has the best match to the experiment. This is true in the sense of size, shape, and peak positions, with even the free N-H and NH_2 modes at 3456 and 3467 cm^{-1} matching the experiment very well. All calculated spectra show a free C-OH band at 3575 cm^{-1}, consistent with the experiment.

Alpha Helical N-H Spectral Signature

Bound N-H modes are observed to be highly dependent on the hydrogen bonding distance. For a given peptide, the order of vibrational modes follows the order of N-H$\cdots$O=C hydrogen bond lengths (longer bond length gives higher frequency vibration). For α-helical conformations, calculated N-H$\cdots$O=C hydrogen bond lengths consistently decrease from the N-terminus ($\approx$2.15 Å) to the C-terminus ($\approx$1.90 Å). This results in a consistent and similar N-H stretching band between 3315-3425 cm^{-1} of ordered N-H modes. In all calculated α-helical structures the N-H group of the third residue is free and found at approximately 3460-3480 cm^{-1}. This consistent pattern of N-H modes produces somewhat of an α-helical fingerprint, remaining very similar for peptides of different numbers of residues. In the Ala_8Na^+ spectrum (Figure 3.39), two modes are found to the red of 3325 cm^{-1} resulting from two short (1.88 and 1.93 Å) N-H$\cdots$O=C hydrogen bonds, while in the lowest-energy globular Ala_9Na^+ structure, N-H$\cdots$O=C distances are not less than 1.99 Å

and no N-H modes are observed below 3361 cm^{-1}. Additionally, this pattern of helical N-H stretching bands is quite sensitive to disruptions in the α-helical structure (especially at the C-terminus). This is well illustrated for the calculated partly-helical $Ala_{11}Na^+$ spectra, where disruption of the helical structure at the C-terminus (favouring a larger extent of coordination of the sodium cation) prevents the short (1.90-1.95 Å) N-H$\cdots$O=C hydrogen bonds at the C-terminal end of the peptide from forming, resulting in an apparent blue shift of the N-H stretching band. For the α-helical peptides, NH_2 modes are consistently found at approximately 3375 cm^1 (symmetric) and approximately 3470 cm^{-1} (asymmetric). NH_2 modes are found over a wider range in the globular conformations, where they are found hydrogen-bound, bound to Na+, or free (asymmetric between 3433-3475 cm^1 and symmetric between 3344-3375 cm^{-1}). Generally, their relatively low intensities and their close proximity to the more intense free N-H modes prevents distinct observation.

3.4 Discussion and Conclusions

Both calculation and experiment clearly support the assignment of α-helical structure to the room-temperature sodiated poly(alanine) peptides, $Ala_{8-12}Na^+$, explored in this work. Consideration of energetics from calculation indicate that for all sizes except Ala_8Na^+, a significant stabilization energy of the α-helical structure comparing to the next lowest-energy structure exists. Calculated results were obtained at multiple levels of theory and while significant spectral differences were found, energetic results remained relatively consistent between methodologies. The main exception was Ala_8Na^+, where **B3LYP-D** calculations indicated the α-helix to be 9.6 kJ mol^{-1} higher in relative free energy (298 K) than the lowest-energy globular structure, although **M06** calculations indicate the α-helix to be the most stable by 32.5 kJ mol^{-1} and RI-CC2/def2-TZVPP/RI-MP2/def2-SVP more stable by 11.9 kJ mol^{-1}.

The fingerprint IRMPD spectrum of Ala_8Na^+ is a fairly conclusive result, where two small bands belonging to the free C=O(7) and C-terminal C=O are seen to the blue of 1750 cm^{-1} and match the calculated spectrum of the α-helix very well and are inconsistent with the calculated globular structures. This strongly suggests that at least a large proportion of the population of ions are α-helical. The positions of the amide II bands in the fingerprint IRMPD spectra are another indication of α-helical structure.

OPO IRMPD spectra in the higher wavelength region (3000-3650 cm^{-1}) provide a wealth of structural information. For $Ala_{8-12}Na^+$, the evidence for α-helical structure resulting from comparison of calculated spectra with experimental spectra can be considered definitive. For Ala_9Na^+, the spectral resolution is somewhat limiting, however comparison with results for other sizes (especially $Ala_{10}Na^+$) allows definitive assignment of α-helical structure. All $Ala_{8-12}Na^+$ peptides have a free C-OH band at 3575 cm^{-1}, a large group of bound N-H modes in the 3300-3400 cm^{-1} region and one free N-H band in the $\approx$3465 cm^{-1} region. For Ala_8Na^+, N-H bands in the 3250-3325 cm^{-1} region appear to be contributions of the lowest-energy globular structure. This is consistent with the energetics, indicating that the transition from globular to helical structure for room temperature sodiated poly(alanine) peptides occurs around Ala_8Na^+.

Through discussion and comparison of $Ala_{8-10}Na^+$ it has been possible to contrast the spectral features of helical and globular peptides and to demonstrate the size dependent transition for room temperature gaseous sodiated poly(alanines) to be at Ala_8Na^+. Additionally, by examining $Ala_{11-12}Na^+$ is was possible to demonstrate a series of mainly α-helical structures and their relative instability in comparison with the fully α-helical structures. The spectral features of partly/mainly helical peptides and the ability to spectrally distinguish even small C-terminal structural variations was also presented. This demonstration of instability of variations to the fully helical structure can be extrapolated to analogous conformations that could be defined for $Ala_{8-10}Na^+$. These results suggest,

that while some particular globular structures are likely to be uniquely low in energy, it is likely that a large series of structures from **helical** $\rightarrow$ **mainly helical** $\rightarrow$ **partly helical** $\rightarrow$ **globular** exists with some gradual increase of energy following this series. Computational results presented here for $Ala_{12}Na^{+}$ identified the first few low-energy disruptions to the fully helical structure, which appear to raise the relative energy by approximately 10-20 kJ mol^{-1}.

Chapter 4

Proton and Sodium Bound Dimers of Poly(alanine) Peptides

4.1 Introduction

Results from the previous section illustrated the α-helical nature of sodiated poly(alanine) peptides in the $Ala_{8-12}Na^+$ range. Earlier studies in the literature, and REMD results in the current work, indicate that while $Ala_{12}Na^+$ is helical, $Ala_{12}H^+$ is not.[134] This is a result of protonation occurring at the N-terminal residue of the peptide, effectively destabilizing the helix. These observations suggest that other ions which interact with the C-terminal residue of an α-helical Ala_{8-12} peptide could also stabilize the structure. An especially interesting example would be the case in which the sodium cation was replaced by a second poly(alanine) peptide that was itself either protonated or sodiated. This complex could offer an additional stabilization to the α-helical structure by providing not only favourable electrostatic interaction with the macrodipole of the helix, but also additional hydrogen bonds between C=O and N-H groups of the two peptides. Comparison of sodiated and protonated hetero-dimers, such as $Ala_{12}Ala_6H^+$ and $Ala_{12}Ala_6Na^+$, could provide insight

into the nature of the interaction between globular and helical conformations (It is assumed that both $Ala_6(H/Na)^+$ ions are globular[134,139]). Additionally, the different sodiated and protonated hetero- and homo-dimers formed between peptides of Ala_6 and Ala_{12} would, in a general sense, be very illustrative of the relative stabilization provided by the interaction between a neutral poly(alanine) peptide and a second cationized poly(alanine) peptide.

During IRMPD experiments of sodiated poly(alanine) peptides (particularly $Ala_{12}Na^+$ (893 m/z)), ions formed by dimerization of smaller poly(alanine) peptides ($(Ala_6)_2H^+$, 889 m/z and $(Ala_6)_2Na^+$, 911 m/z) were observed.[i] The accidental observation of these ions, however, led to the question of whether other dimers could also be seen from the same solutions. After brief experimentation with various solutions and experimental parameters, it was clear that both hetero- and homo-dimers involving Ala_6 and Ala_8 were present after electrospraying the $Ala_{12}Na^+$ solution (this solution was described in detail in the previous chapter of sodiated(polyalanine) peptides). Furthermore, older solutions (those made a few days previously) produced higher intensities of dimers formed from smaller peptides (Ala_6, Ala_8) and lower intensities of the primary Ala_{12} ion. This observation is consistent with hydrolyzation of the larger peptides to smaller peptides in the highly acidic trifluoroacetic acid/trifluoroethanol (TFA/TFE) solvent matrix. In order to obtain dimers between smaller peptides (Ala_6) and larger peptides (Ala_{12}), additional "fresh" Ala_{12} solution was added to the older Ala_{12} solution. From these solutions the following dimer ions were obtained without addition of Na^+ or Ala_6: $(Ala_6)_2H^+$, $(Ala_6)_2Na^+$, $(Ala_{12}Ala_6)H^+$, $(Ala_{12}Ala_6)Na^+$, $(Ala_{12})_2Na^+$. IRMPD results have been obtained for $(Ala_6)_2H^+$.

The above names will be used in the remainder of this section to refer to the dimer ions, where $(Ala_{12}Ala_6)H^+$ implies that the ion is a proton-bound dimer of Ala_{12} and Ala_6 without specifying which peptide is protonated. The form $(Ala_6H^+)Ala_{12}$ would be used, for example, to specify that it is the Ala_6 chain that is protonated. Abbreviations are

[i]These dimer ions were typically much more intense than the $Ala_{12}Na^+$ (893 m/z) ion. Because of their larger intensity and the small mass differences, these ions caused significant difficulty while isolating $Ala_{12}Na^+$ (893 m/z) in the quadrupole.

used to refer to the calculated structures. For $(Ala_6H^+)Ala_{12}$, the form A12A6H_#### is used, in this case still implying that protonation occurs on the Ala_6 peptide and where the labels "####" are again taken from the corresponding REMD indices. For the proton-bound dimers, the labels "a" and "b" will be given to the protonated and neutral chains, respectively. Specification of the C-terminus of the protonated chain would then be labelled as "COOH(a)". The following colours will be associated with elements: **Carbon**, **Oxygen**, **Nitrogen** and **Sodium**.

4.1.1 Results from Previous Studies

Dimers of small alanine peptides have been studied previously using both molecular dynamics simulations and ion mobility measurements.[140] Mainly, the peptides included lysine residues at either the C- or N-terminus to control the conformation of the peptide. A C-terminal protonated lysine interacts favourably with the macrodipole of the helix, stabilizing the structure, while an N-terminal lysine has the opposite effect and favours globular structure. While the peptides examined in the work discussed in this chapter are purely alanine based, consideration of previous results provides some background and basis for discussion and comparison.

Previous studies found multiple low-energy structures for primarily helical, doubly charged dimers composed of two poly(alanine-glycine) peptides, each with a C-terminal lysine (Ac-$(GA)_7$K$\cdots$Ac-A$(GA)_7$K+$2H^+$).[140] These structures were taken from molecular dynamics simulations using the PROSIS package with CHARMM-like potentials. The lowest-energy conformation of this peptidic dimer was proposed to involve two fully helical monomers connected at the C-terminal residues, where each lysine side chain interacts with the C-terminus of the other chain. This conformation leaves an unfavourable interaction between the dipoles of the helices, resulting in a slightly bent conformation of the peptide chains rather than an orientation aligning the central axes of the two helices. An additional

low-energy conformation was found to have an overlapping collinear conformation where the C-terminal end of one peptide interacts with the N-terminal end of the other peptide, however this structure did not match well to their experimental results.[140]

4.2 Computational Results

The computational procedure used to obtain the results in this section was described in detail for the calculation of the sodiated poly(alanine) peptides discussed in the previous section, however no RI-B3LYP-D/def2-TZVPP//RI-B3LYP-D/def2-SVP (abbreviated here **B3LYP-D**, as before) jobs were completed for dimers and only a small selection of **AMOEBA** structures taken from the REMD results were calculated at the M06/6-311+g(d,p)//M06/6-31G(d,p) (abbreviated here **M06**, as before) level. The exclusion of the intermediate **B3LYP-D** calculations was justified by the improved relative accuracy of the **M06** results observed for the sodiated poly(alanine)s in the previous chapter. Additionally, the sizes of the larger dimers (up to $(Ala_{12})_2$ Na^+ - 247 atoms with 3458 basis functions at the **M06** level) become prohibitively large for a large-scale **B3LYP-D** survey as was done for the sodiated poly(alanine)s previously described (20-30 structures were selected from the ≈50 lowest-energy REMD structures and calculated at the **B3LYP-D** level for each peptide size). Due to the enormous computational cost involved in the treatment of such large systems, the computational procedure here was not intended to be as rigorous as the approach used with the poly(alanine)s of the previous chapter. This computational procedure is intended to locate families of the lowest energies structures, however, it is quite possible that the exact lowest-energy structures were not sampled by taking only the three lowest-energy **AMOEBA** structures from the REMD results.

REMD calculations presented in this chapter were completed according to the procedure of the previous chapter without deviation, however only two generations of simulations

were completed (the second using the lowest-energy structure of the first as the starting geometry). All **M06** calculations were completed using Gaussian09 in parallel over 8 CPUs with 15 Gb of memory and with job times of 1-3 weeks for optimizations, 3-7 days for frequencies and 1-3 days for single point energies.

Table 4.1 – Summary of sodiated and protonated poly(alanine) dimers considered at each level of computation. Number of calulated **AMOEBA** structures is for a single REMD simulation.

	AMOEBA	**M06**
$(Ala_6)_2H^+$	532	3
$(Ala_6)_2Na^+$	387	3
$Ala_6(Ala_{12}H^+)$	467	3
$(Ala_6H^+)Ala_{12}$	744	3
$(Ala_{12}Ala_6)Na^+$	166	3
$(Ala_{12})_2Na^+$	842	3

4.2.1 Comparison Between REMD and DFT

Comparison of **AMOEBA** structures and **M06** structures will be made in this section. The discussion from the previous chapter relating to the calculated structures of sodiated poly(alanine) peptides at different levels of theory can be directly applied to the dimers discussed in this chapter. Additionally, consideration of dimers provides comparison of the theoretical models for treatment of more loosely bound structures (two peptides held by intermolecular hydrogen bonds), as well as protonated peptides, in addition to sodiated peptides. In all cases, the **AMOEBA** structures have slightly larger gyration radii, with an increase of about 0.1-0.2 Å. Additionally, the RMSD between **AMOEBA** structures and their corresponding **M06** calculated structures are lower in the case of proton-bound dimers (0.4-0.5Å) than for sodium-bound dimers (0.7-1.0 Å). RMSD and gyration radii are presented in Table 4.2

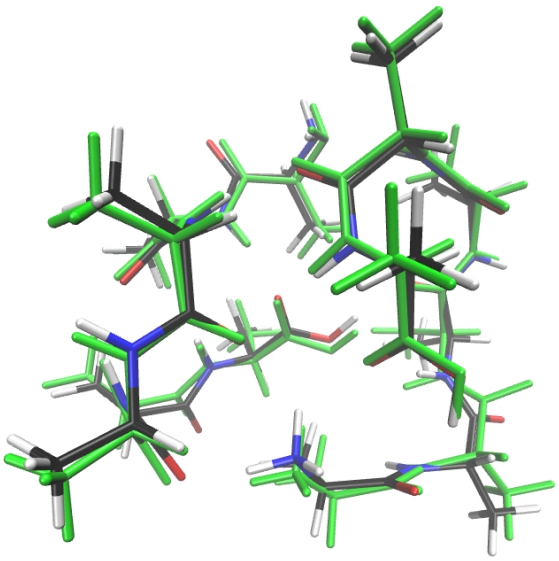

Figure 4.1 – $(Ala_6)_2H^+$ lowest-energy structure, 2A6H_1271. Overlay of calculated **M06** and **AMOEBA** structures.

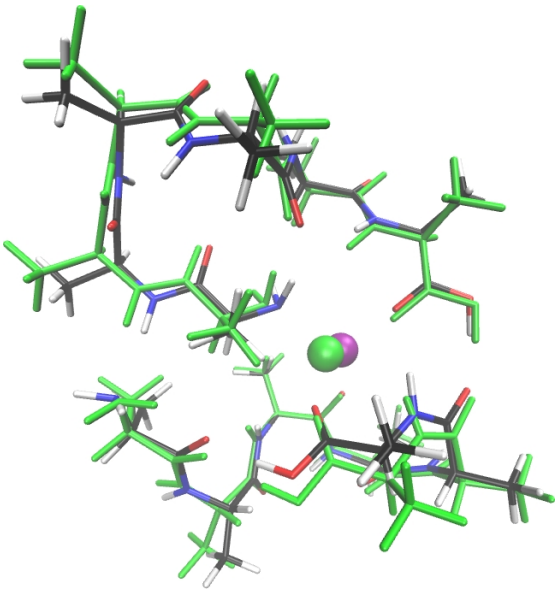

Figure 4.2 – $(Ala_6)_2Na^+$ lowest-energy structure, 2A6Na_3999. Overlay of calculated **M06** and **AMOEBA** structures.

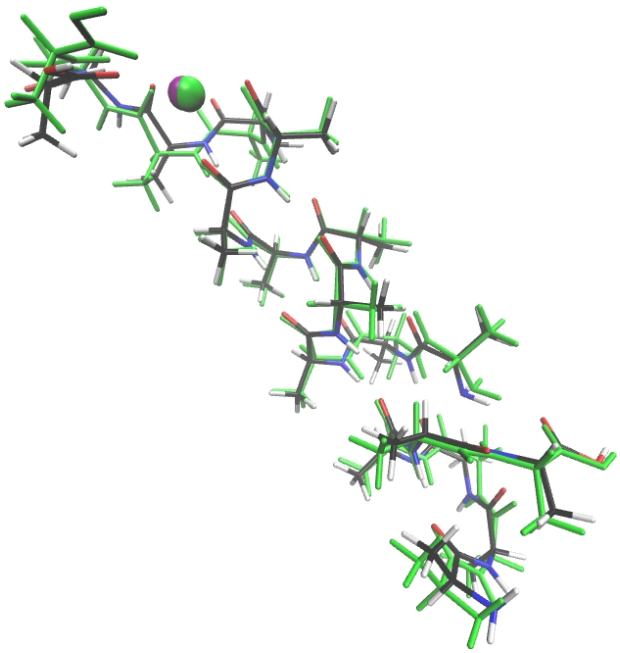

Figure 4.3 – $(Ala_6Ala_{12})Na^+$ lowest-energy structure, A12A6Na_4602. Overlay of calculated **M06** and **AMOEBA** structures.

Table 4.2 – Comparison of radius of gyration (rgyr) and RMSD at the **AMOEBA** and **M06** levels (Å) for the lowest-energy **M06** structure of each dimer.

RMSD	**AMOEBA/M06**	
2A6H_1271	0.41	
A6A12H_4568	0.53	
A12A6H_4371	0.40	
2A6Na_3999	0.68	
A6A12Na_4602	0.66	
2A12Na_6547	1.0	
rgyr	**M06**	**AMOEBA**
2A6H_1271	4.97	5.03
A6A12H_4568	8.38	8.43
A12A6H_4371	6.15	6.31
2A6Na_3999	5.55	5.75
A6A12Na_4602	8.02	8.11
2A12Na_6547	10.09	10.24

4.2.2 Results and Discussion

For each dimer the lowest-energy calculated structure is shown, where the peptide backbone is represented by a ribbon. As an aide in clarity and to distinguish the separate chains more clearly, different colours of ribbons are used for each peptide in the dimer. In cases where structural differences can be clearly and easily illustrated, RMSD-aligned overlays of the two structures are shown. In these overlays, the lowest-energy structure is represented with solid colours and an orange ribbon, while the higher energy structure is represented with slightly faded colours and a green ribbon. For $(Ala_6)_2Na^+$ (Figure 4.6), a different representation is used as the large structural differences could not effectively be represented with a simple overlay. All structural and spectroscopic results in the following sections refer to calculations at the **M06** level, unless otherwise noted.

$(Ala_6)_2H^+$

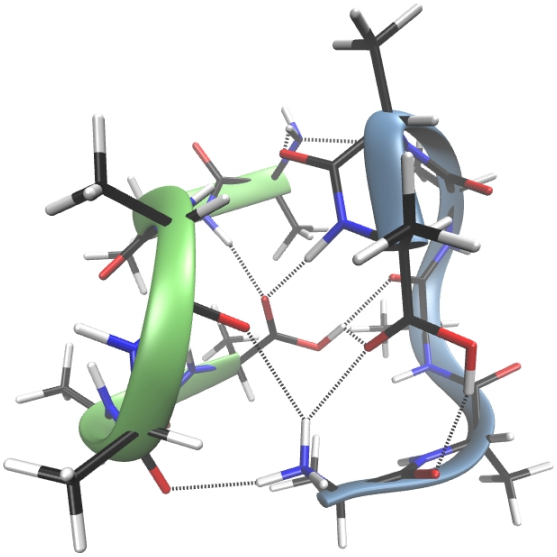

Figure 4.4 – Lowest-energy $(Ala_6)_2H^+$ structure, 2A6H_1271. (**M06**, free energy 298 K)

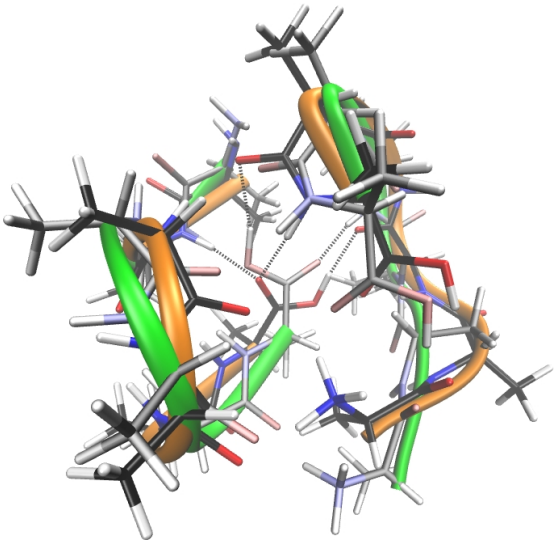

Figure 4.5 – Overlay of lowest-energy $(Ala_6)_2H^+$ structure, 2A6H_1271 (**orange**), with second-lowest-energy structure, 2A6H_0546 (lightly coloured structure, ribbon in **green**). (**M06**, free energy 298 K)

Table 4.3 – Relative electronic (E) and free energies (G, 298 K) for $(Ala_6)_2H^+$ (kJ mol^{-1}).

	M06		M06/6-31G(d,p)	
$\mathbf{(Ala_6)_2H^+}$	E	G	E	G
2A6H_1271	0.0	0.0	0.0	0.0
2A6H_0546	10.3	5.5	12.8	7.9
2A6H_0012	12.1	6.3	14.0	8.3

2A6H_1271 Structural Description

Shown in Figure 4.4 at the **M06** level, 2A6H_1271 is a protonated globular dimer where the two Ala_6 chains are bound together by seven intermolecular hydrogen bonds. The NH_3^+ of the protonated chain interacts with two neighbouring C=O (3b,4b) groups of the neutral chain in addition to the C-terminal C=O of its own chain. The C-OH groups of both the protonated and neutral chains interact with C=O groups (1a,3a respectively) of the protonated chain. Finally, the NH_2 of the neutral chain interacts with N-H(5a) and C=O(5a) of the protonated chain. Both C-OH groups are bound.

2A6H_0546 and 2A6H_0012 Structural Description

M06 calculated structures, 2A6H_0546 and 2A6H_0012 converged to nearly the same geometry (RMSD difference of 0.08 Å between each other and 1.61 and 1.69 Å, respectively, when compared to the lowest-energy structure, 2A6H_1271) and are found to be approximately 10 kJ mol^{-1} higher in both free and electronic energy than the lowest-energy structure, 2A6H_1271. The main structural difference comparing to the lowest-energy structure relates to a rotation of the C-terminus, shown centred in Figure 4.5. This approximately 180 degree rotation of COOH(b) of the neutral chain results in a separate 180 degree rotation of the chain at the fourth residue of the same chain, forming a HOC=O$\cdots$H-N interaction in place of a COOH$\cdots$O=C interaction. However, these rotations only minimally change the shape of the peptide chains, illustrated by comparison of the orange and green ribbons tracing the peptides' backbones.

$(Ala_6)_2Na^+$

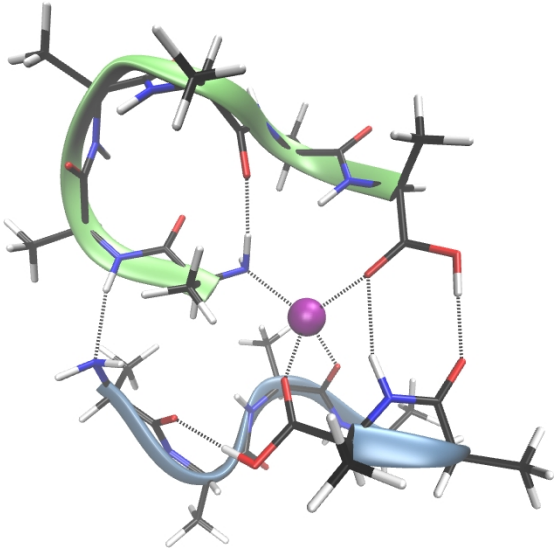

Figure 4.6 – Lowest-energy $(Ala_6)_2Na^+$ structure, 2A6Na_3999. (**M06**, free energy 298 K)

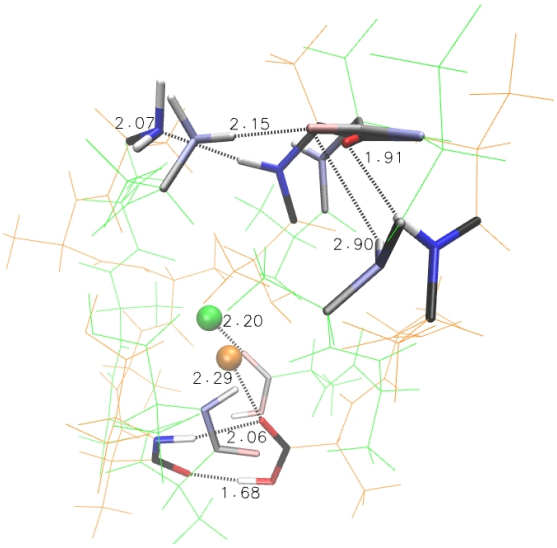

Figure 4.7 – Overlay of lowest-energy $(Ala_6)_2Na^+$ structure, 2A6Na_3999 (fully coloured/**orange** structure), with second-lowest-energy unique structure, 2A6Na_5238 (lightly coloured/**green** structure). (**M06**, free energy 298 K)

Table 4.4 – Relative electronic (E) and free energies (G, 298 K) for $(Ala_6)_2Na^+$ (kJ mol^{-1}).

	M06		M06/6-31G(d,p)	
$(Ala_6)_2Na^+$	E	G	E	G
2A6Na_3999	1.3	0.0	7.0	2.8
2A6Na_6199	0.0	1.1	7.5	0.9
2A6Na_5238	0.1	5.4	0.0	0.0

2A6Na_3999 and 2A6Na_6199 Structural Description

Calculated structures 2A6Na_3999 and 2A6Na_6199 converged to nearly the same geometry and are shown in Figure 4.6 where the RMSD calculated between structures is 0.1 Å. In contrast to the proton-bound dimer of Ala_6, where seven intermolecular hydrogen bonds held the two peptide chains together, the lowest-energy $(Ala_6)_2Na^+$ structure has only three intermolecular hydrogen bonds between the chains, due to coordination of the sodium cation. The three intermolecular interactions involve the C-terminus of the first

(on top, with green ribbon) peptide interacting with N-H(6) and C=O(5) of the second, in addition to the N-terminal nitrogen of the second peptide interacting with N-H(2) of the first peptide. The sodium cation is coordinated by the C-terminal C=O of both peptides, as well as the N-terminus of the first peptide and C=O(3) of the second. Both C-OH groups are bound.

2A6Na_5238 Structural Description

Structure 2A6Na_5238 is calculated to be only 2.8 kJ mol^{-1} higher in relative free energy at the **M06** level. Figure 4.7 illustrates the differences between these two conformations. Firstly, a shift of the N-terminus of one chain breaks an intramolecular $NH_2 \cdots$ H-N interaction (shown at 2.07 Å) while forming a new interaction with an adjacent C=O (2.15 Å). The formation of the interaction between the NH_2 group and the C=O disrupts a previously existing C=O$\cdots$ H-N interaction (shown in Figure 4.7), where the bond distance lengthens from 1.91 to 2.90 Å. Secondly, a shift at the C-terminus of the same chain breaks two interactions between the C-terminus and the other peptide chain, while forming a slightly tighter interaction between the C-terminus and the sodium cation.

$(Ala_6Ala_{12})H^+$

REMD simulations for $(Ala_6Ala_{12})H^+$ must be considered with protonation at both the N-terminus of Ala_{12} and the N-terminus of Ala_6 (since the binding of the proton is specified and its connectivity is not variable within a single simulation). Calculations of the three lowest-energy **AMOEBA** structures, for both $Ala_{12}(A_6H^+)$ and $Ala_6(A_{12}H^+)$, indicate that the lowest-energy structure with protonation of Ala_6 is more favourable by 50.9 kJ mol^{-1} in relative electronic energy and 32.6 kJ mol^{-1} in relative free energy (298 K) at the **M06** level.

Table 4.5 – Relative Electronic (E) and free energies (G, 298 K) for $(Ala_6Ala_{12})H^+$ (kJ mol^{-1}).

	E	G
HA12A6_4568	50.9	32.6
HA6A12_4371	0.0	0.0

$Ala_6(Ala_{12}H^+)$

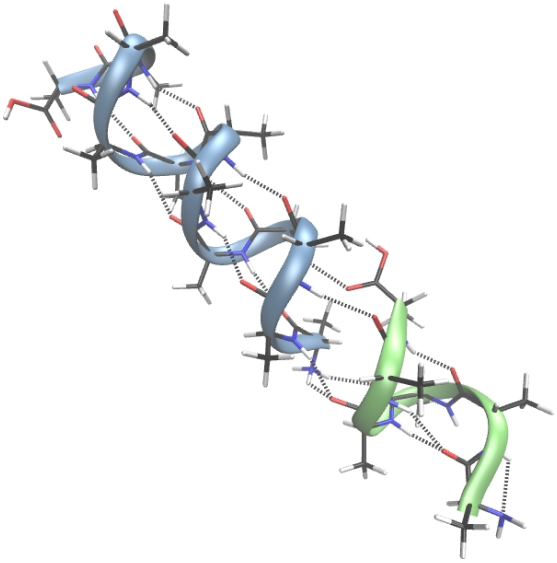

Figure 4.8 – Lowest-energy $Ala_6(A_{12}H^+)$ structure (**M06**, free energy 298 K), HA12A6_4568.

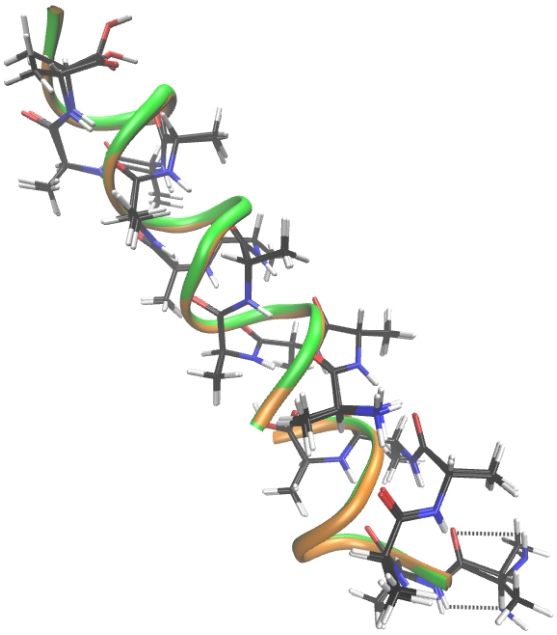

Figure 4.9 – Overlay of lowest-energy $Ala_6(A_{12}H^+)$ structure (**M06**, free energy 298 K), HA12A6_4568 (**orange**), with second-lowest-energy unique structure, HA12A6_8923 (lightly coloured structure, ribbon in **green**).

HA12A6_4568 and HA12A6_5406 Structural Description

Calculated structures HA12A6_4568 and HA12A6_5406 converged very nearly to the same structure during optimization at the **M06** level, with a difference in free energy of 6.4 kJ mol^{-1} and an RMSD of 0.41 Å between structures. While the Ala_{12} chain is protonated (previously discussed to prevent helical structure by unfavourable interaction

Table 4.6 – Relative electronic (E) and free energies (G, 298 K) for $Ala_6(Ala_{12}H^+)$. (kJ mol^{-1})

	M06		M06/6-31G(d,p)	
$Ala_6(Ala_{12}H^+)$	E	G	E	G
HA12A6_4568	0.9	0.0	1.6	0.0
HA12A6_5406	0.0	6.4	0.0	7.2
HA12A6_8923	5.4	20.0	9.9	25.3

with the macrodipole of the helix), interaction with the neutral Ala_6 somewhat disrupts the interaction of the NH_3^+ with the dipole of the helix. This is accomplished, apparently, by the nearly ideal continuation of helical-type hydrogen bonds and backbone structure between the $Ala_{12}H^+$ and Ala_6 chains, giving the appearance of an Ala_{18} helix and placing the charge away from the centre of the structure.

HA12A6_8923 Structural Description

While calculated structure HA12A6_8923 has an RMSD with the lowest-energy structure, HA12A6_4568, of 0.76 Å, the difference in free energy at the **M06** level is 20.0 kJ mol^{-1}. This is mainly the result of a rotation of the N-terminal NH_2 group of the Ala_6 chain interrupting the interaction between the nitrogen of the NH_2(1b) and the N-H(2b). In its place an interaction forms between one proton of NH_2(1b) and C=O(1b). Despite this modification of the structure, the backbone of the helix continues through both chains relatively undisturbed, as illustrated by comparison of the orange and blue ribbons in Figure 4.9.

$(Ala_6H^+)Ala_{12}$

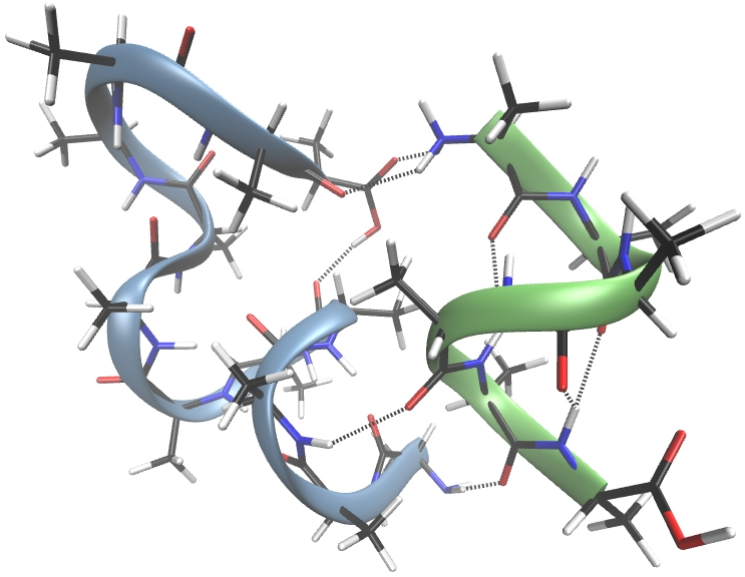

Figure 4.10 – Lowest-energy $(Ala_6H^+)Ala_{12}$ structure, HA6A12_4371. (**M06**, free energy 298 K)

Table 4.7 – Relative electronic (E) and free energies (G, 298 K) for $(Ala_6H^+)Ala_{12}$ (kJ mol^{-1}).

	M06		M06/6-31G(d,p)	
$(Ala_6H^+)Ala_{12}$	E	G	E	G
HA6A12_4371	0.4	0.0	0.0	0.0
HA6A12_8463	0.0	0.5	0.0	0.8
HA6A12_7351	0.4	1.0	0.0	1.0

HA6A12_4371, HA6A12_7351 and HA6A12_7351 Structure Description

Calculated structures HA6A12_4371, HA6A12_7351 and HA6A12_7351 all converged to very similar geometries during **M06** optimization, with an RMSD of less than 0.5 Å calculated between them. Although Ala_6 is protonated, interaction of the C-terminus of Ala_{12} with the protonated N-terminus (NH_3^+) still allows formation of helical structure in Ala_6H^+. Additionally, the neutral Ala_{12} peptide itself is of partial (or distorted) helical structure, however in this case is found laterally aligned to the Ala_6 peptide rather than forming an extended helix as in the previously discussed example of $Ala_6(A_{12}H^+)$.

$(Ala_6Ala_{12})Na^+$

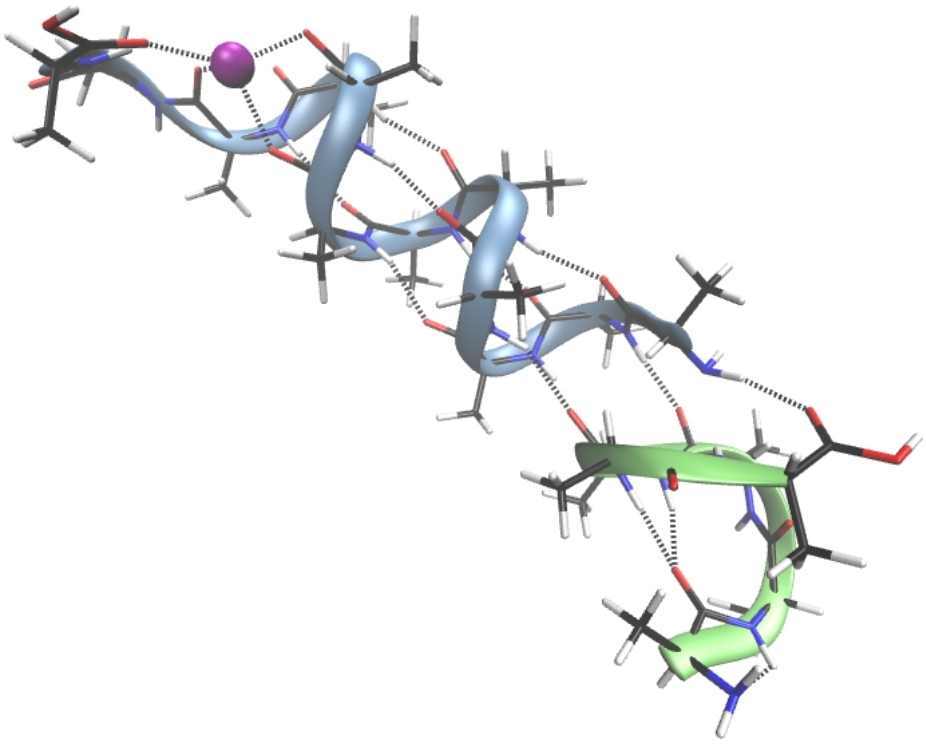

Figure 4.11 – Lowest-energy $(Ala_{12}Ala_6)Na^+$ structure, A12A6Na_4602. (**M06**, free energy 298 K)

Table 4.8 – Relative electronic (E) and free energies (G, 298 K) for $(Ala_{12}Ala_6)Na^+$ (kJ mol^{-1}).

	M06		M06/6-31G(d,p)	
$(Ala_{12}Ala_6)Na^+$	E	G	E	G
A12A6Na_4602	0.4	0.0	0.0	0.0
A12A6Na_1253	0.0	1.6	0.1	2.1
A12A6Na_0168	0.0	1.7	0.1	2.1

A12A6Na_4602, A12A6Na_1253 and A12A6Na_0168 Structural Description

Computed structures A12A6Na_4602, A12A6Na_1253 and A12A6Na_0168 did not result in unique structures after optimization at the **M06** level, having both negligible differences in relative free energy (less than 2.0 kJ mol^{-1}) and in RMSD (not more than

0.29 Å). Although the Ala_6 peptide appears to overall be globular, the side interacting with the helical A_{12} peptide follows the contour of the first four (N-terminal) residues of the helix. Here, four α-helical type hydrogen bonds are formed between the two peptides and a further two helical type hydrogen bonds are formed within Ala_6 itself (between C=O(1)$\cdots$N-H(5,4) and C=O(2)$\cdots$N-H(6)). The sodium cation causes some disruption of the helical structure at the C-terminal end of the Ala_{12} peptide, where it is coordinated by C=O groups of residues 12, 11, 9 and 8. Both C-terminal C-OH groups are free as well as the N-terminal NH_2 of Ala_6.

