UNIVERSITY OF CALIFORNIA
RIVERSIDE

Modeling Low-Barrier Hydrogen Bonds and Solution Effects

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Kelly Lynne Theel

August 2013

Dissertation Committee:

Dr. Gregory J. O. Beran, Chairperson
Dr. Chia-en Angelina Chang
Dr. Eric L. Chronister

The Dissertation of Kelly Lynne Theel is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

Portions of this thesis are reproduced from reference [1] with permission from the PCCP Owner Societies, and from reference [2], copyright 2009, American Chemical Society.

Firstly, thanks to those who set me on a course toward graduate school: Dr. Sabine Goldberg, who has been a constant, and thankfully very accessible, example of everything a scientist could aspire to be; and Dr. Tom Morton, who took me under his wing when I first arrived at UCR, and whose practical advice showed me the path to my dreams.

Secondly, thanks to Dr. Len Mueller and the entire Beran group for support, both technical and moral. Dr. Shuhao Wen's work deriving and coding the multipole translations was invaluable. It's been a privilege working with such a talented group of people.

Special thanks go out to Dr. Greg Beran, whom I chose as an advisor because he has a great personality that I knew I could work with for five years. I was pleased to find that his technical skills are stellar, but most importantly, his constant encouragement and no-pressure management style gave me the freedom to try hard and fail a few times without penalty in his eyes. His supportive stance gave me room to grow, on a number of levels, and has meant more to me than I can express.

Lastly, as usual, I would like to thank my family and friends for their patience. I swear that someday soon I will participate in more than just weddings and funerals.

To my son, Carolus Magnus Theel.

ABSTRACT OF THE DISSERTATION

Modeling Low-Barrier Hydrogen Bonds and Solution Effects

by

Kelly Lynne Theel

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, August 2013
Dr. Gregory J. O. Beran, Chairperson

In the first project, we model low barrier hydrogen bond systems, which have a unique energy profile. Classification of a hydrogen bond as a low barrier hydrogen bond requires assignment of the experimental vibrational spectrum, as well as prediction of the zero point energy. We modeled these systems using a 2D potential energy surface that explored the symmetric and asymmetric stretch of the proton transit. The resulting predicted spectra and zero point energies helped classify three molecules as hydrogen bonds or low barrier hydrogen bonds.

In addition, solvation effects were explored. First, a traditional implicit polarizable continuum model, and standard explicit microsolvation techniques were used to explain why only two out of four isoenergetic gas phase conformers appeared in infrared multiple photon dissociation spectra. Our calculations suggest that this resulted from an experimental artifact. We showed that the two configurations were much more stable in water. Since the ion delivery system was electrospray, it was concluded that the solvent effectively determined the structure of the gas phase ions, since the barrier to isomerization was quite high.

While these simple models were able to explain those experiments, a much more robust and accurate model for solvent effects is needed, especially for systems with nonhomogeneous environments (such as protein binding sites). Solvent structure and features of the environment are completely neglected in both of the rudimentary models. Quantum Mechanical (QM)/ Molecular Mechanical (MM) sampling of solvent configuration space is effective, but extremely expensive, due to the high cost of the repeated QM calculations. Our model samples the configuration space of the solvent using inexpensive, but configurationally accurate, molecular dynamics, then includes the information from several timesteps in one single expensive QM calculation. A mathematically rigorous coarse-graining scheme translates and scales the multipoles, polarizabilities and the dispersions to grid points around the solute region, and then computes the energy of the system in a single QM/MM calculation. The results are extremely promising.

Contents

List of Figures

List of Tables

Chapter 1

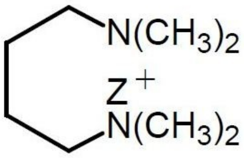
Introduction

1.1 Applications of computational chemistry

This body of work demonstrates how computational chemistry can work hand-in-glove with experimental evidence to gain greater physical insights into complex systems than either could achieve alone. The first work explores what characteristics make molecules good candidates for a class of short, strong hydrogen bonds, which are of great interest in biology and catalysis. The second body of work explores solvation effects, and how to capture the energetics of solvated systems affordably and accurately. Solvation effects have an impact on many chemical systems, and modeling them can be quite challenging, due to the many degrees of freedom in the solvent.

1.2 Modeling low-barrier hydrogen bonds

As an example of how computational chemistry can help shape the approach to a problem, our work in Chapter 2 helps to define characteristics that make molecules good candidates for low barrier hydrogen bond (LBHB) systems, and to confirm their



1a, Z=H

Figure 1.1: Schematic of a molecule with a LBHB. Shown here is the proton between two nitrogens, and so throughout we will refer to the proton transit as the N-H (in the N-H-N) stretch, and movements which change the length of the hydrogen bond as the N-N stretch. It should, however, be understood that this discussion can be generalized to any LBHB.

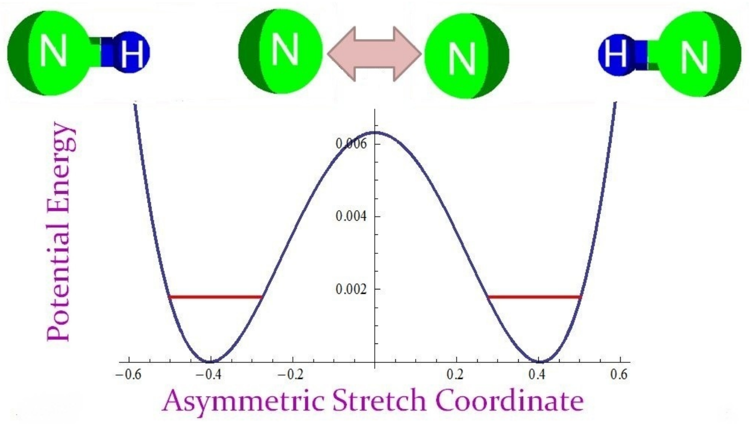


Figure 1.2: Typical hydrogen-bond 1-D PES. The ZPE (in red) is below the energy needed to occupy the midpoint of the transit, consistent with the observation that hydrogen bonds usually are more tightly bound to one of the two electronegative atoms, and don't freely move between them.

experimental assignments. LBHBs are shorter and stronger than traditional hydrogen bonds, and their important role in enzyme catalysis is increasingly recognized [3]. There is also evidence that LBHBs can facilitate spontaneous self-alignment between molecules. [4]

LBHBs possess an interesting energy profile which allows the proton to transit between one of the bounding electronegative atoms to the other. A diagram of a sample molecule with a LBHB is in Figure 1.1, and for ease of reference we will use this to label the motions, although the methods described here would be suitable to model

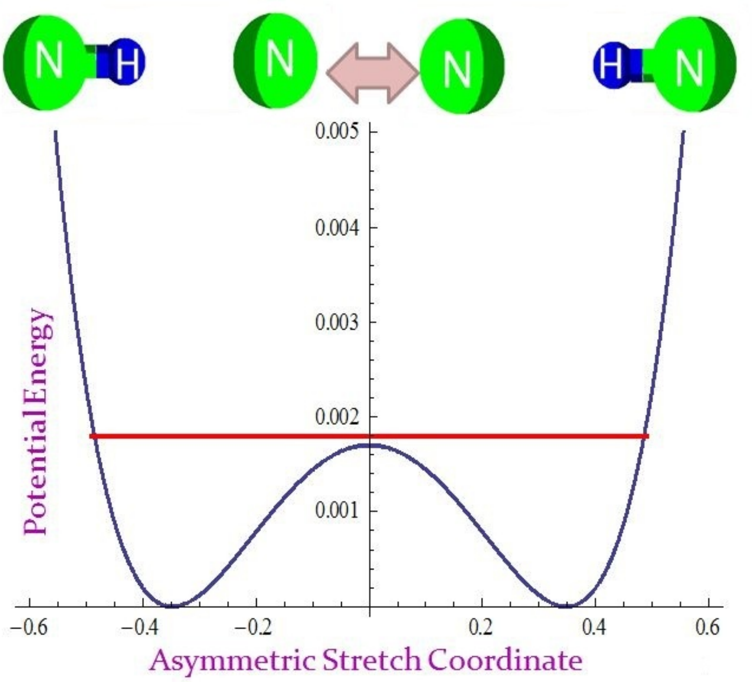


Figure 1.3: LBHB 1-D PES. In contrast with Figure 1.2, the ZPE is very close to the barrier to proton transit. Accordingly, the proton can pass between the two electronegative atoms more freely.

many systems suspected of having a LBHB. We will refer to the proton transit as the N-H movement, and the lengthening or shortening of the hydrogen bond as the N-N movement. Normal hydrogen bond potential energy surfaces (PESs) usually have a deep double well, as shown in Figure 1.2, with the zero point energy (ZPE) lower than the barrier. This is consistent with the fact that in the ground state, a bridging proton is usually associated with one or the other of the electronegative atoms. In contrast, Figure 1.3 shows that the PES of a LBHB differs from a typical hydrogen bond in that the barrier is lower, near the ZPE, which implies that proton transit is accessible from the ground state in LBHBs.

LBHBs are characterized by three key features: (1) a low frequency vibration corresponding to the proton transit, which (2) shifts upon deuteration, and as discussed above, (3) a ZPE near the barrier to transit. So to confirm that a molecule has a LBHB, a particular peak in the vibrational spectra must be assigned to the anharmonic N-H (in N-H-N) stretch, and that wavelength must be seen to change upon deuteration. Also, the ZPE and the barrier to proton transit must be predicted, and the two must be similar. To assign peaks with greater certainty, computational models of the system can predict the spectra associated with a particular motion. Also the ZPE can be calculated, and the barrier to proton transit mapped out with single point energy calculations.

These methods require a wave function ψ, which we find as a solution to the time-independent Schrödinger equation:

$$\hat{H} \left|\psi\right\rangle = E \left|\psi\right\rangle \tag{1.1}$$

The Hamiltonian may be represented as:

$$\hat{H}(r,R) = \hat{T}(R) + \hat{T}(r)\hat{V}(r,R) \tag{1.2}$$

where $\hat{T}(R)$ is the kinetic energy as a function of R, the positions of the nuclei, and $\hat{T}(r)$ is the kinetic energy as a function of r, the positions of the electrons. Similarly, $\hat{V}(r,R)$ is the potential energy of the system as a function of both R and r. The Born-Oppenheimer approximation states that because nuclei are so much more massive than electrons, their motions occur on a different time scale. This allows us to effectively decouple the nuclear potential energy from the electronic potential energy, resulting in:

$$\hat{H}(R,r) = \hat{T}(R) + \hat{T}(r) + \hat{V}(R) + \hat{V}(r) \tag{1.3}$$

The Born-Oppenheimer approximation to the Schrödinger equation allows us to combine the last three terms into a parametric function, $V(r;R)$, where for each R

there is a corresponding set of solutions as r is varied. This is equivalent to saying the electronic energy is a function of the nuclear positions. To predict the spectra associated with a motion, for example a bond stretch, the parametric function $V(r; R)$ can be modeled as a Taylor expansion. The first few terms are shown here in one dimension, $V(x)$, since the interaction between two atoms is a function of the distance x:

$$V(x) = (1/2)kx^2 + (1/6)\gamma_3 x^3 + (1/24)\gamma_4 x^4 ... \tag{1.4}$$

Often it is convenient, and an accurate representation of the PES of the movement, to truncate the series at the first non-zero term. This leaves a function with the form of a harmonic oscillator, which closely resembles the energy profile for a bond stretch in the low-energy region. In Figure 1.4 a harmonic oscillator is shown in green, while the energy profile of the asymmetric stretch as the proton makes its transit is shown in dark blue. It is clear that the harmonic oscillator has an entirely different shape than the double well potential, and is a poor choice to model LBHBs.

Instead, for LBHB systems we must construct a PES and solve the nuclear Schrödinger equation on that PES directly. Doing this for the full many-dimensional PES is computationally intractable. Therefore, we focus on a simplified, 2-D slice of the PES, constructed by varying the N-N length (the symmetric stretch), and the N-H length in N-H-N, which corresponds to the proton moving between the nitrogens (the asymmetric stretch). These two motions are chosen because they are the most telling vibrations in the proton transit. Single point energies are computed at a large number of grid points, especially in low-energy regions. A polynomial fit to the computed PES is then made, and this function of the two coordinates $V(symmetric, asymmetric)$ is used as the $V(r; R)$ in the Schrödinger equation:

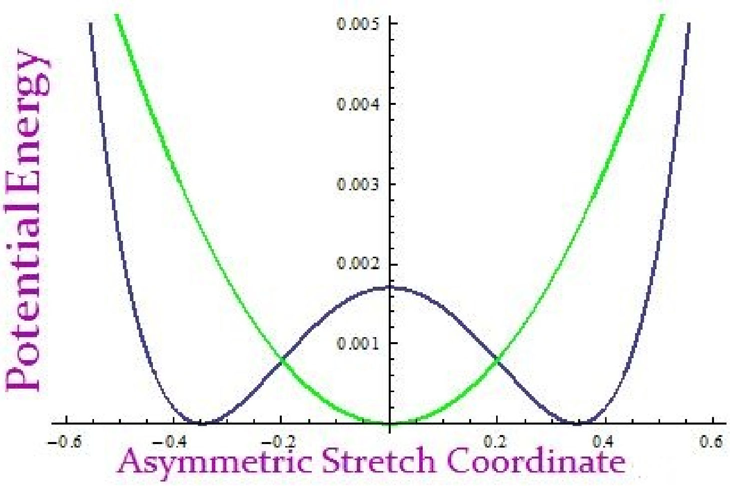


Figure 1.4: Harmonic oscillator, in green, compared to LBHB energy profile, in dark blue. The harmonic oscillator is obviously not a good choice to approximate the LBHB energy, especially in the interesting center region, which corresponds to the proton transit.

$$\hat{H}(R,r) = \hat{T}(R) + \hat{V}(r;R) \tag{1.5}$$

Equation 1.5 can be substituted back into Equation 1.1:

$$[\hat{T}(R) + \hat{V}(r;R)]\,|\psi\rangle = E\,|\psi\rangle \tag{1.6}$$

By the variational principle, any trial wave function that we choose to insert into the Schrödinger equation will yield an energy that is greater than or equal to the actual ground state energy of the system. So, we can variationally optimize a wave function as a linear combination of a well-chosen basis set, such as harmonic oscillator wave functions (ϕ). Assuming N basis functions are used, there will be coefficients $c_1 ... c_N$. To find the variationally optimal solution, we evaluate the nuclear Hamiltonian matrix elements in the harmonic oscillator basis, $\langle\phi_a|\,\hat{H}\,|\phi_b\rangle$, and diagonalize the Hamiltonian matrix. For illustrative purposes, we have taken the determinant of the Hamiltonian minus the energy, below, although modern computer software may use other algorithms to diagonalize the Hamiltonian:

$$det \begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0 \tag{1.7}$$

H_{ab} is $\langle\phi_a| \hat{H} |\phi_b\rangle$, E is the energy which is unknown, and S_{ab} is the overlap matrix which corresponds to the Dirac delta since our basis set is orthonormal. Effectively, this simultaneously solves a set of equations setting the first derivative of $\hat{H} |\psi\rangle - E |\psi\rangle$ with respect to each coefficient equal to zero. This minimizes the difference between $\hat{H} |\psi\rangle$ and $E |\psi\rangle$, which according to the Schrödinger equation, are equal for the actual wave function. So we are optimizing the coefficients to get as close to the actual wave function as our basis set will allow. The energy is also unknown, however, and since we have N equations and $N + 1$ unknowns, we get a set of answers corresponding to N values of E.