$(Ala_{12})_2Na^+$

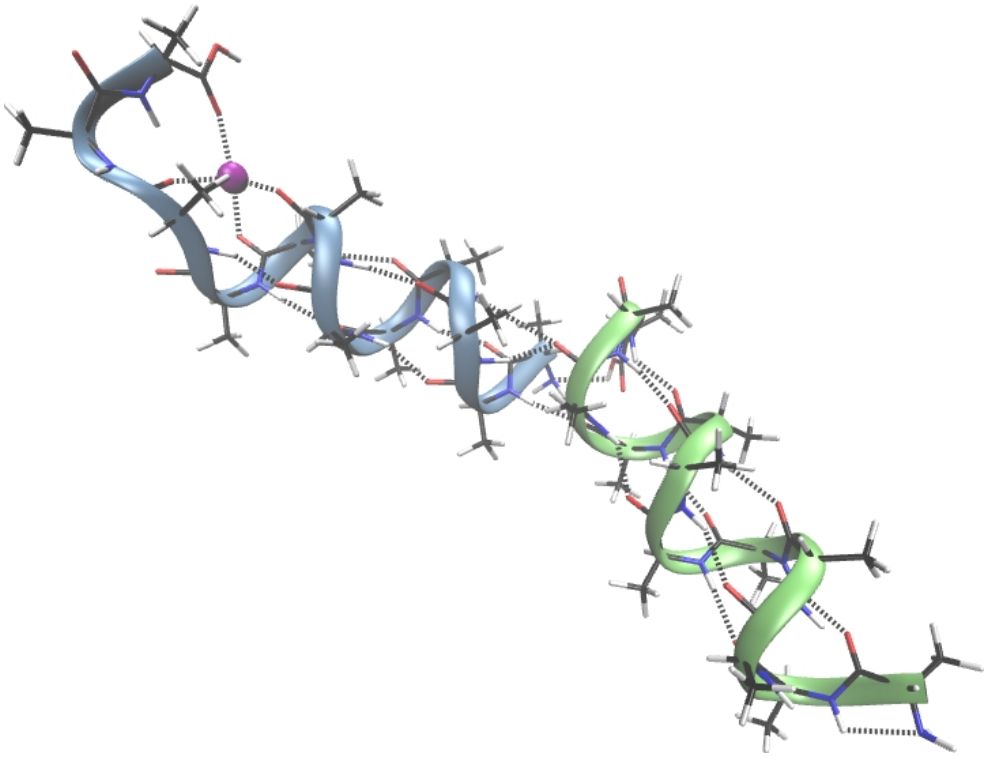

Figure 4.12 – Lowest-energy $(Ala_{12})_2Na^+$ structure, 2A12Na_6547. (**M06**, free energy 298 K)

2A12Na_6547 Structural Description

Calculated structure 2A12Na_6547 is the lowest-energy structure calculated at the

Table 4.9 – Relative electronic (E) and free energies (G, 298 K) for $(Ala_{12})_2Na^+$ (kJ mol^{-1}).

	M06		M06/6-31G(d,p)	
$(Ala_{12})_2Na^+$	E	G	E	G
2A12Na_6547	0.0	0.0	0.0	0.0
2A12Na_7209	2.0	8.0	1.4	7.5
2A12Na_3527	2.0	8.0	1.4	7.5

M06 level by 8.0 kJ mol^{-1} in relative free energy. Both Ala_{12} peptides in this dimer are helical and the connection of the two peptides occurs through the continuation of the helical "backbone" from one peptide through the second. The sodium cation interacts with the Ala_{12} peptide having a free C-terminus and is coordinated by C=O groups of residues 12, 11, 9 and 8. Hydrogen bonds are formed connecting the two peptides in the common α-helical manner and between residues on different peptides (involving C=O and N-H groups on residues four times separated). This forms a structure which somewhat resembles an α-helix of not two separate 12 residue peptides, but a single 24 residue peptide. However, the peptides do not align at 180 degrees, as they did in $Ala_6(Ala_{12}H^+)$, and instead have an angle between them of 130 degrees.

2A12Na_7209 and 2A12Na_3527 Structural Description

Calculated structures 2A12Na_7209 and 2A12Na_3527 are nearly identical and only slightly different from the lowest-energy structure 2A12Na_6547 (Relative free energy difference of 8.0 kJ mol^{-1} and RMSD of 0.29 Å). These structures differ from the lowest-energy structure by a small translation of the sodium-bound C-terminal C=O group.

4.3 Experimental Results

The experimental apparatus was described in the previous chapter and in the introduction of the text.[26,28,29,36,42,55,56,61] Both fingerprint and 3000-3600 cm^{-1} region experiments did

not use the CO_2 laser for fragmentation enhancement.

4.3.1 Fragmentation

IRMPD of $(Ala_2)_6H^+$ resulted in the following fragment ions listed in Table 4.10 below:

Table 4.10 – List of IRMPD fragment ions for $(Ala_6)_2H^+$ (parent ion - 889 m/z).

CLIO		OPO	
889	$(Ala_6)_2H^+$ (Parent)	889	$(Ala_6)_2H^+$ (Parent)
↓		↓	
445	Ala_6H^+	445	Ala_6H^+
356	b^5		

For both experiments, the dominant fragment ion (the only fragment for the OPO experiment) was Ala_6H^+ (445 m/z) resulting from the loss of neutral Ala_6. In the fingerprint region, the higher power of the FEL was sufficient to result in some fragmentation of one of the peptide chains itself, giving the b^5 fragment ion (356 m/z).

4.3.2 Spectral Results and Discussion

$(Ala_6)_2H^+$ 3000-3600 cm^{-1} Region Spectral Assignments and Discussion

Three scans in this region are shown in Figure 4.13, with the red scan extending from 3200-3450 cm^{-1}, the blue from 3300-3400 cm^{-1} and an additional scan (intensities not seen above the noise level) from 3450-3600 cm^{-1}. Each scan was completed using optimized laser intensity for the given region, with higher intensities applied above and below the 3300-3400 cm^{-1} region. The free N-H stretching modes to the blue of the main (primarily bound) N-H stretching band between 3300 and 3400 cm^{-1} were unobserved during these experiments. Additionally the calculated band at 3238 cm^{-1} assigned to the NH_3^+ group is

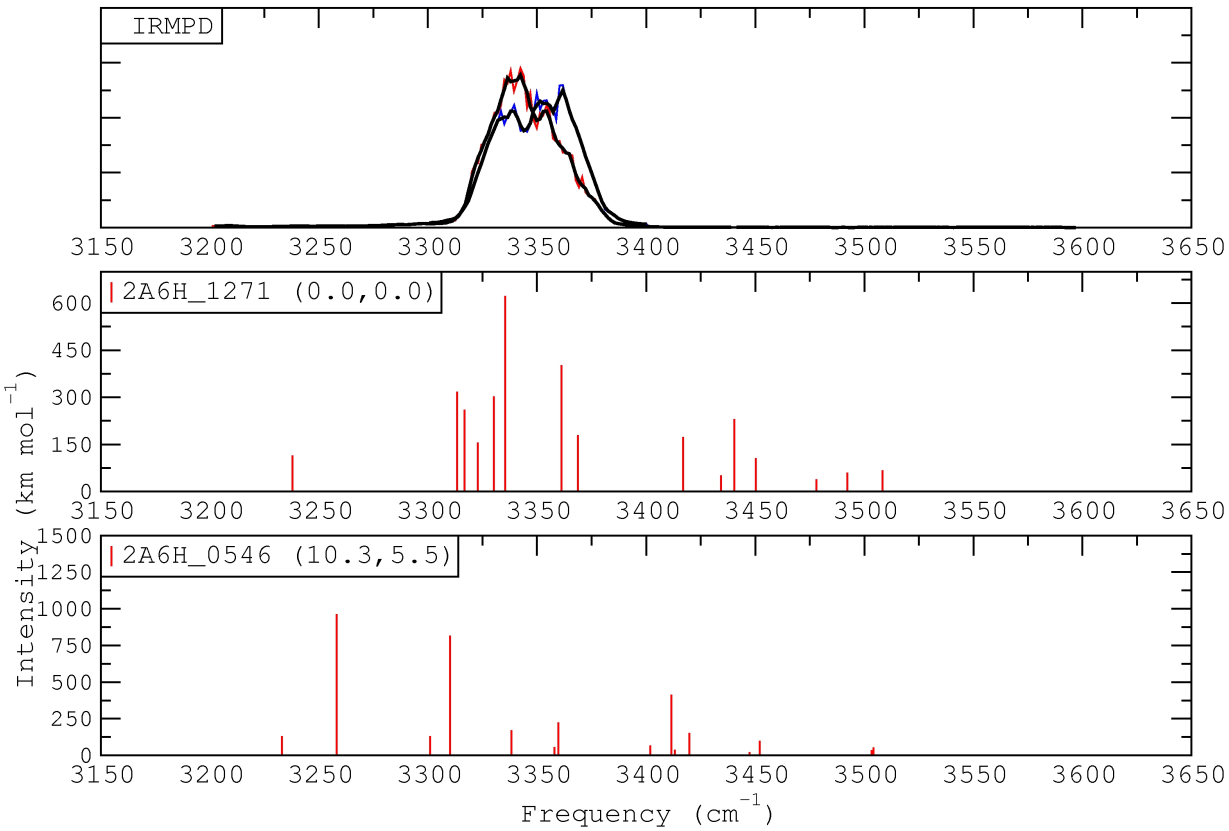

Figure 4.13 – $(Ala_6)_2H^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **M06**, frequencies are scaled by 0.96

unobserved. The absence of these peaks in the experimental spectrum is likely the result of insufficient laser power relative to the lower intensities of these modes. The bound N-H stretching band matches well to experiment in both position and shape, although spectral resolution is not sufficient to discern individual peaks. Consistently with both calculated structures, no free C-OH band (≈3575 cm^{-1}) was observed in the experimental spectrum.

Table 4.11 – $(Ala_6)_2H^+$ 3000-3600 cm^{-1} region M06/6-31G(d,p) scaled (0.96) vibrational assignments (cm^{-1}.

Freq	**Mode**	**Freq**	**Mode**
	2A6H_1271		**2A6H_0546**
3238	NH_3^+ (a)	3233	NH_3^+ (a)
3313	NH_2(b)+NH(3b,5a)	3258	C-OH(a)
3317	NH(3b,5a)	3301	NH(6a,3b)
3323	NH_3^+(a)+C-OH(a)	3310	NH(6a,3b)
3330	NH_3^+(a)+C-OH(a)	3338	NH(2a)
	NH_2(b)+NH(3b,5a,6a)	3358	NH_2(b)
3335	NH_3^+(a)+C-OH(a)	3360	NH(3a)
	NH_2(b)+NH(3b,5a,6a)	3402	NH(5b,6b)
3361	NH(6a)	3412	NH(5b,6b)+NH_3^+(a)
3369	NH(5b)	3413	NH(5b,6b)+NH_3^+(a)
3417	NH(3a)	3420	NH(4a)
3434	NH_2(b)(asym)	3448	NH_2(b) (asym)
3441	NH(2b)	3452	NH(5a)
3450	NH(6b)	3503	NH(2b)
3478	NH(2a)	3504	NH(4b)
3492	NH(4b)		
3508	NH(4a)		

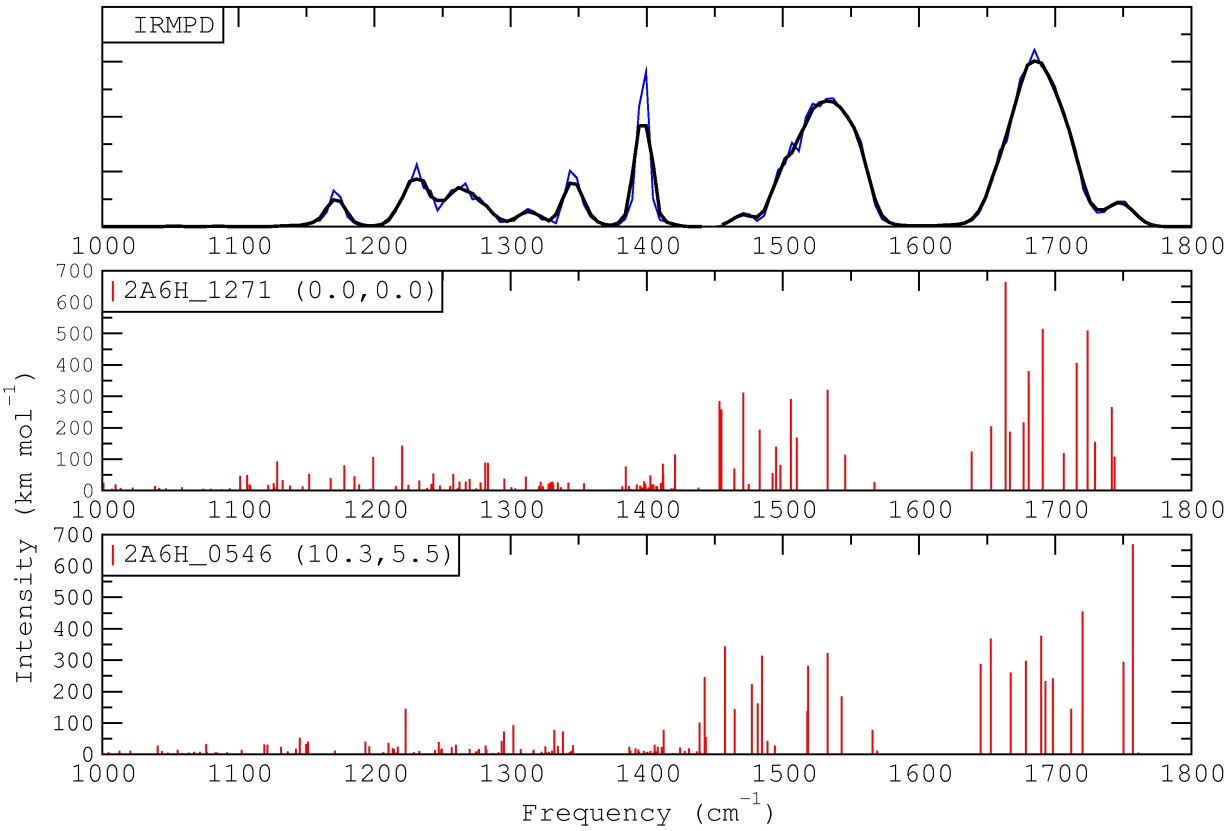

Figure 4.14 – $(Ala_6)_2H^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are scaled arbitrarily. (kJ mol^{-1}, **M06**, frequencies are scaled by 0.95

$(Ala_6)_2H^+$ Fingerprint Region Spectral Assignments and Discussion

The C=O modes of the amide I band farthest to the blue in the calculated spectra are distinct between the two calculated results, where for 2A6H_0546 there is a blue shift and separation between the other C=O modes. While both represent vibration of the C-terminal C=O groups, in 2A6H_0546 the terminal C=O modes are separated by 30 cm^{-1} from the first C=O mode of the main band, while for 2A6H_1271 the terminal C=O modes are largely convoluted with the main band. The line positions of these bands in the calculated spectrum for 2A6H_1271 match slightly better than the calculated positions for 2A6H_0546. For 2A6H_0546, the amide II band is broadened in comparison to 2A6H_1271

and extends continuously from 1400 cm^{-1} to approximately 1575 cm^{-1}. This is inconsistent with the gap that exists in the experimental spectrum between approximately 1425 and 1450 cm^{-1}. The amide II band of 2A6H_1271 is more closely in agreement with the experimental band, however a blue shift of approximately 15 cm^{-1} would be required for ideal matching of the calculated and experimental bands. Bands below 1400 cm^{-1} are in fairly consistent agreement, however due to the large number of bands in this region it becomes difficult to distinguish the two calculated spectra in comparison with the experimental result.

Table 4.12 – $(Ala_6)_2H^+$ fingerprint region M06/6-31G(d,p) scaled (0.95) vibrational assignments.

Freq	**Mode**	**Freq**	**Mode**
2A6H_1271		**2A6H_0546**	
1000	CH_3,CH,NH,CN	1000	CH_3,CH,NH,CN
1469	↓	1454	↓
1468	NH(2b)	1458	NH(4b)
1470	NH(5,4b)	1459	NH(6a)
1480	NH(5,4b)	1473	NH(5a)
1486	NH(4a,3a,2a)	1480	NH(2b)
1490	NH(6a)	1493	NH(4a)
1499	NH(6b)	1497	NH(5b)
1508	NH(5a,3a)	1500	NH(2a)
1511	NH(5a,3a)	1504	NH(6a),COOH
1514	NH(4a)	1510	NH(6a),COOH
1522	NH(2a)	1534	NH(3a,3b)
1526	NH(5b)	1534	NH(3a,3b)
1549	NH(3b)	1549	NH(6b),NH_3^+ (a)
1562	NH_3^+ (a)	1560	NH(6b),NH_3^+ (a)
1584	NH_2(b)	1582	NH_3^+ (a)
1656	NH_3^+ (a)	1586	NH_2(b)
1671	C=O(4b,3b)	1663	C=O(5b)
1681	C=O(3a)	1670	C=O(5a)
1684	C=O(5a,2b)	1685	C=O(2b)
1694	C=O(2b)	1696	C=O(4a)
1698	C=O(1a)	1708	C=O(1b)
1709	C=O(5a,3b)	1711	C=O(1a)
1724	C=O(4a)	1716	C=O(4b)
1734	C=OOH(a)	1730	C=O(2a)
1742	C=O(5b)	1738	C=O(3b)
1747	C=O(1b)	1769	C=OOH(a)
1760	C=O(2a)	1776	C=O(3a),C=OOH(b)
1762	C=OOH(b)	1780	C=O(3a),C=OOH(b)

4.4 Discussion of the Formation of Helical Structures and Conclusions

Table 4.13 – Summary of **M06** structural features for poly(alanine) protonated and sodiated dimers. H-bonds refers to the number of hydrogen bonds connecting monomer units.

Dimer	Structure	H-bonds*	Na^+ - # coord
$(Ala_6)_2H^+$	globular	7	-
$(Ala_6)_2Na^+$	globular	3	4
$Ala_6(Ala_{12})H^+$	fully helical	6	-
$Ala_{12}(Ala_6)H^+$	Ala_6 - helical/Ala_{12} partially helical	4	-
$(Ala_6Ala_{12})Na^+$	Ala_{12} - helical/A_6 partially helical	4	4
$(Ala_{12})_2Na^+$	fully helical	5	4

A summary of general structural properties is provided in Table 4.13. $(Ala_6)_2H^+$ and $(Ala_6)_2Na^+$ are both found to have globular lowest-energy structures, consistent with the assumption that the protonated and sodiated monomers of Ala_6 would also be globular (previous studies suggest that $Ala_{<20}H^+$ and possibly higher are globular[134]). For $(Ala_6)_2H^+$ and $(Ala_6)_2Na^+$ it can then be suggested that the interaction of Ala_6H^+ or Ala_6Na^+ with neutral Ala_6 is not sufficient to reduce the relative energy of the helical conformation below that of the globular conformation. In the globular conformations, intermolecular and intramolecular interactions between the peptides (for $(Ala_6)_2Na^+$, also coordination of the sodium by both peptides) are favoured over intramolecular helical-type hydrogen bonds. In the globular conformation of $(Ala_6)_2Na^+$, both C-terminal COOH groups and one NH_2 group coordinate the sodium cation, which would not be possible in the helical conformation.

Of all dimers discussed, $Ala_6Ala_{12}H^+$ is calculated to have the most ideal α-helical structure. Here, interaction of the protonated NH_3^+ group with the C-terminal C=O of the neutral A_6 chain places the positive charge somewhat on the exterior of the helix (since it is involved in two helical-type hydrogen bonds in the same way as the amidic N-H groups

in the structure, with the "additional" proton pointing outwards) and allows for nearly perfect continuation of the helical structure through both peptides. This conformation appears to limit the unfavourable electrostatic interaction of the positive charge with the macrodipole of the helix, but is still 32.6 kJ mol^{-1} higher in free energy (298 K) at the **M06** level than the lowest-energy conformation of the dimer formed from protonated Ala_6 and neutral Ala_{12}. In the lowest-energy conformation of $Ala_{12}Ala_6H^+$, the protonated Ala_6 chain is calculated to be fully helical, while the neutral Ala_{12} chain has only partial helical character. Isolated neutral Ala_{12} is calculated by **AMOEBA** to be fully helical and it is predictable that interaction of protonated Ala_6 with the C-terminus of neutral Ala_{12} would only increase the favourability of the helical structure. However, interactions between the N-terminus of one peptide and C-terminus of the other peptide is only possible at both ends of the peptides if the neutral Ala_{12} chain breaks its fully helical structure. If Ala_{12} remained fully helical, C- and N-terminal hydrogen bonds, and others, would only exist at one end of the helical Ala_6 peptide, forming an apparently higher energy conformation.

Comparison of $(Ala_6Ala_{12})Na^+$ and $Ala_6(Ala_{12}H^+)$ provides an interesting illustration. For N-terminally protonated Ala_{12}, the helical structure is destabilized by the presence of the positive charge unfavourably interacting with the helical macrodipole. This makes interaction of C=O groups of Ala_6 with the NH_2 and N-H groups of Ala_{12} more favourable and results in the full incorporation of Ala_6 into the helical structure of Ala_{12}. In sodiated Ala_{12}, the macrodipole of the helix is stabilized by the interaction of Na^+ at the C-terminal end of the peptide, making its interaction with the neutral Ala_6 at the N-terminal end weaker. It then becomes more favourable for neutral Ala_6 to form strong hydrogen bonds, both within its own structure and with the N-termninal end of the sodiated Ala_{12}, and is possibly the reason that it adopts a structure which is not truly helical.

The structure of $(Ala_{12})_2Na^+$ is a somewhat clear case, where both monomer units of the peptide are calculated to be independently helical. The sodiated Ala_{12} is stabilized at

both the C-terminus by the sodium cation and at the N-terminus by hydrogen bonds with the neutral Ala_{12}. This suggests that the stabilizing interaction with the macrodipole of the helix is somewhat shared by interactions at both ends of the sodiated Ala_{12}. Although $Ala_{12}Na^+$ is calculated to be completely helical through all 12 residues, the more complete coordination of the sodium possibly occurs because the charge-dipole interaction between the macrodipole of the helix and the sodium cation is weaker as a result of the interaction between the C-terminus of the second Ala_{12} and the N-terminus of the sodiated Ala_{12}. A similar reasoning can be applied to explain the stronger hydrogen bonding connecting the N-terminus of the sodiated Ala_{12} with the C-terminus of the neutral Ala_{12} and can possibly be applied to justify the bent conformation of the two helices. This is in contrast to $Ala_6(Ala_{12}H^+)$, where the two peptides connect at an angle of 180 degrees (see Figure 4.8), possibly favouring the interaction of the two marcrodipoles in absence of other stabilizing interactions.

Chapter 5

Guanidinium Chloride Based Ionic Clusters

5.1 Introduction

Guanidinium is a highly resonance-stabilized cation resulting from protonation of the strong base guanidine (pK_a 13.6) and forms the chloride salt guanidinium chloride (also referred to as guanidine hydrochloride), shown in Figures 5.1, 5.2 and 5.3 below.

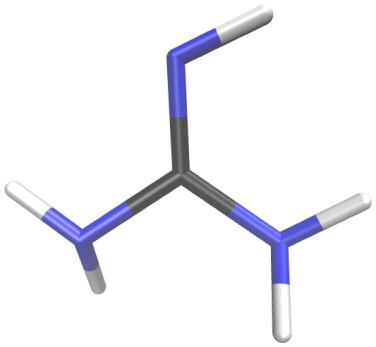

Figure 5.1 – Guanidine

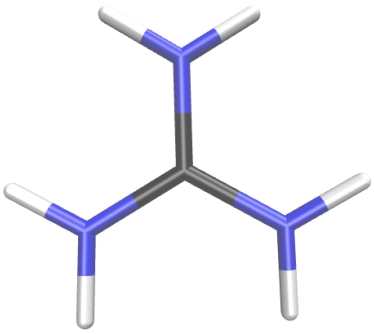

Figure 5.2 – Guanidinium ion

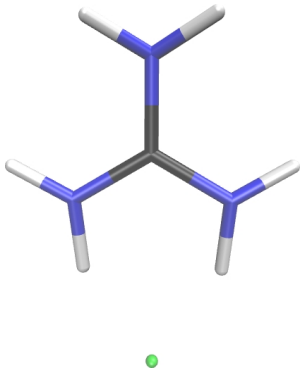

Figure 5.3 – Guanidinium chloride

Guanidinium has significant relevance to biochemical systems. As the side-chain of arginine, guanidine imparts strong basicity, making arginine the most basic amino acid with a pK_a of 12.1. Arginine is an essential amino acid and plays a key role in many biochemical processes such as the urea cycle, creatine production and in the synthesis of agmatine and proteins.[141] Additionally, guanidinium and guanidinium based ions influence and pass through sodium channels.[142] Guanidinium-based toxins tetrodotoxin (TTX) and saxitoxin (STX) are found in many aquatic animals (produced by symbiotic bacteria) and are toxins of extremely high potency due to their function as blocking ligands of voltage-gated Na^+ channels of nerve-cell membranes.[142,143] Guanidinium chloride is also one of the strongest and most widely studied and used protein denaturants, where in a six molar solution nearly all proteins cease to exhibit well ordered secondary structure.[144,145] It has been proposed that the mechanism by which guanidinium chloride causes the denaturation of proteins does not centre around altering properties of the solution, but rather through interaction with the protein itself and especially with regions of α-helical structure.[146]

Guanidinium chloride is also significant as an ionic liquid, with a melting point of 182.3 degrees C[i]. Ionic liquids have potentially tremendous applications as powerful solvents and electrically conducting liquids. As non-molecular solvents, they exhibit low vapour pressures, stability at high temperatures, and are miscible with both polar and non-polar solvents. Most often, as is the case for guanidinium chloride, an organic cation is associated with an inorganic anion.[147] Many guanidinium based ionic liquids have been reported in the literature, employed over a wide range of applications.[148] Some studies have discussed the bulk structure of these guanidinium-based compounds using mainly molecular modelling and dynamics.[149–151] In various guanidinium-based systems they calculated each cation to be coordinated by five to seven anions and the first solvation shell around a given cation to be approximately 5-8 Å from the centre of mass of the cation.

[i]A melting point below 100 degress C is loosely used as the criteria for a salt to be considered an ionic liquid. Nevertheless, 182.3 degrees C is a very low melting point for an ionic compound, and even though guanidinium chloride itself does not meet this definition, many guanidinium based salts do.

The guanidinium ion (Figure 5.2) has six N-Hs that could potentially interact with an anion, allowing for a potentially wide range of gaseous ionic clusters to be formed. The work presented in this chapter involves a computational and spectroscopic survey of a series of guanidinium chloride based gaseous ions. This series of ions differs sequentially by one guanidinium chloride neutral (as shown below), starting with x=1 for both $[\mathrm{Guan}_x\mathrm{Cl}_{(x-1)}]^+$ and $[\mathrm{Guan}_x\mathrm{Cl}_{(x+1)}]^-$.

$$[\mathrm{Guan}_x\mathrm{Cl}_{x-1}]^+ + [\mathrm{GuanCl}] \rightarrow [\mathrm{Guan}_{x+1}\mathrm{Cl}_x]^+$$

and

$$[\mathrm{Guan}_x\mathrm{Cl}_{x+1}]^- + [\mathrm{GuanCl}] \rightarrow [\mathrm{Guan}_{x+1}\mathrm{Cl}_{x+2}]^-$$

Some discussion has existed regarding the planarity of the guanidinium ion in various environments.[152,153] Previous computational studies have examined the orientation of NH_2 groups in guanidinium (their rotation out of the CN_3 plane) and found that rotation by -15 to +15 degrees of the NH_2 groups producing "right- or left-handed propellers" (rotation of the H-N-C-N dihedral) was energetically negligible.[153] Rotations by 15, 10, 5, -5, -10, -15 degrees all resulted in structures that are minima on the potential energy surface, while the exactly planar structure is a transition state with a barrier of 1.0 kJ mol^{-1}.[153] At the RI-B3LYP-D/def2-SVP level, the calculated minimum-energy conformation has an H-N-C-N dihedral angle of 4.2 degrees (referring to the nearly planar angle of rotation of the NH_2 groups out of the CN_3 plane).

5.2 Computational Methodology and Discussion

From a computational perspective, this study is especially challenging as the potential energy surface is expected to be both particularly flat and complex. Additionally, the

number of heavy atoms increases rather quickly with increasing cluster size and necessitates the use of efficient computational procedures. The methodology applied to the sodiated poly(alanine) peptides of the preceding chapter is well suited to the study of these guanidinium chloride based clusters. REMD simulations are appropriate for surveying the potential energy surface, while DFT calculations can be used to refine the selection of structures and to obtain vibrational results. The polarizable **AMOEBA** force field is well equipped to deal with the high density of electric charges present in these ions. Force field parameters for the side chain of arginine were taken for guanidinium, using the parameters describing the carbon and NH_2 of the guanidinium group. As well, optimized parameters for the chloride ion are provided in the standard AMOEBA parameter sets.

The set of geometries obtained from REMD simulations was analyzed and sorted to obtain a selection of low-energy structures to further consider using DFT. Details of the computational procedure used was thoroughly described in the introductory and previous chapter. Similar notation will be used as in the previous sections with the same labels referring to each level of calculation.

(AMOEBA) REMD structures optimized using the AMOEBA force field

(B3LYP-D) RI-B3LYP-D/def2-TZVPP//RI-B3LYP-D/def2-SVP [ii]

(M06) M06/6-311g+(d,p)//M06/6-31g(d,p)

The RI-B3LYP-D/def2-SVP level of theory was used to obtain optimized geometries, vibrational frequencies and thermochemical properties for the $[Guan_{(2-6,9,10,15)}Cl_{(1-5,8,9,14)}]^+$ and $[Guan_{(2-6,10,15)}Cl_{(3-7,11,16)}]^-$ series of ionic clusters (frequencies calculated without the RI approximation)[iii]. Electronic energies at the RI-B3LYP-D/def2-TZVPP level were considered for the RI-B3LYP-D/def2-SVP geometries. Finally, for validation of the chosen

[ii]Frequencies were obtained without the use of the RI approximation.

[iii]Notational forms $[Guan_5Cl_4]^+$, $Guan_5^+$, and G_5^+ refer to the same ion and will have corresponding calculated structures Guan5P_####, where P refers to positive and would be replaced with N if it was referring to the negative ionic clusters.

theoretical models, $[Guan_2Cl]^+$, $[Guan_5Cl_4]^+$ and $[Guan_5Cl_6]^-$ were calculated at the **M06** and **MP2** levels and anharmonic vibrational frequencies were calculated at the B3LYP/6-311+G(d,p) level for $[Guan_2Cl]^+$. In calculated structures the following colours are associated with elements: **Carbon**, **Nitrogen** and **Chlorine**.

5.2.1 Preliminary DFT Results

Preliminary **B3LYP-D** calculations were completed for the smaller sized cationic cluster-ions using manually defined starting geometries. However, as the cluster size increased it became apparent that optimized structures based on manually constructed input structures were still relatively far from the lowest-energy structures on the PES. This was seen by large inconsistencies in binding energies between sizes (energy associated with the sequential addition of one neutral guanidinum chloride), where it would be expected that the binding energies would be fairly consistent and roughly converge as size increased if the lowest-energy structure was located in each case (see Appendix A). Table 5.1 shows the relative electronic energies of the lowest-energy structures (at the **B3LYP-D** level) calculated from manually created input structures and those calculated from input structures taken from REMD results. The relative energy difference increases up to cluster size $[Guan_9Cl_8]^+$, in which case the optimization happened to reach a relatively low-energy local minimum conformation.

All calculated structures discussed in the following sections are based on results taken from REMD simulations since, in all cases, these produced the lowest-energy optimized structures at the **B3LYP-D** level. Additionally, as will be seen in the following sections, there is no common pattern of structures which evolves with increasing cluster size, necessitating the completion of REMD simulations for each size.

Table 5.1 – Relative electronic energies of the lowest-energy manually defined RI-B3LYP-D/def2-SVP optimized structures comparing to the lowest-energy RI-B3LYP-D/def2-SVP optimized structures selected from REMD results (kJ mol^{-1}).

Cluster Size	Energy
9	56.3
6	262.3
5	192.1
4	134.8
3	0.0
2	0.0

5.2.2 REMD Results and Discussion

Many candidate structures resulted from REMD simulations, which followed the procedure previously described in the text. Two to five sequential REMD simulations were completed until the lowest-energy results of two subsequent simulations were consistent. Tables 5.2 and 5.3 below show the number of structures resulting from one REMD simulation as a function of cluster size, as well as the number of REMD (**AMOEBA**) candidate structures that were further considered using DFT methods.

Table 5.2 – List of $[Guan_xCl_{(x-1)}]^+$ structures calculated at each level of theory. The number of **AMOEBA** structures is for a single generation.

	AMOEBA	**B3LYP-D**(OPT+SP)	**B3LYP-D**(Freq)
$[Guan_{15}Cl_{14}]^+$	1660	10	5
$[Guan_{10}Cl_9]^+$	972	8	8
$[Guan_9Cl_8]^+$	928	10	7
$[Guan_6Cl_5]^+$	501	10	4
$[Guan_5Cl_4]^+$	1096	10	4
$[Guan_4Cl_3]^+$	2259	10	6
$[Guan_3Cl_2]^+$	588	10	2
$[Guan_2Cl]^+$	256	10	3

Table 5.3 – List of $[Guan_xCl_{(x+1)}]^-$ structures calculated at each level of theory. The number of **AMOEBA** structures is for a single generation.

	AMOEBA	**B3LYP-D**(OPT+SP)	**B3LYP-D**(Freq)
$[Guan_{15}Cl_{16}]^-$	1260	5	2
$[Guan_{10}Cl_{11}]^-$	1271	5	2
$[Guan_6Cl_7]^-$	1681	5	2
$[Guan_5Cl_6]^-$	976	5	2
$[Guan_4Cl_5]^-$	2645	5	2
$[Guan_3Cl_4]^-$	743	5	2
$[Guan_2Cl_3]^-$	230	5	2

AMOEBA structures were analyzed using energetic and structural considerations. For both $[Guan_xCl_{(x-1)}]^+$ and $[Guan_xCl_{(x+1)}]^-$, the potential energy surface appears to be quite flat and insensitive to certain relatively large structural deviations (this will be discussed more in the following sections). This made selection of geometries from the REMD results significantly more difficult and energetic considerations had to be more strongly relied upon than structural considerations. Sorting of structures using RMSD and gyration radius based considerations, as was done for the sodiated poly(alanine) peptides discussed in the previous chapter, was not possible. Initially, the ten lowest-energy $[Guan_xCl_{(x-1)}]^+$ **AMOEBA** structures were optimized at the RI-B3LYP-D/def2-SVP level and SP electronic energies (def2-TZVPP) were obtained to allow for further elimination of structures based on energetic considerations in advance of the relatively costly frequency calculations. After optimization at the RI-B3LYP-D/def2-SVP level, many of the unique **AMOEBA** structures converged to the same structure (especially for the smaller sized clusters). For the ten lowest-energy **AMOEBA** calculated structures for each cluster, the force field energy never spanned more than 7.5 kJ mol^{-1}. This was found to be somewhat inconsistent with the **B3LYP-D** calculated energies, both in terms of range and ordering (see Table 5.4 for more detail). However, the lowest-energy **B3LYP-D** structure resulting from optimization of the **AMOEBA** structures was in all cases one of the five lowest-energy **AMOEBA** structures in terms of the force field energy. Using these considerations, only

the five lowest-energy **AMOEBA** structures were calculated at the **B3LYP-D** level for $[Guan_xCl_{(x+1)}]^-$. All results are summarized in Table 5.4 below, where the index indicates which of the **AMOEBA** structures produced the lowest-energy **B3LYP-D** structure after optimization (order from lowest to highest energy with a total of 10 for cationic clusters, 5 for anionic clusters) and the range indicates the range of **B3LYP-D** electronic energies over the whole set of **AMOEBA** structures that were optimized at the **B3LYP-D** level.

Table 5.4 – Indices and energy ranges of the lowest-energy **B3LYP-D** structures among **AMOEBA** results. Energy ranges are listed in terms of electronic energies (kJ mol $^{-1}$).

	Index (0-10)	**Range**		**Index (0-5)**	**Range**
$[Guan_{15}Cl_{14}]^+$	5	62.2	$[Guan_{15}Cl_{16}]^-$	3	5.6
$[Guan_{10}Cl_9]^+$	5	49.6	$[Guan_{10}Cl_{11}]^-$	4	34.5
$[Guan_9Cl_8]^+$	2	39.1	$[Guan_6Cl_7]^-$	3	25.1
$[Guan_6Cl_5]^+$	1	0.9	$[Guan_5Cl_6]^-$	5	0.4
$[Guan_5Cl_4]^+$	4	11.7	$[Guan_4Cl_5]^-$	3	0.3
$[Guan_4Cl_3]^+$	1	31.9	$[Guan_3Cl_4]^-$	3	14.7
$[Guan_3Cl_2]^+$	4	0.22	$[Guan_2Cl_3]^-$	1	0.4
$[Guan_2Cl]^+$	3	0.22			

In some cases, the range of **B3LYP-D** energies is fairly large and in others it is very small. This is mainly due to the degree of structural change during optimization at the **B3LYP-D** level, where for some sizes all **AMOEBA** structures converged to essentially the same minimum at the **B3LYP-D** level, while in other cases multiple minima were located from the REMD selection. Intuitively, the range tends to increase as the size of the cluster increases, albeit with some exceptions. Using $[Guan_2Cl]^+$ as an example, structure guan2P_7290 is 1.33 kJ mol^{-1} higher in energy than guan2P_0477 in the **AMOEBA** force field (RMSD of 1.92 Å). guan2P_7290 is shown in Figure 5.4 where one guanidinium ion has one N-H interacting with a nitrogen of the other guanidinium in place of an interaction with the chloride (single point RI-B3LYP-D/def2-SVP energy calculations suggest guan2P_7290 to be 43.5 kJ mol^{-1} higher in electronic energy). At the **B3LYP-D** level, these two **AMOEBA** conformations were optimized to the same conformation with a negligible energy separation (as were all other **AMOEBA** $[Guan_2Cl]^+$ structures, see blue structure

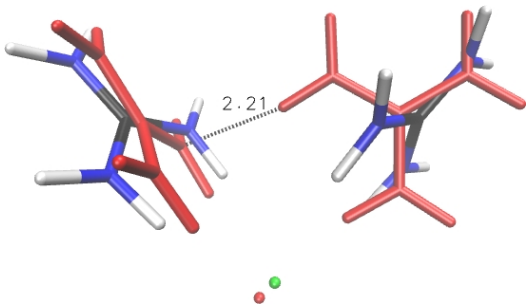

Figure 5.4 – Comparison of **AMOEBA** $[Guan_2Cl]^+$ structures. guan2P_0477 (coloured by elements) is overlaid with guan2P_7290 (**Red**). RMSD between these two structures is calculated to be 1.92 (Å).

in Figure 5.6) illustrated by the very small range of **B3LYP-D** energies in Table 5.4. As will be discussed in more detail later in the text, this example could be an indication of an increased stabilization of N-H$\cdots$N-H hydrogen bonds relative to N-H$\cdots Cl^-$ interactions at the **AMOEBA** level.