The lowest value of E is the ground state eigenvalue, and the corresponding N coefficients are used for the linear combination of harmonic oscillators to produce the eigenstate.

$$\psi = c_1\phi_1 + c_2\phi_2 + ... + c_N\phi_N \tag{1.8}$$

Once a wave function has been found, the ground state energy determines the ZPE, and the excited states are assigned to either the symmetric (N-N) stretch or the asymmetric stretch (N-H in the N-H-N) based on a plot of the wave function. This reduced dimensional model is well suited for working with LBHBs, and proved extremely valuable in identifying them.

It has been theorized that chelated N-H-N systems might have a preference for a low barrier to hydrogen transit if the basicity of the nitrogens is similar and the N-H-N angle is near linear. To test this theory, a large experimental and theoretical collaboration was undertaken. [1] Indeed, our modeling evidence, when combined with these experimental results, supports the theory that LBHBs are more likely to be found in proton bound systems with linear N-H-N conformations and nitrogens with similar basicities.

1.3 Modeling solvent effects

1.3.1 Simple models

In addition to collaborating with experimentalists to model LBHBs, we also helped solve the mystery of why, out of four essentially isoenergetic configurations of dimers, only two were presenting in experimental spectra. Gas phase modeling predicted that four conformers of proton-bound adenine dimers had low enthalpies, within 4 kJ mol^{-1} of each other. One would expect to see all four represented in gas phase spectroscopy. However, Fridgen et al. [2] performed infrared multiple photon dissociation (IRMPD) experiments using electrospray as the source of gas phase proton-bound adenine dimers, and found evidence of only two of the four isomers. Our efforts helped explain why.

Since the experimental source of ions was electrospray, it was prudent to examine solvent effects on the four proton-bound dimers of interest. Two standard quantum mechanical (QM) modeling techniques were used to do this: a polarizable continuum model (PCM), which is an implicit solvent, and microsolvation, an explicit model. PCMs have a featureless electrostatic potential, as well as empirical dispersion and cavitation terms that vary with the surface area of the molecule. [5] Since the terms are empirical, every solvent needs a tailored PCM in which a QM calculation is then run. This simple, affordable model often captures solvent effects remarkably well - however, it has no terms for explicit interaction with solvent. Consequently, this mean field approach will not capture the effects of high-energy localized interactions, such as solute-solvent hydrogen bonding, which might be crucial to the conformation and/or energetics of the system.

Microsolvation involves placing a small number of solvent molecules, in this case water, around the solute molecule. A QM minimization is performed on these best-guess starting configurations, and the resulting low energy configurations are examined. So, while this method is dependent on the chemical intuition of the modeler, often the conformations found yield important insights into the system. Limitations include the fact that solvent, by its very nature, is fluid, and requires extensive sampling of the configuration space to completely capture the energetics. However, since low energy conformations are highly represented, this small sampling is often enough. Also, the interaction between the explicit solvent and the next solvation shell, or even the bulk solvent, is completely neglected, and so some conformational and energetic effects are never explored.

These methods are standard compromises between the high cost of adequately sampling the configuration space of a very large system, such as a molecule in bulk water, and the necessity of capturing the important energetics. In the case of the protonated adenine dimers, both of these methods predicted the same relative energies for the four proton-bound dimers, and it was clear that two were energetically favored over the others when solvated. Unexpectedly, these solvation effects impacted the gas phase measurements, because the barrier to transition between the conformations favored in solution and the other two low-energy conformations found in the gas phase was predicted to be larger than the 120-130 kJ mol^{-1} binding energy of the dimers.

1.3.2 Modeling the structure of the solvent

While in the adenine proton-bound dimer system these incomplete descriptions of the solvent effects proved useful, simple solvent models like these can sometimes fail spectacularly. For example, the Cope elimination can have a rate increase on the order of a million, simply by changing from protic to aprotic solvent, but PCMs predict no change at all. [6–8] One can also envision that for a non-homogeneous environment, such as a protein binding site, sampling more than just the near-neighbor molecules would be necessary since the features of the surroundings, including the solvent itself, would have an impact on shaping the conformation of the solvent structure of the area of interest. To treat it accurately, this type of system would require a much larger explicit solvent region than microsolvation affords.

Acevedo and Jorgensen [9] were able to successfully model the Cope elimination by exploring the configuration space along the reaction coordinate of a periodic box containing a large number of explicit solvent molecules and the reactants. They used mixed QM and molecular mechanic (MM) methods. However, this sort of sampling comes at a high computational cost. The main expense comes from the QM calculations, which are repeated at every time step in the simulation, in most QM/MM schemes. This can amount to tens of thousands of costly QM calculations. Alternatively, inexpensive but conformationally accurate MM can be used to explore the configuration space, but each step still needs a QM evaluation of the energetics.

An even less expensive scheme is to use MM methods to generate conformations, then somehow average the solvent conformations into a single QM calculation that captures the energetics of the all of the sampled configurations. [10–14] This has the benefit of moving the large number of sampling calculations into the MM domain, at little to no impact on the accuracy of the configurations, while still calculating the sensitive energetics with QM. The accuracy of the QM energy calculation depends on the level of theory used to average the solvent configurations. Sanchez et al. [10] used point charges to model all of the solvent effects, while we propose a much more accurate method here.

Since the first solvation shell dominates the solvation energy profile, we will treat that explicitly, using our QM/MM methodology that has been proven to work well in condensed phase systems. [15–17] We sample solvent configurations using inexpensive MD. A large number of well separated time steps are then superimposed and scaled, yielding an averaged environment which we can then interact with the explicit solute region. The solvent effects will be modeled by translating the multipole moments, polarizabilities and dispersion terms, as calculated by our ab initio based force field (AIFF), [16, 18] to gridpoints surrounding the explicit region. Translations are rigorous, and would be exact if the multipole expansion were infinite. The translated properties are then used to calculate electrostatic, polarizability and dispersion energies for the whole system. This coarse graining method effectively captures the structure of the bulk solvent, and interacts that with the explicit region, yielding an energy that is a function of the actual features of the system instead of a mean field.

In Chapter 4 we provide a demonstration of this approach and obtain very encouraging results.

Chapter 2

Vibrations of a chelated proton in a protonated tertiary diamine

2.1 Introduction

2.1.1 Context

In this example of how computational chemistry can aid experimentalists, we were interested in identifying a class of molecules as possible low-barrier hydrogen-bonds (LBHBs). LBHBs have a zero point energy near the barrier to proton transit, as well as a low frequency vibration associated with that proton movement. To accurately predict the characteristics of this system, we generated a two dimensional theoretical potential energy surface, by optimizing the molecule along two coordinates of the multidimensional potential energy surface. This reduced-dimensionality surface was then used to solve the nuclear Schrödinger equation. The resulting wavefunction was used to predict spectra and the zero point energy. Our calculations were extremely helpful in peak assignment, as well as verifying that the zero point energy is indeed near the top of the potential energy barrier. This was instrumental in identifying one of the molecules of interest as a having a low-barrier hydrogen bond, and helped elucidate under what conditions molecules are likely to have LBHBs.

2.1.2 Motivation

Quantum-mechanical implications of a symmetric double-well potential have been widely discussed for many years. [19–23] For molecular systems, a proton moving between two equivalent sites over a sufficiently low barrier ought to have a zero point level close to the barrier top, as schematically indicated by Fig. 1. Such a system should display unusual properties, including (1) having a low frequency vibration corresponding to transit of the H^+ between sites, and (2) having the maximum of its zero point vibrational wavefunction near the top of the potential energy barrier. Recently we published experimental evidence that ion **1a** (Figure 2.2) contains a chelated proton, i.e. one whose average position midway between two basic nitrogens exemplifies the second of the above properties. [24] Here we report vibrational spectroscopy for this ion, consistent with the first of the aforementioned properties, confirming that the isolated monoprotonated ion, as well as its salts, contains a LBHB.

Diamines enjoy greater basicity than their monofunctional analogues. A great deal of study has been devoted to proton sponges, basic nitrogens either attached to (directly or via sp^3 carbons) or incorporated within aromatic rings. In general, the monoprotonated salts of such species have NHN bond angles that deviate substantially from linearity. Salts of these ions often display more than one peak in their solid phase ^{15}N NMR spectra, [25] demonstrating that the strong intramolecular hydrogen bond does not behave like an LBHB.

Horbatenko and Vyboishchikov recently published a theoretical analysis of pro-

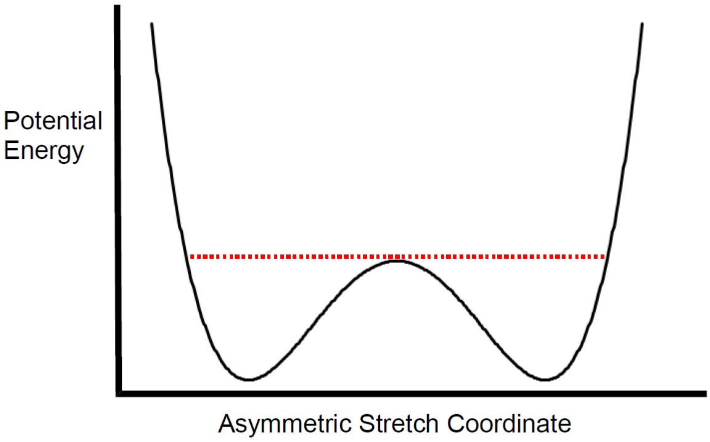
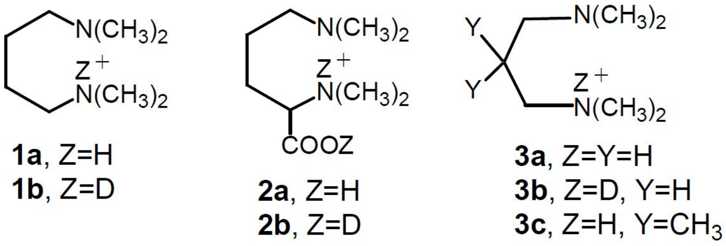


Figure 2.1: Schematic representation of a Low Barrier Hydrogen Bond (LBHB), in which the energy of the zero point level (represented by the dotted line) is comparable to the central energy barrier of the double well potential.



1a, Z=H
1b, Z=D

2a, Z=H
2b, Z=D

3a, Z=Y=H
3b, Z=D, Y=H
3c, Z=H, Y=CH_3

Figure 2.2: Conjugate acid ions from acyclic α, ω-diamines.

tonated proton sponges that have low barriers to H^+ passage from one nitrogen to the other. [26] By calculating 3-dimensional potential energy surfaces for examples that possess a plane of symmetry, they predict a vibrational frequency between 300 and 350 cm^{-1} for a case where the central barrier ranges from 4.5 to 6.5 kJ mol^{-1}. The frequency has a value an order of magnitude lower than that of an ordinary NH bond, demonstrating that transit between nitrogens exerts a profound effect on vibrational frequency. Because the NHN bond angle is far from linear in these proton sponges, considerable mixing occurs between the asymmetric NHN stretch and the in-plane NHN bend.

Monoprotonated ions from diamines form easily within a mass spectrometer, in which they can be isolated. [27] In the present study ions produced from an electrospray source have been trapped in a Fourier Transform Ion Cyclotron Resonance spectrometer (FT-ICR), where irradiation by IR pulses from a free-electron laser induces frequency-dependent dissociation via loss of neutral dimethylamine. After absorbing a single photon at a resonant frequency, the density of vibrational states becomes so great that the ion absorbs subsequent photons rapidly and decomposes unimolecularly. Infrared Resonant Multiple Photon Dissociation (IRMPD) thus provides an action spectrum monitored by mass spectrometry, which reveals the IR absorption profile of an isolated ion. Isotopic substitution permits the assignment of absorption bands to motions involving the bridging H^+.

In parallel with these gas phase experiments, crystalline samples of the monoprotonated salts have been examined by NMR, X-ray crystallography, inelastic neutron scattering (INS), and Raman spectroscopy at high pressures. It is generally accepted that the increased basicity of aliphatic diamines in the gas phase arises from formation of cyclic structures that contain intramolecular strong hydrogen bonds (or proton bridges), as the left hand side of the equation in Figure 2.3 illustrates. The presence of

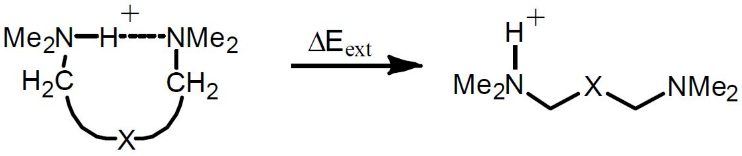


Figure 2.3: Cyclic structures with strong intramolecular hydrogen bonds (proton bridges) have increased basicity.

these intramolecular strong hydrogen bonds has been demonstrated by the solid phase structural studies of the triflate salt of **1a** (X = CH_2CH_2), [24] whose neutral conjugate base is commonly called tetramethylputrescine. Tetramethylputrescinium ions are illustrated in Figure 2.2, along with relevant homologues. Recent solution-phase studies by Pluth, Bergman, and Raymond [28] explored the rates of ring opening, as illustrated in Figure 2.3. They find that chainlength affects the strength of proton bridges. The present work provides evidence showing that, in a case where the two nitrogens have the same basicity and the NHN bond angle is nearly linear, the proton bridge behaves like a low barrier hydrogen bond (LBHB), in which the proton is poised midway between the two nitrogens.

What properties indicate that an intramolecular strong hydrogen bond corresponds to an LBHB? This paper first addresses that question using theory as a prelude to experiment. DFT calculations gauge suitable candidates having a low barrier to proton transit from one nitrogen to the other, a nearly linear NHN bond angle, and identical proton affinities of the two amino groups. It turns out that the ions that give the lowest barriers for proton transit include ion 1 and also have NHN angles close to 180°.

A 2-dimensional potential energy surface (PES) for ion 1 has been constructed. To ensure an accurate and computationally affordable description of the noncovalent interactions governing this PES, a double-hybrid density functional was used. The molecular system of 1 is too large for the complete potential energy surface to be probed, so predictions based on a greatly reduced dimensionality have been explored for comparison with experiment. Taking advantage of the nearly linear NHN alignment, the calculations predict vibrational quantizations by treating 1 as though it were a linear triatomic. Within the limitations of this simplifying assumption, the reduced dimensionality approach computes frequencies and the shifts expected when bridging H^+ is replaced by D^+. Proceeding from theory to experiment, measured IRMPD spectra are compared with calculated band positions, which agree with the theoretical predictions for an LBHB. In the 300-600 cm^{-1} domain, the observed NHN asymmetric stretching band for **1a** matches the value computed from the 2-dimensional PES. The result indicates that the vibrational motion that moves a proton from one nitrogen to the other (an asymmetric NHN stretch) gives rise to a discrete IR absorption. The band corresponds to the expected low frequency proton transit mode in an LBHB, consistent with the inference that LBHBs can occur in proton-bridged, aliphatic diamines. The absence of this spectroscopic signature in other proton-bridged diamines implies that not all such species contain LBHBs. Specifically, the characteristic isotopic shift in the 300-600 cm^{-1} domain is not observed in the asymmetric, bridged diamine 2. Solid phase NMR and X-ray crystallography probe the structure of salts of **1a**. NMR demonstrates that the polycrystalline monoiodide salt exhibits a single, sharp ^{15}N resonance. Crystallography compares experimentally measured distances with those predicted by DFT for the free ion. Our previously reported study of dipolar ^{15}N-^{1}H coupling in the triflate salt situates the bridging proton midway between the two nitrogens, rather than rapidly exchanging from one to the other. [24]

Inelastic neutron scattering (INS) detects a vibration of the monoiodide salt of **1a** close to the same position as seen by IRMPD for isolated **1a** in the gas phase, which again is greatly attenuated by replacement the bridging H^+ with deuterium. The closeness of the band position seen in the solid to that of the isolated ion in the gas phase (as well as to theory) predicts that the NHN asymmetric stretch, if correctly assigned, ought to shift only slightly with applied pressure.

High pressure Raman spectroscopy reveals a vibrational mode for the monoiodide salt of **1a** that agrees with the band observed for the gaseous ion and by INS, which is not observed for the monoiodide salt of the deuterated ion **1b** nor for the diiodide salt of the diprotonated diamine. This position of this Raman absorption depends on applied pressure much less than do nearby vibrations, an outcome consistent with the aforementioned prediction. Taken together, the results demonstrate how theory, isotopic labeling, IRMPD, Raman spectroscopy, and INS can provide an assignment for vibrational motion of a bridging proton within a LBHB.

2.2 Experimental section and computational methods

2.2.1 Computational

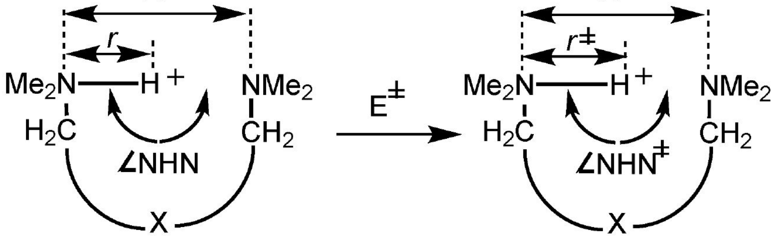


Figure 2.4: Distances and angles for protonated α, ω-diamine equilibrium geometries and for the barrier top corresponding to proton transit.

Reaction coordinates for internal nucleophilic displacement (Figure 2.4 and

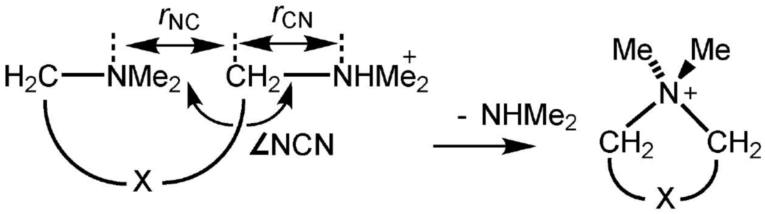


Figure 2.5: Transition state geometry for internal nucleophilic displacement

Figure 2.5) were calculated utilizing constrained geometry optimizations at B3LYP/6-31G(d, p) using a pruned (99, 590) ultrafine grid, with normal mode calculations at stationary points, by means of the Gaussian03 program suite.(CITE GAUSSIAN) Potential energy surfaces (PESs) for proton transit using these geometries were fitted to calculated data points as polynomials and vibrational energy levels computed by solving for the eigenvalues and eigenvectors of the corresponding Hamiltonian matrix expanded in a harmonic oscillator basis within the Mathematica programming environment.

Energetics for proton transit, as per Figure 2.3 and for the 2-dimensional PES in Figure 2.6 were computed using the recently developed double-hybrid B2-P3LYP density functional [29] in the cc-pVTZ basis. [30] Standard density functionals exhibit known deficiencies in treating van der Waals interactions. [31] Double-hybrid functionals [32] attempt to correct for this deficiency by mixing Mller-Plesset (MP2)-like long-range correlation into the functional. The B2-P3LYP functional significantly outperforms B3LYP for systems where non-covalent interactions are important, so it was used for computing energetics. The cc-pVTZ basis is the recommended basis set for use with B2-P3LYP, since the functional's empirical parameters were fitted using this basis.

B2-P3LYP calculations were performed using a developmental version of Q-Chem, version 3.1. [33] Because analytical gradients of the B2-P3LYP functional are currently unavailable, the structures optimized at B3LYP/6-31G(d, p) were used. Two

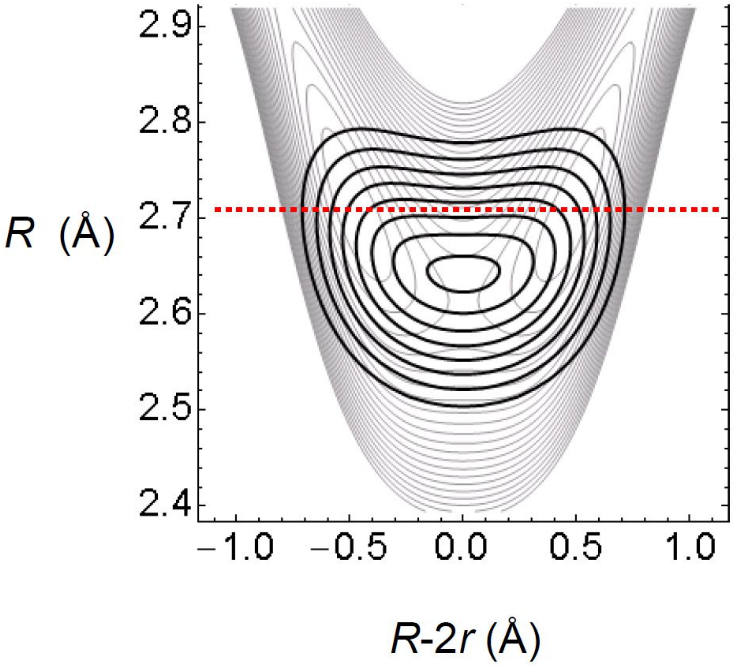


Figure 2.6: 2-Dimensional potential energy surface for the conjugate acid ion of N, N, N, N-tetramethylputrescine fitted to 132 data points calculated at B2-P3LYP/cc-pVTZ//B3LYP/6-31G(d, p) for different values of R and r. The heavy black contours represent the zero point wavefunction for ion **1a** with contour lines at 30, 40, 50, 60, 70, 80, 90, and 98% of maximum amplitude. The dashed line represents a straight cut from one energy minimum to the other.

additional approximations were invoked to accelerate the B2-P3LYP energy calculations. First, just as in the original B2-P3LYP paper, [29] the resolution-of-the-identity (RI) approximation was used, [34] with the auxiliary cc-pVTZ fitting basis, [35] to speed the MP2 portion of the B2-P3LYP calculation. Second, a dual-basis Hartree-Fock (HF)/MP2 calculation [36] was used to approximate the MP2/cc-pVTZ solution by taking the converged HF density matrix from a carefully chosen smaller basis set, projecting the density matrix into the larger cc-pVTZ basis, and taking a single HF iteration in the larger basis before applying the MP2 correction. Each of these approximations typically introduces negligible errors into relative energies.

To generate a 2-dimensional PES, B2-P3LYP/cc-pVTZ energies were computed for >120 geometries, which were generated by fixing the two NH distances over a domain of values and performing constrained B3LYP/6-31G(d, p) optimizations to relax all other degrees of freedom. In the first pass, the two NH bond lengths were varied independently in 0.25 Å increments. Additional grid points were added at spacings of 0.05-0.10 Å or less in low-energy regions. Sampling intervals as tight as 0.01 Å were tested, but the impact on the polynomial fit was insignificant.

The B2-P3LYP data points were fitted within the Mathematica programming environment using the symmetry coordinates $s_1 = R/\sqrt{2}$ and $s_2 = (R - 2r)/\sqrt{2}$ (cf. Figure 2.4) to a 2-dimensional potential energy surface using polynomial coefficients and cross-terms up to order 4 for s_2 with an additional R^6 term for s_1 using non-linear least-squares fitting subroutines. Geometry optimizations did not constrain any angles but nevertheless gave NHN values within 10° of linearity. Having NHN bond angles this close to 180° justified the use of s_1 and s_2 (internal coordinates for a linear triatomic) in constructing the PES and calculating vibrational levels. Energies and wavefunctions for the movement of a proton or deuteron on the 2-dimensional PES were determined by solving the nuclear Schrödinger equation in a 2-dimensional basis of harmonic oscillator eigenfunctions. At least 20 basis functions, centered about 0 for s_2 and the equilibrium point in s_1, were used in each coordinate. The PESs of interest were each fitted to give RMS deviations <0.2 kJ mol^{-1}. The reduced masses for a linear triatomic having a small mass m between two larger masses M are M for s_1 and $mM/(m + 2M)$ for s2. [37] In terms of the integer masses of hydrons and nitrogen, these correspond to 14 Da for s_1, 14/29 Da for s_2 with bridging H^+, and 14/15 Da for s_2 with bridging D^+. Energies and eigenstates were calculated using standard matrix techniques as implemented in Mathematica. Convergence was tested by varying both the number of basis functions and the fundamental frequency of the oscillators. Full dimensional DFT anharmonic calculations were performed using a second-order perturbative evaluation of vibrorotational parameters implemented in the Gaussian program suite. [38]

2.2.2 Gas phase ion spectroscopy and crystallography

N, N, N, N-Tetramethylornithine [39] (the conjugate base of ion 2) was prepared by $NaBH_3CN$ reduction of a mixture of formaldehyde and commercial ornithine and purified by ion-exchange chromatography on a Dowex 50Wx8-400 column, using a procedure analogous to that reported by Tian, Lis, and Kass. [40] Deuterated analogues of N, N, N, N-tetramethylputrescine were synthesized as previously described. [24] The crystalline monoiodide salt of **1a** was prepared by dissolving 5 mmol of commercial N, N, N, N-tetramethylputrescine (N, N, N, N-tetramethyl-1, 4-butanediamine, Aldrich) in 20 mL absolute ethanol and adding 3 mmol of commercial 57% hydriodic acid dropwise. The solution was allowed to cool, and colorless crystals deposited after a few minutes, which were then recrystallized from absolute ethanol and subsequently used for crystallography at 100 K on a CCD X-ray diffractometer system (Mo-radiation, $\lambda = 0.71073$ Å, 50 KV/40 mA power).

Gaseous conjugate acid ions were produced using electrospray ionization of free diamines or their salts from 50:50 methanol:water containing a trace of acetic acid (for protonated ions) or from 50:50 CH_3OD:D_2O containing a trace of CD_3COOD (for deuteronated ions) and introduced into a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR) coupled to the infrared beam line of FELIX, the free electron laser for infrared experiments at the FOM Institute Rijnhuizen. Details of the laser, the FT-ICR, and the multiple photon dissociation (IRMPD) process have described elsewhere. [41–46] Briefly, electrosprayed ions were injected into a home-built 4.7 T FTICR mass spectrometer via a quadrupole deflector and a 1-m long RF octupole ion guide. IRMPD spectra were monitored by expulsion of neutral dimethylamine (loss of 45 Da from protonated ions and of 46 Da from deuteronated ions) as a function of wavelength.

2.2.3 Solid phase NMR and high pressure Raman spectroscopy

Magic-angle-spinning (MAS) solid-state NMR experiments were performed at 9.4 T (^{1}H frequency 400 MHz) on a Bruker AVANCE III spectrometer equipped with a double-resonance 4 mm MAS probe, spinning at a MAS rate of 8 kHz. ^{1}H direct excitation experiments were acquired using 83 kHz $\pi/2$ excitation pulses, and ^{15}N cross-polarization (CP) experiments were acquired using 83 kHz ^{1}H $\pi/2$ excitation and decoupling pulses along with a 2 ms CP spin-lock. During CP the ^{1}H nutation rate was set to 50 kHz and the ^{15}N nutation rate ramped from 38 kHz-46 kHz. For a typical CP spectrum, 4096 complex data points with a dwell of 10 ms (spectral width 50 kHz, total acquisition time 40.1 ms) were acquired with a recycle delay of 5 s for a total experiment time on the order of 17 h. ^{1}H chemical shifts were referenced to TMS and ^{15}N shifts to liquid-NH_3 using solid ammonium chloride (at 39.2 ppm) as a secondary chemical shift reference.