5.2.3 Structural Comparisons Between AMOEBA and B3LYP-D Results

Table 5.5 shows the average RMSD values between **AMOEBA** calculated structures and **B3LYP-D** calculated structures for each cluster. The degree of structural difference is in general lower among the negative clusters than among positive ones, and is the lowest in cases where the structures are somewhat constrained by H$\cdots Cl^-$ interactions such as $[Guan_2Cl_3]^-$ (RMSD - 0.09 Å) comparing to $[Guan_2Cl]^+$ (RMSD - 1.71 Å) (see Figures 5.6 and 5.9). The excess chloride ion and associated negative charge hold the structure together more tightly and occupy more of the available guanidinium interaction sites. This can also be seen by the generally smaller gyration radii of the negative ions than the positive ions of the same size in Table 5.6.

Table 5.5 – RMSD comparison between **AMOEBA** and the corresponding **B3LYP-D** optimized structures presented as averages over all pairs of structures at each cluster size for $[Guan_xCl_{(x-1)}]^+$ and $[Guan_xCl_{(x+1)}]^-$ (Å).

	Average RMSD		**Average RMSD**
$[Guan_2Cl]^+$	1.71	$[Guan_6Cl_5]^+$	1.03
$[Guan_3Cl_2]^+$	0.70	$[Guan_9Cl_8]^+$	1.24
$[Guan_4Cl_3]^+$	1.36	$[Guan_{10}Cl_9]^+$	1.31
$[Guan_5Cl_4]^+$	0.99	$[Guan_{15}Cl_{14}]^+$	1.36
	Average RMSD		**Average RMSD**
$[Guan_2Cl_3]^-$	0.09	$[Guan_6Cl_7]^-$	1.18
$[Guan_3Cl_4]^-$	1.01	$[Guan_{10}Cl_{11}]^-$	1.03
$[Guan_4Cl_5]^-$	0.86	$[Guan_{15}Cl_{16}]^-$	1.01
$[Guan_5Cl_6]^-$	0.72		

Figures 5.5-5.11 on the following pages show structural superpositions of **AMOEBA** and **B3LYP-D** results for the lowest-energy conformations of $[Guan_{2,5,10}Cl_{1,4,9}]^+$ and $[Guan_{2,5,10}Cl_{3,6,11}]^-$ with RMSD values listed in the captions. After visual inspection of the figures, one can see that relatively large differences exist between results at the two levels of theory. This can also be seen from comparisons of bond lengths and gyration radii in tables 5.6-5.10, where the **AMOEBA** structures tend to be more compact than the **B3LYP-D** structures with shorter $H\cdots Cl^-$ distances, although nearly identical C-N and N-H distances.

Table 5.6 – Comparison of gyration radii by level of theory. Listed values are the averages of all calculated structures for each size(Å). For formatting considerations, $[Guan_xCl_{(x-1)}]^+$ is abbreviated as $[Guan]^+$ and $[Guan_xCl_{(x+1)}]^-$ as $[Guan]^-$.

$[Guan]^+$	**B3LYP-D**	**AMOEBA**	$[Guan]^-$	**B3LYP-D**	**AMOEBA**
15	5.69	5.56	15	5.67	5.57
10	5.36	4.83	10	4.75	4.76
9	4.67	4.42	-	-	-
6	4.25	3.98	6	4.10	3.95
5	3.83	3.63	5	3.71	3.80
4	3.72	3.39	4	3.42	3.32
3	3.03	3.11	3	3.02	3.21
2	3.87	2.99	2	2.35	2.40

Table 5.7 – Comparison of average **AMOEBA** $[\text{Guan}_x\text{Cl}_{(x-1)}]^+$ structural parameters for the lowest-energy (**B3LYP-D**) structure for each size of cluster (Å).

	H$\cdots$Cl$^-$	**C-N**	**N-H**	**#H$\cdots$Cl$^-$**	**H$\cdots$Cl$^-$ per GuanCl**
15	2.32	1.34	1.03	52	3.5
10	2.35	1.34	1.03	39	3.9
9	2.33	1.34	1.03	32	3.6
6	2.30	1.34	1.03	20	3.3
5	2.30	1.34	1.03	14	2.8
4	2.24	1.34	1.03	11	2.8
3	2.21	1.34	1.03	6	2.0
2	2.27	1.34	1.03	4	2.0

Table 5.8 – Comparison of average **B3LYP-D** $[\text{Guan}_x\text{Cl}_{(x-1)}]^+$ structural parameters for the lowest-energy (**B3LYP-D**) structure for each size of cluster (Å).

	H$\cdots$Cl$^-$	**C-N**	**N-H**	**#H$\cdots$Cl$^-$**	**H$\cdots$Cl$^-$ per GuanCl**
15	2.39	1.34	1.02	85	5.7
10	2.34	1.34	1.02	53	5.3
9	2.39	1.34	1.02	48	5.3
6	2.28	1.34	1.02	27	4.5
5	2.41	1.34	1.02	23	4.6
4	2.39	1.34	1.02	18	4.5
3	2.43	1.34	1.02	12	4.0
2	2.17	1.34	1.02	4	2.0

Table 5.9 – Comparison of average **AMOEBA** $[\text{Guan}_x\text{Cl}_{(x+1)}]^-$ structural parameters for the lowest-energy (**B3LYP-D**) structure for each size of cluster (Å).

	H$\cdots$Cl$^-$	**C-N**	**N-H**	**#H$\cdots$Cl$^-$**	**H$\cdots$Cl$^-$ per GuanCl**
15	2.34	1.34	1.03	65	4.3
10	2.37	1.34	1.03	42	4.2
6	2.36	1.34	1.03	25	4.2
5	2.42	1.34	1.03	24	4.8
4	2.32	1.34	1.03	15	3.8
3	2.31	1.34	1.03	12	4.0
2	2.46	1.34	1.03	12	6.0

Table 5.10 – Comparison of average **B3LYP-D** $[Guan_xCl_{(x+1)}]^-$ structural parameters for the lowest-energy (**B3LYP-D**) structure for each size of cluster (Å).

	H···Cl⁻	**C-N**	**N-H**	**#H···Cl⁻**	**H···Cl⁻ per GuanCl**
15	2.33	1.34	1.03	85	5.7
10	2.34	1.34	1.03	57	5.7
6	2.36	1.34	1.03	36	6.0
5	2.34	1.34	1.03	30	6.0
4	2.25	1.34	1.03	20	5.0
3	2.4	1.34	1.03	18	6.0
2	2.3	1.34	1.03	12	6.0

H···Cl⁻ interactions are clearly favoured at the expense of NH-H···NH_2 hydrogen bonds at the **B3LYP-D** level comparing to the **AMOEBA** optimized structures. Tables 5.7-5.10 show the significant increase in number of H···Cl⁻ interactions in the **B3LYP-D** structures comparing to the **AMOEBA** calculated structures. Figure 5.5 illustrates this more clearly, where shifts of the guanidinium ions can be seen in the **AMOEBA** structure to form hydrogen bonds between guanidinium ions.

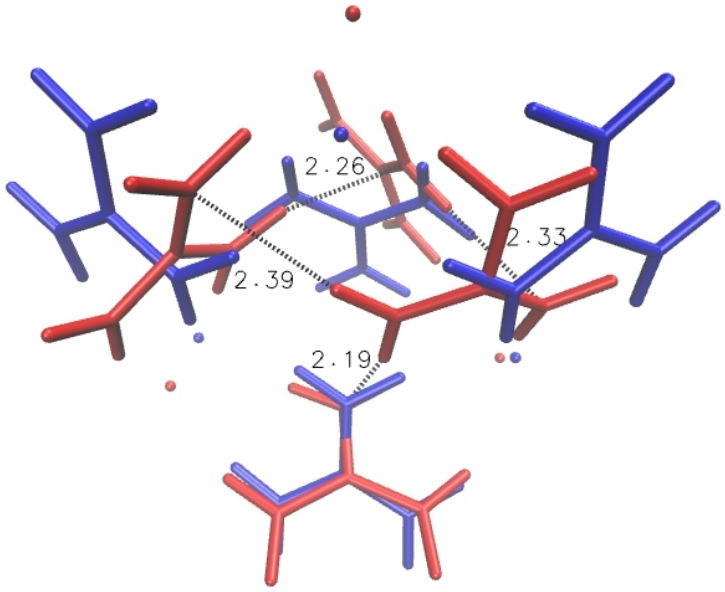

Figure 5.5 – Comparison of the lowest-energy (**B3LYP-D**) $[Guan_4Cl_3]^+$ structure at the **B3LYP-D** and **AMOEBA** levels. Hydrogen bonds between guanidinium ions in the **AMOEBA** structure are highlighted and labelled (guan4P_9932, Å).

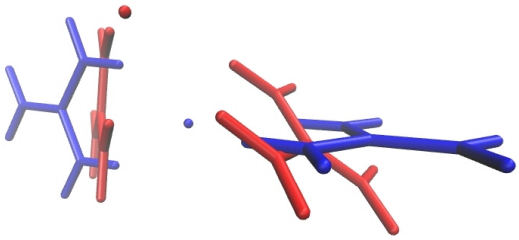

Figure 5.6 – Comparison of the lowest-energy (**B3LYP-D**) $[Guan_2Cl]^+$ structure at the **B3LYP-D** and **AMOEBA** levels (guan2P_0477, RMSD 1.69 Å).

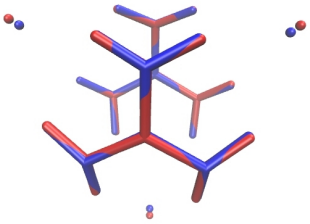

Figure 5.9 – Comparison of the lowest-energy (**B3LYP-D**) $[Guan_2Cl_3]^-$ structure at the **B3LYP-D** and **AMOEBA** levels (guan2N_1493, RMSD 0.09 Å).

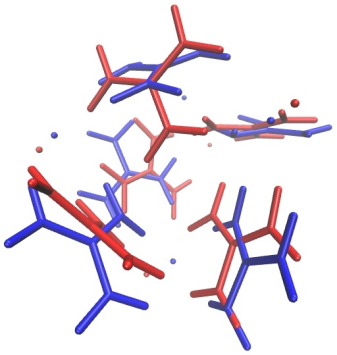

Figure 5.7 – Comparison of the lowest-energy (**B3LYP-D**) $[Guan_5Cl_4]^+$ structure at the **B3LYP-D** and **AMOEBA** levels (guan5P_6236, RMSD 1.25 Å).

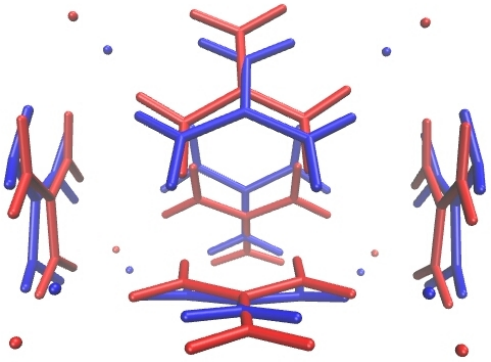

Figure 5.10 – Comparison of the lowest-energy (**B3LYP-D**) $[Guan_5Cl_6]^-$ structure at the **B3LYP-D** and **AMOEBA** levels (guan5N_6254, RMSD 0.72 Å).

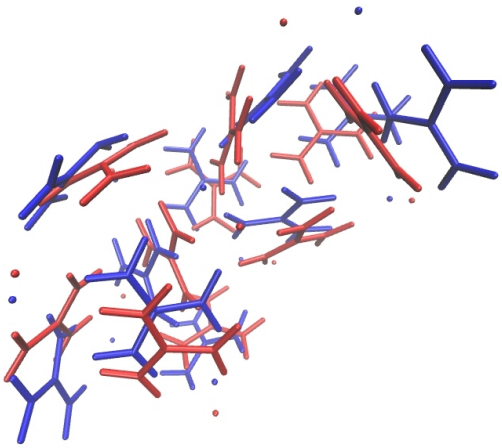

Figure 5.8 – Comparison of the lowest-energy (**B3LYP-D**) $Guan_{10}Cl_9]^+$ structure at the **B3LYP-D** and **AMOEBA** levels (guan10P_1355, RMSD 1.42 Å).

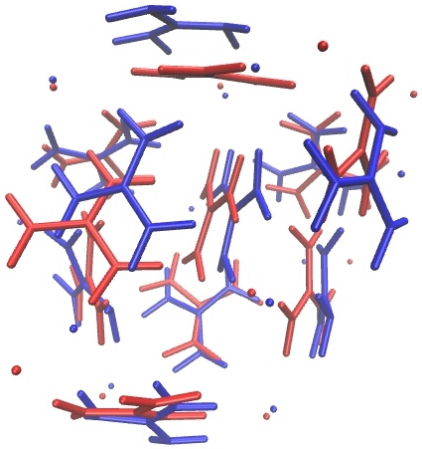

Figure 5.11 – Comparison of the lowest-energy (**B3LYP-D**) $[Guan_{10}Cl_{11}]^-$ structure at the **B3LYP-D** and **AMOEBA** levels (guan10N_8756, RMSD 1.20 Å).

Tables 5.7-5.10 show the number of $H\cdots Cl^-$ interactions per guanidinium for $[Guan_xCl_{x-1}]^+$ cluster increases continually to size fifteen (up to 5.7 interactions) for **B3LYP-D** structures, while for **AMOEBA** structures this value seems to plateau at approximately 3.5 interactions per guanidinium ion. There is not a similar trend for the $[Guan_xCl_{x+1}]^-$ ions since the maximum of six interactions per guanidinium ion is reached at some of the smaller sizes, while for the larger sizes the number of interactions per guanidinium ion converges to 5.7 as for $[Guan_xCl_{x-1}]^+$. A plateau at about 4.3 interactions per guanidinium ion is seen in the **AMOEBA** results. In principle, one could imagine a limitation of six interactions per guanidinium ion if no NH-$H\cdots NH_2$ hydrogen bonds were formed, illustrating again the preference for $H\cdots Cl^-$ interactions in the **B3LYP-D** results. The higher stability of NH-$H\cdots NH_2$ interactions in the **AMOEBA** optimized structures could be a result of the atomic charges on guanidinium nitrogens and hydrogens being slightly too large, resulting in the overestimation of the energy associated with these interactions relative to $H\cdots Cl^-$ interactions.

Literature Comparisons

Guanidinium chloride, the chloride bound dimer of guanidinium ($[Guan_2Cl]^+$) and the guanidinium ion itself have been previously examined.[153–157] Table 5.11 compares calculated bond lengths for gas phase guanidinium ion at the MP2/6-31++G(d),[153] PBE/DNP[155] and **B3LYP-D** (this work) levels and demonstrates only negligible differences (less than 2% in all cases).

Table 5.11 – Bond length comparisons for the guanidinium ion from the literature, in the gas phase (Å).

Reference	MP2/6-31++G(d)[153]	PBE/DNP[155]	**B3LYP-D**
CN1	1.34	1.34	1.34
CN2	1.34	1.34	1.34
CN3	1.34	1.34	1.34
N1H1	-	1.02	1.01
N1H2	-	1.03	1.01
N2H3	-	1.02	1.01
N2H4	-	1.02	1.01
N3H5	-	1.02	1.01
N3H6	-	1.03	1.01

Comparison of structural parameters for guanidinium chloride are shown in Table 5.12. **B3LYP-D** results are consistent with calculated gas phase results from the literature,[153] however $H\cdots Cl^-$ lengths are somewhat shorter and N-H lengths are very slightly longer. Both gas phase results are in agreement with the calculated crystal and aqueous results, however all calculations differ rather largely from the experimental crystal results.

Table 5.12 – Bond length comparisons for guanidinium chloride, from the literature (Å).

phase	crystal	crystal	aqueous	gas	gas
reference	Exp.[154]	PBE[155]	PBE[155]	MP2/6-31+G(d)[153]	**B3LYP-D**
N1H1-Cl	2.35	-	-	2.08	1.97
N2H3-Cl	2.66	-	-	2.08	1.98
CN1	1.32	1.34	1.33	1.33	1.33
CN2	1.33	1.34	1.33	1.33	1.33
CN3	1.33	1.35	1.37	1.37	1.36
N1H1	0.86	1.03	1.11	1.05	1.07
N1H2	0.96	1.03	1.02	1.01	1.01
N2H3	0.73	1.03	1.11	1.05	1.07
N2H4	0.87	1.03	1.02	1.01	1.01
N3H5	1.17	1.03	1.02	1.01	1.01
N3H6	1.17	1.03	1.02	1.01	1.01

Results for $[Guan_2Cl]^+$ at the MP2/6-31+G(d) are shown in Table 5.13.[153] The same minimum-energy structure as was found in the present work, at the **B3LYP-D** level, was identified and structural parameters agree with the **B3LYP-D** results.[153] As was seen previously in the comparisons for guanidinium chloride, results calculated at the **B3LYP-D** level in Table 5.13 show bond lengths to be slightly contracted.

Table 5.13 – Comparison of non-bonded distances and angles for $[Guan_2Cl]^+$ (Å, degrees).

$[Guan_2Cl]^+$	MP2/6-31+G(d)[153]	**B3LYP-D**	Percent Difference,%
H-Cl	2.23	2.17	2.7
C-Cl	3.63	3.57	1.7
Cl-H-N	152.3	151.5	0.5

Using the largest calculated cluster sizes allows for the closest comparison to the crystal structure. $[Guan_{15}Cl_{14}]^+$ and $[Guan_{15}Cl_{16}]^-$ are already much more similar to the crystal structure (shown below in Figures 5.12 and 5.13) with average $H\cdots Cl^-$ lengths of 2.39 and 2.33 Å, respectively, and comparing much more closely to the average crystal $H\cdots Cl^-$ length of 2.39 Å. The guanidinium chloride crystal structure is orthorhombic and in the Pbca space group. Lattice constants are a=9.184, b=13.039, c=7.765 Å, (±0.005 A). The guanidinium ions were reported to be planar within the accuracy of the experiment, and in the C_{3v} point group.[154]

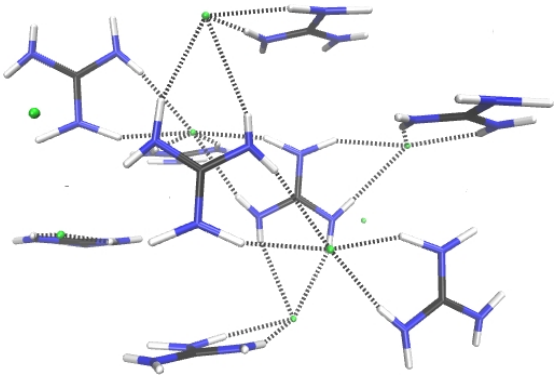

Figure 5.12 – The crystal structure of guanidinium chloride.[154]

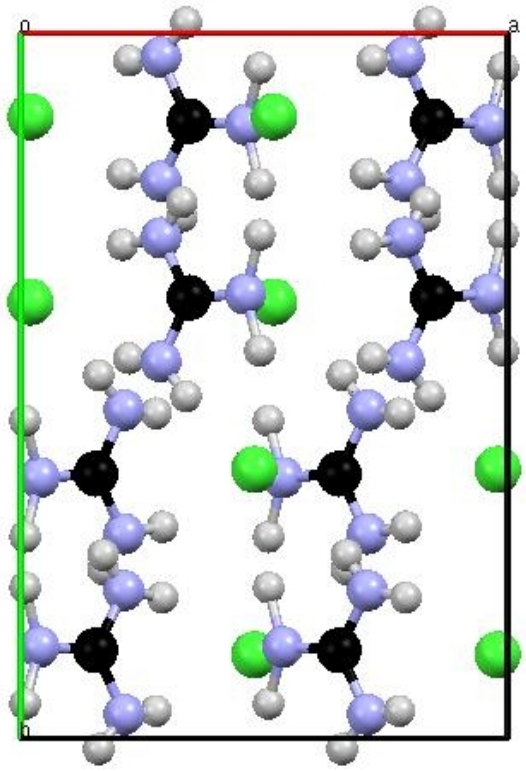

Figure 5.13 – The crystal structure of guanidinium chloride with the unit cell highlighted.[154]

5.3 Vibrational Properties of Guanidinium Chloride Based Clusters

5.3.1 General Description of Vibrational Modes

All clusters of both the anionic and cationic series have many common vibrations and spectral features[iv]. In the fingerprint region, NH_2 scissor modes are found in the ≈1450-1650 cm^{-1} range while C-N stretching modes are found in the ≈1600-1725 cm^{-1}.[155] In the higher energy 3000-3600 cm^{-1} region, symmetric NH_2 stretching modes are in the ≈3000-3350 cm^{-1} range and asymmetric NH_2 stretching modes are in the ≈3200-3575 cm^{-1} region.

[iv] All calculated vibrational frequencies are at the **B3LYP-D** level and scaled by 0.98 in the fingerprint region and 0.96 in the 3000-3600 cm^{-1} (OPO) region unless otherwise specified. Justification of this methodology and scaling will be discussed further in the following sections.

Modes are often strongly coupled and delocalized between guanidinium ions. For example, modes in the 1600 cm^{-1} region are often the combination of C-N stretching modes and NH_2 scissor modes. Coupled modes in the region where symmetric and asymmetric N-H stretches overlap (3200-3300 cm^{-1}) often have groups vibrating both symmetrically and asymmetrically. Additionally, modes are delocalized over many guanidinium ions, such as NH_2 scissoring modes whose motion is synchronized over different NH_2 groups on different guanidinium ions. As a result (and especially for the larger sized clusters) modes become very complex and mixed, making assignments and descriptions of each vibration very difficult.

The range of NH_2 scissor vibrations and C-N vibrations are relatively limited and bonding environments are highly similar between different sizes as well as between conformations of the same size. This results in very similar spectra between both different sizes and conformations of the same sized cluster, especially in the fingerprint region. Positions of modes found in the 3000-3600 cm^{-1} region are much more sensitive to different interaction environments and thus are more distinguishable. However, the large number of modes in the larger sizes ($\gtrapprox$ $[Guan_5Cl_6]^-/[Guan_5Cl_4]^+$) often results in a large and convoluted band over much of the N-H stretching range. In the conformations (such as $[Guan_5Cl_6]^-$, see Figures 5.49, 5.69 and 5.52) where all NH_2s and C-Ns are found in similar environments, one can see a strong narrowing of spectral bands in both the fingerprint region and the 3000-3600 cm^{-1} region. In general, symmetric stretching modes are found at lower wavenumbers than asymmetric modes, with coupled modes having both symmetric and asymmetric character in the overlapping region.

broadening and red shifting. This will be discussed in more detail in the following sections.).

5.3.2 Additional Spectral Considerations

Strongly hydrogen-bound vibrational modes present difficulties to both experimental and theoretical (harmonic) treatment.[60,158–162] Hydrogen bonding, for example, in N-H$\cdots$Cl$^-$, causes a stretching and weakening of the N-H bond and an increase in the anharmonicity. This, along with an increased dipole moment, results in red shifted, broadened bands with increased intensities. In experiment, these modes can be sufficiently broadened that they are practically unobservable. This is often used as a indication of the binding environment of the C-terminus in peptides and proteins, where the absence of a CO-H band is taken as evidence that the C-terminus is strongly hydrogen-bound.[136,137] This can affect symmetric and asymmetric vibrational modes differently. Using as an example N-Hs on different NH_2 groups of the same guanidinium ion interacting with the same chloride ion, if both N- Hs vibrate asymmetrically the effect on the dipole moment is somewhat balanced, however, the symmetric vibration would feel the hydrogen bonding effect additively. In the spectra in the following sections of the text, this consideration will play a dominant role in judging the degree of consistency between experiment and calculation.

Previous studies have made detailed examinations of the vibrational properties for the guanidinium ion from various vibrational spectroscopy experiments in the crystalline form and in solution.[155–157] The isolated planar guanidinium ion has D_{3h} symmetry in the gas phase, however the symmetry is undoubtedly reduced in the crystal structure by interactions with other ions (to C_{3v}).[156] It has been suggested that the D_{3h} selection rules may still be approximately followed if the ions remain relatively planar[156] (this depends on the extent to which each guanidinium ion is influenced by its immediate environment and, if discussed in terms of the present results, would vary significantly between clusters). In the C_{3v} environment, the A_1' and A_2'' symmetry vibrations of the D_{3h} group are of A_1 symmetry and are IR allowed. As well the E$'$ and E$''$ modes of the D_{3h} group are of E symmetry and are also IR active. The strongly absorbing C-N stretching modes, NH_2

scissor modes and N-H stretching modes (symmetric and asymmetric) are contained in both the E′ (D_{3h}) and E (C_{3v}) groups and the C-N (symmetric) stretching modes, NH_2 scissor modes and N-H stretching modes (symmetric) allowed in the A_1 (C_{3v}) but forbidden in the A_1' group (D_{3h}) are only weakly absorbing.[156] In general, these selection rules are not expected to play a large role for the guanidinium chloride clusters discussed here.

5.3.3 Scaling Factors For Calculated Vibrational Modes

It is apparent from Tables 5.14 and 5.15 (and the discussion in the following sections, see Figures 5.62 and 5.63 specifically) that harmonic **B3LYP-D** frequencies require a smaller correction for the fingerprint region than the typically applied 0.96 scaling factor (for the 3000-3600 cm^{-1} N-H stretching region 0.96 will be seen to be appropriate). This will also be seen for **M06**[v] results where an even smaller correction is necessary. For the fingerprint region of both $[Guan_xCl_{x-1}]^+$ and $[Guan_xCl_{x+1}]^-$ ions, a scaling factor of 0.98 was found to be appropriate for **B3LYP-D** modes and 0.99 for **M06** modes. In the 3000-3600 cm^{-1} region, the 0.96 scaling factor is appropriate for both positive and negative ions for both **B3LYP-D** and **M06** modes. In the fingerprint region, there are in general just two bands, one of convoluted NH_2 modes and one of convoluted mainly C-N modes. Comparisons of the experimental and **B3LYP-D** peak centres are shown in Tables 5.14 and 5.15 below for different scaling factors. Further discussion will follow in Section 5.4.6.

[v]See Section 5.4.6 for presentation of **M06** results.

Table 5.14 – Experimental and calculated (**B3LYP-D**) peak centre positions in the fingerprint region for $[Guan_xCl_{x-1}]^+$ (cm^{-1}).

IRMPD	CN str		NH_2 scis	
$[Guan_{15}Cl_{14}]^+$	1675		1575	
$[Guan_{10}Cl_9]^+$	1670		-	
$[Guan_6Cl_5]^+$	1675		1575	
$[Guan_5Cl_4]^+$	1675		1570	
$[Guan_4Cl_3]^+$	1670		1575	
$[Guan_2Cl]^+$	1675		1575	
RI-B3LYP-D	CN str		NH_2 scis	
scale	0.96	0.98	0.96	0.98
$[Guan_{15}Cl_{14}]^+$	1640	1674	1505	1536
$[Guan_{10}Cl_9]^+$	1645	1679	1510	1541
$[Guan_6Cl_5]^+$	1640	1674	1505	1536
$[Guan_5Cl_4]^+$	1645	1679	1515	1547
$[Guan_4Cl_3]^+$	1640	1674	1510	1541
$[Guan_2Cl]^+$	1630	1664	1540	1572

Table 5.15 – Experimental and calculated (**B3LYP-D**) peak centre positions in the fingerprint region for $[Guan_xCl_{x+1}]^-$ (cm^{-1}).

IRMPD	CN str		NH_2 scis	
$[Guan_{10}Cl_{11}]^-$	1675		-	
$[Guan_6Cl_7]^-$	1680		1570	
$[Guan_5Cl_6]^-$	1680		1590	
$[Guan_4Cl_5]^-$	1670		1580	
$[Guan_2Cl_3]^-$	1680		1665	
RI-B3LYP-D	CN str		NH_2 scis	
scale	0.96	0.98	0.96	0.98
$[Guan_{10}Cl_{11}]^-$	1650	1685	1530	1562
$[Guan_6Cl_7]^-$	1640	1674	1540	1572
$[Guan_5Cl_6]^-$	1640	1674	1575	1608
$[Guan_4Cl_5]^-$	1650	1685	1530	1562
$[Guan_2Cl_3]^-$	1660	1694	1535	1567

5.4 Results and Discussion

Each cluster of the $[\text{Guan}_x\text{Cl}_{(x-1)}]^+$ and $[\text{Guan}_x\text{Cl}_{(x+1)}]^-$ series will now be discussed in detail. Theoretical structural descriptions and energetic results for the five lowest-energy **B3LYP-D** structures for each cluster will be followed by experimental spectra and discussion. Bold black spectra represent a three-point moving average of the experimental data points. Assignments are largely approximate to the extent that, in most cases, modes are highly coupled and the associated descriptions are mainly an attempt to describe the dominant characteristics of each band. In cases where negligible difference between computed spectra for different low-energy calculated structures exist, only a single calculated spectrum is presented for comparison with experiment.

5.4.1 Experimental Considerations

The experimental apparatus was described in detail in both the previous chapter and the introductory section of this text, and only details specific to the experiments involving clusters of guanidinium chlorides will be discussed here.

Electrospray ionization was used to generate ions from guanidinium chloride solutions of 10^{-4}-10^{-6} M in acetonitrile. Variation of concentration (and accordingly, the flow rate to maintain appropriate ion current) allowed for some influence of the size distribution of generated ions. Additionally, adjustment of parameters in the mass spectrometer and ion source allowed for more accurate control of the ion distribution (within the ranges of: $[\text{Guan}_{(x=2-20)}\text{Cl}_{x-1}]^+$, $[\text{Guan}_{(x=2-10)}\text{Cl}_{x+1}]^-$). The previously described FT-ICR experimental apparatus was used for the 3000-3600 cm^{-1} region experiments,[26,28,29,36,42,55,56,61] while a modified Bruker Esquire 3000+ quadrupole ion trap[61] was used for experiments generating fingerprint region spectra. As well, the previously described FEL (CLIO beam) and OPO/OPA (also at CLIO) were used as light sources. The main difference between

these mass spectrometers in reference to the IRMPD experiments discussed here, was in the increased flexibility of the irradiation time in the FT-ICR as well as the use of the previously described CO_2 laser as a pump source (not currently possible with the ion trap setup). Spectra in the following sections did not make use of the CO_2 laser except for the two OPO experiments on the negative $[Guan_3Cl_4]^-$ and $[Guan_5Cl_6]^-$ ions. Experimental considerations limited collection of spectra for sizes larger than $[Guan_{15}Cl_{14}]^+$ for positive ions and $[Guan_{10}Cl_{11}]^-$ for negative ions, although as mentioned previously, larger ions were observed.

In many cases, scans obtained under different experimental conditions are shown in the figures. In general, as the laser power (or irradiation time) increases, so does the width of the band (especially noticeable for the 1600-1700 cm^{-1} region). While the relative intensities of different bands within a single scan remain unmodified, the y-axes of different scans have been adjusted for better presentation with each other and with the calculated modes. Especially for the larger clusters, many low intensity modes fill both the 1000-1200 cm^{-1} and the 1500-1600 cm^{-1} regions of the spectrum. While higher laser power experiments (900 mW, 400 ms irradiation time) were completed, it appeared that only the relatively intense modes (those calculated to be above ≈30-40 km mol^{-1}) were ever observed experimentally. Future experiments could consider these two spectral regions using the CO_2 pump laser in addition to the FEL and possibly obtain additional and valuable spectral information.

As will be seen later in the text (see Figures 5.68 and 5.69), the experimental band of the NH_2 modes in the 1500-1650 cm^{-1} region tends to vary rather significantly between different clusters. In order to attribute the cause of this variation to experimental conditions, it is illustrative to consider the different conditions used to obtain spectra for each cluster.

Table 5.16 – Summary of experimental conditions for FEL experiments in the fingerprint region. Beam power is measured at gap 25 mm in mW, irradiation times in ms, $Frag_{eff}$ is listed as the maximum over the full spectral range and calculated absorption intensity (maximum, lowest-energy structure, **B3LYP-D**) is in km mol^{-1}.

$[\mathbf{Guan}_x\mathbf{Cl}_{x-1}]^+$	Power	IR_{time}	#-Attenuators	$Frag_{eff}$	$Intensity_{calc}$
$[Guan_{15}Cl_{14}]^+$	830	200	0	1.2	54
$[Guan_{10}Cl_9]^+$	1000	300	2	0.6	41
$[Guan_6Cl_5]^+$	800	200	0	1.5	34
$[Guan_5Cl_4]^+$	1000	250	0	1.5	50
$[Guan_4Cl_3]^+$	1000	300	0	1.4	73
$[Guan_2Cl]^+$	1050	200	0	2.4	118
$[\mathbf{Guan}_x\mathbf{Cl}_{x+1}]^-$	Power	IR_{time}	#-Attenuators	$Frag_{eff}$	$Intensity_{calc}$
$[Guan_{10}Cl_{11}]^-$	960	200	1	0.26	83
$[Guan_6Cl_7]^-$	940	200	0	1.28	73
$[Guan_5Cl_6]^-$	830	200	0	1.1	33
$[Guan_4Cl_5]^-$	900	200	0	1.26	99
$[Guan_2Cl_3]^-$	940	400	0	0.25	48

Relating the experimental conditions to the observed spectra (see Figures 5.68 and 5.69 for a summary of experimental spectra) provides some explanation to the nature of the observed NH_2 band of 1450-1650 cm^{-1}. For $[Guan_xCl_{x-1}]^+$, significant bands are seen for $[Guan_{2-5}Cl_{1-4}]^+$ where the maximum observed fragmentation efficiencies are also large. For $[Guan_{6,15}Cl_{5,14}]^+$, only very small NH_2 bands are seen in this region, consistent with relatively small calculated absorption intensities, shorter irradiation times and lower laser powers. The complete absence of an NH_2 band for $[Guan_{10}Cl_9]^+$ is clearly not a result of very low absorption intensity, but rather the use of two attenuators which resulted in a significantly lower fragmentation efficiency (likely a fully experimental spectral difference). For $[Guan_xCl_{x+1}]^-$, the only spectrum with a significant NH_2 band between 1450-1650 cm^{-1} is $[Guan_6Cl_7]^-$, where the maximum fragmentation efficiency was also the highest of the anionic clusters. $[Guan_{2,5,10}Cl_{3,6,11}]^-$ have very small NH_2 bands likely due to both low absorption intensities in this region and the lower fragmentation conditions of the scans. $[Guan_4Cl_5]^-$ has only a very small NH_2 band in the experimental spectrum, however

in comparison to $[Guan_6Cl_7]^-$ in Table 5.16, this is not easily justified by experimental conditions or weak absorption intensities.

5.4.2 Fragmentation

Details of IRMPD fragment ions are seen in Table 5.17. The table is divided into two parts with experiments for $[Guan_xCl_{x-1}]^+$ detailed in the top section and experiments for $[Guan_xCl_{x+1}]^-$ seen in the bottom section. In the bottom right side of the top section, the fragmentation of $[Guan_2]^+$ and $[Guan_2H]^+$ are described. For formatting considerations, names are abbreviated. For example, $[G_{15}]^+$ represents $[Guan_{15}Cl_{14}]^+$. Labels G_2H^+ and GH^+ represent the proton bound dimer of guanidine and the guanidinium ion, respectively.

Fragmentation in these clusters mainly occurs by unit loss of guanidinium chloride neutrals[vi]. In the smaller positive ions, loss of HCl, or the equivalent of HCl + a unit number of neutral guanidinium chloride(s) is observed, leaving the proton bound dimer of guanidine. Guanidinium, 60 m/z, is also seen as a fragment ion from the smaller positive clusters. In the negative ion series, the same pattern of fragmentation is observed, however these non-chloride-containing fragment ions are not seen. For the 3000-3650 cm^{-1} region experiments, positive ions required only the use of the OPO/OPA with approximately 1.0 second irradiation times, while the negative ions necessitated the previously described CO_2 laser in addition to the OPO/OPA. For the fingerprint region experiments, only the FEL was used, requiring approximately 200-400 ms irradiation times.

[vi]In some cases, mass differences between sequential (differing by one guanidinium chloride neutral) fragment ions do not appear to correspond exactly to 95 m/z. This is due to the observation of incomplete isotopic massifs resulting from the low intensity of the largely ^{37}Cl and entirely ^{35}Cl peaks associated with the $^{35}Cl/^{37}Cl$ ratio of 3.13.

Table 5.17 – Fragmentation resulting from IRMPD for $[Guan_xCl_{x-1}]^+$, $[Guan_xCl_{x+1}]^-$ and $Guan_2H^+$ (m/z). Values listed under the column "loss" refer to the size of cluster (the number of guanidinium ions) corresponding to the fragment masses listed in the neighbouring column. Values in the row labelled "parent" are the masses of the parent ions labelled below these mass values.

parent	1402-1392		921-921		541-535		444-440		347-343	
loss	$[G_{15}]^+$		$[G_{10}]^+$		$[G_6]^+$		$[G_5]^+$			$[G_4]^+$
14	1303	9	826	5	446	4	347	3		252
	1301		824		444		345			250
	1299		822		442		343	2		157
13	1208	8	731		440		157			155
	1206		729	4	349	2	155	G_2H^+		119
	1204		727		347		153	GH^+		60
12	1111		725		345					
	1109									
11	1016									
	1014									
10	921									
	919									
	917									
9	826									
	824						parent	155/157		119
	822						loss	$[G_2]^+$		$[G_2H]^+$
8	729						G_2H^+	119	GH^+	60
	727						GH^+	60		

parent	995-985		611-605		516-510		419-415		229-225	
loss	$[G_{10}]^-$		$[G_6]^-$		$[G_5]^-$		$[G_4]^-$			$[G_3]^-$
9	900	5	516	4	418	3	324	1		131
	898		514		416		322			129
	896		512		414		320			
	894		510	3	319	2	229			
	892	4	419		323		227			
	890		417		321		225			
8	801		415	2	226					
	799	3	322		224					
	797		320							
7	706	2	227							
	704		225							
	702									
	700									

5.4.3 Results - $[\mathbf{Guan}_x\mathbf{Cl}_{(x-1)}]^+$

$[\mathbf{Guan_2Cl}]^+$

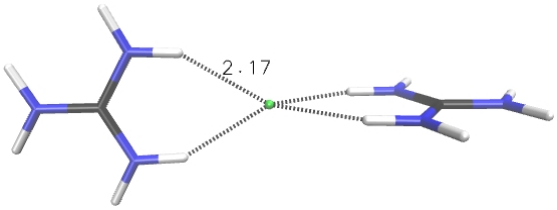

Figure 5.14 – guan2P_0477, the lowest-energy **B3LYP-D** $[Guan_2Cl]^+$ structure.

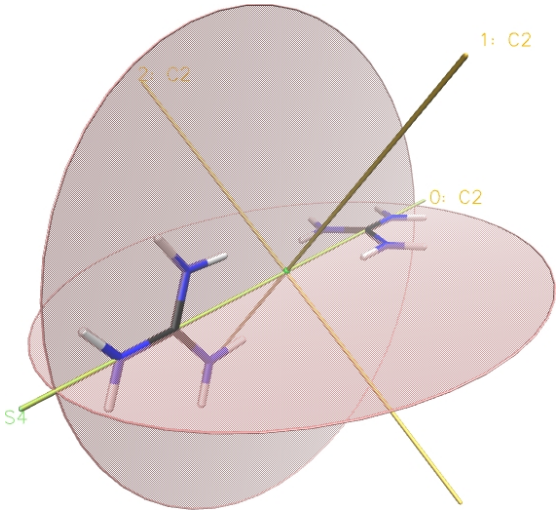

Figure 5.15 – guan2P_0477, with symmetry features illustrated (D_{2d})

Table 5.18 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) and dihedral angles (degrees) for $[Guan_2Cl]^+$ (kJ mol^{-1}).