Raman scattering studies of salts of **1a**, **1b**, and diprotonated tetramethylputrescinium diiodide were carried out under high pressure using a diamond anvil cell (DAC). Polycrystalline **1a** was purified by three successive recrystallizations from absolute ethanol containing a slight excess of free tetramethylputrescine. Samples of **1b** were prepared by bubbling hydrogen iodide gas through a solution of excess tetramethylputrescine in ethanol-*O*-*d* followed by three recrystallizations from ethanol-*O*-*d* containing excess of free diamine, as above. The diprotonated diiodide salt was prepared by adding excess aqueous 57% hydriodic acid to an ethanolic solution of tetramethylputrescine, collecting the precipitated crystals, and recrystallizing from ethanol. Stoichiometry of all samples was assayed gravimetrically by precipitation of AgI from addition of ethanolic $AgNO_3$ to weighed aliquots in ethanol. ^{15}N natural abundance solid phase NMR showed that the sample of **1a** exhibited a very small peak (approximately 2% the area of the peak from the monoprotonated salt) corresponding to an impurity of the diprotonated diamine. The same lot of polycrystalline **1a** was used for DAC Raman, NMR, and INS experiments. Deuterated salt **1b** was freshly recrystallized from ethanol-*O*-*d* before each experiment.

The Raman spectrum of the sample was obtained by using a Dilor XY modular Raman spectrometer. The excitation source was the 514 nm line of an argon ion laser or the 647 nm line of a krypton ion laser. The excitation light was directed into the DAC sample chamber, and the backscattered (180°) light collected and sent through a double subtractive monochromator and detected by a liquid nitrogen-cooled CCD. The sample pressure was determined by measuring the fluorescence shift of the R1 emission of ruby placed inside the sample chamber of the DAC. [39] The peak positions of the Raman bands were found by fitting the Raman spectra with Gaussian line shape functions. Some overlapping bands were fit with composite Gaussian functions.

2.2.4 Inelastic neutron scattering

Inelastic neutron scattering spectra were obtained on the inverse geometry Filter Difference Spectrometer at the Lujan Center at Los Alamos National Laboratory. [47] Approximately 500 mg of the monoiodide salt of monoprotonated (**1a**) tetramethylputrescine was loaded in an aluminum sample container under helium atmosphere for data collection at 10 K. As noted above, integration of the solid phase NMR showed this sample to contain 2% of the diprotonated diamine. The monoprotonated diamine salt was subsequently recrystallized three times from ethanol-*O*-*d* containing a slight excess of neutral tetramethylputrescine to exchange the H-bond proton with deuterium, yielding the monodeuteronated (**1b**) monoiodide salt. An equivalent INS spectrum was then collected with the aim of comparing the two spectra reproduced in Figure 2.15. Subtraction (not shown) eliminated all bands, except for those strongly coupled to displacements of the H-bond proton.

2.3 Results

The polycrystalline salts of tetramethylputrescine display characteristic solid phase MAS NMR spectra. The monoprotonated d_{16} monotriflate $[(CD_3)_2NCD_2CH_2CH_2CD_2NH(CD_3)_2{}^+\ CF_3SO_3^-]$ exhibits the spectrum reproduced in Figure 2.7. The resonance assigned to the bridging H^+ lies 13 ppm downfield of the aliphatic hydrogens. The natural abundance ^{15}N resonances of the monotriflate and monoiodide salts of undeuterated **1a** and of the diprotonated tetramethylputrescine-d_0 diiodide salt all show sharp (FWHM = 10 Hz) peaks in the solid phase MAS NMR at room temperature. The chemical shifts of the mono- and diiodide salts are well separated, 40.0 and 47.3 ppm, respectively. The monoprotonated monotriflate salt exhibits a simlarly sharp ^{15}N peak and has a chemical shift of 36 ppm in the solid phase NMR.

Singly charged conjugate acids of tertiary diamines have been investigated because only one proton is attached to nitrogen, which can be exchanged for a deuteron. DFT was used to select the ion most suitable for vibrational spectroscopy. The singly charged ions $\mathbf{1-3}$ were surveyed computationally and the experimental spectra of gaseous **1** (as a LBHB candidate) compared with those of **2a**, **2b**, **3a**, and **3b** (as controls). Four criteria were used to select ions for experimental study of LBHB properties:

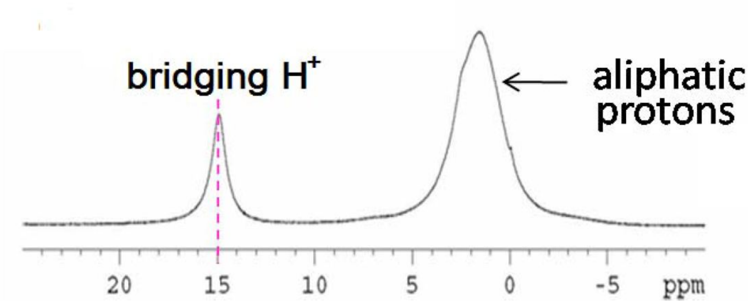


Figure 2.7: MAS ^{1}H NMR of the monotriflate salt of monoprotonated tetramethylputrescine-d_{16}. The dashed line indicates the chemical shift of the bridging proton.

Table 2.1: Geometrical parameters (Å and degrees) corresponding to the scheme in Figure 2.4 for internal H^+ transit within singly protonated aliphatic diamines calculated at B3LYP/6-31G(d, p), along with electronic energy barriers ($E^\ddagger$ in kJ mol^{-1}) based on single points calculated at B2-P3LYP/cc-pVTZ//B3LYP/6-31G(d, p). Symmetry of the geometry at the barrier top is indicated in the right-hand column.

Chainlength X	r	R	<NHN	$E^\ddagger$	$r^\ddagger$	$R^\ddagger$	$<NHN^\ddagger$	Symm
CH_2 (**3a**)	1.089	2.713	155°	7.19	1.308	2.575	160°	C_s
CMe_2 (**3c**)	1.092	2.698	155°	6.94	1.304	2.569	160°	C_s
CH_2CH_2 (**1**)	1.114	2.708	171°	3.63	1.306	2.608	173°	C_2
$CH_2CH_2CH_2$	1.111	2.734	183°	4.30	1.312	2.624	183°	C_s
$CH_2CMe_2CH_2$	1.115	2.715	182°	3.63	1.308	2.616	182°	C_s
$(CH_2)_4$	1.080	2.895	191°	12.5[a]	1.328	2.657	188°	C_1

[a] Starting from the more stable tautomer in the unsymmetrical double-well potential; $E^\ddagger = 11.4$ kJ mol^{-1} starting from the less stable tautomer.

(1) the two amino groups should have equal basicities (2) the cyclic geometry should involve linear or nearly linear NHN bond angles (which is why ion **3** was selected as a control); (3) the bridged diamines should exhibit low barriers for proton transit between the two nitrogens; and (4) a relatively low barrier for loss of Me_2NH via internal S_N2 is required for the IRMPD experiments. Figure 2.4 depicts relevant geometrical features of proton transit, with the corresponding values from B3LYP/6-31G(d, p) geometry optimizations for a homologuous series summarized in Table 2.1. Table 2.2 summarizes energies and geometries for decompositions of these cations via the equation found in Figure 2.3 followed by the scheme found in Figure 2.5.

Table 2.2: Geometrical parameters (Å and degrees) corresponding to Figure 2.5 for internal nucleophilic displacement of dimethylamine in monoprotonated aliphatic diamines calculated at B3LYP/6-31G(d, p) along with electronic energy changes corresponding to the equation given in Figure 2.3 (ΔE_{ex}) and for Figure 2.5 (ΔE_{SN2} and the barrier height $\Delta E^{\ddagger}{}_{SN2}$) in kJ mol^{-1} based on single points calculated at B2-P3LYP/cc-pVTZ//B3LYP/6-31G(d, p).

Chain X	ΔE_{ex}	ΔE^{a}_{SN2}	$\Delta E^{\ddagger}{}_{SN2}$	<NHN	r_{NC}	r_{CN}
CH_2 (**3a**)	64	159	161	178°	1.98	2.13
CMe_2 (**3c**)	54	134	140	182°	2.01	2.11
CH_2CH_2 (**1**)	78	94	128	192°	2.13	2.12
$CH_2CH_2CH_2$	63	57	120	198°	2.16	2.08
$CH_2CMe_2CH_2$	65	48	125	197°	2.16	2.07
$(CH_2)_4$	68	88	138	199°	2.18	1.92

^{a}Not corrected for basis set superposition error.

The diamine selected for comparison with theory, **1**, corresponds to a symmetric double well along the asymmetric stretch coordinate, whose geometry possesses an NHN angle within 10° of linearity. Shorter chains (such as in **2**) have NHN bond angles too small for the reduced dimensionality treatment described below. Diamine **1** has a low calculated proton transit barrier between the tertiary amino groups. The ring size for **1** also facilitates internal nucleophilic displacement via the pathway portrayed in the scheme in Figure 2.5. The ease with which intramolecular displacement can occur allows this reaction to be used for detection of IR absorption of the gaseous ion via action spectroscopy. The calculations described below suggest that the NHN stretch for **1** should occur between 500 and 550 cm^{-1}. The calculated barrier for the scheme found in Figure 2.5 thus requires the absorption of approximately 20 photons or less at that frequency to reach threshold for expelling dimethylamine.

It is important that the NHN angles be nearly linear for two reasons. First, the lowest energy trajectory may be inaccessible if the strong hydrogen bond has an angle much less than 180°. Perrin and coworkers [48, 49] have shown that H^+ transit in the conjugate acid of 1, 8-bis(dimethylamino)napthalene corresponds to equilibration of two isomers of equal stability, even though the calculated electronic energy barrier is low. Majerz and Oluvsson [50, 51] have more recently pointed out the strong influence of the bond angle on the facility with which proton transit can take place. In a bent NHN geometry, the lowest barrier on the energy surface may not be accessible for low-lying vibrational states. In other words, for highly nonlinear strong hydrogen bonds, the effective barrier might not correspond to the picture in Fig. 1, even if that potential does represent the minimum energy pathway.

Second, a theoretical analysis (for comparison with experiment) requires evaluation of anharmonic motion. For molecules as large as **1**, which lacks a plane of symmetry, it is necessary, at present, to treat the system using a greatly reduced dimensionality. This paper considers the NHN asymmetric stretch and the NN symmetric stretch as though they were a linear triatomic. This approach neglects bending motions.

2.3.1 Predicted potential energy surfaces for 1

DFT locates two low energy equilibrium conformations of the isolated ion **1a**. The calculations place the less stable one 12 kJ mol^{-1} higher, an enthalpy difference that means the more stable conformation (which does not lie far from possessing C_2 symmetry) represents >99% of the population at room temperature. X-ray crystallography confirms the predominance of the more stable conformation in crystalline salts. [24]

The simplest way to make a prediction for the strong hydrogen bond in **1a** is to calculate a 1-dimensional potential energy curve for the more stable conformation, as previous investigations have done. [19–22] The next level of approximation requires a 2-dimensional potential for that conformation, which provides a good estimate of bond distances. [24] In both models, vibrational quantizations are computed by assuming linear NHN (even though this constraint was not imposed on the geometry optimizations), with the corresponding reduced masses for stretching modes of a linear triatomic. [37]

As Table 2.1 indicates, the NHN angle for monoprotonated diamine **1a** increases by only 2° as the proton passes over the central barrier. Nevertheless, the NN distance at the barrier top, $R^{\ddagger}$, has a value 0.1 Å shorter than does R at the equilibrium geometry. For this reason, a 2-dimensional potential energy surface has been calculated for ion **1** as a function of R - 2r (corresponding to the motion of the hydrogen treated as an asymmetric stretch) and R (corresponding to the NN stretch). Dividing these by $\sqrt{2}$ gives the symmetry coordinates s_2 and s_1 used to compute anharmonic frequencies. As Table 2.1 summarizes, approximation as a linear equilibrium geometry becomes invalid for shorter homologues such as ions **3**. Published crystal structures of longer homologues [24] indicate that the crystalline salts examined to date exhibit disorder or dimer formation that interfere with solid phase investigations.

The 2-dimensional potential energy surface was computed for ion **1** using B3LYP/6-31G(d, p) optimized geometries at various fixed values of the NH distances, as described in the experimental section, with single point energies computed at B2-P3LYP/cc-pVTZ for each data point. This functional and basis set have performed well both in general noncovalent interaction benchmarks [29] and in predicting the binding energies of the proton-bound dimers of adenine. [52–54] In contrast, B3LYP consistently underestimates such interactions [29, 55] (regardless of inclusion of diffuse functions in the basis set). The 2-dimensional potential energy surface for **1** is illustrated in Figure 2.6, along with a contour plot for the zero point wavefunction of **1a**. Predicted zero-point energies, anharmonic frequencies, and isotope effects resulting from the 2-dimensional surfaces will be discussed below in the context of the experimental results.

2.3.2 Experimental studies

The validity of the 2-dimensional potential energy surface can be gauged by comparing the predicted N-N distance R with the experimental value. The expectation value <R> for the zero-point vibrational wavefunction of **1a** for the surface depicted in Figure 2.6 has a value of 2.643 Å. The value from X-ray crystallography of the monotriflate salt of **1a** is R = 2.659 Å. [24] Because the structure of the cation in the crystalline salt is perturbed by the counterion, the monoiodide salt was prepared, as well, for which X-ray crystallography at 90 K gives the value R = 2.647 Å. The calculated value of <R> thus agrees well with the crystallographic N-N distance for the monoiodide salt of **1a**.

The top panel in Figure 2.8 reproduces the IRMPD spectra of the gaseous ions **1a** (m/z 145 in the black upper trace, electrosprayed from 50:50 methanol:water containing a trace of acetic acid) and **1b** (m/z 146 in the red lower trace, electrosprayed from 50:50 CH_3OD:D_2O containing a trace of CD_3COOD) in the 600-1800 cm^{-1} domain. As noted above, IR absorption was monitored by the loss of neutral dimethylamine observed using FT-ICR (m/z 145 –> m/z 100 for the black trace, m/z 146 –> m/z 100 for the red trace). Most of the observed bands match peaks predicted by DFT anharmonic calculations [38] shown in the lower panel, with assignments indicated by the dashed lines. The most intense IR absorption band (around 1500 cm^{-1}) in Figure 2.8 comes from HCH scissor modes. The position of this band does not shift in changing from **1a** to **1b**, even though the intensity of the IRMPD absorptions (the fractional dissociation represented by the y-axis) diminishes. This consequence of deuteration appears general and can be ascribed to the increasing heat capacity of the deuterated ion.