$[\mathbf{Guan_2Cl}]^+$	**E**	**G**	**Dihedral ($\mathbf{N_{(1)}C_{(1)}C_{(2)}N_{(2)}}$)**
guan2P_0477	0.0	0.0	97.7
guan2P_4470	0.01	-	93.9
guan2P_3348	0.02	-	86.5
guan2P_7624	0.1	0.8	103.6
guan2P_1338	0.2	1.2	107.7

Guanidinium ions in the lowest-energy structures of $[Guan_2Cl]^+$ form nearly perpendicular planes (97.7 degrees). The chloride anion is coordinated by four hydrogens in all of the calculated structures with the main differences between these nearly identical structures being the dihedral angle between the planes containing the guanidinium ions. The relative energy seems to be relatively insensitive to the angle between these planes (in the table listed as "Dihedral $N_{(1)}C_{(1)}C_{(2)}N_{(2)}$") with a ten degree variation (97.7 to 107.7 degrees) only affecting the relative free energy by 1.2 kJ mol^{-1}. All four interactions with

the chloride anion are 2.17 Å. This structure is approximately of the D_{2d} point group with two C_2 axes perpendicular to the principle C_2 axis and two reflection planes.

Fingerprint region- $[Guan_2Cl]^+$

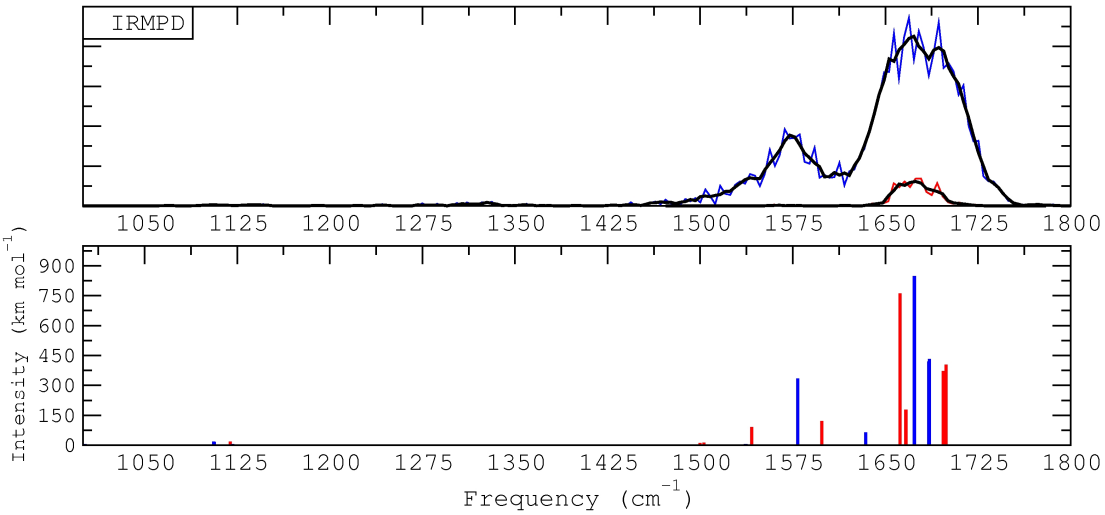

Figure 5.16 – $[Guan_2Cl]^+$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. **B3LYP-D** frequencies are scaled by 0.98 and in **red** and B3LYP/6-311+G(d,p) anharmonic frequencies are scaled by 1.011 and in **blue**.

In Figure 5.16, the harmonic **B3LYP-D** calculated $[Guan_2Cl]^+$ spectrum matches very well to both of the experimental scans, with the peak centres of the two C-N stretches (at approximately 1700 and 1660 cm^{-1}) matching the experimental position very closely. A small dip in intensity is seen in the blue scan corresponding to the region of zero intensity between the two calculated C-N modes. The calculated modes at 1700 cm^{-1} correspond to C-N motion (of the nitrogens of the bound NH_2 groups) perpendicular to the principle axis (C-Cl-C), while the modes at 1660 cm^{-1} correspond to the C-N motion along the principle axis of the C-N containing the free NH_2. At 1625 cm^{-1} is the fully symmetric scissor mode (in reference to synchronization and phase of the motion) of all four bound NH_2 groups around the chloride ion, however the calculated intensity of this mode is nearly zero and it is not seen in the spectra. The mode at 1600 cm^{-1} is the same vibration but with the motion of NH_2 groups on different guanidinium ions being out of phase with each other. At

1550 cm^{-1} and at 1540 cm^{-1} (vanishingly low intensity) are the in-phase and out-of-phase motions of the two free NH_2 groups (one on each guanidinium), respectively. Finally two low intensity modes at 1500 cm^{-1} (not visible in spectrum) correspond to the asymmetric motion of bound NH_2 groups on each guanidinium ion. The anharmonic spectrum offers an even better match to the experimental spectrum with line positions of both the C-N and NH_2 vibrations matching very well. The most notable improvement is obtained from the blue shift of the NH_2 modes comparing to the corresponding harmonic modes. This correction to the NH_2 modes from the anharmonic analysis is important to note here as a similar error in harmonic spectra is seen for many other clusters. Although anharmonic analysis was not completed for any other sizes due to the enormous computational expense, the correction seen here for $[Guan_2Cl]^+$ suggests that a similarly correcting blue shift of the NH_2 modes in anharmonic spectra could possibly result for other clusters as well.

3000-3600 cm^{-1} - $[Guan_2Cl]^+$

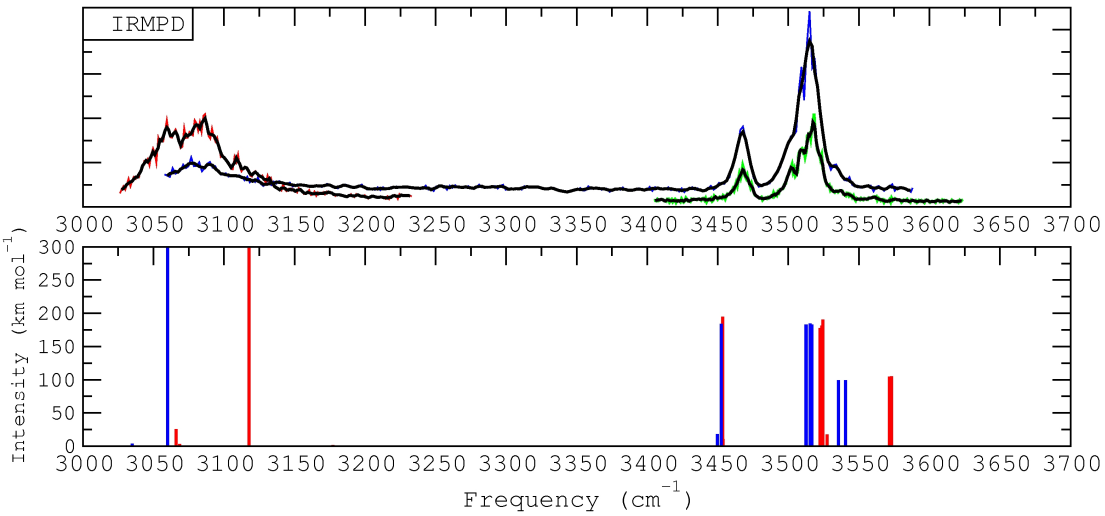

Figure 5.17 – $[Guan_2Cl]^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. **B3LYP-D** frequencies are scaled by 0.96 and in **red** and B3LYP/6-311+G(d,p) anharmonic frequencies are scaled by 1.011 and in **blue**.

The experimental 3000-3600 cm^{-1} region (OPO) spectrum of $[Guan_2Cl]^+$ is shown in Figure 5.17 along with the lowest-energy calculated result. The experimental spec-

trum is composed of three separate scans, one of the entire spectral range from 3000-3600 cm^{-1} (**blue**), a second scan of the 3400-3600 cm^{-1} (**green**) and finally a third scan (**red**, with increased IRMPD time aiming to more definitively observe the symmetric stretching modes) of the 3000-3200 cm^{-1} region. Together, these scans provide a complete spectral description for this cluster with good resolution and signal to noise ratio. Twelve N-H stretching modes result from the twelve N-H bonds found in this cluster, however grouping of these modes into six degenerate pairs, with two of these pairs themselves degenerate, results in a significantly simplified spectrum with only five distinct calculated bands. Two symmetric N-H stretching modes at 3070 cm^{-1} correspond to vibration of the four bound NH_2 groups around the chloride, with groups on the same guanidinium moving out of phase (maxima and minima of motion correspond). A very intense mode at 3120 cm^{-1} corresponds to the same mode, but movement of NH_2 groups on each guanidinium ion are in phase (degenerate modes is of nearly zero intensity and not seen). The last two symmetric N-H stretches at 3455 cm^{-1} correspond to motion of the two free NH_2 groups (one on each guanidinium). Asymmetric stretches are found still slightly higher, four at 3520 and two at 3570 cm^{-1}, with the lower frequency band corresponding to the bound NH_2 groups and the higher frequency band to each of the free NH_2 groups. Comparison between the experimental and calculated spectra immediately results in a couple of rather obvious issues. Firstly, the symmetric N-H stretches are generally of rather weak intensity relative to the calculated results, including the most intense mode (with a calculated intensity of 4000 km mol^{-1}) at 3120 cm^{-1} which is barely observed in principle. This is likely a result of poor harmonic modelling of these strongly hydrogen-bound modes which are heavily broadened and red shifted. Additionally, some discrepancy remains in the comparison of the 3450-3600 cm^{-1} region and no obvious conclusions can be made relating the three harmonic bands to the two experimental bands. Consideration of the anharmonic modes suggests that the two calculated bands of bound N-H stretches are convoluted in the single experimental band in this region. Scaling of the anharmonic modes by 1.011 results in

an approximately 40 cm^{-1} blue shift in the 3000-3600 cm^{-1} region and significantly improves the agreement with experiment. While little precedent or explicit explanation can be provided to justify such a scaling of the anharmonic spectrum, it is worth noting that previous examples of B3LYP/6-311+G(d,p) anharmonic frequencies, relating to a similar type of highly-electrostatic complex, in some cases also demonstrated slightly red shifted vibrations.[163,164]

$[Guan_3Cl_2]^+$

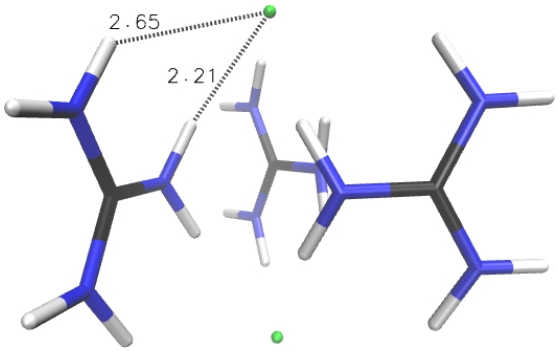

Figure 5.18 – guan3P_9671, the lowest-energy **B3LYP-D** $[Guan_3Cl_2]^+$ structure.

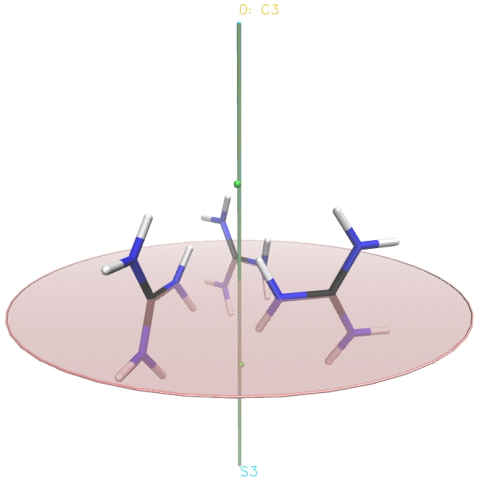

Figure 5.19 – guan3P_9671 with symmetry features illustrated (C_{3h}).

Table 5.19 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_3Cl_2]^+$ (kJ mol^{-1}).

$[Guan_3Cl_2]^+$	E	G
guan3P_9671	0.2	0.0
guan3P_8714	0.0	0.2
guan3P_9279	0.01	-
guan3P_8262	0.02	-
guan3P_0170	0.1	0.7

In all calculated structures listed in the table above, the three guanidinium ions are situated around the vertically aligned chloride anions. In the **B3LYP-D** calculated

lowest free energy structure, guan3P_9671, the three carbon atoms are situated at the vertices of an equilateral triangle. Additionally, the nitrogens of the NH_2 groups also form an equilateral triangle, as do the hydrogens on top of the nitrogens and the hydrogens below the nitrogens. The inward facing NH_2 group has one hydrogen interacting with the Chloride at the top of the structure and one at the bottom, each at 2.21 ±0.01 Å, for all three guanidinium ions. The inward facing hydrogens on both the top and bottom NH_2 groups also interact with the chloride anions at 2.65 ±0.01 Å. for all three guanidinium ions. The C-N bonds involving the interacting NH_2 groups on each guanidinium ion are exactly contained in a plane. Structural differences between calculated structures listed in Table 5.19 are fairly large relative to the very small energy differences, and are mostly associated with tilting upwards or downwards of guanidinium ions relative to the plane of the carbon atoms.

$[Guan_4Cl_3]^+$

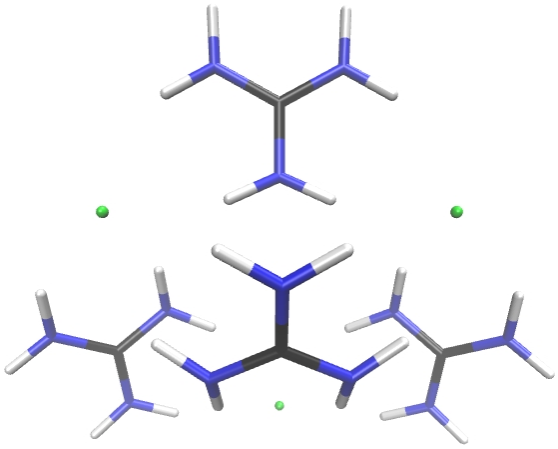

Figure 5.20 – guan4P_9932, the lowest-energy **B3LYP-D** optimized $[Guan_4Cl_3]^+$ structure from **AMOEBA** results (C_s).

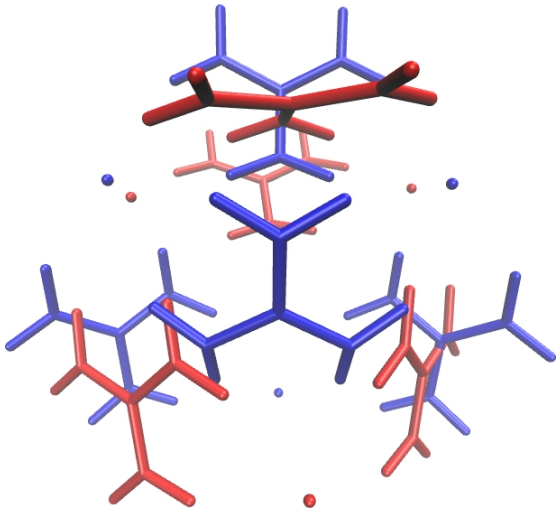

Figure 5.21 – Overlay of **guan4P_9932** with second-lowest-energy unique $[Guan_4Cl_3]^+$ structure from **AMOEBA** results **guan4P_9500**.

The lowest-energy structure, guan4P_9932, is in the C_s point group with a plane

Table 5.20 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_4Cl_3]^+$ (kJ mol^{-1}).

$[\mathbf{Guan_4Cl_3}]^+$	**E**	**G**
guan4P_9932	0.0	0.0
guan4P_0939	0.3	1.1
guan4P_8834	0.2	1.9
guan4P_0797	1.1	2.4
guan4P_9500	23.3	6.6
guan4P_9932_symm	-0.9	-2.7

of reflection situated vertically, perpendicular to the plane of the page as seen in Figure 5.20. The top guanidinium ion turns slightly out (into the page) of the approximate plane formed by the others. The central guanidinium ion has six $H\cdots Cl^-$ interactions (although at longer distances) while the other three guanidinium ions have only four each. guan4P_9500 is shown in Figure 5.21 in an overlay with guan4P_9932 because it is the only structure included in Table 5.20 having significant structural differences. In guan4P_9500, two of the guanidinium ions are roughly perpendicular to the plane of the other two, reducing the total number of $H\cdots Cl^-$ interactions by two and raising the relative free energy of this structure by 6.6 kJ mol^{-1} in comparison to guan4P_9932 (the difference is smaller in terms of free energy than in terms of electronic energy due to the larger entropy associated with the less constrained structure of guan4P_9500). Other low-energy calculated structures differ mainly by very small structural variations which result in slight losses of symmetry in reference to guan4P_9932 and guan4P_9932_symm. At the bottom of Table 5.20, structure guan4P_9932_symm is listed to be -2.7 kJ mol^{-1} lower in free energy than guan4P_9932. guan4P_9932_symm differs from guan4P_9932 through an increased degree of symmetry around the central guanidinium ion, bringing it into the C_{3v} point group as illustrated in Figure 5.22 below.

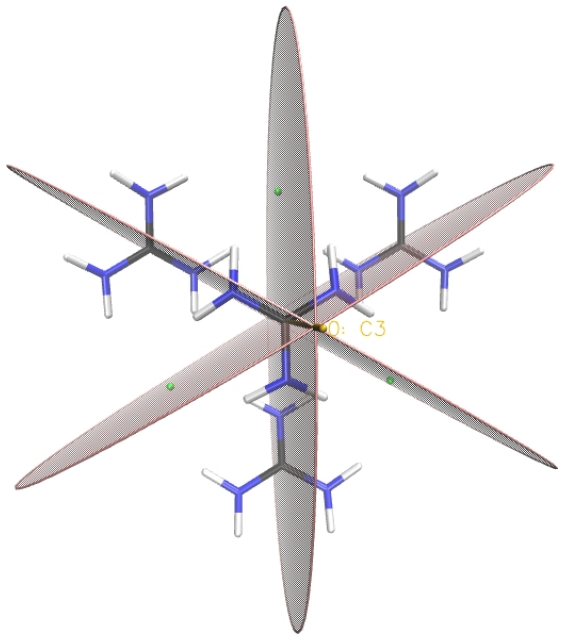

Figure 5.22 – Lowest-energy **B3LYP-D** $[Guan_4Cl_3]^+$ structure guan4P_9932_symm (C_{3v}). This higher symmetry structure was not located by **B3LYP-D** optimization of selected **AMOEBA** structures.

This structure was not originally obtained from **B3LYP-D** optimization of REMD identified **AMOEBA** structures and was only obtained after an MP2/6-31+g(d,p) reference calculation identified the higher symmetry minimum. guan4P_9932_symm resulted, following re-optimization of the MP2/6-31+g(d,p) structure at the **B3LYP-D** level, where it both remained a minimum on the PES and retained its higher symmetry structure.

Fingerprint region- $[Guan_4Cl_3]^+$

In the fingerprint spectrum of $[Guan_4Cl_3]^+$, calculated modes for the lowest-energy structure, guan4P_9932_symm, match the peak centre of the lower intensity (red) scan very closely. These modes, as well as two modes slightly higher in frequency at 1690 cm^{-1} correspond to delocalized C-N stretching modes, with the higher frequency modes belonging mainly to the more loosely bound central guanidinium ion. A slightly larger distribution of modes is calculated for guan4P_9932 (red line spectrum) than for guan4P_9932_symm, however this is not experimentally distinguishable. Calculated NH_2 scissor modes (delo-

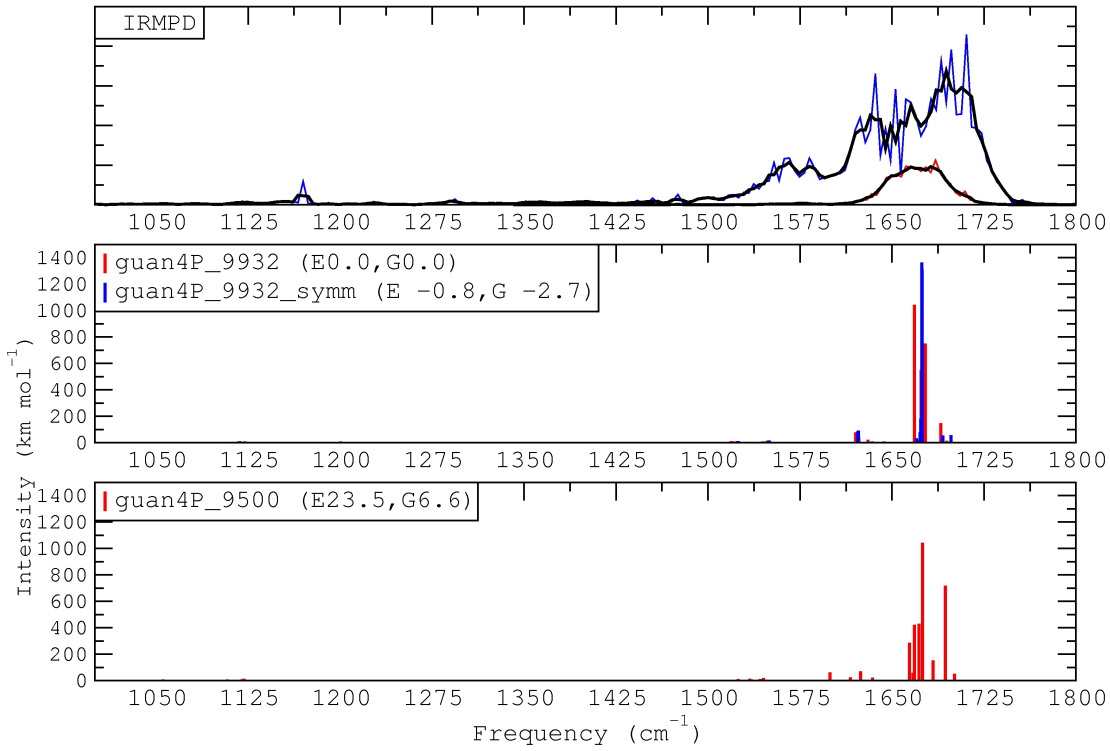

Figure 5.23 – $[Guan_4Cl_3]^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.98)

calized and of very low calculated intensity) start at 1520 cm^{-1} and extend to 1640 cm^{-1} where again the highest frequency modes are calculated to correspond mainly to motion of the NH_2 groups on the central and most loosely bound guanidinium ion. The NH_2 scissor modes do not appear to match perfectly and a gap between 1545-1620 cm^{-1}, calculated to have zero intensity, is nearly centred on the experimental NH_2 band. An approximately 20-30 cm^{-1} blue shift would need to be applied to bring the calculated modes into alignment with the experimental band.

$[Guan_5Cl_4]^+$

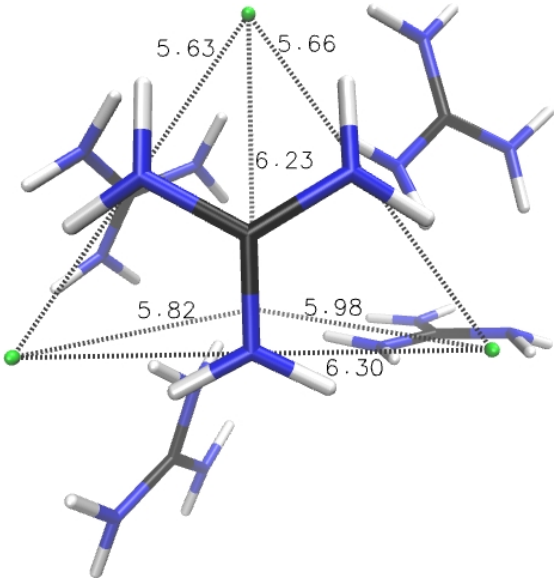

Figure 5.24 – guan5P_6236, the lowest-energy **B3LYP-D** $[Guan_5Cl_4]^+$ structure. Distances between and positions of chloride anions are highlighted.

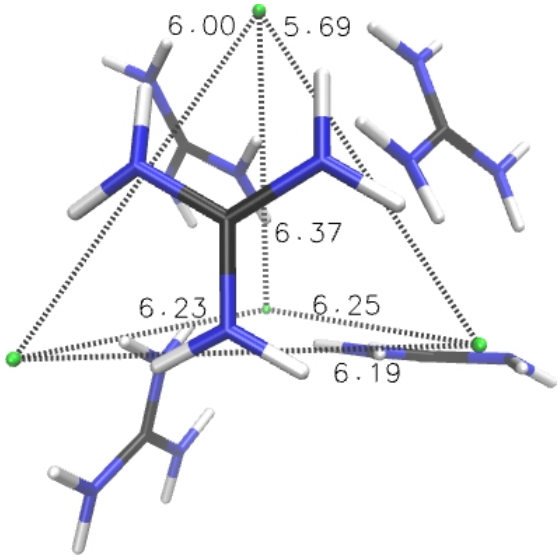

Figure 5.25 – guan5P_3998, the second-lowest-energy **B3LYP-D** $[Guan_5Cl_4]^+$ structure. Distances between and positions of chloride anions are highlighted.

Table 5.21 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_5Cl_4]^+$ (kJ mol^{-1}).

$[Guan_5Cl_4]^+$	E	G
guan5P_6236	0.0	0.0
guan5P_1905	11.3	-
guan5P_3998	11.7	0.8
guan5P_3839	11.0	2.0
guan5P_4358	11.5	2.3

In the structure guan5P_6236, two guanidinium ions have six $H\cdots Cl^-$ coordinations, while two others have four and one has three. The chloride ions form a tetrahedron, where two of the guanidinium ions occupy sides of the pyramid (the ones with six $H\cdots Cl^-$ coordinations) with their NH_2 groups each coordinated to two chloride ions, as seen in Figure 5.24 above. The guanidinium ions which form only four $H\cdots Cl^-$ coordinations are "outside" of the pyramid, and their NH_2 groups connect chloride ions on only one edge of

the pyramid formed by the chloride ions. The second-lowest-energy structure guan5P_3998 can be seen in Figure 5.25. Both structures have a hydrogen bond from one NH_2 of the right-side guanidinium to the guanidinium below of 2.23 Å. These structures differ by the position of the guanidinium on the left side of the cluster. In guan5P_6236, the left guanidinium is situated on the left face of the pyramid interacting with all three chloride ions. In guan5P_3998, this same guanidinium does not form a face of the pyramid but is rotated outwards and only interacting with the two chloride ions forming the rear edge of the pyramid. This structural difference results in three chlorides being coordinated by six N-H groups and one by five N-H groups for guan5P_6236, while for guan5P_3998, only two chlorides are coordinated by six N-H groups, one by five and the last by only four N-H groups (corresponding to the front left chloride in Figure 5.25). These two structures are calculated to be very similar in relative free energies, however it will be seen in the following sections that the calculated spectrum of guan5P_3998 is observed to be in better agreement with experiment.

Fingerprint region- $[\mathbf{Guan_5Cl_4}]^+$

The calculated spectrum for guan5P_6236, is consistent with both $[Guan_5Cl_4]^+$ experimental scans shown in Figure 5.26. The calculated C-N stretching modes in the 1660-1715 cm^{-1} region correspond very well to the width and position of the experimental scan. A dip in experimental intensity at 1690 cm^{-1} matches very closely to a similar feature in the calculated guan5P_6236 spectrum. The band comprised of NH_2 scissor modes centred at 1550 cm^{-1} in the calculated spectrum is only very slightly red shifted from the centre of the experimental band and still contained, although of very weak intensity. The spectrum of the second-lowest-energy conformation, guan5P_3998, is also consistent with experiment. Between the two calculated spectra, slight differences exist in the position of the modes around 1625 m^{-1} and the shape of the mainly C-N band, centred at approximately 1675 cm^{-1}. While it is difficult to make specific conclusions about which calculated spectrum is

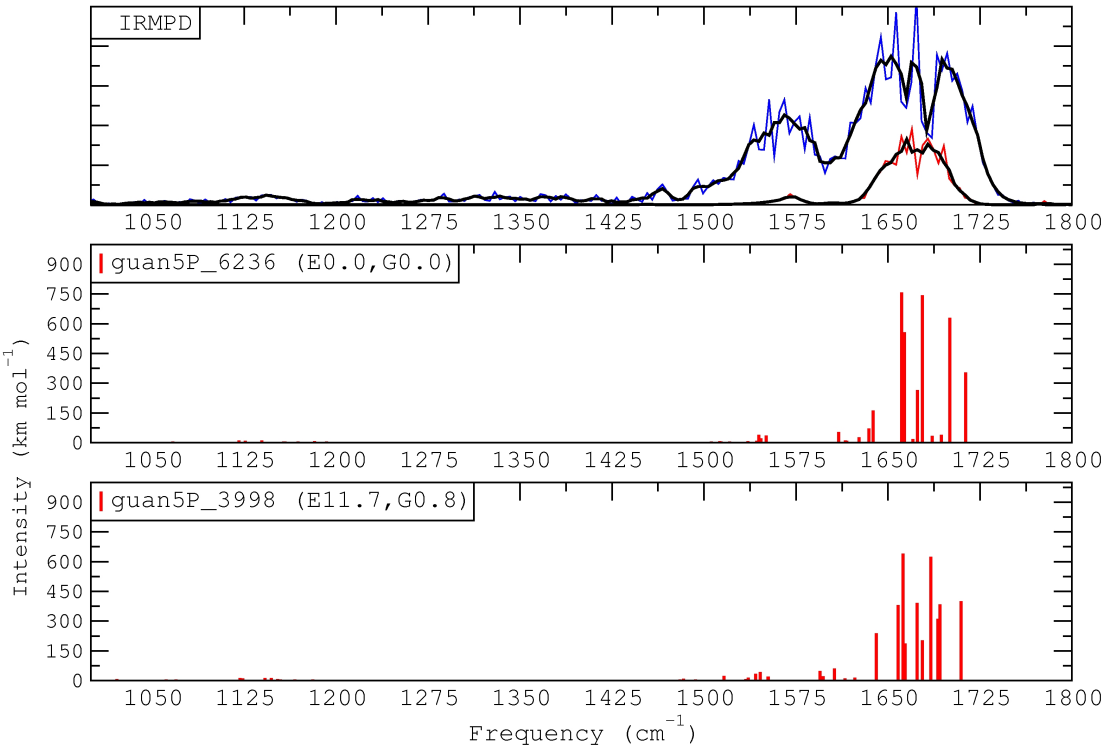

Figure 5.26 – $[Guan_5Cl_4]^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.98)

more consistent with the experimental spectrum the 3000-3600 cm^{-1} region spectrum in the next section will support the assignment of guan5P_3998.

3000-3600 cm^{-1} - $[Guan_5Cl_4]^+$

Calculated and experimental spectra in the 3000-3600 cm^{-1} region for $[Guan_5Cl_4]^+$ are shown in Figure 5.27. The calculated symmetric N-H stretching modes starting at approximately 3050 cm^{-1} and continuing until 3300 cm^{-1}, are again not evident in the experimental spectrum. The two calculated spectra shown, although calculated for structures which are very close in energy, differ significantly in the 3400-3500 cm^{-1} region. Here, there is a region of low intensity for guan5P_3998 and a region of maximum intensity for guan5P_6236. It is clear that the experimental spectrum is more consistent with the calculated spectrum for guan5P_3998 as the maximum intensities for the asymmetric

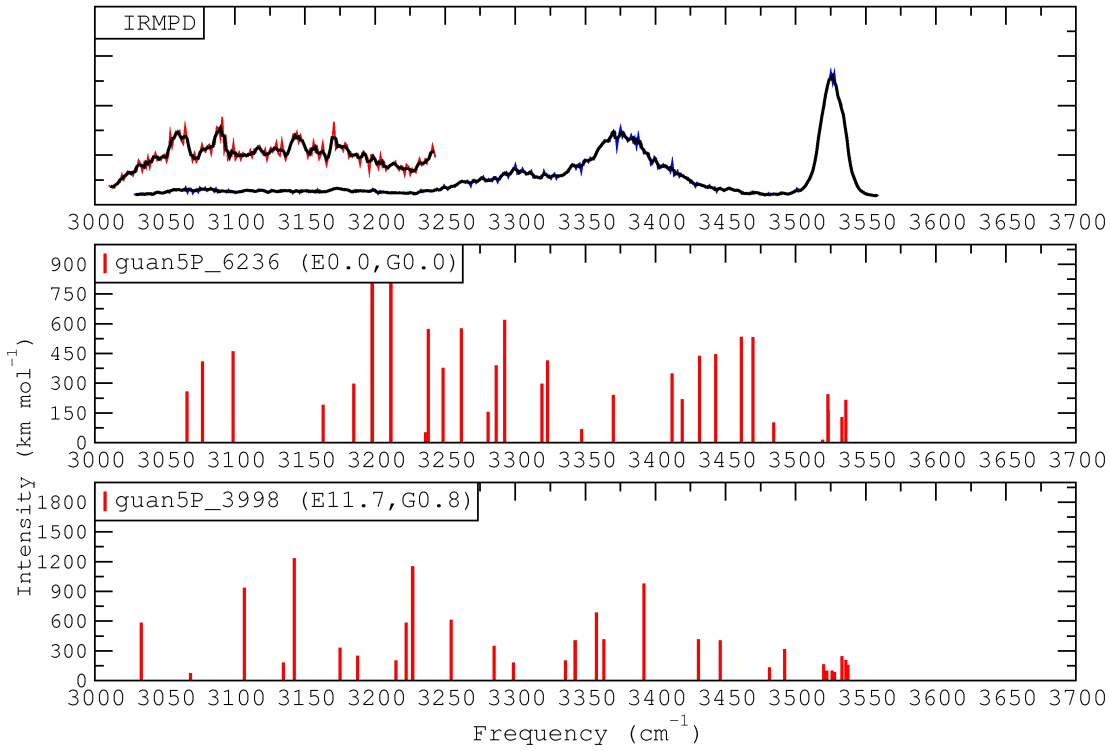

Figure 5.27 – $[Guan_5Cl_4]^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.96)

N-H stretches at approximately 3375 cm^{-1} and 3525 cm^{-1} match nicely (approximate assignments are discussed in the following paragraph). The symmetric stretching modes for guan5P_3998 are calculated to end at 3300 cm^{-1} and the asymmetric modes begin at 3335 cm^{-1}. The red scan in the experimental spectrum is a scan with longer irradiation time in the region of mainly bound and symmetric N-H stretching modes. Very small peaks in this experimental scan are difficult to distinguish from noise, but possibly correspond to some of the symmetric stretching vibrations which are again strongly hydrogen-bound and thus highly red shifted and broadened. This discrepancy will be demonstrated to be less pronounced for **M06** calculated modes in a later section (5.4.6) of the text discussing different computational methodologies for vibrational analysis.

The strong qualitative difference in computed spectra allow $[Guan_5Cl_4]^+$ to be used as

an illustration of the position of N-H modes in a general context. The main structural differences between calculated structures guan5P_6236 and guan5P_3998 relate to the position of one guanidinium ion (on the top-left in Figures 5.24 and 5.25) where in the calculated lowest-energy structure (guan5P_6236, lower by 11.7 kJ mol^{-1} in electronic energy and 0.8 kJ mol^{-1} in free energy) it forms the face of the pyramid and in guan5P_3998 it forms only the rear edge of the structure. This results in guan5P_3998 having eight free asymmetric N-H stretches (3537-3519 cm^{-1} with an outlying mode at 3481 cm^{-1}) and guan5P_6236 having only six free asymmetric N-H stretches (3536-3518 cm^{-1} with an outlying mode at 3484 cm^{-1})[vii]. While the difference in number of free and bound asymmetric N-H modes would not be expected to affect this band (the ranges of the bands are nearly identical for both structures, and resolution is not sufficient to distinguish individual modes), it does affect the bound asymmetric stretching modes rather significantly. For the guanidinium ion that differs between the two structures (top left in the figure), there are a larger number of bound asymmetric N-H modes which are also more highly coupled and fairly significantly blue shifted in guan5P_6236 comparing to guan5P_3998. This results in the range of bound asymmetric N-H stretches for guan5P_6236 to be between 3414-3469 cm^{-1} and for guan5P_3998 to be between 3335-3445 cm^{-1}, bringing the calculated spectrum of guan5P_3998 into much closer agreement with the experimental one. In guan5P_3998, where the top left guanidinium ion has only four N-H$\cdots Cl^-$ interactions (and the front left Cl^- has only four N-H interactions in total), the N-H$\cdots Cl^-$ interactions are shorter and stronger resulting in a clear red shifting of the band relative to guan5P_6236 where they are weaker, longer and more coupled between different guanidinium ions. The top left guanidinium ion in guan5P_6236 has an average length of 2.46 Å over its six N-H$\cdots Cl^-$ bonds, while the average N-H$\cdots Cl^-$ length in the whole cluster is 2.41 Å. In guan5P_3998, the top left guanidinium ion has an average length of 2.29 Å over its four N-H$\cdots Cl^-$ bonds,

[vii]The outlying single free asymmetric stretch for both calculated structures corresponds to the NH_2 group where one N-H is interacting with a chloride and one is free and the nitrogen of the NH_2 is hydrogen-bound to N-H of a different guanidinium ion.

comparing to an average length of 2.30 Å over all N-H$\cdots$Cl$^-$ bonds in the cluster.

[Guan$_6$Cl$_5$]$^+$

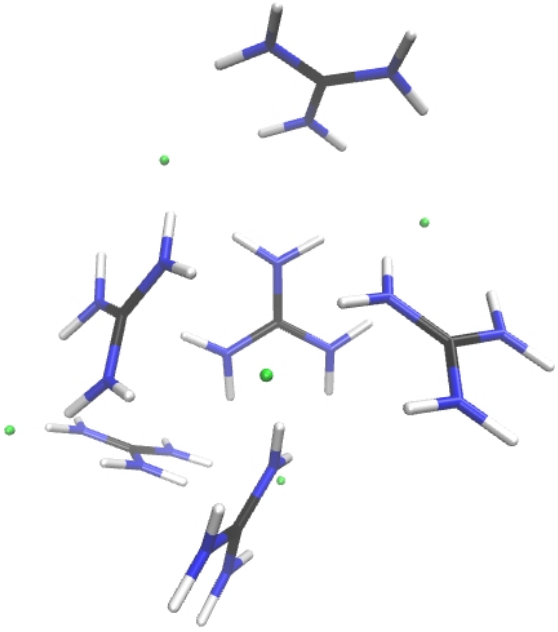

Figure 5.28 – guan6P_6555, the lowest-energy **B3LYP-D** [Guan$_6$Cl$_5$]$^+$ structure.

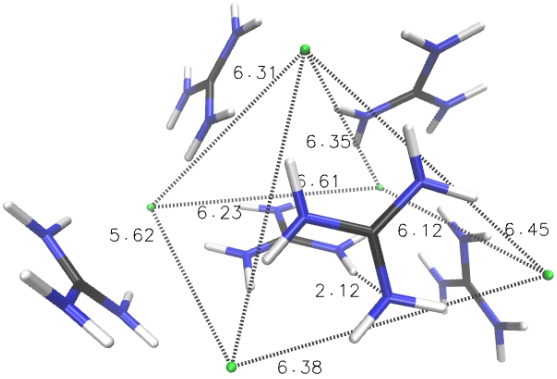

Figure 5.29 – guan6P_6555 with distances between and positions of chloride anions highlighted.

Table 5.22 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for [Guan$_6$Cl$_5$]$^+$ (kJ mol^{-1}).

[Guan$_6$Cl$_5$]$^+$	**E**	**G**
guan6P_6555	0.3	0.0
guan6P_6727	0.4	-
guan6P_0182	0.0	0.4
guan6P_1539	0.9	1.3
guan6P_3879	0.7	1.4

The lowest-energy structure calculated for [Guan$_6$Cl$_5$]$^+$, guan6P_6555, is somewhat similar to the lowest-energy [Guan$_5$Cl$_4$]$^+$ calculated structure, except that here the chloride ions form a mildly distorted square pyramid structure. The larger square base accommodates a hydrogen bond of 2.12 Å, which is formed between two guanidinium cations. Two

guanidinium ions are coordinated six times to chloride anions and the others are coordinated four times.

Fingerprint region - $[Guan_6Cl_5]^+$

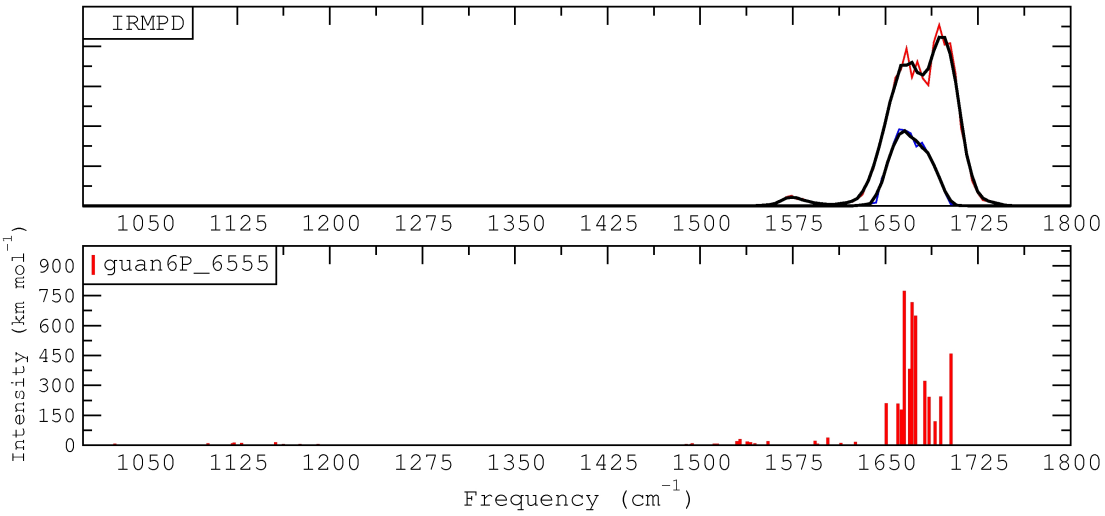

Figure 5.30 – $[Guan_6Cl_5]^+$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.98)

In the experimental spectrum of $[Guan_6Cl_5]^+$, the 1650-1700 cm^{-1} region of C-N stretching modes is well reproduced by calculation, where not only the width and position but also the shape of the peak is reproduced. The farthest blue and higher intensity mode in the calculated spectrum corresponds to a narrow maximum in the experimental spectrum. Experimentally, the band of NH_2 scissor modes is almost not seen, however this very small band is within 15 cm^{-1} of calculated NH_2 scissor modes seen at 1595 cm^{-1}.

3000-3600 cm^{-1} - $[Guan_6Cl_5]^+$

Calculated and experimental spectra for $[Guan_6Cl_5]^+$ are presented in Figure 5.31. The strongly hydrogen-bound symmetric stretching modes which are again heavily broadened and red shifted (and thus not observed in the experimental spectrum) are contained in the 3100-3250 cm^{-1} region of the calculated spectra. Mainly asymmetric bands in the

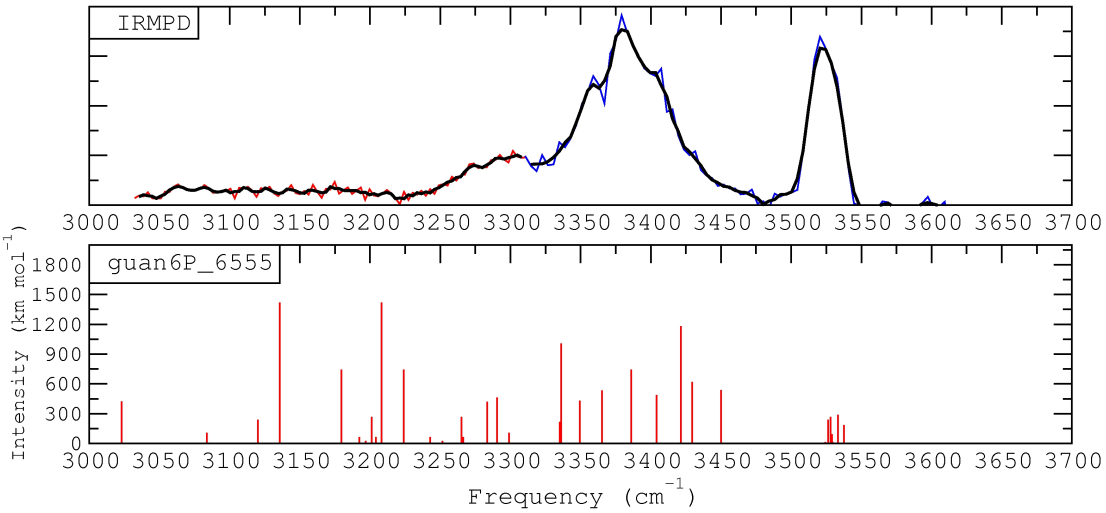

Figure 5.31 – $[Guan_6Cl_5]^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.96)

3300-3450 cm^{-1} region, as well as a drop in intensity followed by the bands of the asymmetric and free N-H modes, all match well to the experimental scan. Comparing to the 3000-3600 cm^{-1} region of $[Guan_5Cl_4]^+$ in Figure 5.27, a clear emergence of three main vibrational bands can be observed. Going from the red end of the spectrum to the blue end, the free symmetric band, bound asymmetric band and finally the free asymmetric band can be easily distinguished (see also Figure 5.70).

$[\mathbf{Guan_9Cl_8}]^+$

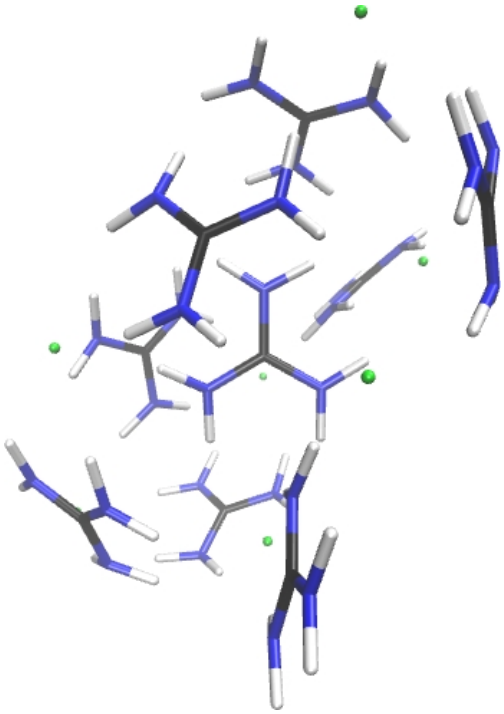

Figure 5.32 – guan9P_2896, the lowest-energy **B3LYP-D** $[Guan_9Cl_8]^+$structure.

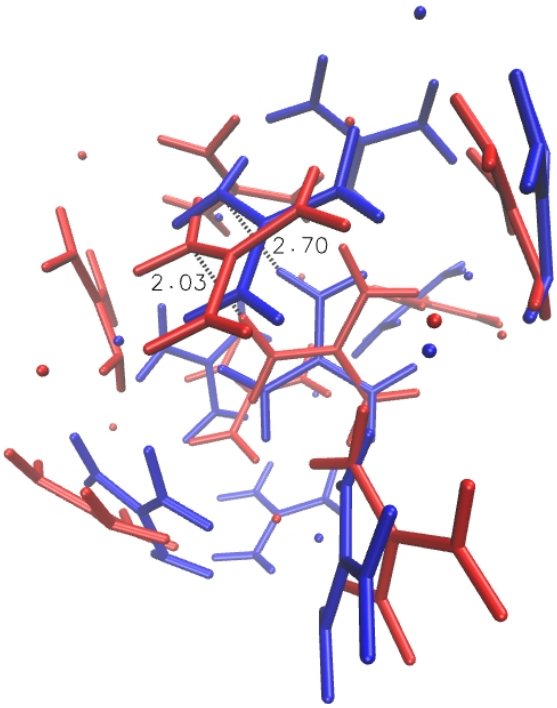

Figure 5.33 – Overlay of **guan9P_2896** with second-lowest-energy $[Guan_9Cl_8]^+$ structure, **guan9P_2179**.

Table 5.23 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_9Cl_8]^+$ (kJ mol^{-1}).

$[\mathbf{Guan_9Cl_8}]^+$	**E**	**G**
guan9P_2896	0	0.0
guan9P_2179	6.5	7.7
guan9P_9857	2.1	9.7
guan9P_5581	2.9	10.8
guan9P_5238	11.3	-

The lowest-energy structure calculated for $[Guan_9Cl_8]^+$, guan9P_2896, is sufficiently large to accommodate one guanidinium (centred in Figure 5.32) inside of the cluster, where it is almost entirely surrounded by other guanidinium ions. In Figure 5.33, guan9P_2179 is shown in red in an overlay with guan9P_2896. The structures are largely similar, however

a shift at one site forming a hydrogen bond (2.03 Å) between two NH_2 groups in the top right corner of the structure, raises the free energy by 7.7 kJ mol^{-1}.

$[Guan_{10}Cl_9]^+$

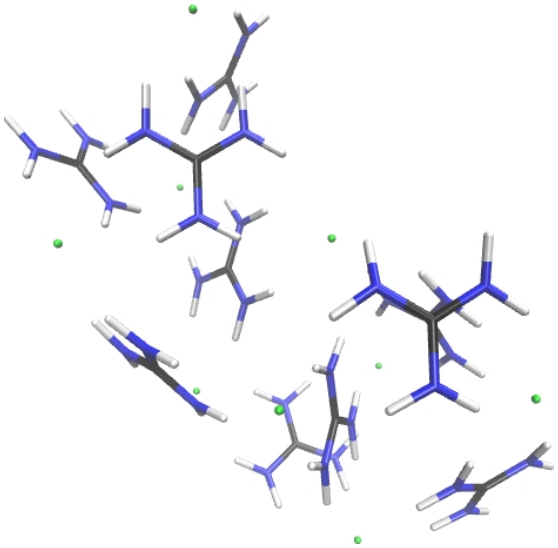

Figure 5.34 – guan10P_1355, the lowest-energy **B3LYP-D** $[Guan_{10}Cl_9]^+$ structure.

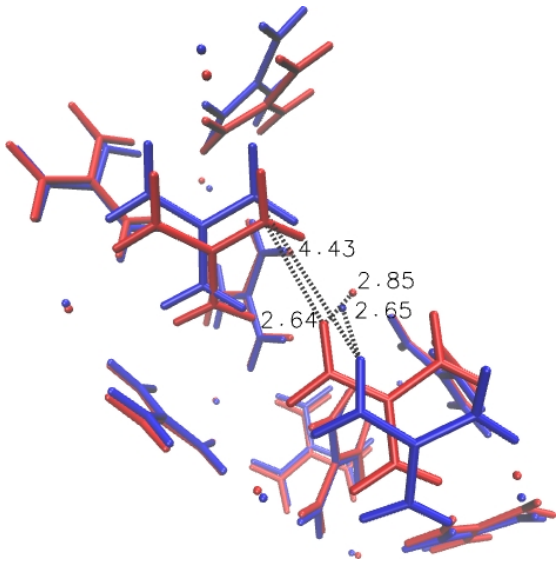

Figure 5.35 – Overlay of **guan10P_1355** with second-lowest-energy unique $[Guan_{10}Cl_9]^+$ structure, **guan10P_8290**.

Table 5.24 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_{10}Cl_9]^+$ (kJ mol^{-1}).

$[Guan_{10}Cl_9]^+$	**E**	**G**
guan10P_1355	0.2	0.0
guan10P_2103	0.0	2.1
guan10P_0006	0.5	3.9
guan10P_8290	13.0	12.0
guan10P_6763	49.4	66.1

guan10P_1355, the lowest-energy structure calculated for $[Guan_{10}Cl_9]^+$, also has one guanidinium inside of the cluster, shown above in the centre of Figure 5.34. The higher energy structure being used for comparison has a deviation relating to formation of a hydrogen bond between two NH_2 groups. This can be seen in Figure 5.35 where the H$\cdots$$Cl^-$ bond distance for the blue structure is 2.65 Å, and the NH-H$\cdots$$NH_2$ distance is

4.43 Å, while in the red structure the H$\cdots$Cl$^-$ distance is 2.85 Å, while the NH-H$\cdots$NH$_2$ distance is 2.64 Å.

Fingerprint region - [Guan$_{10}$Cl$_9$]$^+$

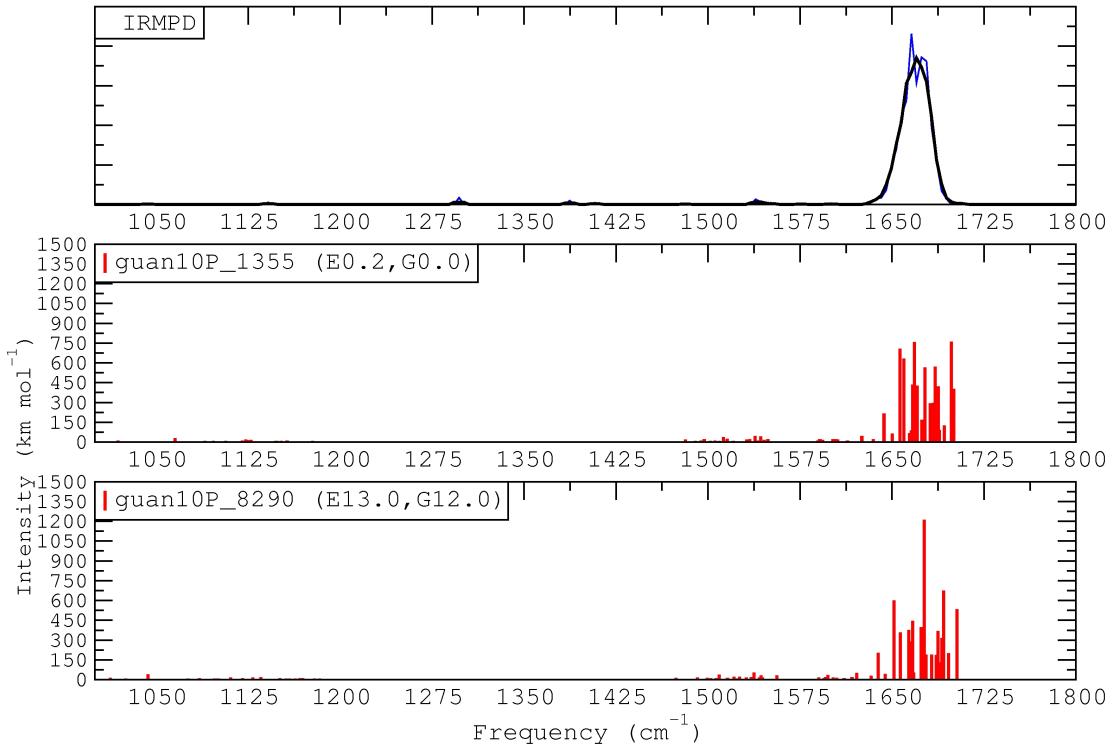

Figure 5.36 – [Guan$_{10}$Cl$_9$]$^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.98)

The experimental spectrum for [Guan$_{10}$Cl$_9$]$^+$ consists of a relatively narrow C-N band centred at approximately 1675 cm^{-1} and the NH$_2$ band at ≈1600 cm^{-1} is not observed, although some very small intensity is seen between 1500-1600 cm^{-1}. The absence of this band is most likely due to the experimental conditions, where an insufficient laser intensity was applied. The calculated C-N band between 1645-1700 cm^{-1} for guan10P_1355, the lowest-energy calculated structure for [Guan$_{10}$Cl$_9$]$^+$, match well to the C-N stretching band, while the most blue C-N mode, at 1698 cm^{-1} lies just outside of the experimental band.

3000-3600 cm^{-1} - [Guan$_{10}$Cl$_9$]$^+$

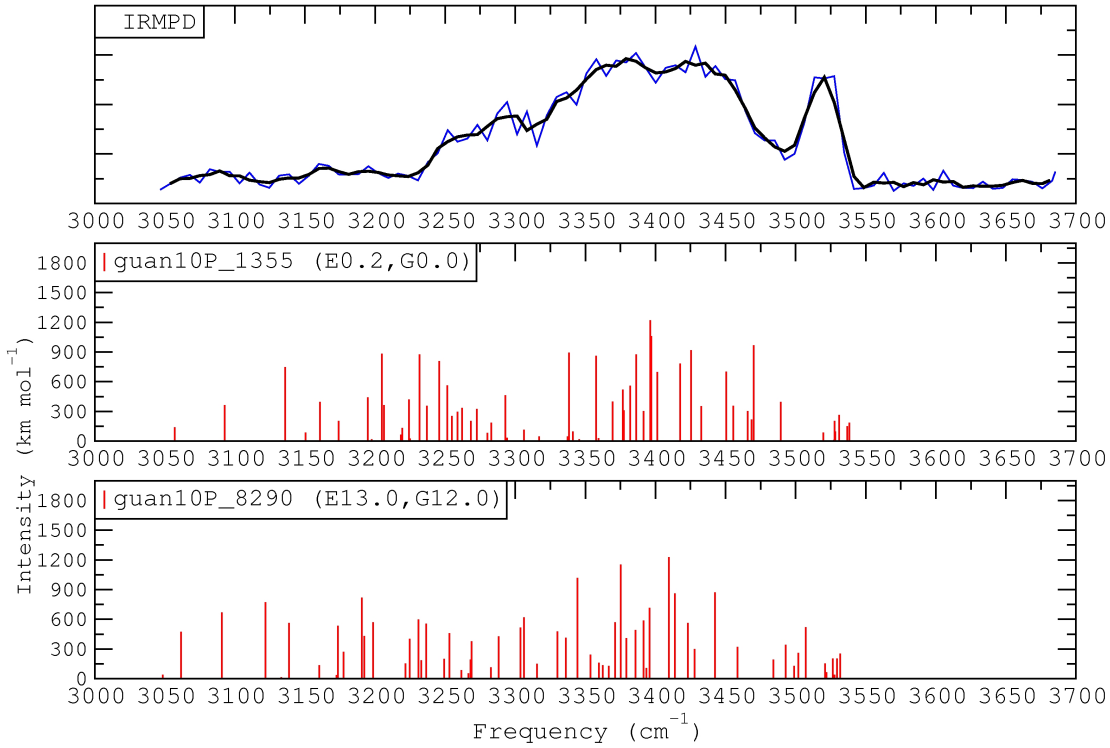

Figure 5.37 – [Guan$_{10}$Cl$_9$]$^+$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.96)

Calculated and experimental spectra for [Guan$_{10}$Cl$_9$]$^+$ are shown in Figure 5.37. Although this experimental spectrum has a worse signal to noise ratio, the main features are very similar to the spectrum in Figure 5.31 for [Guan$_6$Cl$_5$]$^+$. Again, the bound symmetric stretching modes are not treated properly within the harmonic approximation and these intense modes in the calculated spectrum are red shifted and broadened and seemingly unobserved experimentally. The mostly asymmetric bands in the 3350-3450 cm^{-1} region are again followed by a drop in intensity and bands of the asymmetric and free N-H modes. These features all match very well to the two calculated spectra, which are again for nearly identical structures. The structural shift to accommodate an additional hydrogen bond in the calculated structure guan10P_8290 (see Figure 5.35) seems to result in slight differences

in both the 3470 and 3300 cm^{-1} regions. While these modes are fairly coupled and difficult to assign, the observed spectral differences in these regions are associated with vibrations of guanidinium ions affected by this structural variation. Comparing again to the 3000-3600 cm^{-1} region of $[Guan_5Cl_4]^+$ and $[Guan_6Cl_5]^+$ in Figures 5.27 and 5.31, a clear emergence of three main vibrational bands (the free symmetric band, bound asymmetric band and finally the free asymmetric band) can again be noted (see Figure 5.70).

$[Guan_{15}Cl_{14}]^+$

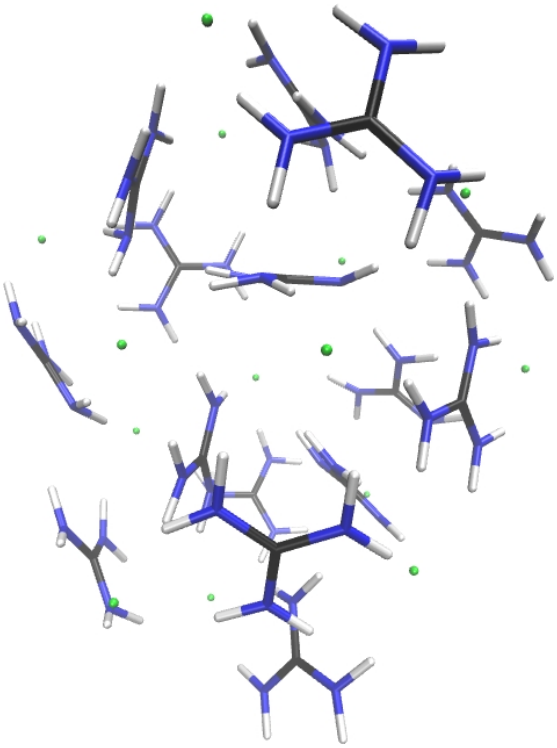

Figure 5.38 – guan15P_8412, the lowest-energy **B3LYP-D** $[Guan_{15}Cl_{14}]^+$ structure.

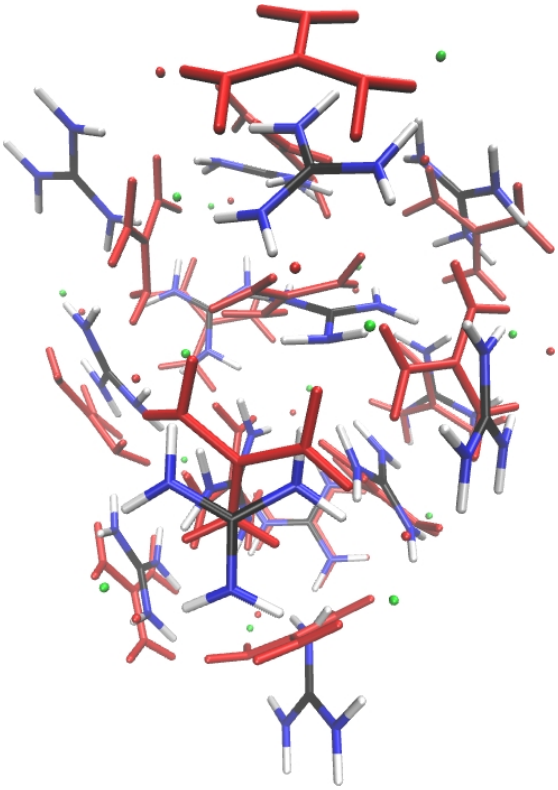

Figure 5.39 – Overlay of third lowest-energy structure **guan15P_7256** (coloured by elements) with second-lowest-energy **B3LYP-D** $[Guan_{15}Cl_{14}]^+$ structure, **guan15P_9579**.

The calculated lowest-energy $[Guan_{15}Cl_{14}]^+$ structures have three guanidinium ions contained by the exterior cage-like structure of guanidinium ions. The three lowest-energy

Table 5.25 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_{15}Cl_{14}]^+$ (kJ mol^{-1}).

$[\mathbf{Guan_{15}Cl_{14}}]^+$	**E**	**G**
guan15P_8412	0.0	0.0
guan15P_9579	6.3	11.4
guan15P_7256	22.7	11.8
guan15P_5440	23.1	12.6
guan15P_6413	28.0	21.9

conformations, spanning 11.8 kJ mol^{-1} in free energy (298 K), differ mainly by the orientation of two guanidinium ions, one on the top left of the structure and one on the very bottom of the structure (Figure 5.39). In the third lowest-energy structure, guan15P_7256, both of these guanidinium ions have two N-Hs (on different NH_2 groups) free and pointing outwards, away from the centre of the structure. In the second-lowest-energy structure, guan15P_9579, neither guandinium ion is orientated to have these free outward-pointing N-H groups, while in the lowest-energy structure, guan15P_8412, one guandinium ion at the bottom of the structure (Figure 5.38) has two free N-Hs. The presence of these free N-H groups in guan15P_7256 and guan15P_8412 could possibly result in a distinct feature in an experimental spectrum in the 3000-3600 cm^{-1} region, potentially allowing these two conformations to be distinguished from guan15P_9579.

Fingerprint region - $[Guan_{15}Cl_{14}]^+$

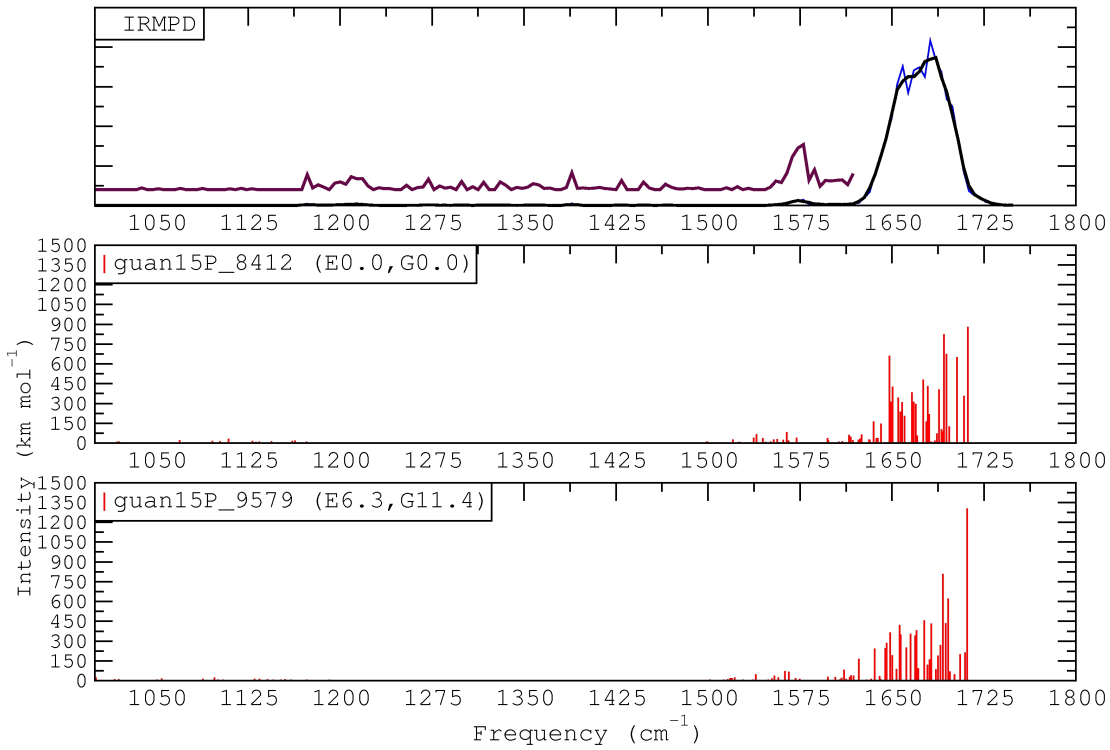

Figure 5.40 – $[Guan_{15}Cl_{14}]^+$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points and the **purple** line is a six-times magnification. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.98)

The experimental spectrum of $[Guan_{15}Cl_{14}]^+$ consists of a C-N band centred at approximately 1675 cm^{-1} and a weakly observed NH_2 band just below 1600 cm^{-1} (again, the low relative intensity is likely only due to light fragmentation conditions in the experiment). The region of C-N modes in the calculated spectrum of guan15P_8412 between 1645-1710 cm^{-1} matches well with the experimental band, while the NH_2 scissor modes are found slightly red shifted (10-15 cm^{-1}), starting at 1560 cm^{-1} in the experiment. For guan15P_8412, low intensity calculated bands at 1630 and 1635 cm^{-1} correspond to the NH_2 motions of the guanidinium ions having free outward-pointing N-H groups (see structural descriptions of guan15P_8412, guan15P_9579 and guan15P_7256 above), but are convoluted with the much larger mainly C-N band.

5.4.4 Results - $[Guan_xCl_{(x+1)}]^-$

$[Guan_2Cl_3]^-$

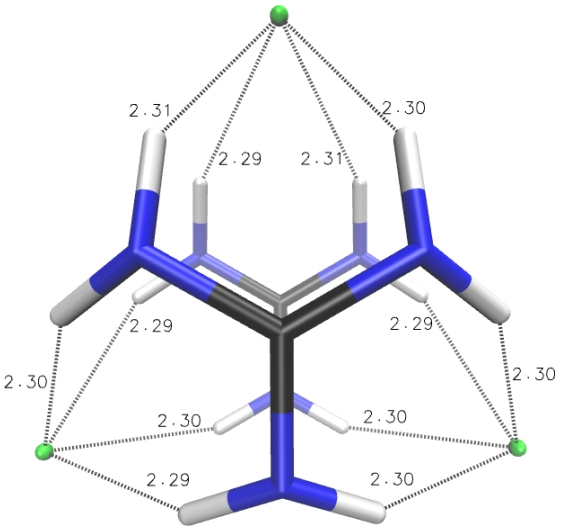

Figure 5.41 – guan2N_1493, the lowest-energy **B3LYP-D** $[Guan_2Cl_3]^-$ structure.

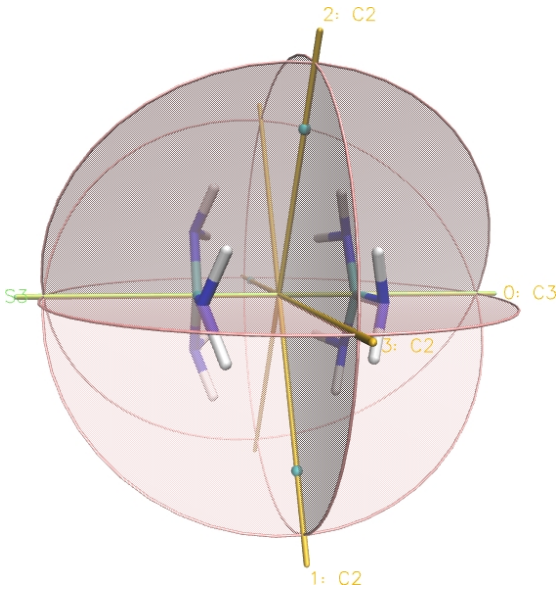

Figure 5.42 – guan2N_1493 with symmetry features illustrated (D_{3h}).

Table 5.26 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_2Cl_3]^-$ (kJ mol^{-1}).

$[Guan_2Cl_3]^-$	E	G
guan2N_1493	0.00	0.0
guan2N_1392	0.02	-
guan2N_0989	0.03	-
guan2N_1474	0.04	-
guan2N_1461	0.41	0.1

$[Guan_2Cl_3]^-$ forms a highly symmetrical, compact and tightly bound cluster, presented in Figures 5.41 and 5.42 above. Each guanidinium forms six $H\cdots Cl^-$ interactions at 2.3 Å. The interactions with chloride draw inwards on the NH_2 groups, resulting in quite non-planar guanidinium ions in which NH_2 groups are rotated 22 degrees out of the CN_3 plane. This structure, in the D_{3h} point group, has two mirror planes, three C_2 axes as well

as a principle C_3 axis containing and connecting carbon atoms on both guanidinium ions. All five **B3LYP-D** calculated structures converged to structures essentially identical to the one shown above.

Fingerprint region - $[Guan_2Cl_3]^-$

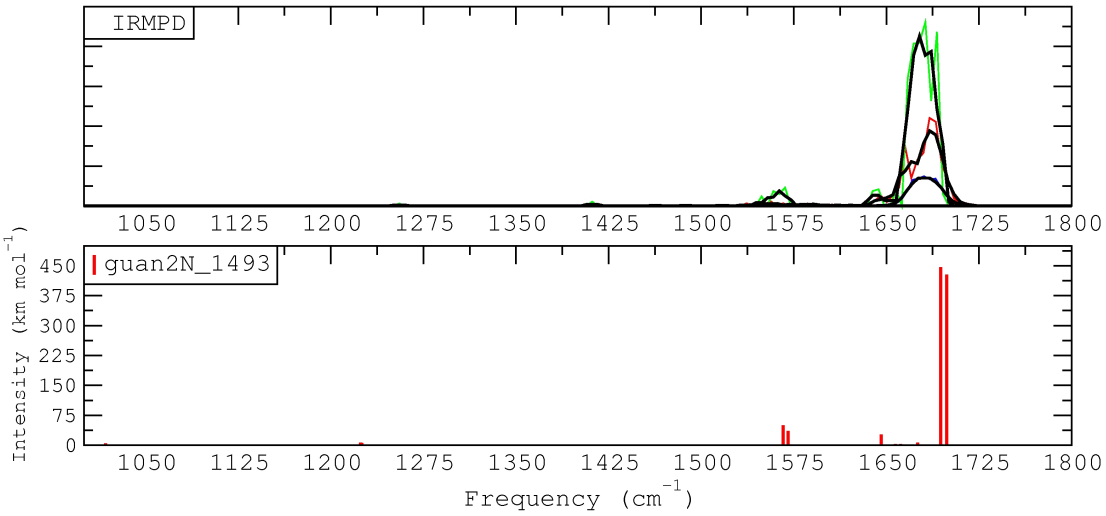

Figure 5.43 – $[Guan_2Cl_3]^-$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.98)

The calculated $[Guan_2Cl_3]^-$ spectrum matches very well to all three of the experimental scans. All modes are strongly coupled between guanidinium ions and the largest peak represents C-N modes moving concertedly in phase in both guanidinium ions. These modes are calculated to be at 1695 and 1700 cm^{-1}, while calculated C-N modes where the vibrations of each guanidinium ion are out of phase, are found at ≈1660 cm^{-1} with nearly zero intensity. The small intensity centred at 1565 cm^{-1} is attributed to coupled NH_2 scissor and C-N stretching modes, where corresponding groups on each guanidinium ion move in-phase (calculated at 1565 and 1570 cm^{-1}). The analogous out-of-phase modes are found at 1545 and 1550 cm^{-1} with nearly zero intensity. Finally, purely NH_2 scissor modes are calculated at 1646 and 1675 cm^{-1}. The lower-frequency band corresponds to motion on different guandinium ions being out-of-phase, and is observed experimentally

at 1640 cm^{-1}, while the higher intensity band corresponds to the analogous in-phase motion and is convoluted within the much larger band of C-N stretches. All band positions match very well between the experimental spectrum and the calculated spectrum for the lowest-energy structure, guan2N_1493. In this structure (see Figure 5.42) every NH_2 and C-N environment is identical, resulting in the very narrowly distributed modes seen in both experimental and calculated spectra.

$[Guan_3Cl_4]^-$

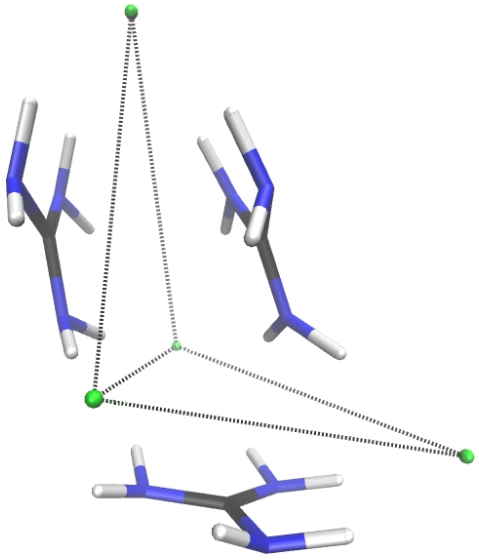

Figure 5.44 – guan3N_6935, the lowest-energy **B3LYP-D** $[Guan_3Cl_4]^-$ structure.

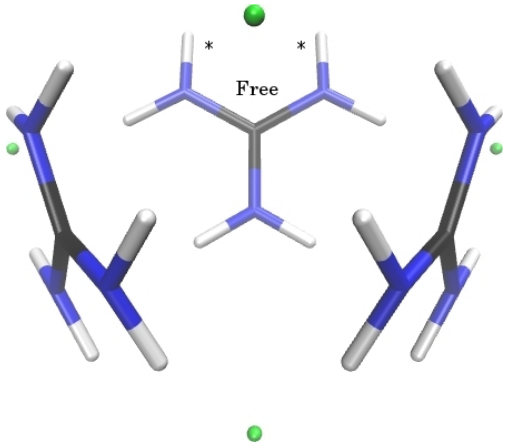

Figure 5.45 – guan3N_6589, the second-lowest-energy **B3LYP-D** $[Guan_3Cl_4]^-$ structure (C_s).

Table 5.27 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_3Cl_4]^-$ (kJ mol^{-1}).

$[Guan_3Cl_4]^-$	**E**	**G**
guan3N_6935	0.0	0.0
guan3N_6589	14.5	22.0
guan3N_8794	14.5	-
guan3N_8553	14.5	-
guan3N_6436	14.6	-

Of the five calculations for $[Guan_3Cl_4]^-$ listed in Table 5.27, all converged to only one

of two unique structures. The four highest energy structures converged to guan3N_6859, the symmetric (C_s point group) structure in Figure 5.44. This structure is 22.0 kJ mol^{-1} higher in energy than guan3N_6935, the lowest-energy structure, in Figure 5.45. guan3N_6935 is quite similar to guan2N_1493, the lowest-energy $[Guan_2Cl_3]^-$ structure (Figure 5.41), but accommodates an additional guanidinium chloride neutral. Each guanidinium ion has six H$\cdots$Cl$^-$ interactions, compared to the symmetric guan3N_6859 structure where there are two free N-Hs (with "*" in Figure 5.45) and two NH-H$\cdots$NH$_2$ hydrogen bonds. This structure is of C_s symmetry with a mirror plane in the vertical position and perpendicular to the plane of the page.

3000-3600 cm^{-1} region - $[Guan_3Cl_4]^-$

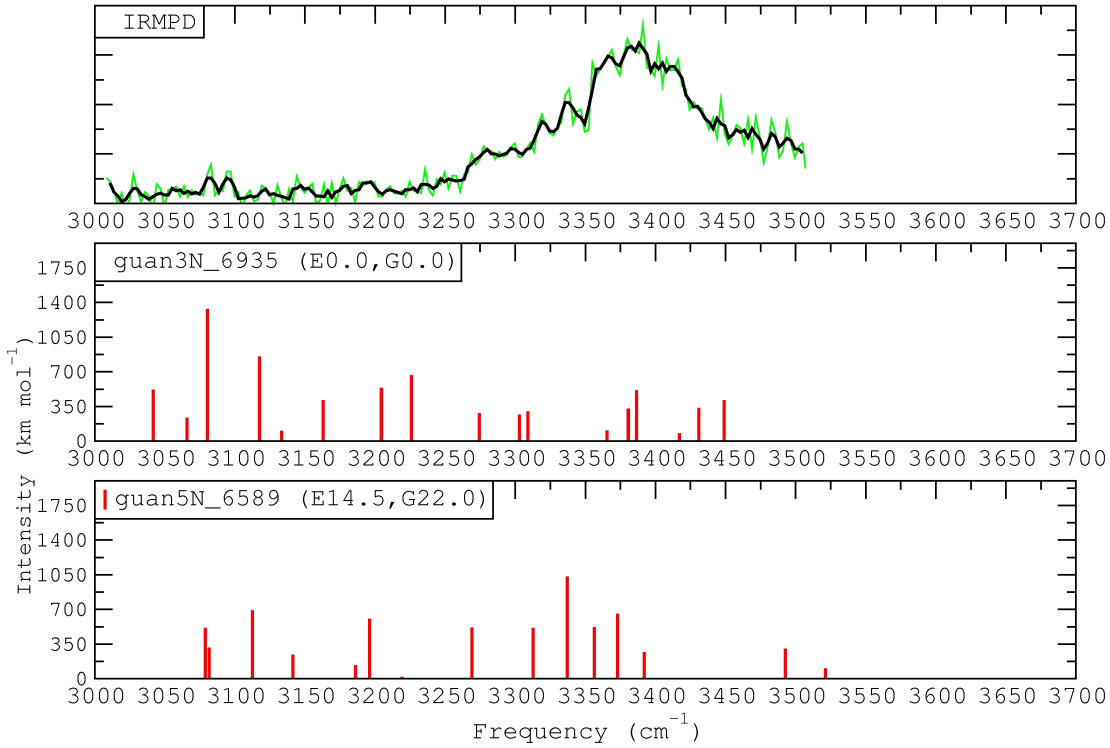

Figure 5.46 – $[Guan_3Cl_4]^-$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.96)

Symmetric N-H stretching modes are calculated to be in the 2985-3305 cm^{-1} range

and asymmetric stretching modes in the 3225-3450 cm^{-1} range for the lowest-energy structure, guan3N_6935. Unfortunately, due to unintended interruption of the experiment, this scan does not go beyond 3500 cm^{-1}. However, the range of 3226-3449 cm^{-1} appears to match the experimental result quite closely. The second-lowest-energy structure, guan3N_6589, has calculated asymmetric N-H stretches in the 3270-3520 cm^{-1} range, potentially being a better match to the experiment since the intensity at 3500 cm^{-1} in the experimental spectrum has not fully dropped to the baseline. Additional experiments would need to be completed to confirm structural assignment of this cluster. From an energy standpoint, guan3N_6935 is significantly favoured, being 22.0 kJ mol^{-1} lower in free energy at the **B3LYP-D** level.