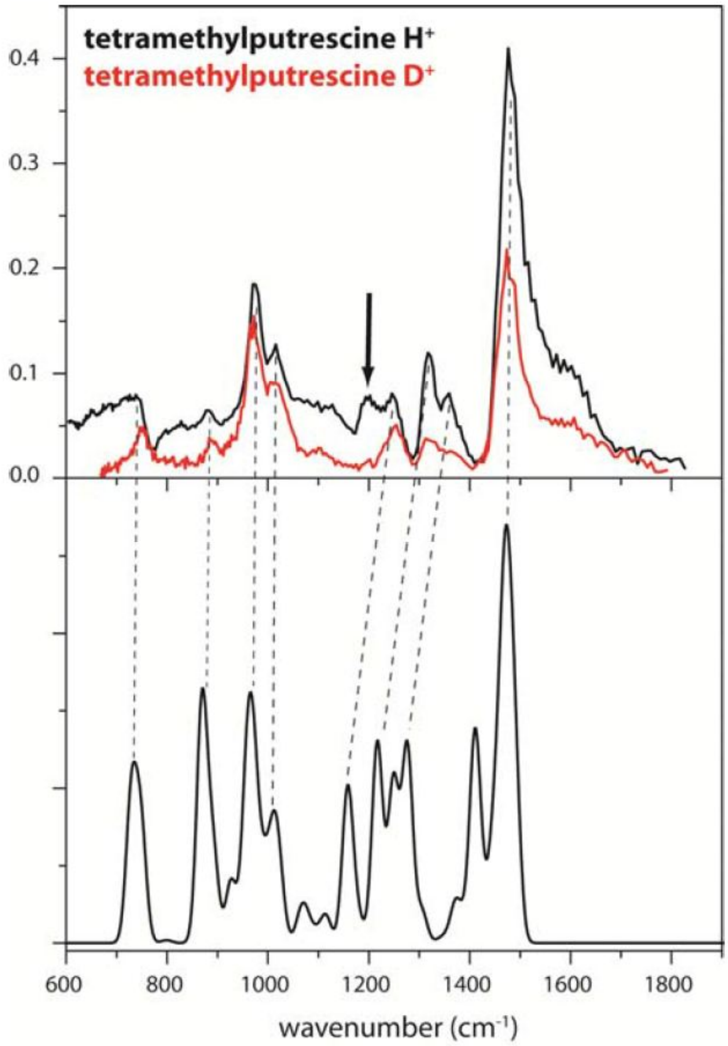


Figure 2.8: Experimental IRMPD spectra of ions **1a** (upper panel, upper trace) and **1b** (upper panel, lower trace), in which the y-axis indicates the fractional yield of dimethylamine loss, compared with an anharmonic normal modes calculation for **1a** [38] (using harmonic absorption cross sections and 30 cm^1 FWHM Gaussian lineshape). The solid arrow near 1200 cm^{-1} in the experimental spectra in the upper panel indicates a band that disappears upon deuteration, which is provisionally assigned to an NHN bending mode. The lower panel shows the IR absorptions predicted by a B3LYP/6-31G(d,p) anharmonic calculation on **1a**, from which three bands (the NHN asymmetric stretch at 961 cm^{-1} and the two NHN bends at 1598 and 1640 cm^{-1}) have been deleted. Dashed lines indicate matches assigned between the IRMPD and the calculated spectrum.

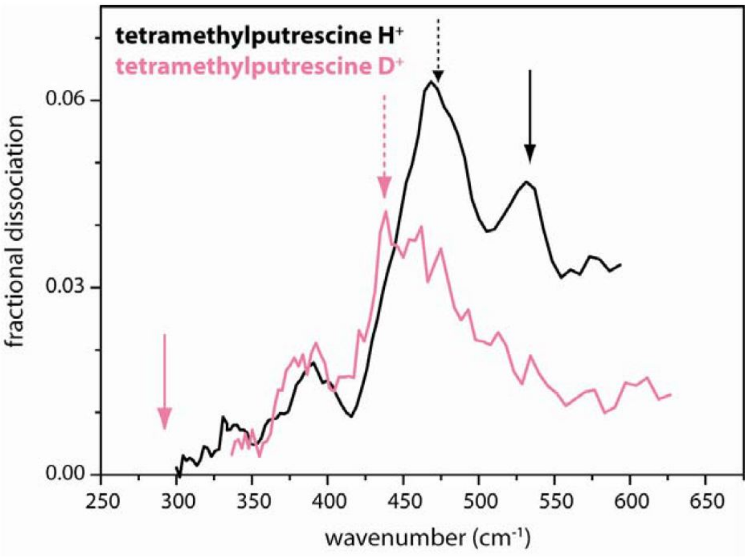


Figure 2.9: IRMPD spectrum of ions **1a** (black trace) and **1b** (pale trace) in the long wavelength domain. Arrows correspond to frequencies calculated from the 2-dimensional surface in Figure 2.6: dashed arrows for NN stretching modes (**1a** at 515 cm^{-1}, **1b** at 456 cm^{-1}); the solid arrow for the asymmetric NHN stretch predicted for **1a** at 528 cm^{-1}; and the solid arrow for the position of the asymmetric NDN stretch at 282 cm^{-1}, which lies outside the experimentally accessible frequency range.

2.3.3 NHN Asymmetric stretching vibrations

The IR spectra of **1a** and **1b** are reproduced in Figure 2.8 and Figure 2.9. The free electron laser requires readjustment to scan at frequencies below 600 cm^{-1}, so the spectra in Figure 2.9 were recorded separately from those in Figure 2.8. One band in the 300-600 cm^{-1} region of Figure 2.9 displays a major shift of peak position as a consequence of isotopic substitution. Upon replacement of the bridging H$^+$ with D$^+$, the band near 540 cm^{-1} greatly diminishes, while the bands near 390 and 460 cm^{-1} do not move appreciably, although a new peak appears around 440 cm^{-1}. The vibration that vanishes is taken to represent motion of the bridging proton.

Low frequency motions of the alkyl chain also occur in this region. To take these into account, Gaussian03 B3LYP/6-31G(d, p) anharmonic calculations [38] for motions of the methylene chain in this domain give frequencies, which can be compared with those predicted by a harmonic normal modes calculation [27] (511 vs. 503 cm^{-1}; 497 vs. 483 cm^{-1}; 446 vs. 447 cm^{-1}; 345 vs. 348 cm^{-1} for the four most intense peaks). The small differences between harmonic and anharmonic calculations suggest the accuracy of those values. Setting them aside, the vertical, solid black arrow indicates the position of the asymmetric NHN vibration of the bridging proton of **1a** predicted from the 2-dimensional surface (symmetry coordinate s_2) in Figure 2.6 at 528 cm^{-1}. When bridging H^+ is substituted by bridging D^+, the predicted peak position shifts to 282 cm^{-1} (outside the domain of the free electron laser scan), as indicated by the solid red arrow.

Harmonic calculations predict a much higher value for the NHN asymmetric stretch. The value from a normal modes computation (2027 cm^{-1}) shifts greatly in the Gaussian03 anharmonic calculation (961 cm^{-1}). [38] Neither value agrees with the prediction from the 2-dimensional surface in Figure 2.6. As Figure 2.8 shows, the experimental peaks between 800 and 1000 cm^{-1} exhibit no shift upon deuteration. Moreover, nearly all of the experimental peak positions of **1a** match those of the calculated spectrum in the lower panel (from which the 961 cm^{-1} band has been omitted). Hence, assigning the band near 540 cm^{-1} to the NHN asymmetric stretch seems plausible. Additional evidence below corroborates that assignment.

Comparison of ions **1** and **2** provides one type of control. The sort of isotopic shift in going from ion **1a** to **1b** is not seen in the conjugate acid ions (**2a** and **2b**) from tetramethylornithine, $Me_2NCH_2CH_2CH_2CH(NMe_2)COOH$, whose dimethylamino groups have unequal basicities. This basicity difference ought to prevent the strong intramolecular hydrogen bond from behaving like a LBHB. The gas-phase mid-IR spectrum of the protonated amino acid (MH^+ ion) is reproduced in Figure 2.10 panel A. The lowest frequency absorption detected by IRMPD occurs near 600 cm^{-1}. As Figure 2.10 panel B shows, **2a** exhibits no other bands in the 300-600 cm^{-1} region. Does

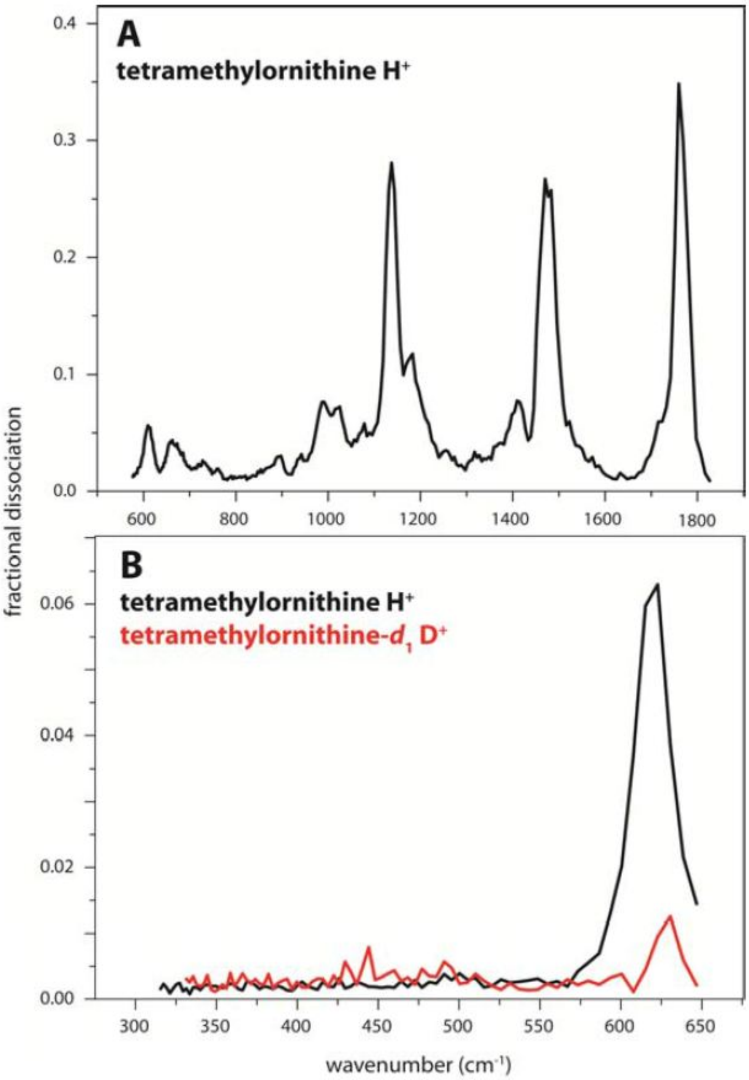


Figure 2.10: IRMPD spectrum of monoprotonated N, N, N, N-tetramethylornithine, **2a**, in the 600-1800 cm^{-1} domain (A) and a comparison of **2a** with the d_2-ion **2b** in the 300-600 cm^{-1} domain (B).

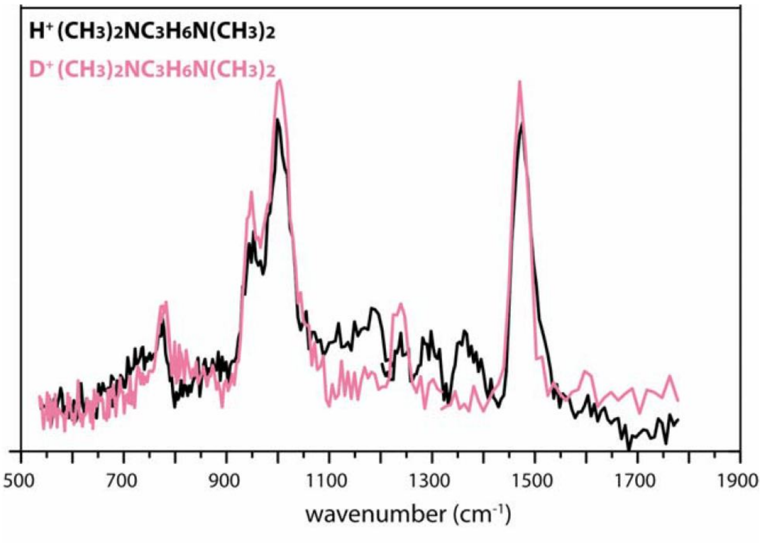


Figure 2.11: IRMPD spectra of N, N, N, N-tetramethyl-1, 3-propanediamine (black trace with bridging H^+; pale trace with bridging D^+). Note that there are at least 3 bands present in the black trace that are absent in the pale trace.

that vibration involve an exchangeable hydrogen?

Panel B in Figure 2.10 compares the IRMPD spectra of the MH^+ (**2a**, m/z 189, upper trace) and MD^+ (**2b**, m/z 191, lower trace) ions in the low frequency domain. The latter ion contains 2 deuteria, because the carboxyl group also exchanges with solvent. As in the case of ion **1**, the MD^+ spectrum (lower) is much less intense than the MH^+ spectrum (upper), but the principal band in the 300-600 cm^{-1} domain does not change position. Small features appear around 450 cm^{-1} in the deuterated spectrum, but they do not appear to be significantly above background and may be noise.

Since no shift of band position takes place, the 600 cm^{-1} band in **2** does not involve the exchangeable hydrogens. Nor do any other absorptions appear at lower frequencies. The example of **2** serves as a control for the method of isotopic substitution for assessing vibrations assigned to the bridging proton.

What happens if the NHN bond deviates too much from linearity? Making the bond angle much less than 180° need not prevent proton bridges from behaving like

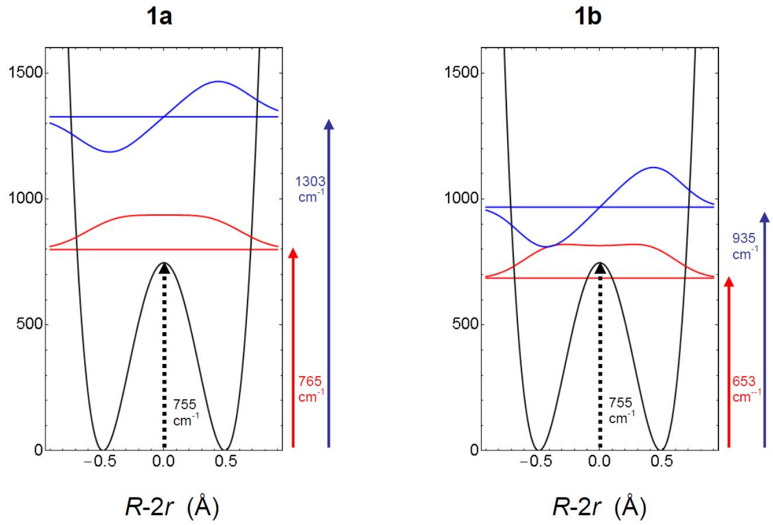


Figure 2.12: Potential energy curves (y-axes in cm^{-1}) corresponding to the straight cut from minimum to minimum across the 2-dimensional surface indicated by the dashed line in Figure 2.6 for the conjugate acid ions of N,N,N,N-tetramethylputrescine with bridging H^+ (**1a**) and bridging D^+ (**1b**). The $\nu = 0$ and $\nu = 1$ energy levels and the corresponding cuts through the wavefunctions show the probability that the bridging hydron lies close to the midpoint between the nitrogens in the zero point level.