$[Guan_4Cl_5]^-$

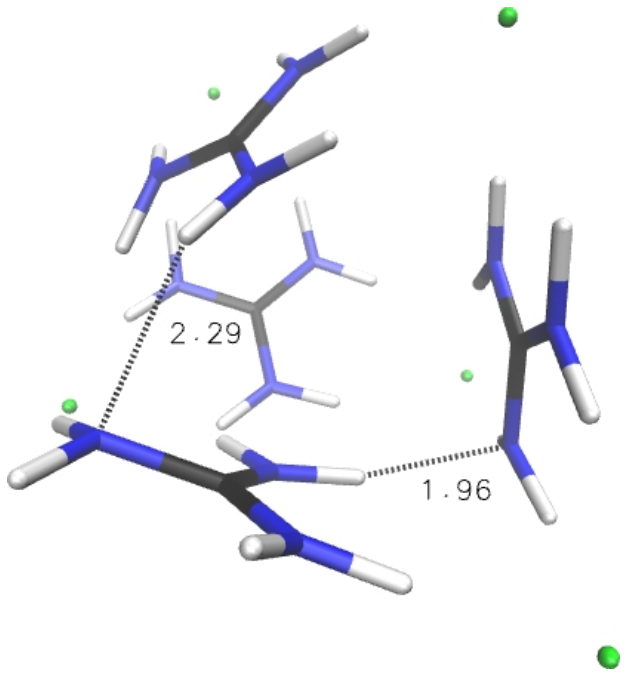

Figure 5.47 – guan4N_3607, the lowest-energy **B3LYP-D** $[Guan_4Cl_5]^-$ structure.

All five **AMOEBA** calculated structures of $[Guan_4Cl_5]^-$ converged to one common structure after DFT geometry optimization and are listed in Table 5.28. Two intermolecular hydrogen bonds are formed between guanidinium ions of 1.96 and 2.29 Å, illustrated

Table 5.28 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_4Cl_5]^-$ (kJ mol^{-1}).

$[Guan_4Cl_5]^-$	E	G
guan4N_3607	0.0	0.0
guan4N_4855	0.1	-
guan4N_5867	0.1	-
guan4N_3182	0.1	-
guan4N_2841	0.3	0.1

in Figure 5.47.

Fingerprint region - $[Guan_4Cl_5]^-$

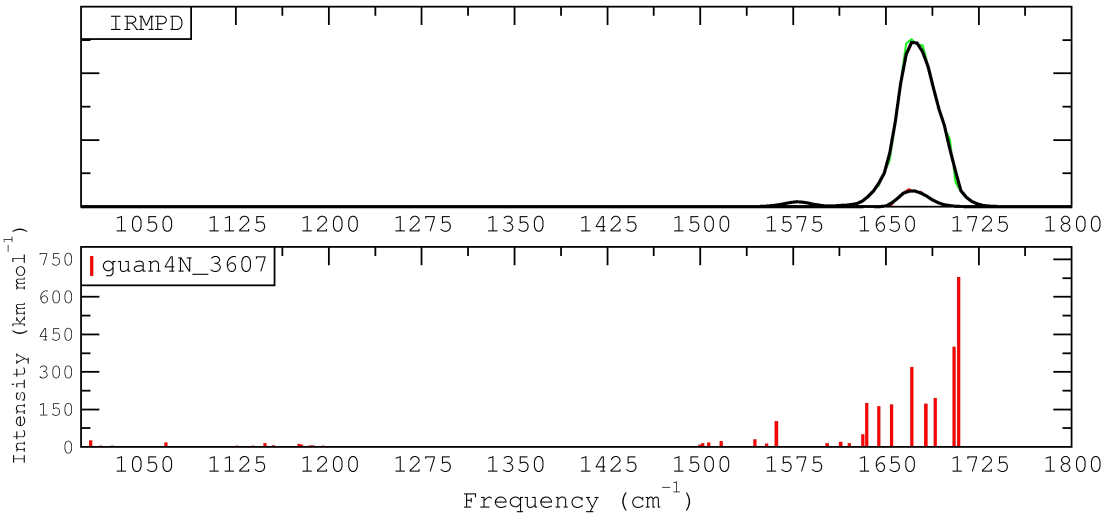

Figure 5.48 – $[Guan_4Cl_5]^-$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.98)

In the fingerprint spectrum of $[Guan_4Cl_5]^-$, calculated modes from 1655 cm^{-1} to 1710 cm^{-1} correspond to C-N stretching modes, while modes between 1500 and 1645 cm^{-1} correspond to NH_2 scissor modes. Already at this size of cluster it is very difficult to discuss assignments of individual vibrations as most modes are coupled and delocalized over different guanidinium ions. Experimental peak centre positions match well to computed line positions (guan4N_3607), except for a small (approximately 10 cm^{-1}) red shift of the

1500-1575 cm^{-1} modes. Comparing to $[Guan_2Cl_3]^-$, modes are more widely distributed due to a wider range of bonding environments.

$[Guan_5Cl_6]^-$

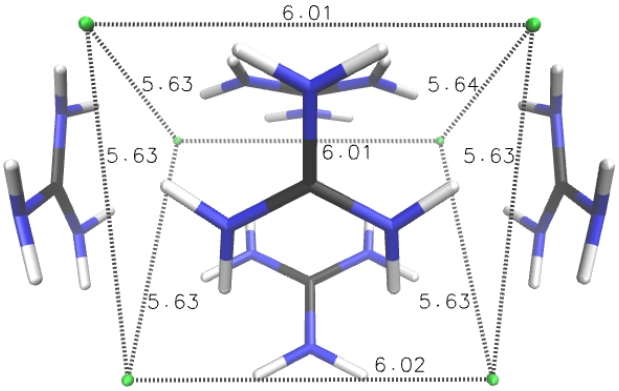

Figure 5.49 – guan5N_6254, the lowest-energy **B3LYP-D** $[Guan_5Cl_6]^-$ structure.

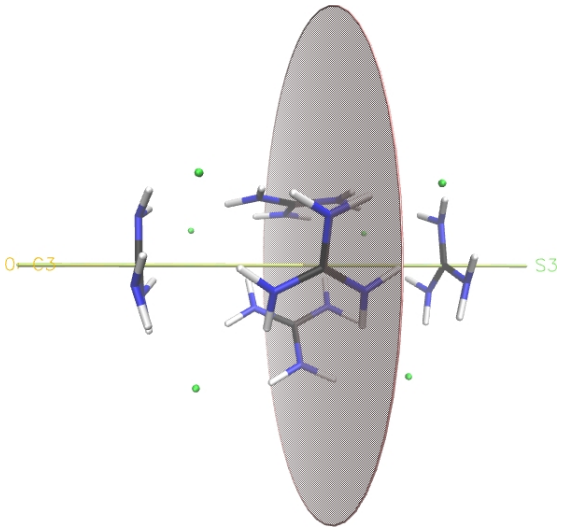

Figure 5.50 – guan5N_6254 with symmetry features illustrated (C_{3h}).

Table 5.29 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_5Cl_6]^-$ (kJ mol^{-1}).

$[Guan_5Cl_6]^-$	E	G
guan5N_6254	0.0	0.0
guan5N_8044	0.2	-
guan5N_6893	0.2	-
guan5N_7687	0.4	-
guan5N_6890	0.4	1.8

All calculations propose that $[Guan_5Cl_6]^-$ adopts a very interesting structure contained by a triangular prism. Although a triangular prism has D_{3h} symmetry, the three guanidinium ions on the joining (side) faces are not centred on the faces and remove the three mirror planes that would otherwise be present in a prismatic structure. In the C_{3h} point group, guan5N_6254 has a principle C_3 axis of rotation and a mirror plane containing the three carbon atoms of the joining faces. All five calculated structures converged to the

conformation of guan5N_6254 during geometry optimization. All of the 30 available N-H sites are interacting with chloride, with an average distance of 2.34 Å.

Fingerprint region - $[Guan_5Cl_6]^-$

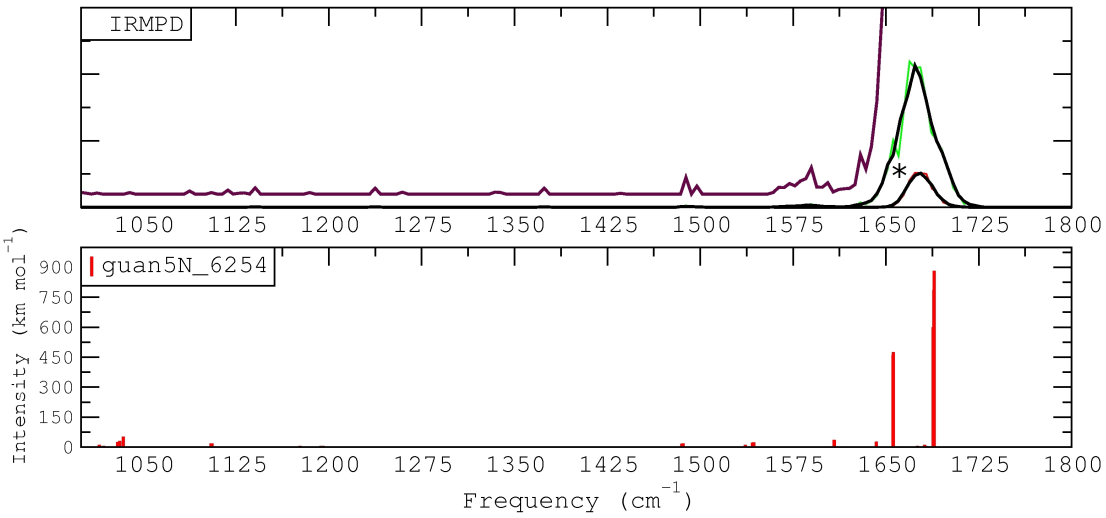

Figure 5.51 – $[Guan_5Cl_6]^-$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points and the **purple** line is a six-times magnification. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.98)

The calculated and experimental spectra for $[Guan_5Cl_6]^-$ both demonstrate sharp and tightly bunched modes resulting from a highly-symmetric structure with few variations in bonding environments and high levels of mode coupling and delocalization. The most intense calculated band centred at 1690 cm^{-1} (mainly C-N scissoring modes) matches both experimental peak positions very well and the second most intense peak (to the red at 1655 cm^{-1}) is convoluted and seemingly only observed as a small shoulder of the more intense absorption (indicated in the green experimental spectrum with "*"). NH_2 modes are weakly observed between 1470 and 1640 cm^{-1}.

3000-3600 cm^{-1} region - [Guan$_5$Cl$_6$]$^-$

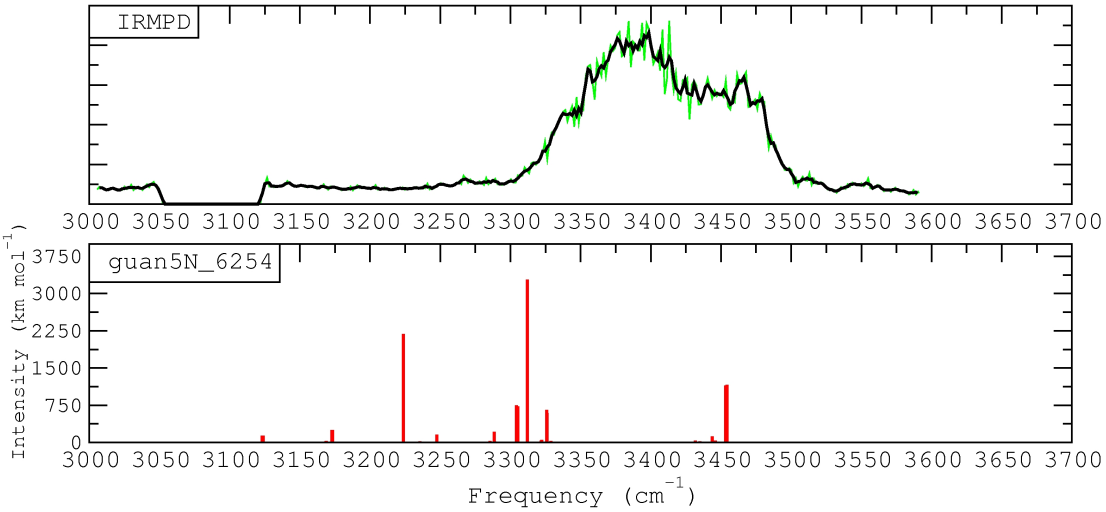

Figure 5.52 – [Guan$_5$Cl$_6$]$^-$ 3000-3600 cm^{-1} region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.96)

Referring to the central and front guanidinium ion in Figure 5.49, one can see that there exist different types of N-H$\cdots$Cl$^-$ interactions. The four N-Hs pointing upwards and interacting with chlorides of the top corners of the structure have an average H$\cdots$Cl$^-$ length of 2.30 Å, while the bottom two downward facing N-Hs interacting with the chlorides at the bottom corners of the structure have an average H$\cdots$Cl$^-$ of 2.56 Å. The other two guanidinium ions on the central sides are equivalent, resulting in a total of six equivalent weakly bound N-Hs with an average H$\cdots$Cl$^-$ of 2.56 Å. The coupled N-H stretching modes of these six N-Hs are found closely grouped at ≈3450 cm^{-1} and match well to the experimental spectrum. All other N-H modes (3125-3325 cm^{-1}) in this structure correspond to more strongly hydrogen-bound groups and are likely to be broadened and red shifted, as previously seen for these types of modes in other clusters. These modes are also calculated to be highly delocalized over the whole cluster, with many containing both symmetric and asymmetric character. Experimental scans of [Guan$_5$Cl$_6$]$^-$ again do not contain bands corresponding to the heavily red shifted and broadened bound symmetric N-H stretching

modes harmonically calculated to be of high intensity (≈3225 cm^{-1}). Calculated structure, guan5N_6254, has asymmetric N-H stretching modes in the range of 3235-3455 cm^{-1} appearing to match poorly to both the range of the experimental band and the position. A gap exists in the calculated modes between 3330-3430 cm^{-1} which is not observed in the experimental spectrum. The drop in intensity between 3050 and 3125 cm^{-1} is a result of a temporary experimental malfunction. Again, further discussion of these discrepancies will follow in section 5.4.6, where results at different levels of theory are able to offer some possible explanations to the disagreement between the textbfB3LYP-D modes and the experimental spectrum.

$[Guan_6Cl_7]^-$

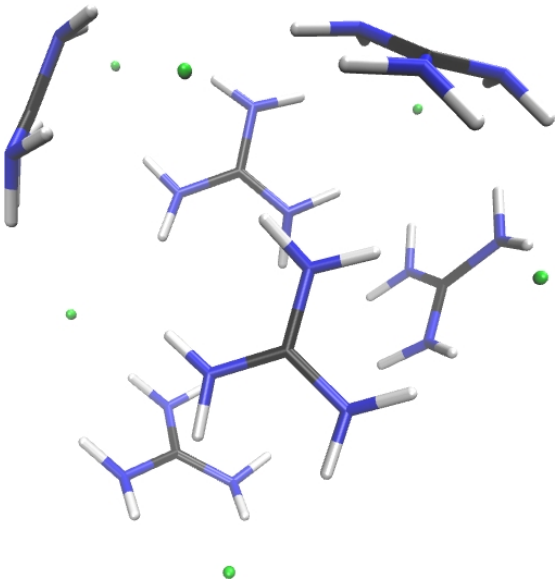

Figure 5.53 – guan6N_6159, the lowest-energy **B3LYP-D** $[Guan_6Cl_7]^-$ structure.

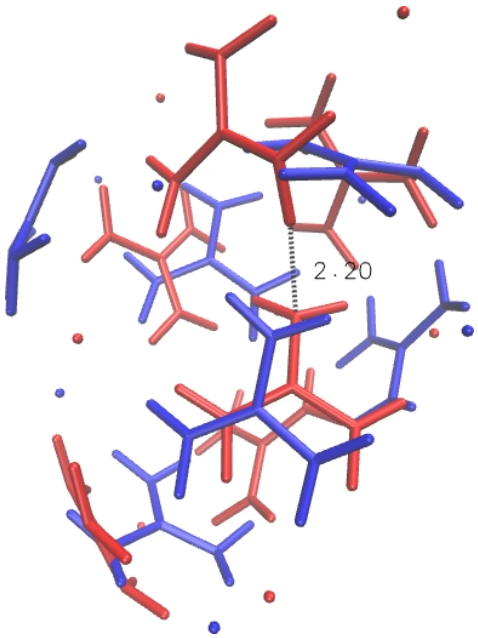

Figure 5.54 – Overlay of **guan6N_6159** with the second-lowest-energy unique **B3LYP-D** $[Guan_6Cl_7]^-$ structure, **guan6N_3088**.

The four lowest-energy structures calculated for $[Guan_6Cl_7]^-$ converged to the same structure shown in Figure 5.53, guan6N_6159. In this structure, all N-H protons are interacting with chloride ions. The fifth-lowest-energy structure was optimized to a unique

Table 5.30 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_6Cl_7]^-$ (kJ mol^{-1}).

$[Guan_6Cl_7]^-$	E	G
guan6N_6159	0.0	0.0
guan6N_3100	0.1	-
guan6N_9319	0.1	-
guan6N_7002	0.1	-
guan6N_3088	25.1	26.5

structure where not all the N-H protons interact with chloride, but an NH-H$\cdots$$NH_2$ hydrogen bond of 2.20 Å, as seen above in Figure 5.54, is formed and the relative free energy is increased by 26.6 kJ mol^{-1}.

Fingerprint region - $[Guan_6Cl_7]^-$

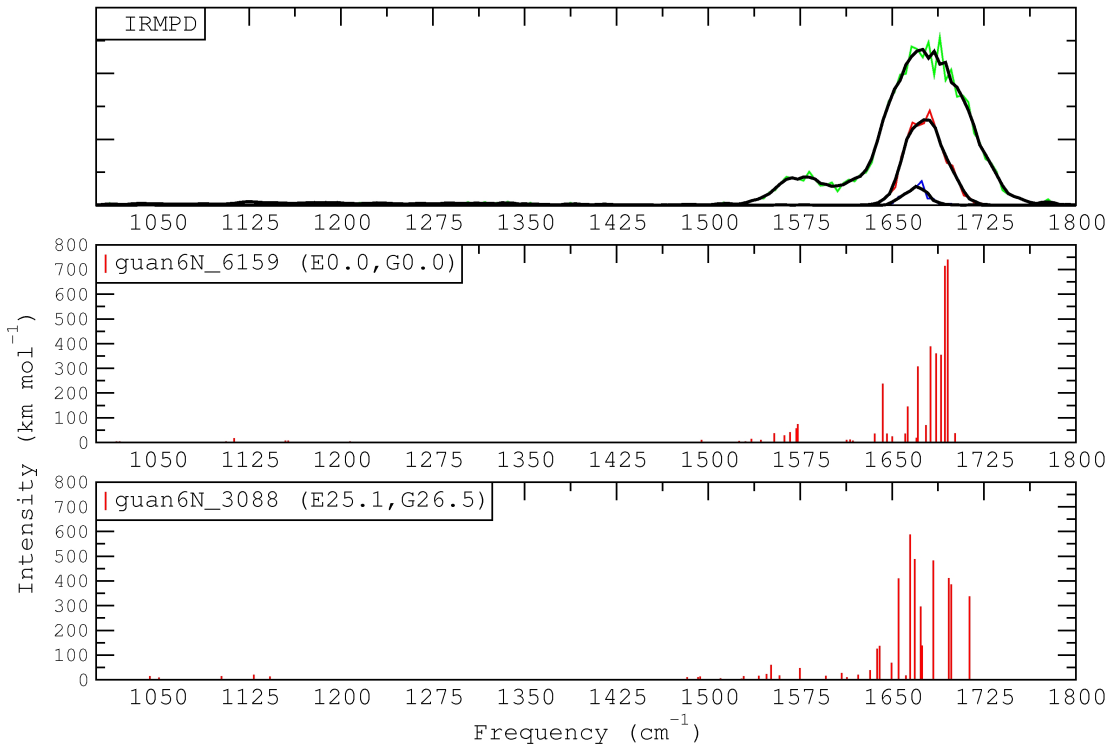

Figure 5.55 – $[Guan_6Cl_7]^-$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.98)

The computed line positions for the lowest-energy $[Guan_6Cl_7]^-$ structure, guan6N_6159, again appear to match very well to the experimental spectra. Calculated C-N stretching modes are in the 1640-1700 cm^{-1} region and the NH_2 scissor modes in the 1495-1650 cm^{-1} region.

$[Guan_{10}Cl_{11}]^-$

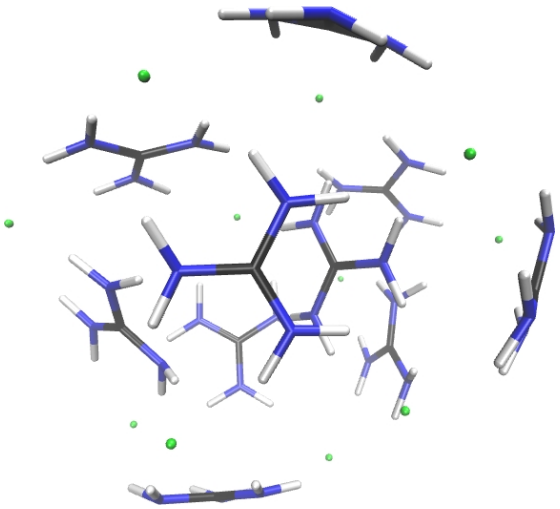

Figure 5.56 – guan10N_8756, the lowest-energy **B3LYP-D** $[Guan_{10}Cl_{11}]^-$ structure.

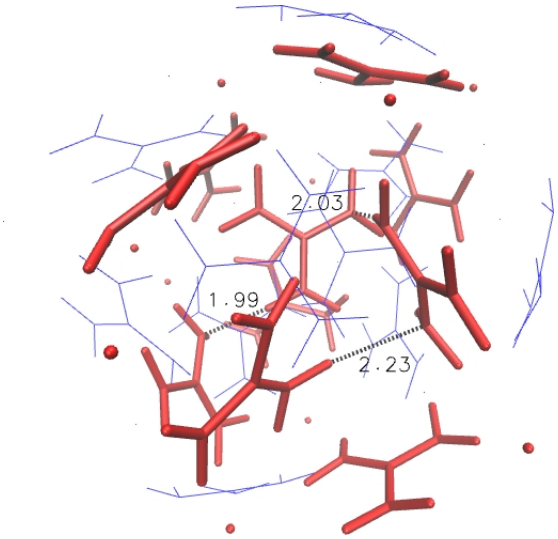

Figure 5.57 – Overlay of **guan10N_8756** with the second-lowest-energy unique **B3LYP-D** $[Guan_{10}Cl_{11}]^-$ structure, **guan10N_7884**.

Table 5.31 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_{10}Cl_{11}]^-$ (kJ mol^{-1}).

$[Guan_{10}Cl_{11}]^-$	**E**	**G**
guan10N_8756	0.0	0.0
guan10N_2476	0.3	-
guan10N_7884	34.5	27.4
guan10N_6669	34.5	-
guan10N_0818	34.5	-

Two unique structures resulted from the five **B3LYP-D** optimizations for $[Guan_{10}Cl_{11}]^-$ and are shown in the Figures above. guan10N_8756 is the lowest-energy structure calculated, while guan10N_7884 is 27.4 kJ mol^{-1} higher in relative free energy as a result of

the shift of the a guanidinium ion "inside" of the structure forming three NH-H$\cdots$NH$_2$ hydrogen bonds at 1.99, 2.03 and 2.23 Å.

Fingerprint region - $[Guan_{10}Cl_{11}]^-$

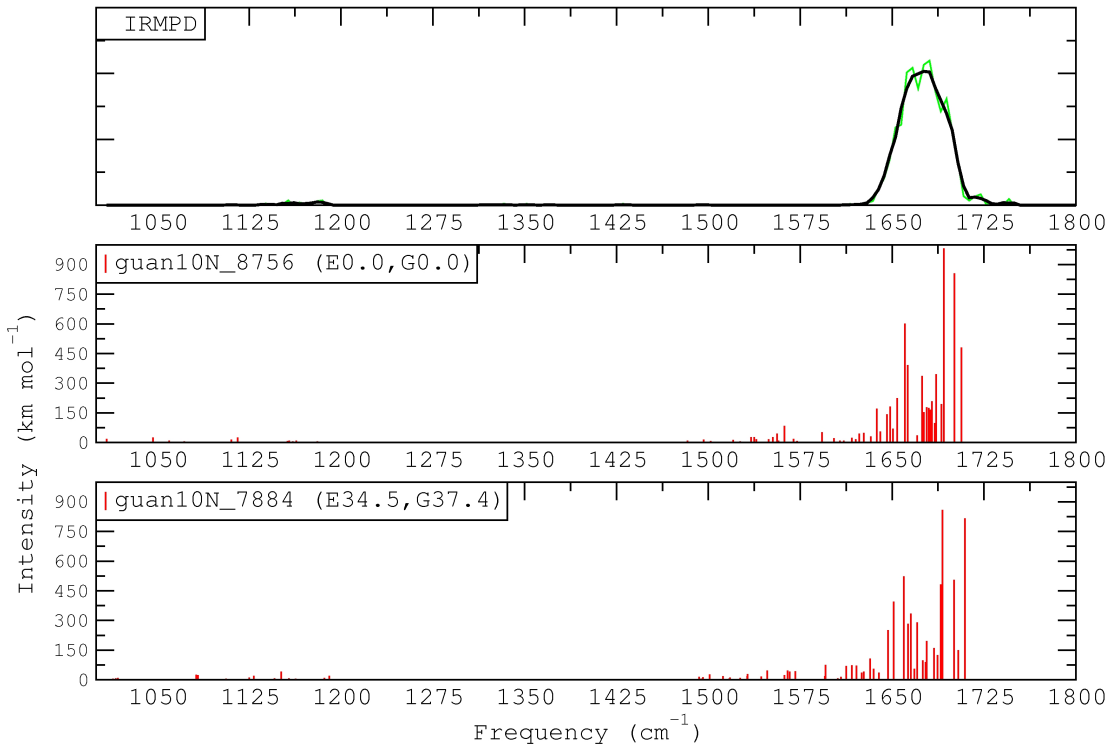

Figure 5.58 – $[Guan_{10}Cl_{11}]^-$ fingerprint region IRMPD experimental and calculated spectra. Relative electronic (E) and free (G, 298 K) energies are provided in parentheses. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrarily scaled. (kJ mol^{-1}, **B3LYP-D**, frequencies scaled by 0.98)

The experimental and calculated spectra for $[Guan_{10}Cl_{11}]^-$ (guan10N_8756) are very similar to those of $[Guan_6Cl_7]^-$ and $Guan_4Cl_5]^-$. This similarity is likely because none of these clusters have any symmetric or specific structural features which would narrow the distribution or increase bunching of modes (associated with highly similar bonding environments). C-N stretches are in the region of 1705-1640 cm^{-1} and NH$_2$ scissor modes are between 1640-1485 cm^{-1}.

$[Guan_{15}Cl_{16}]^-$ Structural Description

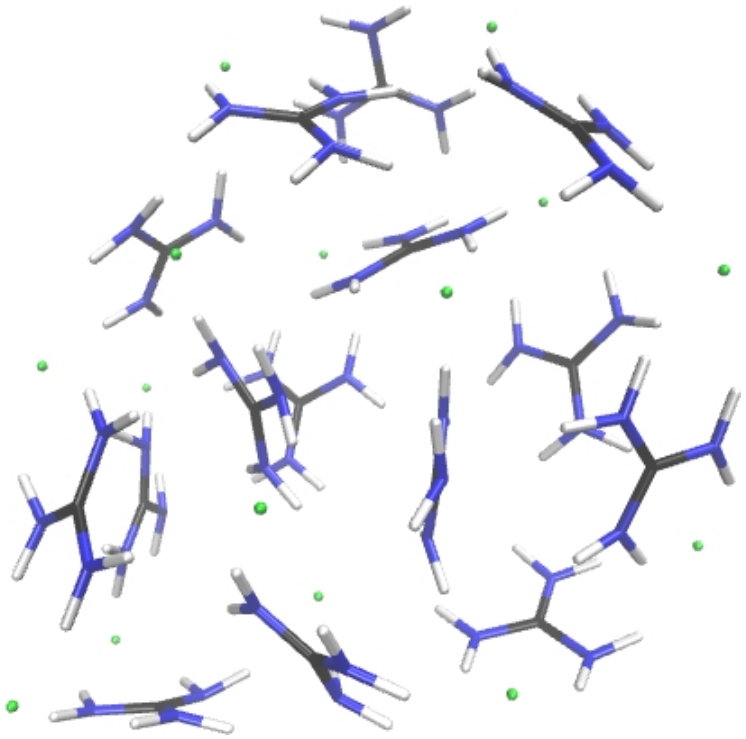

Figure 5.59 – guan15N_1414, the lowest-energy **B3LYP-D** $[Guan_{15}Cl_{16}]^-$ structure.

Table 5.32 – Relative **B3LYP-D** electronic (E) and free energies (G, 298 K) for $[Guan_{15}Cl_{16}]^-$ (kJ mol^{-1}).

$[Guan_{150}Cl_{16}]^-$	E	G
guan15N_1414	5.0	0.0
guan15N_3221	0.0	1.0
guan15N_0422	5.4	-
guan15N_7570	5.6	-
guan15N_5804	5.6	-

All geometries obtained at the **B3LYP-D** level for $[Guan_{15}Cl_{16}]^-$ are very similar having only minor structural variations, as illustrated by the range of relative energies in Table 5.32. Similar to $[Guan_{15}Cl_{14}]^+$, in this structure there are three enclosed guanidinium ions.

5.4.5 Effects of the Bonding Environment on Position of N-H Stretching Modes

The dependence of mode position on the bonding environment can be more generally illustrated using a set of structures with large and contrasting structural differences. Using $[Guan_5Cl_4]^+$ as an example (see also Figure 5.27 and the associated text for a complimentary discussion), an additional structure has been defined and optimized at the **B3LYP-D** level based on an extended chain of guanidinium and chloride ions, guan5P_chain. As well, a structure somewhat in between guan5P_chain and the compact guan5P_3998 conformation (the conformation suggested by 3000-3600 cm^{-1} region experiments), guan5P_b, was defined and optimized at the **B3LYP-D** level for the purpose of connecting the progression of spectral features from the compact guan5P_3998 structure to the extended guan5P_chain structure. These structures are shown in Figures 5.60 and 5.61. Guan5P_chain has an extended conformation where each chloride is interacting with only four N-H groups. The three middle guanidinium ions have four N-H$\cdots$$Cl^-$ interactions while the two end guanidinium ions have only two each. This conformation has a larger number of free N-H groups, as well as shorter average N-H$\cdots$$Cl^-$ interactions (2.20 Å). Guan5P_b is a more extended conformation than guan5P_3998, where three chlorides are coordinated five times and one six times, with an average NH$\cdots$$Cl^-$ distance of 2.32 Å. In guan5P_3998, one chloride is coordinated four times, one five times and two six times at an average N-H$\cdots$$Cl^-$ distance of 2.30 Å.

Spectral comparisons for guan5P_3998, guan5P_b, and guan5P_chain with experiments are presented in Figures 5.60 and 5.61. In the fingerprint region, the C-N band of guan5P_chain does not match the experimental band as well as for guan5P_3998 in both the shape and position of the band, with the experimental band extending 40 cm^{-1} to the red of the calculated modes. As well, a tight bunching of modes is observed, presumably a result of the limited number of very similar bonding environments found in this

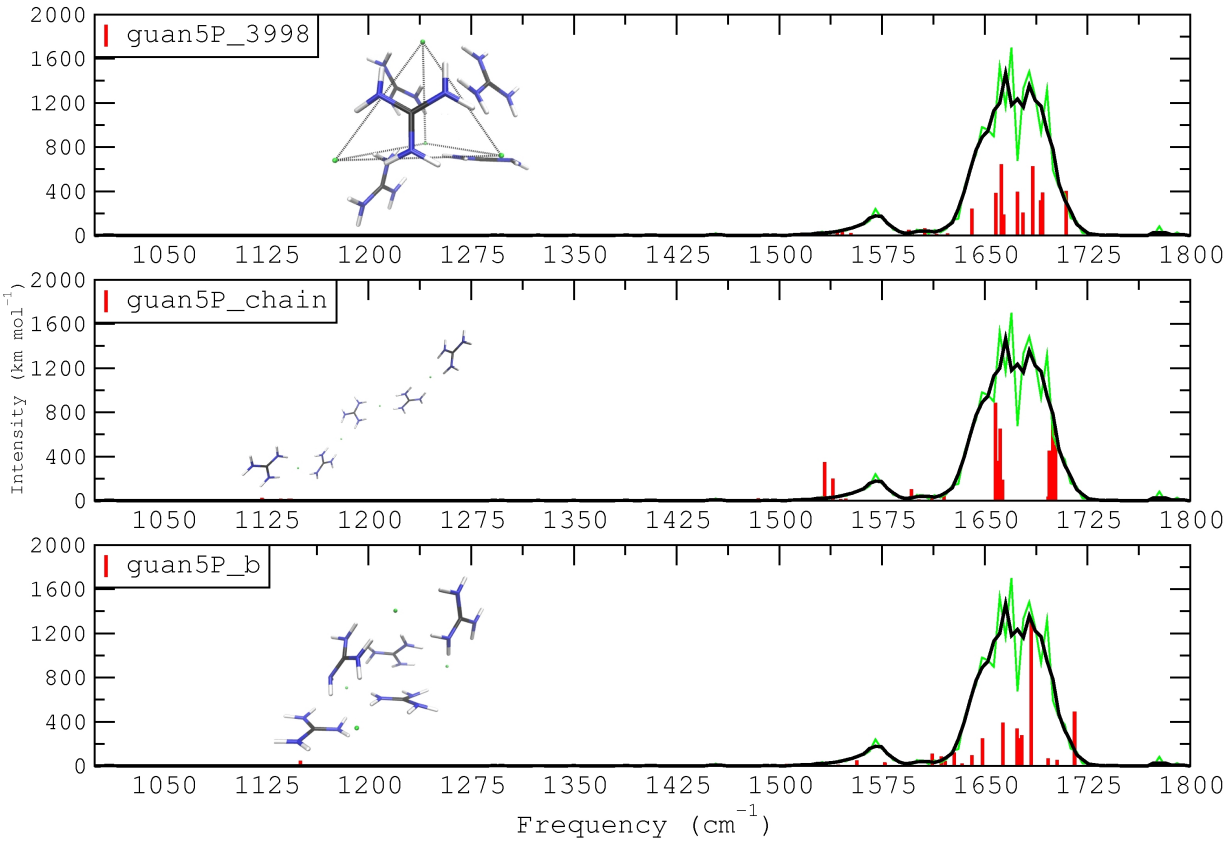

Figure 5.60 – Spectral comparisons of higher energy $[Guan_5Cl_4]^+$ conformations in the 3000-3600 cm^{-1} region. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axes) are arbitrarily scaled. **B3LYP-D** modes are scaled by 0.98.

conformation. guan5P_b matches significantly better, however the calculated C-N modes most to the blue are slightly outside of the experimental band and the overall shape of the band does not match as well as guan5P_3998. While the NH_2 modes of guan5P_b centred around 1575 cm^{-1} match well to the experiment, it was observed that harmonically calculated NH_2 modes were red shifted by approximately 35 cm^{-1} from the corresponding anharmonically calculated modes (see Figure 5.16). If this were the case for $[Guan_5Cl_4]^+$, it would imply that the NH_2 modes of guan5P_3998, if calculated anharmonically, would be in better agreement than those of guan5P_b.

In the 3000-3600 cm^{-1} region, the free asymmetric band at approximately 3540 cm^{-1} is found in all conformations in a very narrow range and all calculated bands match reason-

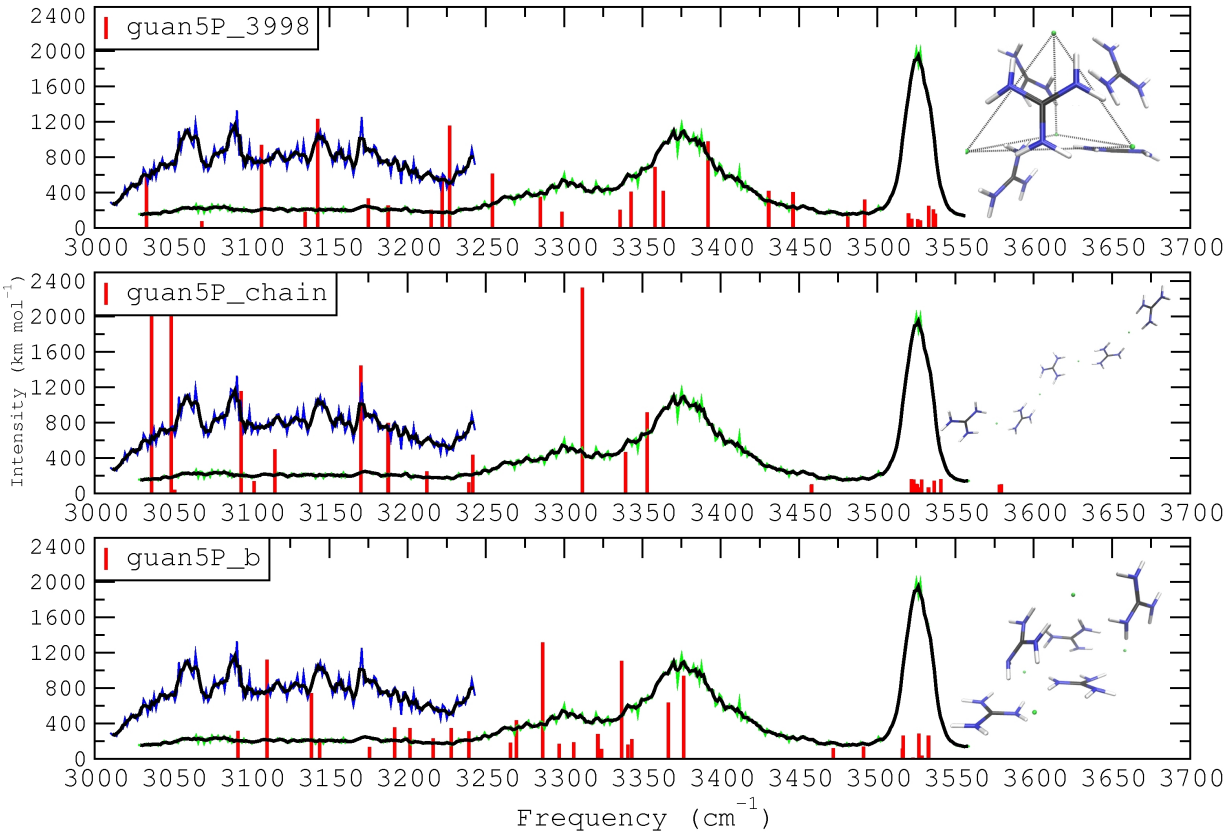

Figure 5.61 – Spectral comparisons of higher energy $[Guan_5Cl_4]^+$ conformations in the fingerprint region. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axes) are arbitrarily scaled. **B3LYP-D** modes are scaled 0.96.

ably well to the experimental band. In the 3375-3450 cm^{-1} region there is clear disagreement between calculated spectra, of both guan5P_chain and guan5P_b, and the experimental spectrum. The bound asymmetric band in guan5P_3998 between 3335-3445 cm^{-1}, accounts for this region very well, reducing the likelihood of more extended structures for $[Guan_5Cl_4]^+$. This region was earlier observed to be diagnostic while comparing low-energy conformations of $[Guan_5Cl_4]^+$ (guan5P_3998 and guan5P_6236). N-H$\cdots$Cl$^-$ interactions which were shorter and stronger resulted in a clear red shifting of the band, indicating a spectral region that is very sensitive to bonding environment.

5.4.6 Comparison of DFT Functionals for Calculating Vibrational Modes

In the figures below, harmonic (B3LYP/6-311+G(d,p), **B3LYP-D** and **M06**) and anharmonic (B3LYP/6-311+G(d,p)) results are compared. Anharmonic results provide an estimate of the magnitude of error associated with the use of the harmonic approximation. **M06** results, while not assumed in advance to be an improvement upon the **B3LYP-D** results, provide an additional reference. Anharmonic frequency calculations are very computationally costly and currently only possible for the smallest of cases examined here (see introduction of the text for a more detailed discussion of vibrational anharmonicity). Anharmonic frequencies were obtained at the B3LYP/6-311+G(d,p) level for $[Guan_2Cl]^+$ and are compared to harmonic results of the same level of theory in Figure 5.62. Intensities of the anharmonic modes are taken from the corresponding modes of the harmonic vibrational analysis. 3000-3600 cm^{-1} region harmonic modes are presented scaled, however the fingerprint region is presented unscaled as the magnitude of the correction is much smaller.

Fingerprint region spectra of $[Guan_2Cl]^+$ are presented in Figure 5.62. A small scaling factor (approximately 0.98 for **B3LYP-D** and 0.99 for M06) would be necessary for full agreement with the experimental spectrum. There is a relatively large discrepancy between the harmonic B3LYP/6-311+G(d,p) spectrum and the **B3LYP-D** spectrum. As the dispersion correction is not expected to have a large effect in cases with few non-bonded interactions (see Figure 5.65 and 5.66), it is likely that the difference is a basis set related effect. Anharmonic frequencies match very well to the experimental spectrum when scaled by 1.011, especially improving the agreement between the calculated NH_2 scissor modes and the experimental band.

In the 3000-3600 cm^{-1} region, application of a larger scaling factor to the harmonic modes is obviously necessary. The set of experimental bands centred around 3500 cm^{-1} is composed of only two distinct peaks, while the corresponding calculated spectra (at

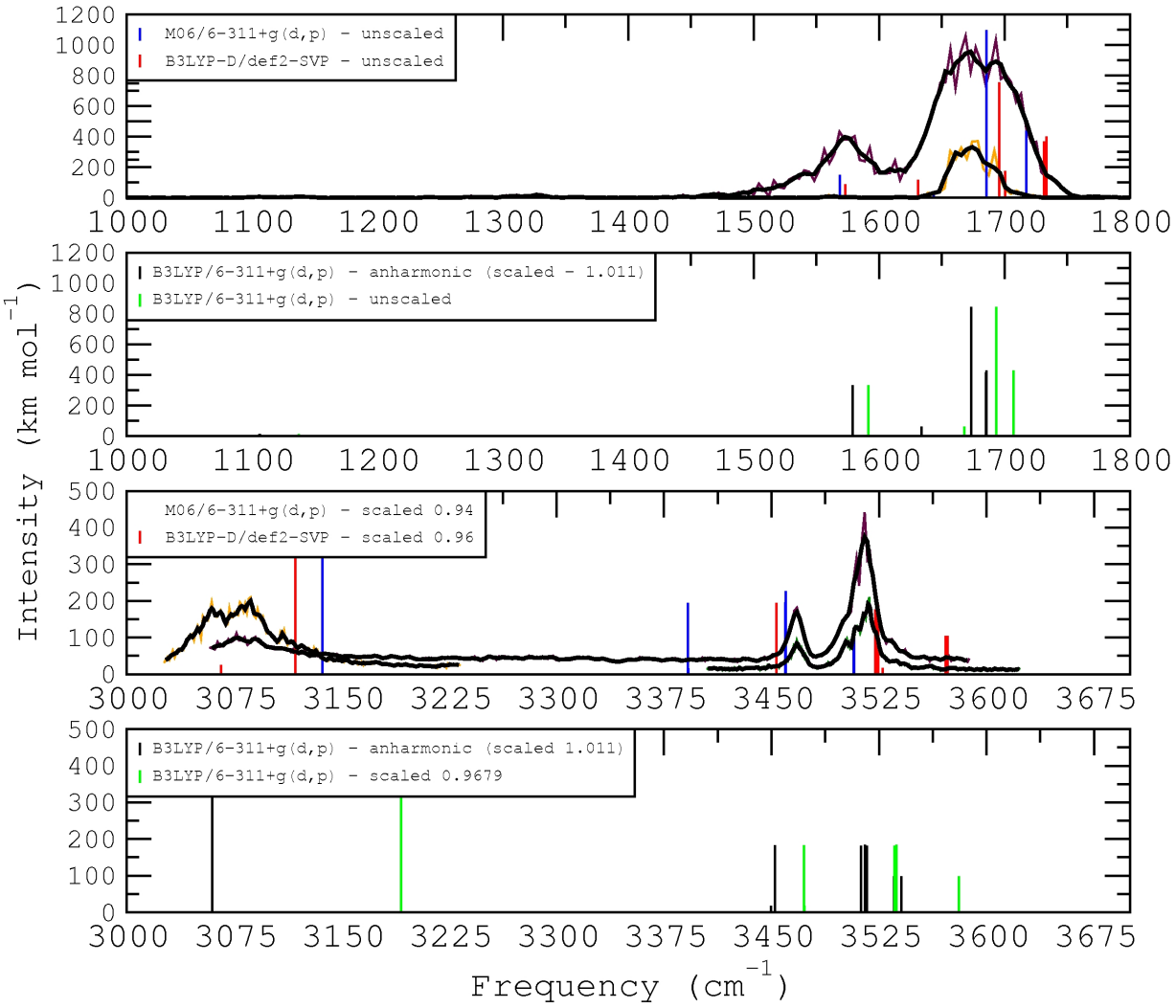

Figure 5.62 – Comparison of $[Guan_2Cl]^+$ experimental IRMPD spectra and calculated frequencies at various levels of theory for the lowest free energy (298 K) **B3LYP-D** structure.

all levels) appear to have three groups of modes (corresponding, in order of blue to red, to the asymmetric stretch of both N-Hs of the free NH_2 (anharmonic - 3540 cm^{-1}), coupled stretching modes of the four N-Hs (angled slightly away from the chlorides) of the NH_2 groups interacting with chloride (anharmonic - 3510 cm^{-1}), and finally the symmetric stretching of the free NH_2 groups (anharmonic - 3450 cm^{-1})). It is likely that anharmonic modes at 3510 and 3540 cm^{-1} are convoluted and both correspond to the larger band in the experimental spectrum at 3515 cm^{-1}, while the smaller band in the experimental spectrum at 3465 cm^{-1} corresponds to the anharmonic modes at 3450 cm^{-1}. For the har-

monic spectra, scaling (**B3LYP-D** - 0.96, B3LYP/6-311+G(d,p) - 0.9679 and **M06** - 0.94) improves agreement with the experimental spectrum, however they are still significantly more discrepant than the anharmonic spectrum. Even after application of these scaling factors to the 3000-3600 cm^{-1} region (see Figure 5.62) somewhat of a dilemma remains. M06/6-311+G(d,p) calculated modes would suggest that the symmetric free NH_2 stretches are unobserved experimentally while the two groups of modes farthest to the blue correspond well to the experiment. Contrastingly, the **B3LYP-D** and B3LYP/6-311+G(d,p) harmonic modes suggest just the opposite, that the asymmetric free NH_2 stretches are unobserved experimentally and the two sets of modes at ≈3510 and 3450 cm^{-1} match well to the experiment.

A similar comparison of experimental and calculated spectra for $[Guan_5Cl_4]^+$ (guan5P_3998) and $[Guan_5Cl_6]^-$ (lowest-energy structure) is shown in Figure 5.63. The main difference in the calculated fingerprint region spectra of $[Guan_5Cl_4]^+$ is the width of the distribution of the C-N band, where the **B3LYP-D** and MP2/6-31G(d) bands are significantly wider than the **M06** band. However, the experimental band is wider than all three of the calculated bands making it difficult to judge which is a better match. To a smaller extent, an analogous difference in widths exists for calculated bands of $[Guan_5Cl_6]^-$. For both cases, it appears that a scaling factor of approximately 0.99 is necessary for **M06** calculated modes, 0.98 for **B3LYP-D** modes and 0.943 for MP2/6-31G(d) calculated modes and, if applied, would bring all calculated spectra into qualitative agreement with the experimental spectra.

In the 3000-3600 cm^{-1} region comparisons, scaled (0.96) **M06** modes in the 3200-3350 cm^{-1} region are both more closely bunched and blue shifted in comparison to the scaled (0.96) **B3LYP-D** modes bringing them into better agreement with the experimental spectra for both $[Guan_5Cl_4]^+$ and $[Guan_5Cl_6]^-$. While scaled **B3LYP-D** results are in general similar, there are quite noticeable discrepancies in line positions in comparison

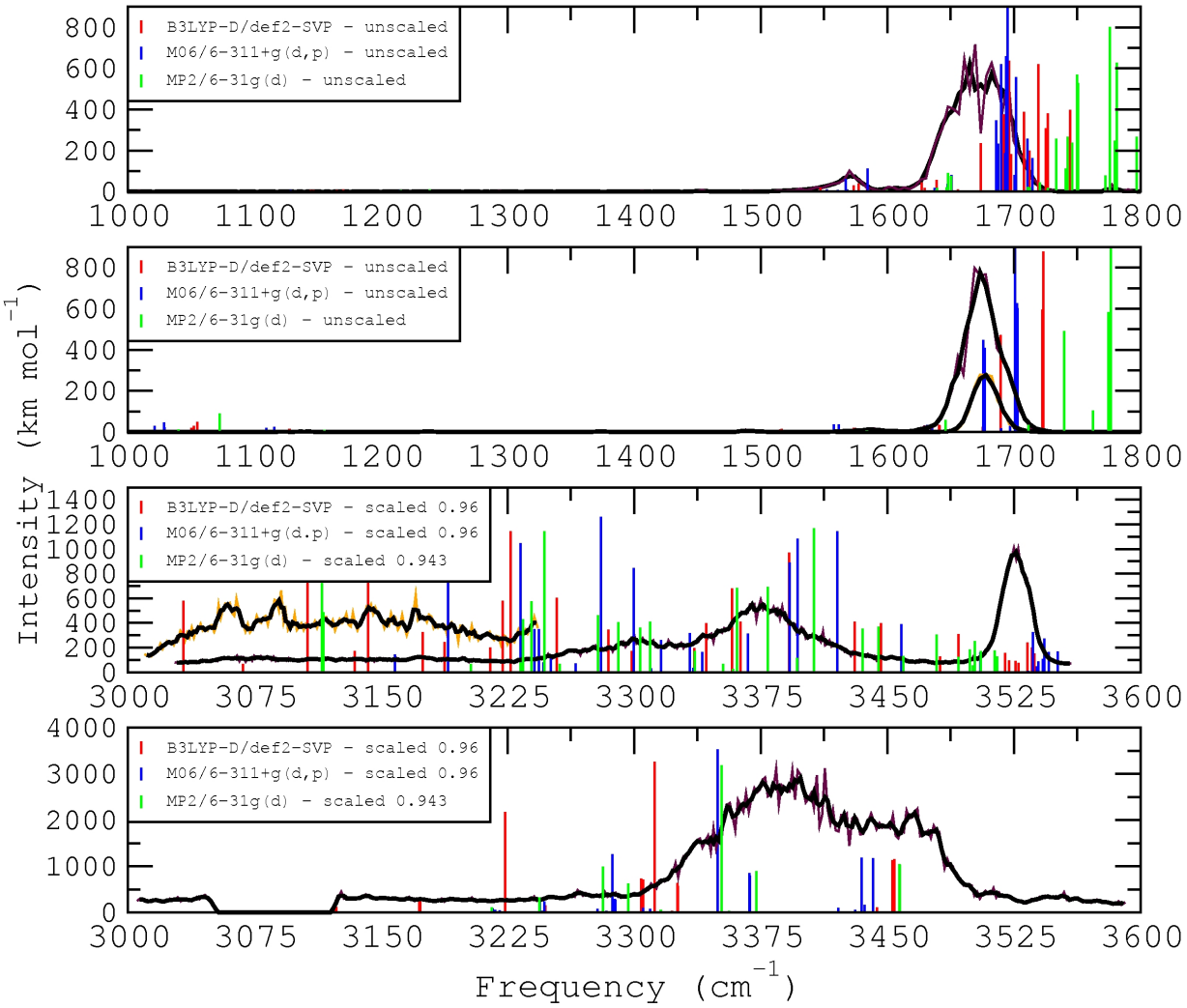

Figure 5.63 – Comparison of $[Guan_5Cl_4]^+/[Guan_5Cl_6]^-$ experimental IRMPD spectra and calculated frequencies at various levels of theory for the lowest free energy (298 K) **B3LYP-D** structures.

to experimental results. Most notable is the 3450-3500 cm^{-1} region for $[Guan_5Cl_4]^+$ and the 3300-3450 cm^{-1} region for $[Guan_5Cl_6]^-$. However, the free asymmetric N-H stretching modes in $[Guan_5Cl_4]^+$ (the farthest blue modes in both calculated spectra) are in good agreement with the experimental band (in this case, slightly worse for the **M06** band). For the 3000-3600 cm^{-1} region, the consistency of the scaled **B3LYP-D** modes, in reference to the free N-H modes for $[Guan_5Cl_4]^+$ and the least tightly bound N-H modes for $[Guan_5Cl_6]^-$, would seem to indicate that the scaled **M06** modes are in significantly better agreement with experiment than the scaled **B3LYP-D** modes. In this spectral region,

MP2/6-31G(d) calculated modes do not appear to offer better agreement with experiment than the **M06** results. For $[Guan_5Cl_4]^+$, the agreement with experiment in the ≈3450-3525 and <3250 cm^{-1} regions is moderately worse than the **M06** modes. For $[Guan_5Cl_6]^-$, MP2/6-31G(d) modes are very similar to the **M06** modes, except in the ≈3450 cm^{-1} region where the MP2/6-31G(d) modes are slightly blue shifted in comparison to the **M06** modes and in better agreement with the experiment.

5.4.7 Empirical Dispersion Correction

Figures 5.64, 5.65 and 5.66 show comparisons of spectra at the **B3LYP-D** and RI-B3LYP/def2-SVP levels for $[Guan_2Cl]^+$ and $[Guan_2Cl_3]^-$. Only minor and inconsistent peak shifts are seen in both the fingerprint and the 3000-3600 cm^{-1} regions. For Guan2P_0477, the average shift is -0.9 cm^{-1} with a maximum shift of 13.7 cm^{-1}, while for Guan2N_1493, the average shift is 1.8 cm^{-1} with a maximum of 15.7 cm^{-1}. From these results, it can be concluded that none of the difficulties in modelling frequencies calculated at the **B3LYP-D** level result from the use of the empirical dispersion correction. Additionally, the effect of the dispersion correction seems to be quite small in reference to positions of the vibrational modes, possibly because interactions are largely electrostatic for these small clusters. The lowest-energy $[Guan_9Cl_8]^+$ spectrum was also calculated with and without the dispersion correction, since it is possible that as the number of hydrogen bonds and van der Waals contacts between guanidinium ions increase, dispersive interactions could play a larger role. No experimental results were obtained for $[Guan_9Cl_8]^+$, however if general comparisons to $[Guan_{10}Cl_9]^+$ are made, the results corrected for dispersion have better spectral agreement.

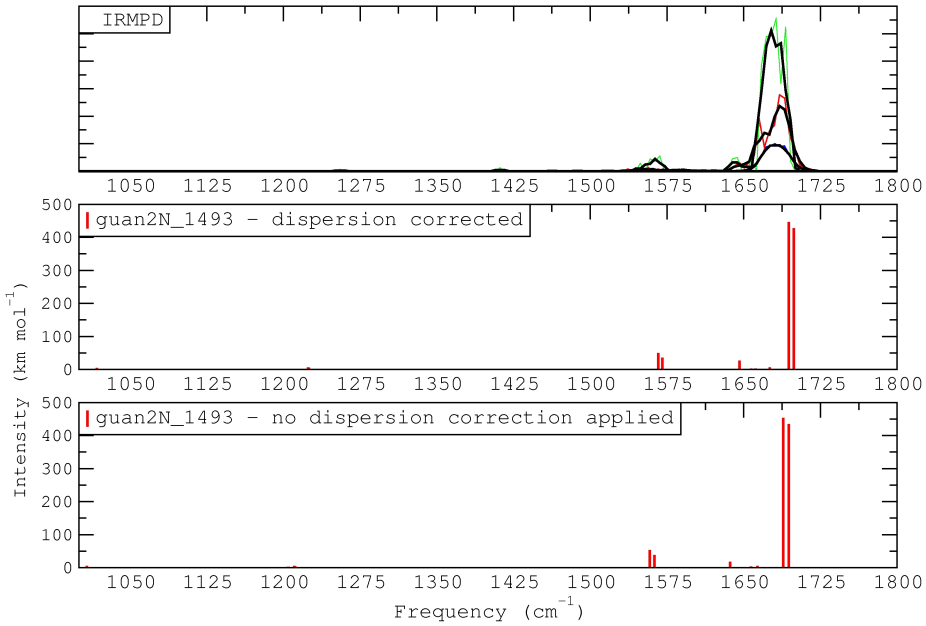

Figure 5.64 – Comparison of $[Guan_2Cl_3]^-$ fingerprint region calculated spectra with and without an empirical dispersion correction (**B3LYP-D** and RI-B3LYP/def2-SVP, frequencies scaled by 0.98).

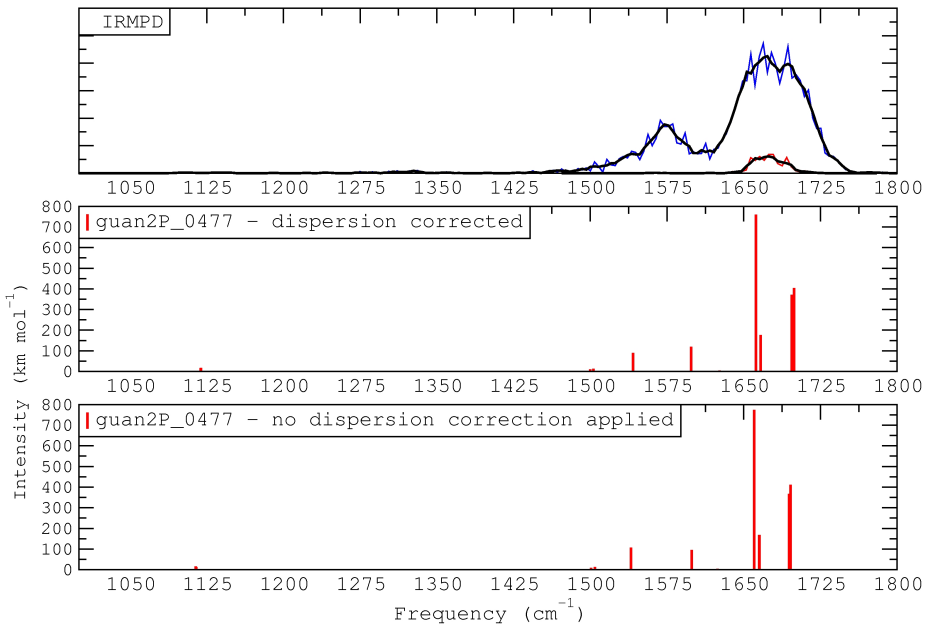

Figure 5.65 – Comparison of $[Guan_2Cl]^+$ fingerprint region calculated spectra with and without an empirical dispersion correction. (**B3LYP-D** and RI-B3LYP/def2-SVP, frequencies scaled by 0.98)

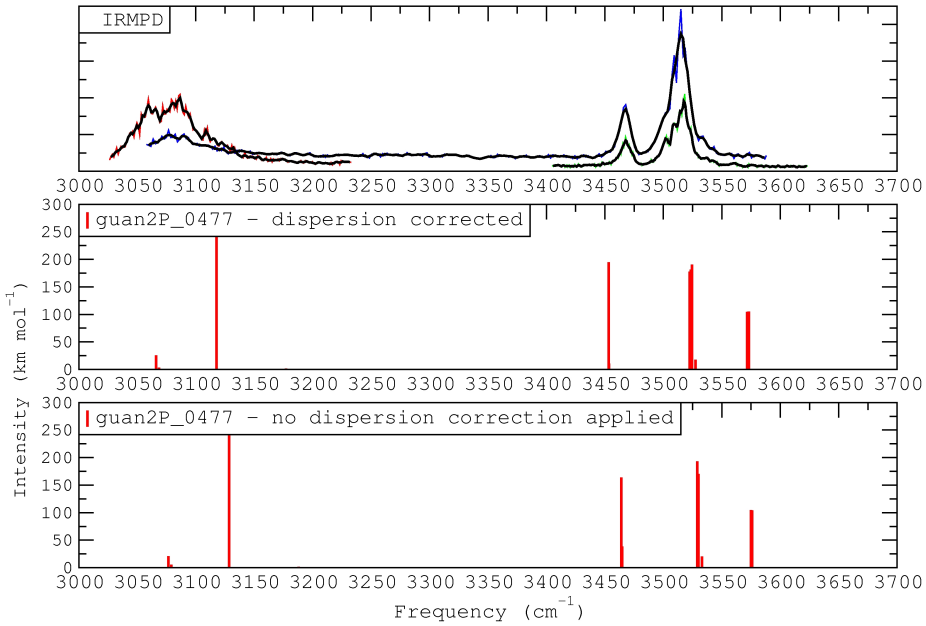

Figure 5.66 – Comparison of $[Guan_2Cl]^+$ 3000-3600 cm^{-1} region calculated spectra with and without an empirical dispersion correction. (**B3LYP-D** and RI-B3LYP/def2-SVP, frequencies scaled by 0.98)

5.5 General Comments and Conclusions

In the fingerprint region, general spectral features appear to be relatively similar for all clusters examined (see Figures 5.68 and 5.69), however it is possible to note a few trends. For all $[Guan_xCl_{x-1}]^+$ and $[Guan_xCl_{x+1}]^-$, the mainly C-N stretching band is found approximately at 1625-1725 cm^{-1}, however, a narrowing of the C-N stretching band is seen in cases where highly uniform structures exist (such as $[Guan_2Cl_3]^-$ and $[Guan_5Cl_6]^-$). This trend is not seen as clearly for $[Guan_xCl_{x-1}]^+$ where structures tend to be both somewhat less tightly bound and have less uniform N-H environments. From calculation, the C-N stretching band does appear to be somewhat unique between the smaller clusters (≤ 5), while for larger clusters a dense group of modes fills the 1625-1725 cm^{-1} region. This progression likely reaches the limit of the crystal structure rather quickly (perhaps already by $[Guan_{10}Cl_9]^+$). Figure 5.67 below illustrates this in the 1625-1725 cm^{-1} region as well as in the higher N-H stretching region above 3000 cm^{-1}.

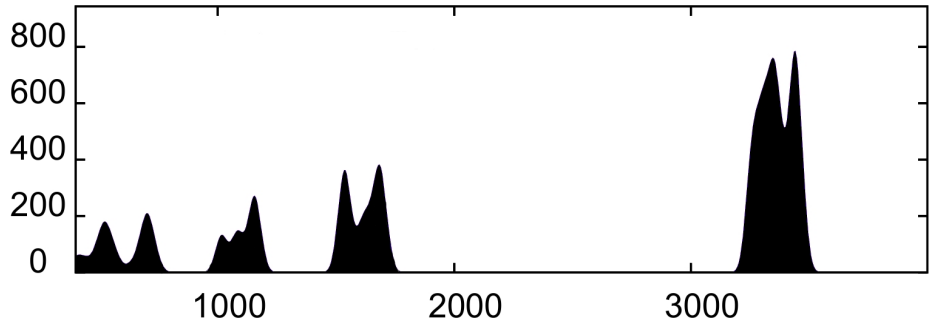

Figure 5.67 – Calculated spectrum of crystalline guanidinium chloride.[155] (PBE/DNP, cm^{-1})

A similar convergence of band saturation can be observed in the 3000-3600 cm^{-1} region. Here, spectra for the smaller sizes are more unique and assignment of even detailed structural features can be made. This was clearly seen for $[Guan_5Cl_4]^+$ where two very unique structures, differing only by a small deviation in the position and coordination of one guanidinium ion, could be easily distinguished. However, a progression can be seen going from $[Guan_5Cl_4]^+$ to $[Guan_6Cl_5]^+$ to $[Guan_{10}Cl_9]^+$ with a likely limit fairly similar to what is observed for $[Guan_{10}Cl_9]^+$ (again, see Figure 5.67). In Figure 5.70, all spectra for clusters larger than $[Guan_2Cl]^+$ have generally similar bands in the 3335-3445 cm^{-1} region, likely resulting from similarly compact-type structures. In these clusters, the bound symmetric stretches are unobserved as a result of broadening, however the free symmetric band (*approx*3225-3325 cm^{-1}), bound asymmetric band (≈3325-3450 cm^{-1}) and the free asymmetric band (≈3500-3550 cm^{-1}) are observed at consistent positions. Conversely, $[Guan_2Cl]^+$ has a structure corresponding to an extended type of conformation. This results in the different spectral features in this region, where the free symmetric stretch of the two completely non-interacting NH_2 groups is uniquely found just above 3450 cm^{-1}.

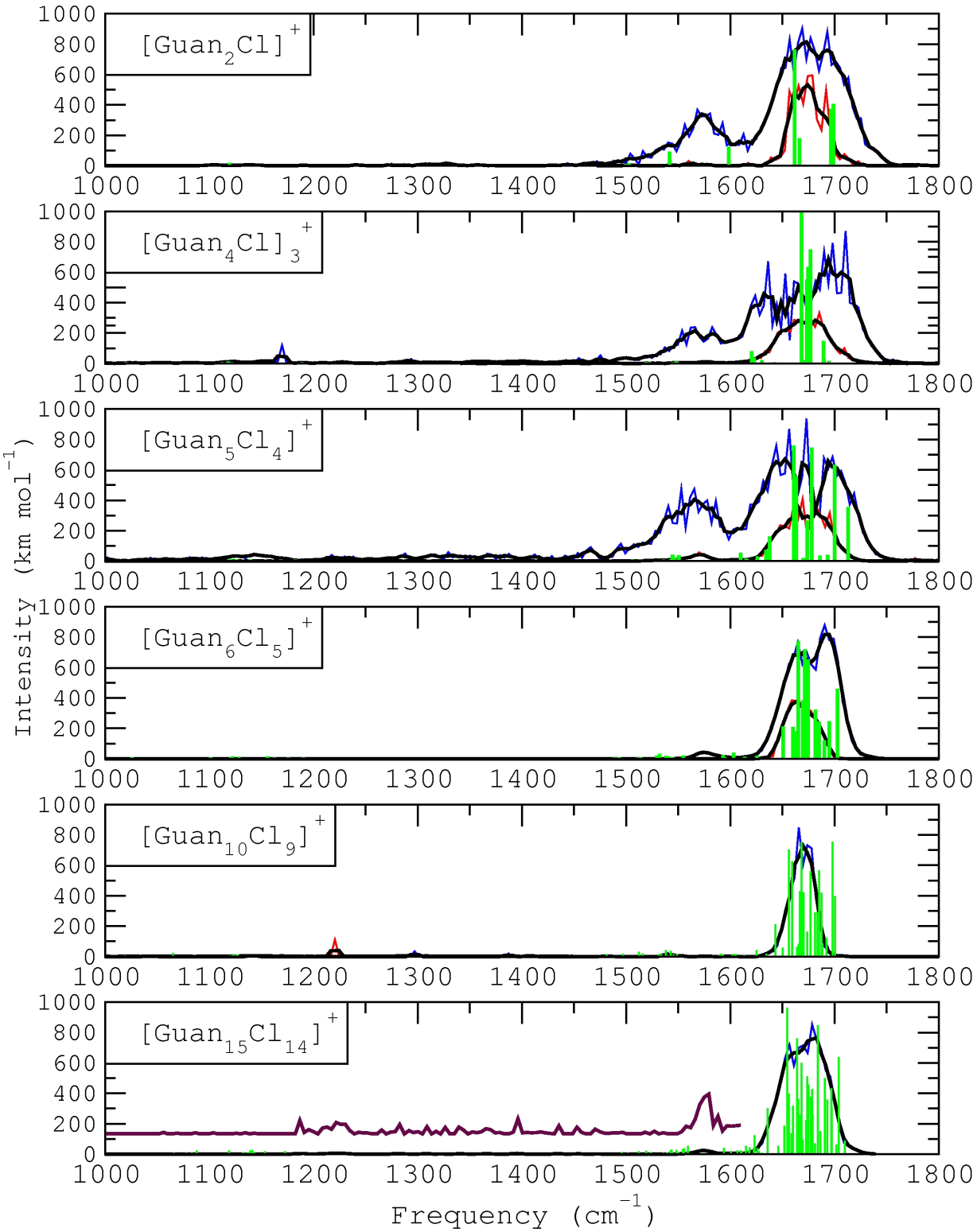

Figure 5.68 – $[\text{Guan}_x\text{Cl}_{x-1}]^+$ fingerprint experimental and scaled (0.98) **B3LYP-D** calculated spectra. **Black** lines are a three-point moving average of the experimental data points and the **purple** line is a six-times magnification. Experimental fragmentation efficiencies (Y-axes) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.98)

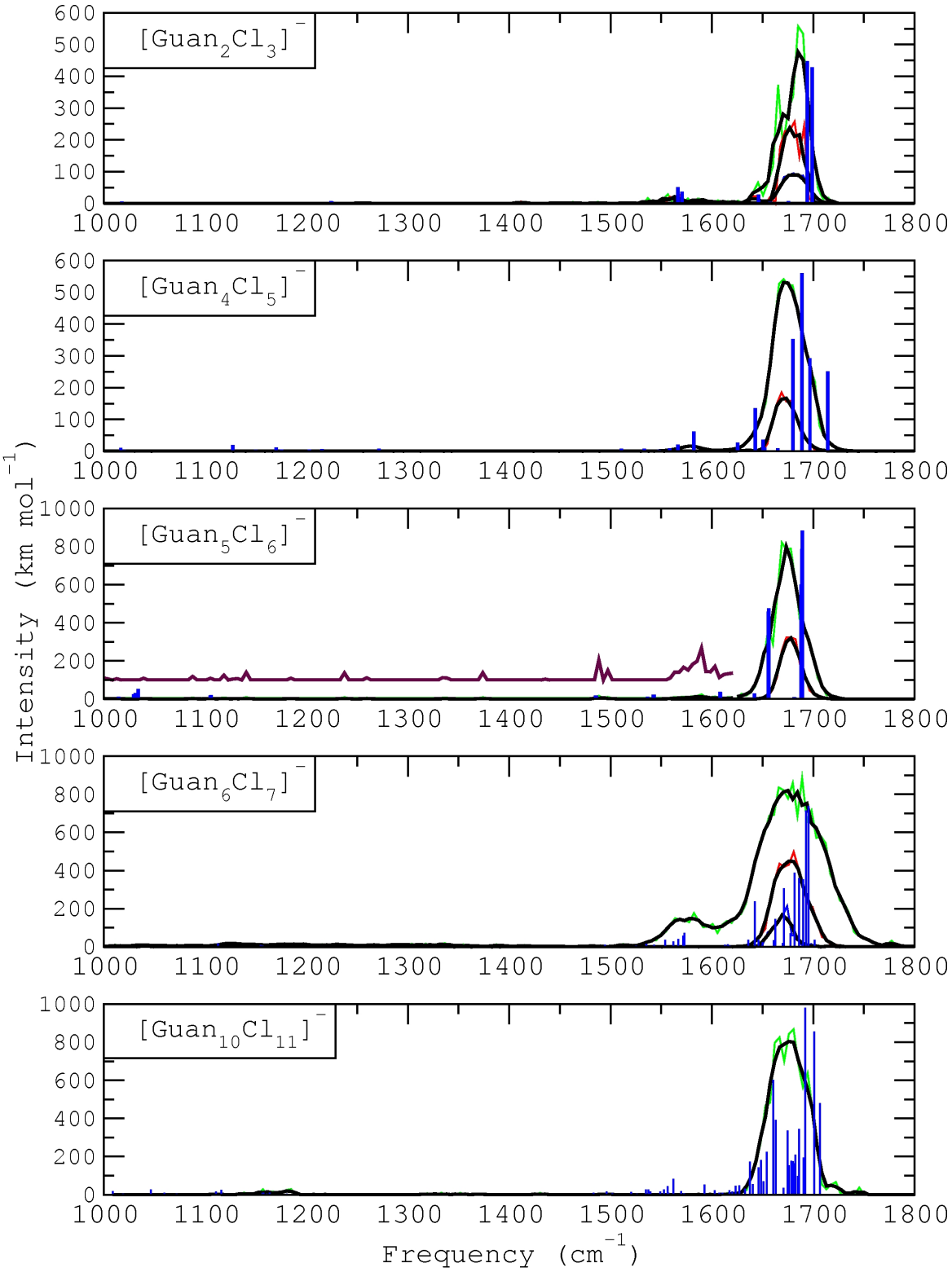

Figure 5.69 – $[\text{Guan}_x\text{Cl}_{x+1}]^-$ fingerprint experimental and scaled (0.98) **B3LYP-D** calculated spectra. **Black** lines are a three-point moving average of the experimental data points and the **purple** line is a six-times magnification. Experimental fragmentation efficiencies (Y-axes) are arbitrarily scaled. (**B3LYP-D**, frequencies scaled by 0.98).

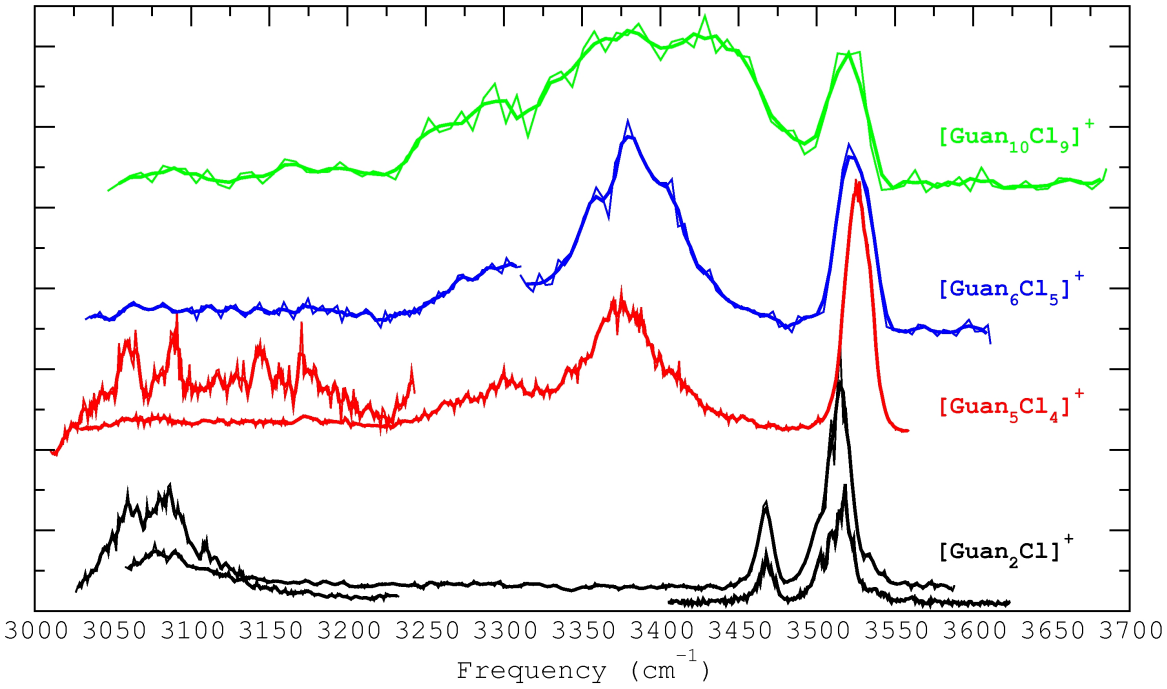

Figure 5.70 – $[\text{Guan}_x\text{Cl}_{x-1}]^+$ 3000-3600 cm^{-1} region experimental and spectra. **Bold** lines are a three-point moving average of the experimental data points.

For structures of increasing size, there are few structural trends before clusters containing approximately nine guanidinium ions are considered, where the presence of internal ions contained by a cage of other ions begins. It is possible to speculate that increasing cluster size would lead to larger cages and shells of solvation of the internal guanidinium ions, where at some point the environment of the internally contained guanidinium ions would be identical to the crystal or ionic liquid environment. Guanidinium-based ionic liquids have been reported in the literature to have first solvation shells around a given cation of approximately 5-8 Å.[149–151] This compares relatively closely to the approximate 7-8 Å separating chloride ions on opposite sides of the internally contained guanidinium ions in the larger clusters discussed here (for example, $[\text{Guan}_{10}\text{Cl}_{11}]^-$) and possibly suggests the formation of the first solvent shell in clusters of approximately this size.

While both clusters of 15 guanidinium ions (cationic and anionic clusters) do not begin to approach the crystal structure in terms of conformations, they do begin to have

generally similar structural features, both with the crystal structure and with each other[viii]. Average N-H$\cdots$Cl$^-$ bond lengths are 2.39 Å (average of 5.7 interactions per guanidinium), 2.33 Å (average of 5.7 interactions per guanidinium) and 2.39 Å (average of 6 interactions per guanidinium) for $[Guan_{15}Cl_{14}]^+$, $[Guan_{15}Cl_{16}]^-$ and the crystal structure, respectively. The shorter average N-H$\cdots$Cl$^-$ length for $[Guan_{15}Cl_{16}]^-$ is somewhat intuitive since the surplus Cl$^-$ (and negative charge) holds the structure together somewhat more tightly.