LBHBs, but it invalidates predictions based upon treating it as a linear triatomic. The experimental spectra of **3a** and **3b** reproduced in Figure 2.11 display more peaks that disappear in the mid-IR (presumably shifted to where they cannot be observed) than do diamines with longer methylene chains. As Table 2.1 summarizes, the NHN strong hydrogen bonds for diamines with 3-carbon chains are substantially bent, and, as a consequence, no effort has been made to predict its vibrations at reduced dimensionality. At least 3 bands in **3a** appear to vanish in the experimental spectrum of **3b**, around 1200 cm^{-1}, 1300 cm^{-1}, and between 1350 and 1400 cm^{-1}. While no assignment is put forth for these bands, this control demonstrates that deuteration can, in some cases, lead to more pronounced changes than those observed in Figure 2.8 and Figure 2.9.

The reduced dimensionality approach used here gives a potential energy surface (Figure 2.1 and Figure 2.12) from which NN symmetric stretching vibrations (symmetry coordinate s_1) are calculated at the same time as the NHN asymmetric stretches. On the one hand, computations based on the 2-dimensional potential energy surface give a value of 515 cm^{-1} for ion **1a**. On the other hand, standard normal modes calculations [27] shows several vibrations that possess NN symmetric stretch character. The normal modes calculations predict frequency shifts <5 cm^{-1} in going from bridging H^+ to bridging D^+, while computations based on the surface in Figure 2.6 (which do not take motions of the methylene chain into account) predict that the NN symmetric stretch should shift to 456 cm^{-1}. The IR absorption maximum in Figure 2.9 around 470 cm^{-1} does not lie too far from the aforementioned predictions for the NN stretch, which is indicated by a black dashed arrow. More importantly, the apparent shift of the experimental maximum to 440 cm^{-1} corresponds to predictions based on the reduced dimensionality treatment (which is indicated by the red dashed arrow). The 2-dimensional potential energy surface also predicts the zero point energy for the asymmetric and symmetric vibrations taken together to undergo an isotopic shift 765 –> 653 cm^{-1} for **1a** –> **1b** (Figure 2.12).

2.3.4 NHN Bending vibrations

Another feature virtually disappears from Figure 2.8 upon replacement of the exchangeable H^+ by D^+: the band around 1200 cm^{-1}. The change in mass of the bridging hydron presumably moves this band to lower frequency, where it is obscured by other bands. Consequently, the 1200 cm^{-1} vibration is inferred to involve a substantial contribution from the bridging proton to its reduced mass. The adjacent band near 1250 cm^{-1}, which does not move upon deuteration, probably corresponds to CH_3 rocking motions calculated at 1158 cm^{-1}. The lower panel in Figure 2.8 shows the B3LYP/6-31G(d, p) anharmonic vibrational spectrum [38] of **1a**, from which the NHN bending peaks (calculated at 1598 and 1640 cm^{-1}) and the asymmetric stretch of the bridging proton at 961 cm^{-1} have been deleted. It would be tempting to surmise that the shoulder around 1600 cm^{-1} in the IRMPD spectrum comes from NHN bending vibrations, but the fact that the shoulder does not move upon deuteration argues against this assignment.

As noted above, anharmonic DFT calculations [38] predict that NHN bending vibrations in **1a** should occur around 1600 cm^{-1}. The shoulder around 1600 cm^{-1} in Figure 2.8 does not shift upon replacing the bridging H^+ with D^+, so it seems unlikely that this feature corresponds to a vibration in which the bridging proton moves. Deuteration of all the CHs in **1a** (giving a d_{20} ion at m/z 165) shifts the most intense feature to a pair of bands near 1100 and 1200 cm^{-1} (see Supporting Information). This shift reveals a broad band centered around 1550 cm^{-1} with a full width a half-maximum (FWHM) of 100 cm^{-1}, which is superimposable upon the shoulder in Figure 2.1. Assignment of this feature, which appears to be common to **1a**, **1b**, and the monoprotonated d_{20} diamine, remains in question.

As pointed out above, the shorter homologue whose spectra are reproduced in Figure 2.11, displays at least three bands in the 1000-1400 cm^{-1} domain that appear to vanish upon deuteration. For this ion, the central barrier to proton transit is higher (7.0 kJ mol^{-1}) and the NHN bond angle deviates further from linearity (155°) than in **1**. These three bands of ion **3a** could correspond to the NHN asymmetric stretch and the NHN bends, but modeling the effects of anharmonicity in ions with NHN angles far from 180° lies beyond the scope of current efforts.

2.3.5 Inelastic neutron scattering

The traces in Figure 2.13 reproduce INS data for the monoiodide salts of **1a** and **1b**. Unlike the IR spectra, these traces do not exhibit features above noise level in

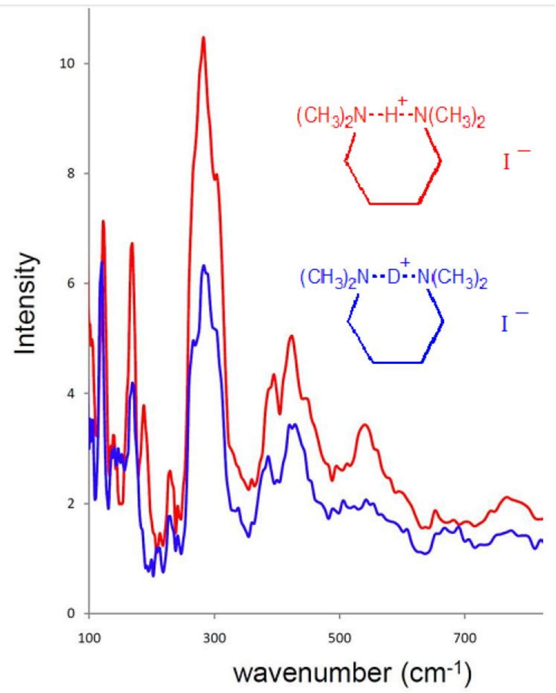


Figure 2.13: Inelastic Neutron Scattering (INS) spectrum of the monoiodide salt of monoprotonated tetra-methylputrescinium (**1a**, upper trace) and of the monoiodide salt of monodeuteronated tetramethylputrescinium (**1b**, lower trace).

the 600-1400 cm^{-1} domain. Because neutron scattering depends on the strong interaction with atomic nuclei, INS intensity scales with motion amplitude. For that reason, higher frequency vibrations, which dominate vibrational spectra observed using electromagnetic radiation, exhibit smaller cross sections than do low frequency modes that have large changes in amplitude upon vibrational excitation.

Two bands essentially disappear when the bridging proton of **1a** is replaced by a deuteron in **1b**: one band at 525 cm^{-1} and the other at 190 cm^{-1}. The latter may be a phonon band, which will not be further discussed here. The 525 cm^{-1} band coincides closely with one of the bands that disappears upon deuteration in the IRMPD spectrum of gaseous **1a** and occurs near the calculated frequency indicated by the solid black arrow in Figure 2.9.

At first it appears remarkable that the frequency calculated by a 2-dimensional calculation matches not only the gas-phase IR spectrum but also the vibration seen in the monoiodide salt. INS also brings to light a band in the same vicinity as carbons are successively deuterated in the monotriflate salt, [56] too, but other vibrations obscure it. The monoiodide salt provides a less congested domain.

Assuming that the 540 cm^{-1} band in the gas-phase IR spectrum and the 525 cm^{-1} band observed by INS both arise from the asymmetric NHN stretch means that either (1) motion of the bridging proton from one nitrogen to the other remains unaffected by the presence of the iodide counterion and (2) the Gruneisen parameter for that vibration is small; or else (3) that several effects (changes in the harmonicity of the potential energy curve, interactions between ions within the lattice, coupling among vibrations, etc.) fortuitously cancel. Pressure-dependent Raman spectroscopy provides a way to test the second of these three hypotheses.

2.3.6 Pressure-dependent Raman spectroscopy

Figure 2.14 reproduces the diamond anvil cell (DAC) Raman spectra of protonated (**1a**) and deuteronated (**1b**) monoiodide salts for ascending and descending pressures. Harmonic DFT calculations predict the Raman activity for chain vibrations. Published experiments [24] demonstrate that the C_2 geometry represents a probable structure of **1a** in its zero point level (even though the C_2 structure lies at the top of a low barrier). The 1-dimensional cut from minimum to minimum, represented by the dashed line in Figure 2.6, gives a slice across the $\nu = 0$ wavefunction for **1a** in Figure 2.12. As can be seen, large amplitude motions in the zero point complicate predictions. The most straightforward estimate of the contribution from chain motion assumes a linear combination of vibrations for the C_2 geometry and of the structure corresponding to the bottom of the double well (the equilibrium geometry).

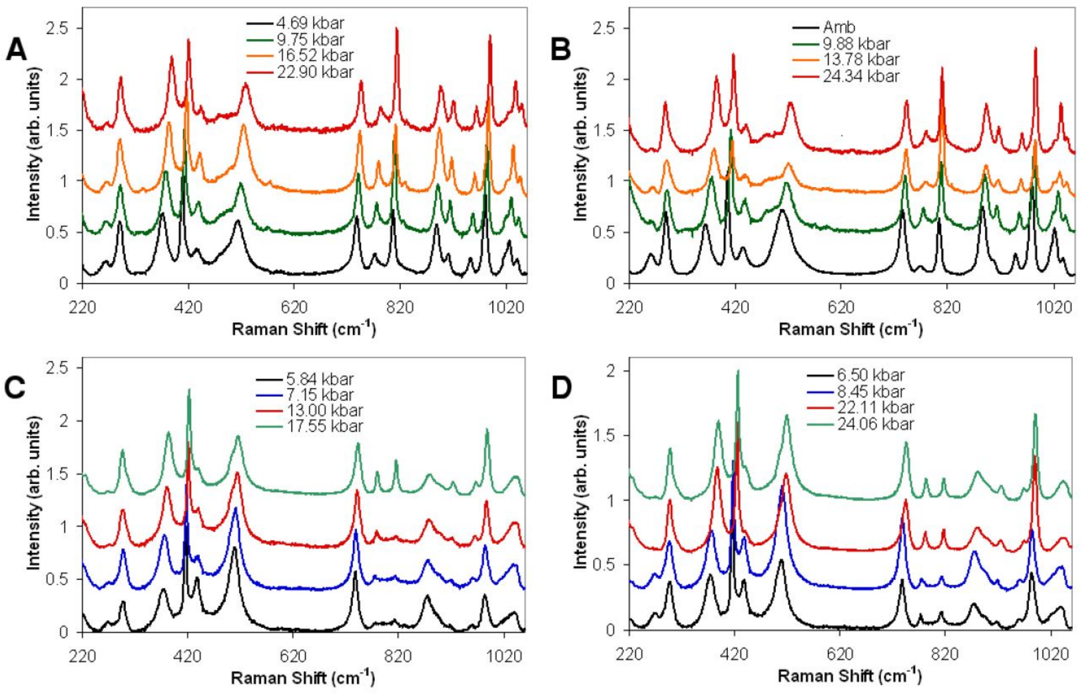


Figure 2.14: Raman spectra as a function of pressure in a diamond anvil cell (DAC): (A) Monoiodide salt of monoprotonated tetramethylputrescinium (**1a**) with increasing pressure; (B) Monoiodide salt of monoprotonated tetramethylputrescinium (**1a**) with decreasing pressure; (C) Monoiodide salt of monodeuteronated tetramethyl-putrescinium (**1b**) with increasing pressure; (D) Monoiodide salt of monodeuteronated tetramethyl-putrescinium (**1b**) with decreasing pressure.

As it happens, the chain vibrations with the greatest Raman activity for **1a** do not differ too greatly between the equilibrium and C_2 geometries. The largest contributor in the 300-700 cm^{-1} domain for the C_2 structure corresponds to torsions about the pair of CH_2-CH_2 single bonds related by symmetry, having a predicted frequency of 567 cm^{-1}. Me-N-Me bends accompany that motion, as well as concomitant symmetrical distortions of the methylene chain. DFT predicts the next most intense Raman bands in that domain to occur at 471 cm^{-1} and 362 cm^{-1}. All correspond to the A representation, having very low predicted IR intensities, with predicted Raman activities in the ratio 1:0.17:0.12. The weaker vibrations correspond to a torsion about the central CH_2-CH_2 single bond and to torsions about all of the single bonds (except for the N-CH_3 bonds), again with concomitant distortions of the methylene chain. For the equilibrium geometry, DFT predicts the two most intense Raman bands in the 300-700 cm^{-1} domain to occur more closely together, at 511 and 497 cm^{-1}. These correspond to torsions about one or the other of the non-central CH_2-CH_2 single bonds, with concomitant Me-N-Me bends and distortions of the methylene chain. DFT predicts the third most intense band to occur at 345 cm^{-1}. Relative to the 567 cm^{-1} band of the C_2 geometry these three bands have predicted Raman activities of 0.26, 0.14, and 0.08, respectively, but should also exhibit substantial IR intensities. Those predictions for **1a** do not change substantially in the monodeuterated analogue **1b**.

The bands seen in the DAC Raman spectra at 530, 430, and 370 cm^{-1} at or not far from atmospheric pressure in Figure 2.14 seem consistent with the calculated band positions. Figure 2.15 expands the DAC Raman spectra in the 425-575 cm^{-1} domain. Comparison of the mono-iodide salts of **1a** (solid trace) and **1b** (dashed trace) near 13 kbar with the atmospheric pressure spectrum of **1a** (dotted trace) reveals a broad band in the fully protonated salt that emerges only with increasing pressure. This

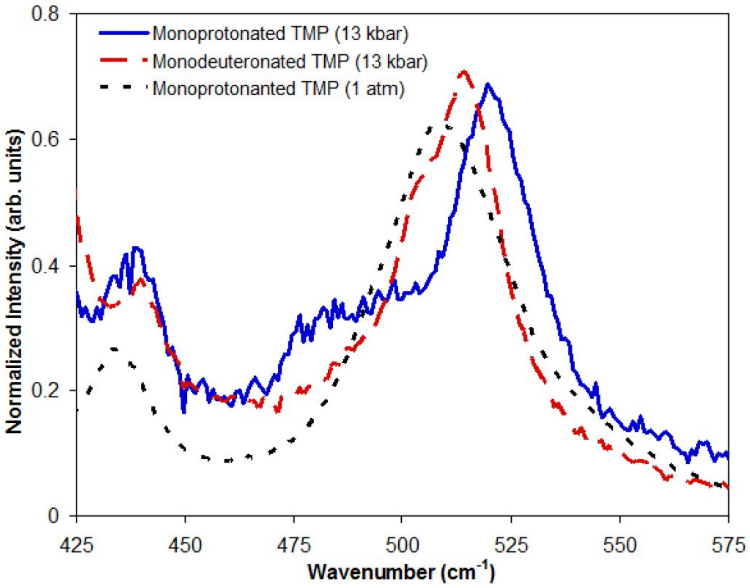


Figure 2.15: Comparision of the monoiodide salt of monoprotonated tetramethylputrescinium (1a, solid trace) with the monoiodide salt of monodeuteronated tetramethylputrescinium (1b, dashed trace) Raman spectra near 13 kbar. Dotted curve shows the trace for monoprotonated tetramethylputrescinium (1a) at atmospheric pressure.

broad band in **1a**, which extends from about 475 to 510 cm^{-1} in the solid blue trace, overlaps a strong band (provisionally assigned to the aforementioned CH_2-CH_2 torsions, which moves up in frequency from 510 cm^{-1} in the dotted trace to 525 cm^{-1} in the solid trace). At atmospheric pressure, this strong band appears to obscure the broad band, which becomes visible in **1a** only with the expected shift of the torsional vibration to higher frequency as pressure increases. The dashed trace for the monodeuterated salt **1b** exhibits virtually the same behavior for the CH_2-CH_2 torsions, but the broad band is absent. Figure 2.16 displays DAC Raman spectra for diprotonated and dideuterated diiodide salts, which provides another control, demonstrating the absence the broad band in those salts at any pressure.

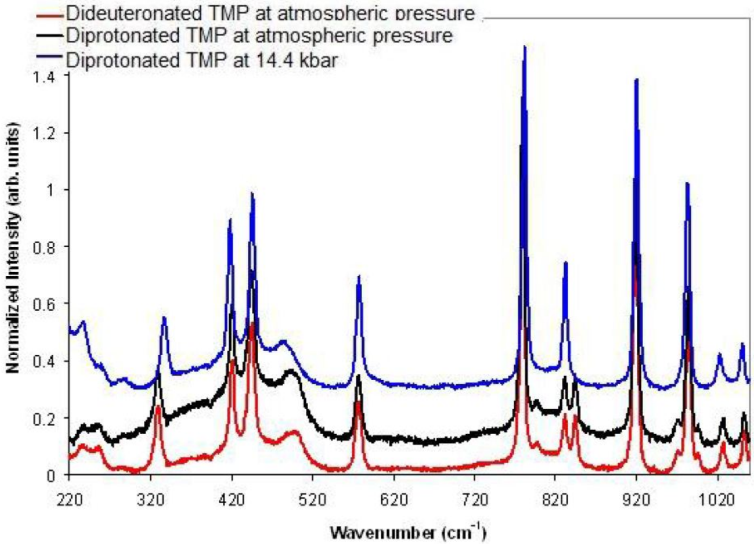


Figure 2.16: Raman spectra as a function of pressure in a DAC of diiodide salts of diprotonated and dideuteronated tetramethylputrescinium. Middle trace shows the spectrum for diprotonated tetramethylputrescinium diiodide at atmospheric pressure. Bottom trace shows the spectrum for dideuteronated tetramethylputrescinium diiodide at atmospheric pressure. Topmost trace shows the spectrum for diprotonated tetramethylputrescinium diiodide at 14.4 kbar.

2.4 Discussion

The spectroscopy of proton-bound dimers and internally proton-bridged ions has long been a subject of interest for both chemists and biochemists. [24, 48, 49, 57–82] In condensed phases the infrared spectra of these species tend to exhibit broad bands, [57–64] while narrower absorptions are reported for gaseous ions. [71–77] One aspect of the gas phase spectra that warrants attention is the effect of anharmonicity on strong hydrogen bonds. Anharmonicity not only results in bands at much lower frequencies than would be predicted by DFT normal modes calculations (e.g. <400 cm^{-1} for the proton-bound dimer of ammonia, $H_3NH^+ \cdots NH_3$ [77–79], but it also causes other peaks to shift from their expected positions (e.g. the internally proton-bridged monoanion of phthalic acid [46]). Accurate accounting for band positions in small ions has required a complete analysis of all internal degrees of freedom (e.g. for the Zundel ion, $H_2OH^+ \cdots OH_2$ [80–82]), but such a rigorous approach is not yet feasible for larger organic species. The present work explores the extent to which consideration of only two anharmonic internal degrees of freedom can describe vibrations of internally hydrogen bonded cations, such as **1a**, in which the NHN bond angle is nearly linear.

The experimentally observed band near 540 cm^{-1} for gaseous **1a** is assigned to the NHN asymmetric stretching vibration of a low barrier hydrogen bond (LBHB) on the basis of solid phase structural data, [24] deuterium substitution, theoretical calculations, and vibrational spectra of the crystalline salts. Figure 2.12 displays a cut across the computed 2-dimensional PES in Figure 2.6 from minimum to minimum at constant R (as represented by the dashed line in Figure 2.6). Zero point levels in these 1-dimensional representations correspond to the combined energies of the symmetric and asymmetric stretches predicted for linear NHN and NDN triatomic systems. In the case of the proton-bridged ion **1a**, the potential energy maxima for proton transit along these cuts lie below the zero point levels, and the vibrational wavefunctions (as shown) are flat in the region of the barrier tops. For the deuteron-bridged ion **1b** (which has a zero point level slightly below the barrier top) the zero point wavefunctions are slightly dished in the region of the barrier tops.

The control experiments summarized in Figure 2.10 and Figure 2.11 exhibit features that suggest the special properties of ion **1**, in which the NHN bond angle is not far from 180°. All of the ions exchange their nitrogen-bound protons completely in D_2O solution. An ion with a very nonlinear NHN bond angle, such as **3a** (which may or may not contain a LBHB), displays three bands between 1100 and 1400 cm^{-1} that disappear when the bridging H^+ is replaced by D^+. This shift between **3a** and **3b** differs from the behavior of ion **1**, in which only one band between 1100 and 1400 cm^{-1} domain vanishes upon deuterium substitution. Turning to the low-frequency domain, when the two nitrogens have unequal basicity (as in ion **2**), the only band observed between 300 and 600 cm^{-1} for the perprotio analogue does not change its position upon deuterium substitution. By contrast, ion **1a** displays low-frequency vibrations that shift when H^+ is replaced by D^+.

As Figure 2.9 summarizes, the positions of NN stretching (symbolized by dashed arrows) and of asymmetric NHN stretching modes (symbolized by solid arrows) can be fitted by using a 2-dimensional potential energy surface. This approach does not consider NHN bending motions. The bands provisionally assigned to bending modes (indicated by a black arrow in Figure 2.8 near 1200 cm^{-1}) appear far below where they are predicted, even by DFT anharmonic normal modes calculations. [38] At the same time, HCH scissor modes of **1a** in the 1200-1400 cm^{-1} domain occur at frequencies approximately 90-100 cm^{-1} higher than predicted. This behavior is reminiscent of the difference between the IRMPD spectra of the terephthalate and phthalate monoanions: the former exhibits a very good match between DFT-calculated and experimental band positions, while the latter (which has a proton bridge between the unionized and ionized carboxylic groups) shows poor agreement. [46]

The consequences of anharmonicity have been discussed more fully in the context of the proton-bound dimer of ammonia, $N_2H_7^+$. [77–79] Selective substitution of the bridging hydron cannot be achieved in that ion, so the present study with di-tertiary amines permits experiments that are not possible with $N_2H_7^+$. The band assigned to the NHN asymmetric stretching mode of **1a** occurs at a low frequency, but it is higher than reported in gaseous $N_2H_7^+$ (for which a fundamental at 374 cm^{-1} has been observed), and are also 100-150 cm^{-1} higher than the corresponding a_{2u} mode predicted for $N_2H_7^+$ by theory using multidimensional surfaces. [77–79] As in the case of $N_2H_7^+$, the band provisionally assigned to NHN bending - 1200 cm^{-1} for **1a** - occurs at higher frequencies than the asymmetric NHN stretch.

Unlike $N_2H_7^+$, the conjugate acid ions of the tertiary diamine examined here do not possess threefold axes of symmetry. Yet, as in the case of $N_2H_7^+$, coupling between the NHN asymmetric and the NN stretching modes takes place, as signaled by the effect of changing the bridging hydron from H^+ to D^+ in **1**. Isotopic substitution not only moves the former vibration out of range, but it also shifts the latter by 30-50 cm^{-1} to the red, as predicted on the basis of the 2-dimensional PESs calculated herein. No absorption has been assigned to the pure NN stretch in the reported IRMPD of $N_2H_7^+$, but the predicted values [77–79] do not lie far from the observed band positions assigned to that mode in Figure 2.9 and Figure 2.14.

The structure of ion **1a** also gives good agreement with theory. The experimental NN distance (from the X-ray structure of the monoiodide salt), R = 2.653 Å, has the same value, within experimental error, as the expectation value <R> based on Figure 2.6. This is considerably shorter than the NN distance in the proton-bound dimer of ammonia, 2.765 Å, when $N_2H_7^+$ is encapsulated within a calixarene and has unequal NH distances [83] (a distance that lies between the values predicted by theory [77–79]).

Available data do not unambiguously indicate the positions expected for bending vibrations of LBHBs between cationized tertiary amines. DFT calculations predict that bending frequencies should go up and NHN stretching frequencies should go down as LBHBs become stronger. For nearly linear LBHBs (such as **1a**) the NHN and NN stretches ought to occur at much lower frequencies than the NHN bends. Full dimensional treatments of $N_2H_7^+$ indicate that the NHN bends for that ion (the degenerate pair ν_{xy}) occur above 1500 cm^{-1}, consistent with experiment [77–79] (a frequency much higher than seen for the NHN asymmetric stretch). If the bands marked by arrows in Figure 2.8 and Figure 2.11 are correctly assigned to NHN bending vibrations, then their occurrence at frequencies below 1400 cm^{-1} must be due to coupling with motions of the methyl groups or the methylene chain. In such a case, treating those bands as vibrations of a triatomic system would prove inadequate.

2.5 Conclusions

Not all proton bridges necessarily form low-barrier hydrogen bonds, but the structural features of salts of monoprotonated tetramethylputrescine, gas-phase vibrational spectroscopy of isolated ions, and the Raman and INS described here present a coherent body of data consistent with LBHB behavior. While molecular symmetry may play a role, the LBHB characteristics presented by ion **1a** seem to require only that the two nitrogens possess the same basicity, a near-linear geometry, and a low barrier to proton transit accompanied by minimal motion of the nitrogens. The spectra in Figure 2.10 for tetramethylornithine conjugate acid ions, **2**, which exhibit no shift of the lowest observed band upon deuteration, illustrate that when the two amino groups have sufficiently different basicities, LBHB behavior is not seen.

Strong hydrogen bonding can shift NH stretching vibrations to lower frequencies in the gas phase without necessarily broadening their absorption profiles, but not all such cases represent LBHBs. [84] Results reported here provide benchmarks for vibrational spectroscopy of linear, cationic LBHBs. The quality of fit between experiment and calculations for low frequency NN and NHN in-line vibrations suggests that theory may ultimately be useful for evaluating isotope effects that result from changing the bridging hydron from H^{+} to D^{+}. Substantial equilibrium isotope effects have been invoked in context of enzymology, where LBHBs have been discussed. [22, 85–87] Additional spectroscopic studies may permit such isotope effects to be predicted with accuracy.

Chapter 3

The Structure of the Protonated Adenine Dimer by Infrared Multiple Photon Dissociation Spectroscopy and Electronic Structure Calculations

3.1 Introduction

3.1.1 Context

This work demonstrates the importance of modeling solvent effects. The infrared multiple photon dissociation (IRMPD) spectrum of electrosprayed adenine proton-bound dimers were recorded in the gas-phase. While gas-phase B3LYP/6-31+G(d,p) calculations indicate that the four lowest isomers are essentially isoenergetic, comparisons of the experimental and predicted IR spectra suggest that only two of the four isomers are observed in the experiment.

However, computed solvation effects, as modeled using both a polarizable continuum model (PCM) and microsolvation with five explicit water molecules, preferentially stabilize these two observed isomers, consistent with the interpretation of the IRMPD spectra. This work shows that for these small species the solvent-phase structure is preserved. It also demonstrates the potential danger of using gas-phase calculations to predict the structures of gaseous ions born in solution, such as those from an electrospray source.