While it is difficult to make predictions about how different ions (for example, an amino acid) would be contained within the "cage" of guanidinium ions, exploration of clusters of this type could begin to probe the properties of guanidinium chloride as a denaturant. Molecular dynamics simulations, as well as infrared spectroscopy experiments of such amino acid or peptide containing guanidinium ion cages would be an extremely interesting continuation of the results presented here. Contrasting the current results with results from clusters formed with an amino acid or peptide and only one or a few guanidinium ions would also be very illustrative, perhaps directly probing the energetics associated with secondary structure in peptides.

Generally, it can be concluded from the presented comparisons of computational methodologies that the **B3LYP-D** level of theory offers a cost-effective approach to the calculation of gaseous guanidinium chloride ionic clusters. The **M06** level of theory in many cases offers a significant improvement to the **B3LYP-D** level at a moderate increase in computational cost[ix]. MP2/6-31G(d) frequencies did not appear to offer a clear improvement over **M06** frequencies while being an additional 2-3 times more computationally expensive. The anharmonic B3LYP/6-311+G(d,p) spectrum, although significantly more computationally expensive, offered crucial insight required for band assignments in the N-H stretching region of $[Guan_2Cl]^+$.

[viii] Likely, consideration of clusters of significantly higher order would be necessary before conformations of guanidinium ions at the core of the structure would be similar to that of the crystal structure.

[ix] **M06** jobs were ≈2-4 times more expensive, however different computational resources were used. Additionally, M06 jobs were calculated using Guassian 09, while **B3LYP-D** jobs were calculated using Turbomole 6.1.

APPENDICES

Appendix A

Energetic Properties of Guanidinium Chloride Clusters

A.1 Binding Energy Associated with Sequential Neutral Guanidinium Chloride Addition

The total interaction energy (equivalent to the negative of the energy required to separate to isolated ions, shown in reactions below) of $[Guan_2Cl]^+$ at the MP2/6-31+g(d) level is reported to be -630.9 kJ mol^{-1}.[153] This compares relatively well to the -665.1 kJ mol^{-1} calculated at the **B3LYP-D** level of theory. Additionally their reported -457.4 kJ mol^{-1} for the interaction energy of Guanidinium Chloride (neutral) compares to the value calculated in this work of -493.5 kJ mol^{-1}.[153]

$Guan^+ + Guan^+ + Cl^- \rightarrow [Guan_2Cl]^+$ (**B3LYP-D**: -665.1 kJ mol^{-1})

$Guan^+ + Cl^- \rightarrow [GuanCl]$ (**B3LYP-D**: -493.5 kJ mol^{-1})

For the series of guandinium chloride clusters discussed up to this point, an observation that the binding energies are relatively consistent between sizes would offer a general indication that the lowest-energy structures were located for each size. Some deviation from a fully smooth progression is predictable since, necessarily, more stable conformations can be formed for some sizes than others (for example, $[Guan_5Cl_6]^-$ forms an obviously favourable completely closed structure, and the binding energy in this case is large).

Binding energies for the $[Guan_xCl_{(x-1)}]^+$ series of ions are listed in the table below. Values are quite consistent between sizes and centred around approximately E - 160 kJ mol^{-1}, H - 150 kJ mol^{-1} and G - 100 kJ mol^{-1}.

Table A.1 – $[Guan_xCl_{(x-1)}]^+$ **B3LYP-D** binding energies associated with sequential addition of guanidinium chloride neutral (kJ mol^{-1}, 298K).

	$[Guan_2]^+$	$[Guan_3]^+$	$[Guan_4]^+$	$[Guan_5]^+$	$[Guan_6]^+$	$[Guan_{10}]^+$
ΔE_{rxn}	-171.5	-140.2	-181.5	-140.8	-163.6	-175.7
ΔH^o_{rxn}	-162.2	-122.4	-169.2	-129.8	-153.7	-165.2
ΔG^o_{rxn}	-119.1	-57.8	-122.5	-75.4	-101.7	-114.4

The free energy associated with binding guanidinium ion and neutral guanidinium chloride to form $[Guan_2Cl]^+$ is high both in respect to electronic and free energies. This structure is formed with a relatively low entropic cost, demonstrated by the relative independence of the energy on the dihedral orientation of the guanidinium ions. This indicates that the guanidinium ions are relatively close to free rotors and not strongly hindered (see Table 5.18). Additionally, the short (2.17Å) $H\cdots Cl^-$ bonds indicate strong interactions with the chloride ion. For $[Guan_3Cl_2]^+$, somewhat of the opposite is true and the free binding energy is somewhat lower since this conformation is highly constrained. $[Guan_4Cl_3]^+$ adopts a fairly symmetrical but loosely bound conformation where chloride ions are highly coordinated, resulting in the largest binding energy (free energy) of the positive ion series. The $[Guan_5Cl_4]^+$ structure is similar to the $[Guan_4Cl_3]^+$ structure however there the additional guanidinium chloride cannot fully be accommodated in this structure, visible by the sharp

decrease in binding energy relative to the previous size. In sizes larger than $[Guan_5Cl_4]^+$ it is difficult to relate structural features so directly to binding energies, however, one can imagine that as size increases sufficiently, addition of neutral guanidinium chloride will not so strongly effect the rest of the structure and converge to a limit in the binding energy (extrapolating this to infinity leads to the crystal or ionic liquid).

Binding energies for the $[Guan_xCl_{(x+1)}]^-$ series of ions are listed in the table below. Values are centred around approximately E - 175 kJ mol^{-1}, H - 160 kJ mol^{-1} and G - 105 kJ mol^{-1}.

Table A.2 – $[Guan_xCl_{(x+1)}]^-$ **B3LYP-D** binding energies associated with sequential addition of guanidinium chloride neutral (kJ mol^{-1}, 298K).

	$[Guan_2]^-$	$[Guan_3]^-$	$[Guan_4]^-$	$[Guan_5]^-$	$[Guan_6]^-$
ΔE_{rxn}	-244.2	-125.5	-169.8	-219.1	-151.7
ΔH^o_{rxn}	-221.6	-113.4	-163.2	-200.9	-136.0
ΔG^o_{rxn}	-158.2	-60.1	-110.7	-140.2	-85.9

$[GuanCl_2]^-$ and $[Guan_5Cl_6]^+$ have very symmetrical and closed structures, where all interaction sites are occupied and the structures are highly constrained. This results in a high electronic binding energy but relatively low free energy of binding.

Appendix B

Additional IRMPD Results

B.1 Chloride-Bound Dimers of Trimethylammonium and Dimethylammonium - Fingerprint IRMPD Spectra

Alkylammonium adduct ions are commonly observed ion-molecule complexes in mass spectrometry, occurring both as a result of contamination (often the case in biochemical analysis) and as intentionally observed ion-molecule reactions.[6,165] During experiments involving trimethylammonium adducts of various amino acids and peptides, an ion of 155 m/z was repeatedly observed. Consideration of the trimethylammonium source as the chloride salt led to the deduction that this ion was the chloride-bound dimer of trimethylammonium. A similar ion was observed during experiments involving dimethylammonium (at 125 m/z) and also corresponded to the chloride-bound dimer, however analogous ions were not observed during experiments using methylammonium chloride.

The chloride-bound dimer (PDB) of trimethylammonium is shown in Figure B.1 and the chloride-bound dimer of dimethylammonium is shown in Figure B.2. These dimers are

formed by two protonated methylamines (tri-, or di-) and a chloride anion producing an overall positively charged ionic complex, as in the reaction schemes and figures below:

Trimethylammonium (TMA^+) + Trimethylammonium Chloride ((TMA^+)-Cl) →
Trimethylammonium Chloride-Bound Dimer $(TMA^+)_2$-Cl

Dimethylammonium (DMA^+) + Dimethylammonium Chloride ((DMA^+)-Cl) →
Dimethylammonium Chloride-Bound Dimer $(DMA^+)_2$-Cl

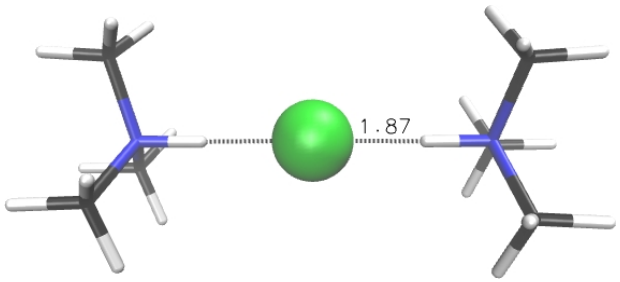

Figure B.1 – Lowest-energy $(TMA^+)_2$-Cl structure (**B3LYP-D**, free energy 298 K). Dihedral angle C-N-N-C - 17.1 degrees.

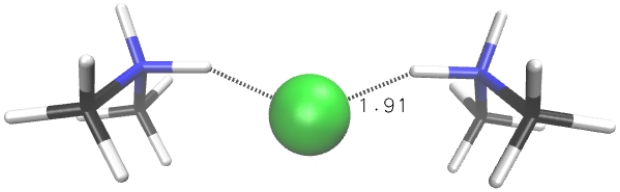

Figure B.2 – Lowest-energy $(DMA^+)_2$-Cl structure (**B3LYP-D**, free energy 298 K). Dihedral angle C-N-N-C - 1.2 degrees.

Binding energies (in terms of electronic energies (E), enthalpies (H) and free energies (G)) for the formation of the chloride-bound dimers shown above are listed in Table B.1, followed by fingerprint IRMPD spectra and tables of vibrational assignments.

Table B.1 – $(TMA^+)_2$-Cl and $(DMA^+)_2$-Cl **B3LYP-D** binding energies (kJ mol^{-1}, 298K).

	ΔE_{rxn}	ΔH^o_{rxn}	ΔG^o_{rxn}
$(TMA^+)_2$-Cl	-164.5	-154.0	-98.3
$(DMA^+)_2$-Cl	-161.7	-147.0	-103.0

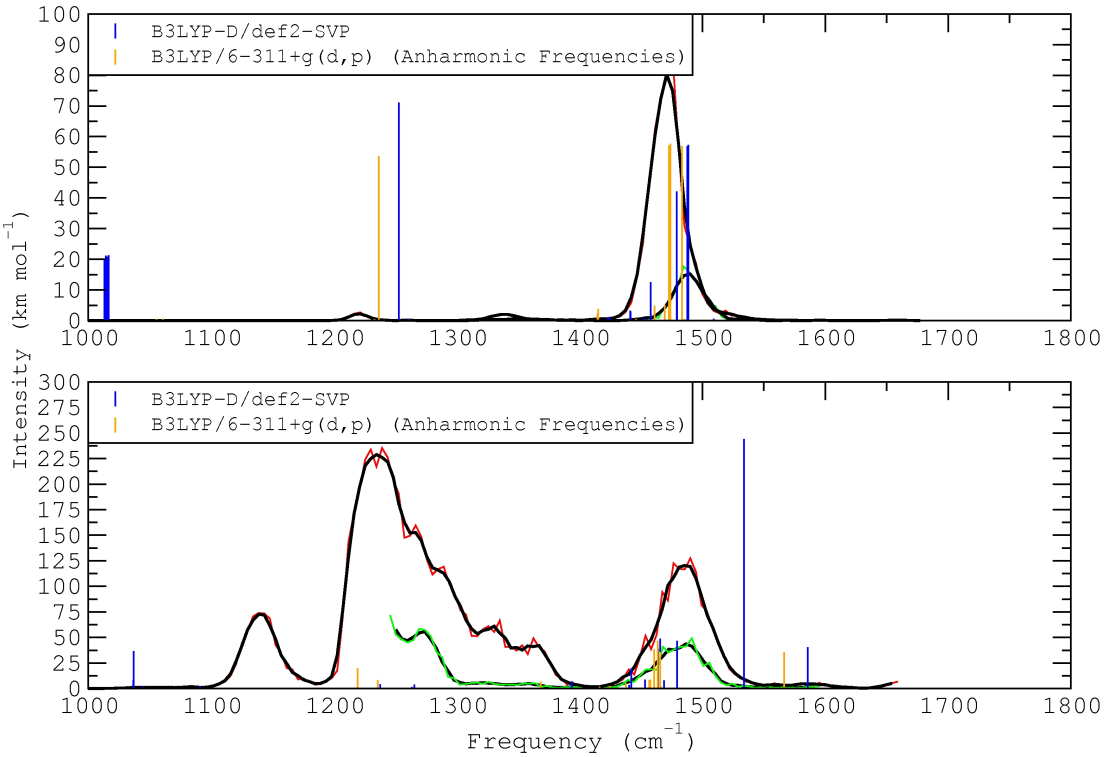

Figure B.3 – $(TMA^+)_2Cl$ and $(DMA^+)_2Cl$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrary scaled. (**B3LYP-D** frequencies scaled by 0.98)

Table B.2 – $(TMA^+)_2Cl$ and $(DMA^+)_2Cl$ unscaled **B3LYP-D** fingerprint region spectral descriptions (cm^{-1}).

$(TMA^+)_2$-Cl	
1479-1423	Mainly CH_3 vibrations
1510-1488	Coupled $NH\cdots Cl^-$ and CH_3 vibrations
$(DMA^+)_2$-Cl	
1238-1265	Mainly CH_3 vibrations
1394-1488	Coupled $NH_2\cdots Cl^-$ and CH_3 vibrations
1586-1534	Mainly $NH_2\cdots Cl^-$ vibrations

Reasonable spectral agreement is obtain from both unscaled harmonic (**B3LYP-D**) and anharmonic (B3LYP/6-311+G(d,p)) calculated modes for the chloride-bound dimer of trimethylammonium, $(TMA^+)_2Cl$, where the anharmonic modes are slightly red shifted from the harmonic modes and in better agreement with the experimental result. The very

small peak in the experimental spectrum just above 1300 cm^{-1} cannot be accounted for by any modes in the calculated spectrum and is possibly not a real peak. The line positions of the anharmonic calculated spectra for the dimethylammonium chloride-bound dimer are mostly in agreement with the experimental result and slightly better than those of the harmonic calculation. However, in both calculated spectra the region between 1100-1200 cm^{-1} is empty and the experimental intensity in this region cannot be accounted for using these calculated results. A different ion of the same mass which could account for these spectral differences is the ion-molecule complex of the proton bond dimer of dimethylammonia with hydrogen chloride ($DMAmmoniaH^+ \cdots HCl$). However, this complex was calculated to be 82.1 kJ mol^{-1} higher in electronic energy at the **B3LYP-D** level. Many calculations were completed over a wide range of methodologies and conformations, however no improvement or ability to account for the experimental bands (both for dimethylammonium and trimethylammonium) was achieved. Obtaining duplicate experiments with less saturated bands would likely help to account for some of the discrepancies between the experimental and calculated results.

B.2 Proton-Bound Dimer of Guanidine - Fingerprint IRMPD Spectrum

The proton-bound dimer (PDB) of guanidine is shown in Figure B.4. In this proposed structure, two NH_2 groups of guanidinium form two hydrogen bonds with the nitrogen of the guanidine imine group. This structure was found to be 72.0 kJ mol^{-1} lower in electronic energy at the **B3LYP-D** level than a conformation where the same two NH_2 groups of guanidinium interact not with the imine group of guanidine, but with the two NH_2 groups of guanidine.

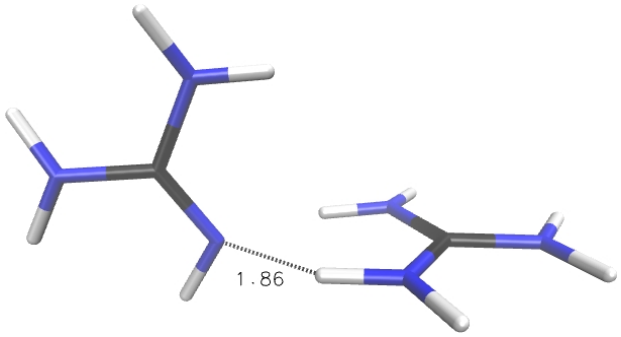

Figure B.4 – Lowest-energy Guanidine$_2$H$^+$ structure (**B3LYP-D**, electronic energy).

Binding energies for the formation of the PDB of guanidine (the dimer of guanidinium and guanidine), as in the reaction below, are listed in Table B.3.

$$\text{Guanidine} + \text{GuanidineH}^+ \text{ (Guanidinium)} \rightarrow \text{Guanidine}_2\text{H}^+$$

Table B.3 – Guanidine$_2$H$^+$ binding energies (kJ mol^{-1}, 298K).

	ΔE	ΔH^o	ΔG^o
RI-B3LYP-D/def2-SVP	-151.7	-146.0	-102.1
B3LYP-D	-128.3	-122.5	-78.7

The fingerprint region experimental spectrum is seen on the following page with vibrational assignments and the scaled (0.98) **B3LYP-D** calculated spectrum, which is in excellent agreement with the experiments.

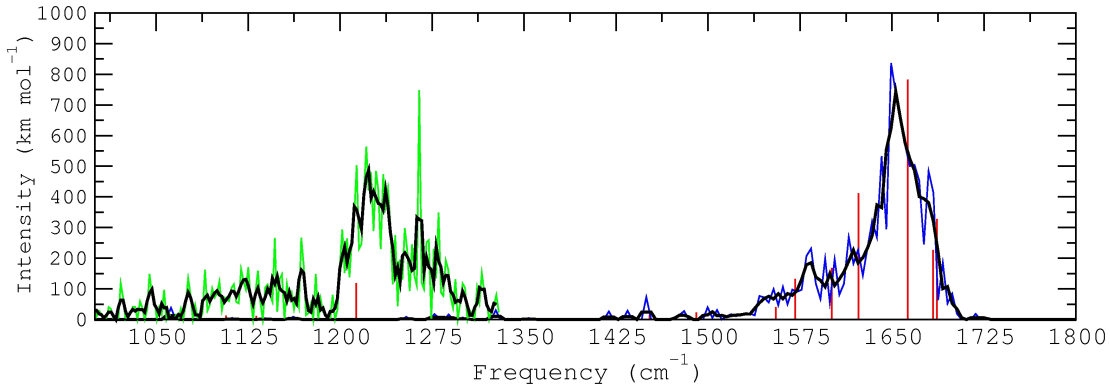

Figure B.5 – Guanidine$_2$H$^+$ fingerprint region IRMPD experimental and calculated spectra. **Black** lines are a three-point moving average of the experimental data points. Experimental fragmentation efficiencies (Y-axis) are arbitrary scaled. (**B3LYP-D** frequencies scaled by 0.98)

Table B.4 – Guanidine$_2$H$^+$ scaled (0.98) **B3LYP-D** fingerprint spectral assignments (cm^{-1}).

1687	C-N modes (guanidinium interacting NH_2s with imine)
1684	C-N modes (guanidinium interacting NH_2s with imine)
1663	Imine, C-N modes (guanidinium free NH_2)
1623	C-N modes (guanidine free NH_2)
1601	NH_2 (guanidine free NH_2 guanidinium - NH_2s interacting with imine)
1571	NH_2 (guanidine free NH_2)
1555	NH_2 (guanidinium - free NH_2)
1490	C-N (guanidine)
1452	C-N (guanidinium)
1213	Imine NH_2

References

[1] Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J. *Nature* **2009**, *460*, 823–830. 1

[2] Dunbar, R. C.; Polfer, N. C.; Oomens, J. *J. Am. Chem. Soc.* **2007**, *129*, 14562–14563. 1, 2

[3] Jarrold, M. F. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179–207. 1

[4] Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G. W.; Tate, C. G.; Schertler, G. F. X. *Nature* **2008**, *454*, 486–491. 1, 2

[5] Baer, T.; Dunbar, R. C. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 681 – 693. 2

[6] Wu, R.; McMahon, T. B. *J. Mass Spectrom.* **2008**, *43*, 1641–1648. 2, 73, 263

[7] Oomens, J.; Sartakov, B. G.; Meijer, G.; von Helden, G. *Int. J. Mass Spectrom.* **2006**, *254*, 1 – 19. 6, 10

[8] Polfer, N. C. *Chem. Soc. Rev.* **2011**, *40*, 2211–2221. 6, 10

[9] Dunbar, R. C. *J. Am. Chem. Soc.* **1971**, *93*, 4354–4358. 6

[10] Dunbar, R. C. *J. Am. Chem. Soc.* **1971**, *93*, 4354–4358. 6

[11] Freiser, B. S.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1974**, *96*, 6260–6266.

[12] Okumura, M.; Yeh, L. I.; Myers, J. D.; Lee, Y. T. *J. Phys. Chem.* **1990**, *94*, 3416–3427. 12, 22

[13] Watson, C. H.; Zimmerman, J. A.; Bruce, J. E.; Eyler, J. R. *J. Phys. Chem.* **1991**, *95*, 6081–6086. 22, 23, 24

[14] Peiris, D. M.; Cheeseman, M. A.; Ramanathan, R.; Eyler, J. R. *J. Phys. Chem.* **1993**, *97*, 7839–7843.

[15] Shin, S. K.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1990**, *112*. 6

[16] Levine, I. N. *Quantum Chemistry*, 5th ed.; Prentice Hall: New Jesery, 2005. 6, 41

[17] Polfer, N. C.; Oomens, J. *Mass Spec. Rev.* **2009**, *28*, 468–494. 9

[18] Lopez, V.; Marcus, R. A. *Chem. Phys. Lett.* **1982**, *93*, 232 – 234. 9

[19] Grant, E. R.; Schulz, P. A.; Sudbo, A. S.; Shen, Y. R.; Lee, Y. T. *Phys. Rev. Lett.* **1978**, *40*, 115–118. 9

[20] Lehmann, K. K.; Scoles, G.; Pate, B. H. *Annu. Rev. Phys. Chem.* **1994**, *45*. 10

[21] Laskin, J.; Futrell, J. H. *Mass Spec. Rev.* **2005**, *24*, 135–167.

[22] Black, J. G.; Yablonovitch, E.; Bloembergen, N.; Mukamel, S. *Phys. Rev. Lett.* **1977**, *38*, 1131–1134.

[23] Mukamel, S.; Jortner, J. *J. Chem. Phys.* **1976**, *65*, 5204–5225.

[24] Eyler, J. R. *Mass Spec. Rev.* **2009**, *28*, 448–467.

[25] Fridgen, T. D. *Mass Spec. Rev.* **2009**, *28*, 586–607.

[26] MacAleese, L.; Maître, P. *Mass Spec. Rev.* **2007**, *26*, 583–605. 100, 166, 198

[27] Schulz, P. A.; Sudb, A. S.; Krajnovich, D. J.; Kwok, H. S.; Shen, Y. R.; Lee, Y. T. *Annu. Rev. Phys. Chem.* **1979**, *30*, 379–409. 10

[28] Maître, P.; Caër, S. L.; Simon, A.; Jones, W.; Lemaire, J.; Mestdagh, H.; Heninger, M.; Mauclaire, G.; Boissel, P.; Prazeres, R.; Glotin, F.; Ortega, J.-M. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2003**, *507*, 541 – 546; Proceedings of the 24th International Free Electron Laser Conference and the 9th Users Workshop. 10, 100, 166, 198

[29] Lucas, B.; Gregoire, G.; Lemaire, J.; Maître, P.; Ortega, J.-M.; Rupenyan, A.; Reimann, B.; Pierre Schermann, J.; Desfrancois, C. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2659–2663. 10, 100, 166, 198

[30] Oomens, J.; van Roij, A. J. A.; Meijer, G.; von Helden, G. *Astro. Phys. J.* **2000**, *542*, 404. 10

[31] Valle, J. J.; Eyler, J. R.; Oomens, J.; Moore, D. T.; van der Meer, A. F. G.; von Helden, G.; Meijer, G.; Hendrickson, C. L.; Marshall, A. G.; Blakney, G. T. *Rev. Sci. Instrum.* **2005**, *76*, 023103. 10

[32] Oomens, J.; Polfer, N.; Moore, D. T.; van der Meer, L.; Marshall, A. G.; Eyler, J. R.; Meijer, G.; von Helden, G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1345–1348. 11

[33] March, R. E. *J. Mass Spectrom.* **1997**, *32*, 351–369. 12

[34] Vasicek, L.; Ledvina, A.; Shaw, J.; Griep-Raming, J.; Westphall, M.; Coon, J.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1105–1108; 10.1007/s13361-011-0119-7. 12

[35] Altinay, G.; Citir, M.; Metz, R. B. *J. Phys. Chem. A* **2010**, *114*, 5104–5112; PMID: 20353200. 12

[36] Aleese, L. M.; Simon, A.; McMahon, T. B.; Ortega, J.-M.; Scuderi, D.; Lemaire, J.; Maître, P. *Int. J. Mass Spectrom.* **2006**, *249-250*, 14 – 20; Chava Lifshitz Memorial Issue. 12, 100, 166, 198

[37] Brodbelt, J. S.; Wilson, J. J. *Mass Spec. Rev.* **2009**, *28*, 390–424. 12

[38] Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282 – 283. 12

[39] Marshall, A. G.; Hendrickson, C. L. *Annu. Rev. Anal. Chem.* **2008**, *1*, 579–599. 12

[40] Rainville, S.; Thompson, J. K.; Pritchard, D. E. *Science* **2004**, *303*, 334–338. 12

[41] Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spec. Rev.* **1998**, *17*, 1–35. 12

[42] Bakker, J. M.; Besson, T.; Lemaire, J.; Scuderi, D.; Maître, P. *J. Phys. Chem. A* **2007**, *111*, 13415–13424; PMID: 18052260. 15, 24, 100, 166, 198

[43] Kebarle, P.; Verkerk, U. H. *Mass Spec. Rev.* **2009**, *28*, 898–917. 17

[44] de la Mora, J. F. *Anal. Chim. Acta* **2000**, *406*, 93 – 104. 18

[45] Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287–2294. 18

[46] Gamero-Castano, M.; de la Mora, J. F. *J. Chem. Phys.* **2000**, *113*, 815–832. 18

[47] Hogan, C. J.; Carroll, J. A.; Rohrs, H. W.; Biswas, P.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 369–377. 18

[48] Roberson, C. W.; Sprangle, P. *Phys. Fluids B* **1989**, *1*, 3–42. 18

[49] Prazeres, R.; Ortega, J. M.; Glotin, F.; Jaroszynski, D. A.; Marcouillé, O. *Phys. Rev. Lett.* **1997**, *78*, 2124–2127. 19

[50] Deacon, D. A. G.; Elias, L. R.; Madey, J. M. J.; Ramian, G. J.; Schwettman, H. A.; Smith, T. I. *Phys. Rev. Lett.* **1977**, *38*, 892–894.

[51] Oepts, D.; van der Meer, A. F. G.; van Amersfoort, P. W. *Infr. Phys. Tech.* **1995**, *36*, 297 – 308; Proceedings of the Sixth International Conference on Infrared Physics. 19

[52] CLIO; *Centre Laser Infrarouge d'Orsay*; `http://clio.lcp.u-psud.fr/clio_eng/FEL.html`; 2012. 19

[53] Giordmaine, J. A.; Miller, R. C. *Phys. Rev. Lett.* **1965**, *14*, 973–976. 20

[54] Guyer, D.; *OPO User Manual*; LaserVision; WA, USA. 20, 21

[55] Bakker, J. M.; Sinha, R. K.; Besson, T.; Brugnara, M.; Tosi, P.; Salpin, J.-Y.; Maître, P. *J. Phys. Chem. A* **2008**, *112*, 12393–12400. 21, 100, 166, 198

[56] Bakker, J. M.; Salpin, J.-Y.; Maître, P. *Int. J. Mass Spectrom.* **2009**, *283*, 214 – 221; A Collection of Invited Papers Dedicated to Michael T. Bowers on the Occasion of his 70th Birthday. 100, 166, 198

[57] Sinha, R. K.; Nicol, E.; Steinmetz, V.; Maître, P. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 758 – 772. 21, 23, 24, 100

[58] Altinay, G.; Metz, R. B. *Int. J. Mass Spectrom.* **2010**, *297*, 41 – 45; Special Issue: Ion Spectroscopy. 22

[59] Jr., W. K. M.; Szczepanski, J.; Pearson, W. L.; Powell, D. H.; Dunbar, R. C.; Eyler, J. R.; Polfer, N. C. *Int. J. Mass Spectrom.* **2010**, *297*, 131 – 138; Special Issue: Ion Spectroscopy. 22

[60] Scuderi, D.; Bakker, J.; Durand, S.; Maître, P.; Sharma, A.; Martens, J.; Nicol, E.; Clavaguéra, C.; Ohanessian, G. *Int. J. Mass Spectrom.* **2011**, *308*, 338 – 347. 23, 24, 57, 100, 195

[61] Maître, P.; Lemaire, J.; Scuderi, D. *Phys. Scr.* **2008**, *78*, 058111. 24, 100, 166, 198

[62] Chernushevich, I. V.; Thomson, B. A. *Anal. Chem.* **2004**, *76*, 1754–1760; PMID: 15018579. 26

[63] Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem.* **1976**, *10*, 1–19. 29

[64] Friesner, R. A. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6648–6653. 30

[65] Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688. 30

[66] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217. 30

[67] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718. 30

[68] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564; PMID: 20136072. 31, 33

[69] Ponder, J. W.; *TINKER Spoftware Tools for Molecular Design*; 2001. 31, 39

[70] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566. 31

[71] Kaminský, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774–1788. 33, 34

[72] Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. *Proc. Natl. Acad. Sci.* **2008**, *105*, 6290–6295. 33

[73] van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407. 33

[74] Thole, B. *Chem. Phys.* **1981**, *59*, 341 – 350. 33

[75] Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quantum Chem.* **2007**, *107*, 1390–1395. 34

[76] Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497–1506.

[77] Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3450–3462. 34

[78] Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6679–6685. 34

[79] Ädcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589–1615; PMID: 16683746. 34

[80] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons Inc., 2004. 34, 41, 55, 56, 58

[81] Beeman, D. *J. Comp. Phys.* **1976**, *20*, 130 – 139. 35, 37

[82] Levitt, M.; Meirovitch, H.; Huber, R. *J. Mol. Biol.* **1983**, *168*, 617 – 620. 35, 37

[83] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690. 38

[84] Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141 – 151. 38, 39

[85] Penev, E. S.; Lampoudi, S.; Shea, J.-E. *Comput. Phys. Commun.* **2009**, *180*, 2013 – 2019. 39

[86] Penev, E.; Ireta, J.; Shea, J.-E. *J. Phys. Chem. B* **2008**, *112*, 6872–6877; PMID: 18476737. 39

[87] Semrouni, D.; Balaj, O.; Calvo, F.; Correia, C.; Clavaguéra, C.; Ohanessian, G. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 728–738; 10.1016/j.jasms.2010.01.029. 39, 49, 71, 115

[88] Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons Ltd.: England, 2007. 41, 55, 56, 58

[89] Szabo, A.; Ostlund, N. *Modern Quantum Chemistry - Introduction to Advanced Electronic Structure Theory*, revised ed.; Dover Publications, Inc.: Mineola, New York, 1996.

[90] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89. 41

[91] Nooijen, M.; Sabin, J. R.; Brndas, E., Eds.; Advances in Quantum Chemistry, Vol. 56; Academic Press, 2009; pp 181 – 216. 46

[92] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. 48

[93] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100. 48

[94] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[95] Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200 – 206. 48

[96] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789. 48

[97] Zhao, Y.; Truhlar, D. *Theor. Chem. Acc.* **2008**, *120*, 215–241; 10.1007/s00214-007-0310-x. 48

[98] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126–13130; PMID: 17149824. 48

[99] Karton, A.; Tarnopolsky, A.; Lamere, J.-F.; Schatz, G. C.; Martin, J. M. L. *J. Phys. Chem. A* **2008**, *112*, 12868–12886; PMID: 18714947. 48

[100] Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799. 49

[101] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473. 49

[102] Černý, J.; Jurečka, P.; Hobza, P.; Valdés, H. *J. Phys. Chem. A* **2007**, *111*, 1146–1154; PMID: 17253667. 49

[103] Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657–2664. 50

[104] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305. 51

[105] Kendall, R. A.; Früchtl, H. A. *Theor. Chem. Acc.* **1997**, *97*, 158–163; 10.1007/s002140050249. 53

[106] Vahtras, O.; Almlöf, J.; Feyereisen, M. *Chem. Phys. Lett.* **1993**, *213*, 514 – 518.

[107] Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340; 10.1007/s002140050269. 53

[108] Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143 – 152. 53

[109] Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183. 53

[110] Hellweg, A.; Httig, C.; Hfener, S.; Klopper, W. *Theor. Chem. Acc.* **2007**, *117*, 587–597; 10.1007/s00214-007-0250-5. 53

[111] James B. Foresman, A. F. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed.; Guassian, Inc.: Pittsburgh, PA, 1996. 53

[112] Möller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

[113] Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503 – 506.

[114] Head-Gordon, M.; Head-Gordon, T. *Chem. Phys. Lett.* **1994**, *220*, 122 – 128. 53

[115] Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256–4266. 55

[116] Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409 – 418. 55

[117] Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154–5161. 55

[118] Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 2872–2887. 57

[119] Bouteiller, Y.; Poully, J. C.; Desfrancois, C.; Gregoire, G. *J. Phys. Chem. A* **2009**, *113*, 6301–6307; PMID: 19435330.

[120] Andersson, M. P.; Uvdal, P. *J. Phys. Chem. A* **2005**, *109*, 2937–2941; PMID: 16833612.

[121] Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683–11700. 57

[122] Semrouni, D.; Clavaguéra, C.; Dognon, J.-P.; Ohanessian, G. *Int. J. Mass Spectrom.* **2010**, *297*, 152 – 161; Special Issue: Ion Spectroscopy. 57

[123] *TURBOMOLE V6.1 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from*
`http://www.turbomole.com`. 60

[124] Frisch, M. J.; et al.; *Gaussian 09 Revision A.1*; Gaussian Inc. Wallingford CT 2009. 60

[125] Humphrey, W.; Dalke, A.; Schulten, K.; *VMD – Visual Molecular Dynamics*; 1996. 60

[126] Schaftenaar, G.; Noordik, J. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123–134; 10.1023/A:1008193805436. 60

[127] Pace, N. C.; Scholtz, M. J. *Biophys. J.* **1998**, *75*, 422–427. 61

[128] Kohtani, M.; Kinnear, B. S.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 12377–12378. 61, 62

[129] Wei, Y.; Nadler, W.; Hansmann, U. H. E. *J. Chem. Phys.* **2007**, *126*, 204307. 62

[130] Hol, W. G. *Prog. Biophys. Mol. Biol.* **1985**, *45*, 149 – 195. 63

[131] Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1999**, *121*, 3494–3501. 63

[132] Polfer, N. C.; Oomens, J.; Dunbar, R. C. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2744–2751. 73

[133] Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci.* **1951**, *37*, 205–211. 74

[134] Hudgins, R. R.; Mao, Y.; Ratner, M. A.; Jarrold, M. F. *Biophys. J.* **1999**, *76*, 1591 – 1597. 100, 151, 152, 173

[135] Arrondo, J. L. R.; Muga, A.; Castresana, J.; ni, F. M. G. *Prog. Biophys. Mol. Biol.* **1993**, *59*, 23 – 56. 108

[136] Stearns, J. A.; Seaiby, C.; Boyarkin, O. V.; Rizzo, T. R. *Phys. Chem. Chem. Phys.* **2009**, *11*, 125–132. 112, 195

[137] Snoek, L.; Robertson, E.; Kroemer, R.; Simons, J. *Chem. Phys. Lett.* **2000**, *321*, 49 – 56. 112, 195

[138] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406. 115

[139] Vaden, T. D.; de Boer, T. S. J. A.; Simons, J. P.; Snoek, L. C.; Suhai, S.; Paizs, B. *J. Phys. Chem. A* **2008**, *112*, 4608–4616; PMID: 18444632. 152

[140] Kaleta, D. T.; Jarrold, M. F. *J. Phys. Chem. A* **2002**, *106*, 9655–9664. 153, 154

[141] Tapiero, H.; Mathé, G.; Couvreur, P.; Tew, K. *Biomed. Pharmacothe.* **2002**, *56*, 439 – 445. 178

[142] Penzotti, J. L.; Fozzard, H. A.; Lipkind, G. M.; Jr., S. C. D. *Biophys. J.* **1998**, *75*, 2647 – 2657. 178

[143] Hwang, D.; Noguchi, T.; Taylor, S. L., Ed.; Advances in Food and Nutrition Research, Vol. 52; Academic Press, 2007; pp 141 – 236. 178

[144] Arakawa, T.; Timasheff, S. N. *Biochemistry* **1984**, *23*, 5924–5929. 178

[145] Smith, J. S.; Scholtz, J. M. *Biochemistry* **1996**, *35*, 7292–7297. 178

[146] Camilloni, C.; Rocco, A. G.; Eberini, I.; Gianazza, E.; Broglia, R.; Tiana, G. *Biophys. J.* **2008**, *94*, 4654 – 4661. 178

[147] Vidal, L.; Riekkola, M.-L.; Canals, A. *Anal. Chim. Acta* **2012**, *715*, 19 – 41. 178

[148] Shang, Y.; Li, H.; Zhang, S.; Xu, H.; Wang, Z.; Zhang, L.; Zhang, J. *Chem. Eng. J.* **2011**, *175*, 324–329. 178

[149] Liu, X.; Zhang, S.; Zhou, G.; Wu, G.; Yuan, X.; Yao, X. *J. Phys. Chem. B* **2006**, *110*, 12062–12071. 178, 254

[150] Liu, X.; Zhou, G.; Zhang, S.; Wu, G.; Yu, G. *J. Phys. Chem. B* **2007**, *111*, 5658–5668.

[151] Liu, J.; Wang, F.; Li, Z.; Zhou, J.; Chen, J.; Xia, C. *Struct. Chem.* **2011**, *22*, 1119–1130; 10.1007/s11224-011-9807-y. 178, 254

[152] Raczyńska, E. D.; Cyrański, M. K.; Gutowski, M.; Rak, J.; Gal, J.-F.; Maria, P.-C.; Darowska, M.; Duczmal, K. *J. Phys. Org. Chem.* **2003**, *16*, 91–106. 179

[153] Rozas, I.; Kruger, P. E. *J. Chem. Theory Comput.* **2005**, *1*, 1055–1062. 179, 190, 191, 192, 259

[154] Haas, D. J.; Harris, D. R.; Mills, H. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **1965**, *19*, 676–679. 191, 192, 193

[155] Todorova, T.; Kröcher, O.; Delley, B. *J. Mol. Struct.* **2009**, *907*, 16 – 21. 190, 191, 193, 195, 251

[156] Angell, C. L.; Sheppard, N.; Yamaguchi, A.; Shimanouchi, T.; Miyazawa, T.; Mizushima, S. *Trans. Faraday Soc.* **1957**, *53*, 589–600. 195, 196

[157] Sension, R. J.; Hudson, B.; Callis, P. R. *J. Phys. Chem.* **1990**, *94*, 4015–4025. 190, 195

[158] Yamashita, T.; Takatsuka, K. *J. Chem. Phys.* **2007**, *126*, 074304. 195

[159] Jung, J. O.; Gerber, R. B. *J. Chem. Phys.* **1996**, *105*, 10332–10348.

[160] Kim, J.; Lee, H. M.; Suh, S. B.; Majumdar, D.; Kim, K. S. *J. Chem. Phys.* **2000**, *113*, 5259–5272.

[161] Chaban, G. M.; Xantheas, S. S.; Gerber, R. B. *J. Phys. Chem. A* **2003**, *107*, 4952–4956.

[162] Low, G. R.; Kjaergaard, H. G. *J. Chem. Phys.* **1999**, *110*, 9104–9115. 195

[163] Marta, R. A.; Ph.D. thesis; University of Waterloo; 2009. 207

[164] Marta, R. A.; Wu, R.; Eldridge, K. R.; Martens, J. K.; McMahon, T. B. *Int. J. Mass Spectrom.* **2010**, *297*, 76 – 84; Special Issue: Ion Spectroscopy. 207

[165] Stefansson, M.; Sjberg, P. J. R.; Markides, K. E. *Anal. Chem.* **1996**, *68*, 1792–1797. 263