3.1.2 Motivation

Adenine (6-aminopurine, $C_5H_5N_5$) plays a crucial role in biochemistry. For example, it is a main component of adenosine triphosphate (ATP), NAD+, and nucleic acids to name a few. Nucleic acids have been used in constructing materials [88] such as nanomachines, nanoscaffolds, and DNA computers. [89] DNA bases themselves have gained interest as components of self-assembling structures. For example, 2-D sheetlike nanostructures containing doubly protonated adenine have self-assembled from low-pH aqueous solutions containing adenine, hydrogen halide, and iodide. These structures have been isolated and characterized by X-ray crystallography. [90]

Adenine itself has been the topic of numerous quantum mechanical molecular structure studies to explore geometries, proton affinities, protonation, polyhydration, tautomerism, the structure of the amino group in terms of planarity, and the mechanism of double proton transfer in adenine-containing complexes. [91–101] The presence of rare or noncanonical tautomeric forms of DNA bases is closely related to mispairing of purines and pyrimidines, causing spontaneous point mutations. [102–104] Therefore,

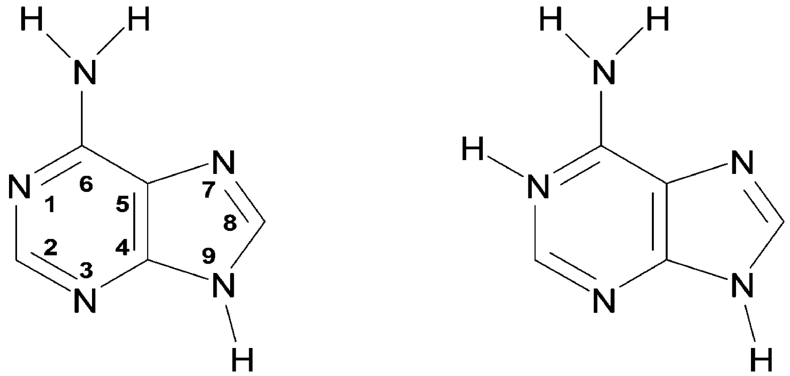


Figure 3.1: Adenine dimer, Scheme 1.

there is an increasing interest to investigate the structure and tautomerism of the DNA bases. The N9H tautomer of adenine was calculated to be the most stable (Figure 3.1, Scheme 1) in the gas phase by some 30-35 kJ mol^{-1}. [105, 106] Protonation of the N9H tautomer at N1 is found to be thermodynamically preferred in the gas phase. [107] Gu et al. [91] explored the water-assisted intramolecular proton transfer in the tautomers of adenine using density functional theory (DFT, B3LYP/6-311G(d,p)) calculations and found that the high-energy imino form of the tautomers of adenine are stabilized by about 8-12 kJ mol^{-1} in the presence of water due to enhancement of the conjugated π electron system found in the imino form of adenine. Because of its larger dipole moment, the N7H tautomer of adenine was found to be preferentially stabilized in the presence of water, where it lies only about 16 kJ mol^{-1} above the N9H tautomer. [91] A larger concentration of the N7H tautomer, then, would be expected in aqueous solutions and in biological systems than indicated by gas-phase calculations. A water molecule was also found to reduce the energy barrier to the tautomerization by acting as both a proton donor and acceptor, thereby assisting the proton transfer isomerization. [91] The tunneling effect in the intramolecular proton transfer in adenine was investigated using the parabolic barrier approximation and one-dimensional model. [108,109] The tunneling rates were calculated to be 1010 times larger than the classical one for the gas phase and 103-104 times larger for the water-assisted process. In addition, the assignment of the bands in the infrared spectra of argon-and nitrogen-matrixisolated adenine and its ^{15}N isotopomers (substituted at N9 or N7 positions) indicated that only the N9H tautomer of adenine was identified. [110] The results of these studies represent the solid phase since the source of matrix-isolated adenine is from heating solid adenine in vacuum and entraining the vapor in matrix gas prior to being condensed on a cryogenic substrate.

Proton-transfer reactions are important in biological systems since they can lead to mutations. [103] It is therefore important to investigate the changes in chemical properties of bases due to protonation. Many experimental and theoretical studies have been conducted on protonated nucleic acid bases and related compunds [111–116] due to the remarkable effect protonation has on conformational structure. [117] For example, base protonation has been implicated in the transformation of the B to Z structures of DNA. [118, 119] Knowledge of the protonation site is also essential for the design of some drugs which regulate the activity of a selected gene by stabilizing the triple helix formed between the target base sequence and an oligonucleotide. [120] The triple helix structure inhibits transcription resulting in its therapeutic effect. [121] Russo et al. [112] have performed gradient-corrected density functional computations with triple-ζ-type basis sets to determine the preferred protonation site and the absolute gas-phase proton affinities of the most stable tautomers of the DNA bases and for the first time predicted the gas phase basicity order among four DNA nucleic acid bases to be

$$\text{guanine} > \text{cytosine} > \text{adenine} > \text{thymine}$$

.

A better redistribution of electron density was found in the most favorable protonated species. The nature of the highest occupied molecular ortbial (HOMO), molecular electrostatic potential, and charge distribution together can explain the effects that stabilize the most stable protonated structures. Proton affinity values of 873.6 (T), 958.6 (C), 944.7 (A), and 963.6 (G) kJ mol^{-1} at 298 K were obtained, in a fair agreement with available experimental data.

The small energy differences between the adenine tautomers and among their conjugate acids make the identification of the most stable tautomers theoretically challenging. Relative tautomer stabilities are highly sensitive to changes in the treatment of electron-electron correlation and/or the one particle basis. [99, 112] Therefore, experimental techniques that target alternative properties are potentially very useful. IR spectroscopy is obviously a powerful technique to study adenine cluster systems; however, other spectroscopic techniques (such as photoexcitation) have been widely used to elucidate changes in the electronic structure and dynamics of adenine due to the presence of a proton. For instance, Marian et al. [122] produced protonated adenine ions by electrospray, stored and cooled them in a Paul trap, and dissociated them using resonant photoexcitation with nanosecond UV laser pulses. By comparing their photofragmentation spectra with computed vertical excitation spectra, it was determined that protonation mainly occurs at the N1 position of an N9H tautomer of adenine, with a possible contribution from the N3 protonated N7H tautomer, which lies only 1.9 kJ mol^{-1} higher in energy.

IRMPD spectroscopy [123–125] provides more direct evidence of cation structure. For instance, Atkins et al. [126] have used IRMPD spectroscopy in the N-H/O-H stretching region along with electronic structure calculations to determine that the lowest energy structure of sodium-bound glycine dimers consists of two symmetric bidentate ligands. Similarly, for the aliphatic amino acid protonbound homodimers composed of glycine, alanine, and valine as well as the heterogeneous alanine/glycine mixture an iondipole complex in which the N-protonated amino acid is bound to the carbonyl oxygen of the second monomer was found to be the dominant structure. [126, 127] In another recent study, the oxazalone structure was confirmed for the b_2^+ ion produced from collision-induced dissociation of the protonated AGG peptide while the protonated cyclic dipeptide is a diketopiperazene. [128] Much less work has been done on DNA bases and related compounds. Salpin et al. [129] identified the lowest energy structures of the protonated pyrimidic bases using IRMPD spectroscopy in the 900 to 2000 $\mathrm{cm^{-1}}$ region. [130] Lithium cationized complexes of thymine and uracil have also been studied in the gas phase by IRMPD spectroscopy in the N-H/O-H stretching region. [131] On the basis of a combination of experimental and theoretical data, it was found that the lithium cation in both thymine and uracil complexes most likely bind to O4 to form linear Li^+-bound dimers. Hydration of these Li^+-bound dimers resulted in significant structural changes to enable strong interbase hydrogen bonding, similar to that in the Watson-Crick model of DNA.

In recent theoretical work by Liu et al., [96] nine stable protonbound adenine dimers, $(C_5H_5N_5)_2H^+$, formed from the N9H tautomer of adenine and the N1 protonated N9H tautomer were found. In some of the proton-bound dimers, the proton partially or completely transferred from the protonated adenine to the neutral. While one might expect the proton-bound dimer to consist of the most stable neutral and protonated monomer forms, Hud and Morton [132] demonstrated that the four protonbound dimers composed of the N9H tautomer protonated at N1 and the N7H tautomer of adenine are by far the lowest energy proton-bound dimers and that the four isomers have fairly similar binding energies. In the present work the structure of the adenine proton-bound dimer is explored by combination of theoretical and IRMPD techniques. By comparison of the experimental IRMPD spectra with theoretical predictions of the vibrational spectra for various isomers, we hypothesize that only one or two isomers are observed experimentally. While our predicted gas-phase thermochemistry data cannot explain this observation, we perform calculations that suggest that the experimentally observed isomers are thermodynamically favored in solution. Therefore, we propose that only two of the most stable gas-phase dimers are prevalent in solution and that these solution-stable dimers are predominantly the ones that are electrosprayed and observed in these gas-phase experiments.

3.2 Methods

3.2.1 Experimental Section

The details of coupling the ApexQe Bruker Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer with a 25 Hz Nd:YAG pumped Laservision optical parametric oscillator/amplifier (OPO/OPA) laser have been presented previously. [133] Adenine proton-bound dimers were electrosprayed from 5 mM solutions of adenine in 18 M Millipore water which had been slightly acidified with a few drops of 1 mM HCl solution. Protonated adenine dimer (m/z 271) was isolated in the ICR cell by standard ejection techniques. Absorption of the infrared laser light resulted in dissociation of the proton-bound dimer, which was monitored by a change in mass of the parent ion. The IR laser scan rate was 0.5 cm^{-1} s^{-1} with irradiation times of 2.0 s. This corresponds to a step size of 1 cm^{-1} between points in the IRMPD spectra. IRMPD efficiency is defined as the negative of the natural logarithm of precursor ion intensity divided by the sum of the fragment and precursor ion intensities.

3.2.2 Computational Section

Optimized structures, dimer interaction energies, harmonic vibrational frequencies, and statistical thermodynamic quantities of the various proton-bound adenine dimer isomers were computed using the B3LYP density functional and the 6-31+G(d,p) basis set in the Gaussian 03 software package. [134] Single point MP2/6-311++G(2d,p) calculations were also done on the B3LYP/6-31+G(d,p) structures. Thermodynamic quantities employing the MP2 electronic energies and the B3LYP thermal corrections are reported as MP2/6-311++G(2d,p)//B3LYP/6-31+G(d,p) energies. In addition, we refine the predicted energies using the recently developed double-hybrid B2P3LYP density functional [29] with the cc-pVTZ basis. [30] Standard density functionals exhibit known deficiencies in treating van der Waals interactions, [31] and double-hybrid functionals [32, 135] attempt to correct for this deficiency by mixing in Moller-Plesset (MP2)-like long-range correlation into the functional. The B2P3LYP functional significantly outperforms B3LYP for systems where noncovalent interactions are important, so we use it here. The cc-pVTZ basis is the recommended basis set for use with B2P3LYP, since the empirical parameters in the functional were fit using this basis. To compensate for basis set superposition errors (BSSE), the standard counterpoise correction [136] was also considered. All B2P3LYP calculations were performed using a developmental version of Q-Chem, version 3.1. [33] Because analytical gradients of the B2P3LYP functional are currently unavailable, we utilize B3LYP/6-31+G(d,p) structures and harmonic frequencies (for computing thermodynamics).

Two additional approximations are invoked to accelerate the B2P3LYP energy calculations. First, just as in the original B2P3LYP paper, we use the resolution-of-the-identity (RI) approximation [34,137,138] and the auxiliary cc-pVTZ fitting basis [35] to speed the MP2 portion of the B2P3LYP calculation substantially. Second, we utilize a dual-basis Hartree-Fock (HF)/MP2 calculation. [36] In this approach, the HF/cc-pVTZ solution is approximated by taking the converged HF density matrix from a carefully chosen smaller basis set, projecting the density matrix into the larger cc-pVTZ basis and taking a single HF iteration in the larger basis. By use of these approximations, the energy of a 5-water-molecule-solvated adenine dimer structure (1044 basis functions in the cc-pVTZ basis) can be computed in about 6 h on a single processor on a modern workstation. At the same time, these two approximations introduce negligible additional errors into the relative energies.

Two separate strategies were used to address aqueous solvation effects on the proton-bound dimers. First, PCM calculations [139] at the B3LYP/6-31+G(d,p) level are performed to address bulk solvation effects. Adenine proton-bound dimer gas-phase structures were reoptimized and frequencies computed in the presence of the polarizable continuum. Second, gas-phase calculations (with no PCM model) employing up to 5 explicit water solvent molecules around the proton-bound dimers were used to investigate specific adenine-solvent interactions.

3.3 Results and Discussion

3.3.1 Computed Structures and Thermochemistry of Adenine Proton-Bound Dimers

The four lowest-energy structures of the adenine proton-bound dimer are shown in parts A-D in Figure 3.2. These structures were first presented by Hud and Morton. [132] They are planar and best described as N3 or N1 protonated N9H adenine tautomers interacting with the N7H neutral tautomer of adenine. At the B3LYP level, they lie

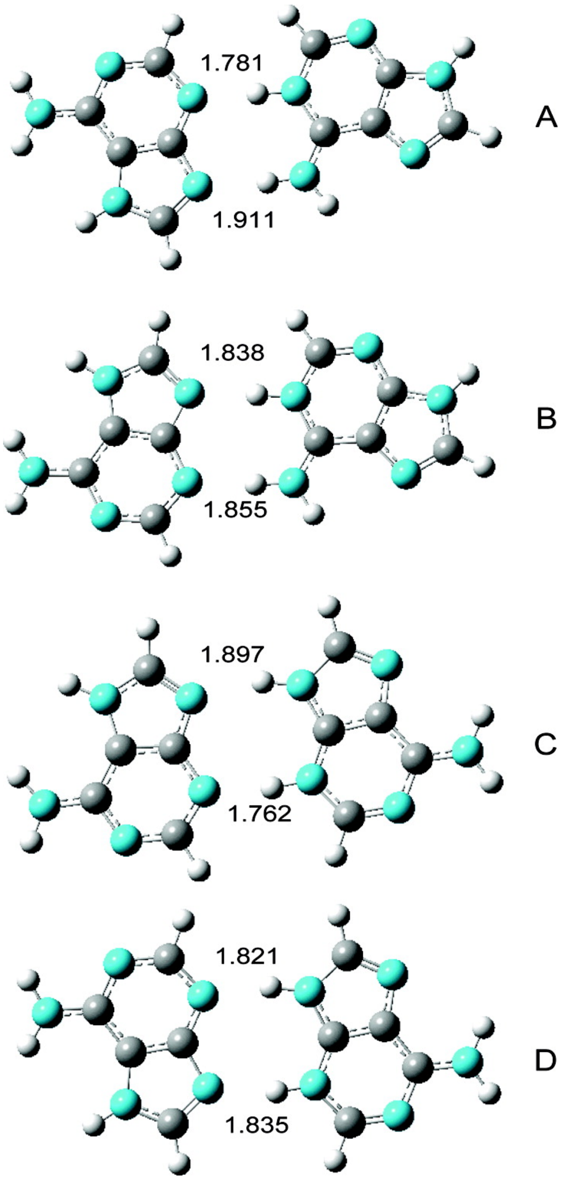


Figure 3.2: B3LYP/6-31+G(d,p) structures of the four lowest-energy proton-bound adenine dimers.

some 20 kJ mol^{-1} lower in enthalpy than the proton-bound dimers stemming from the N9H tautomers (E-J in Figure 3.3). [132]

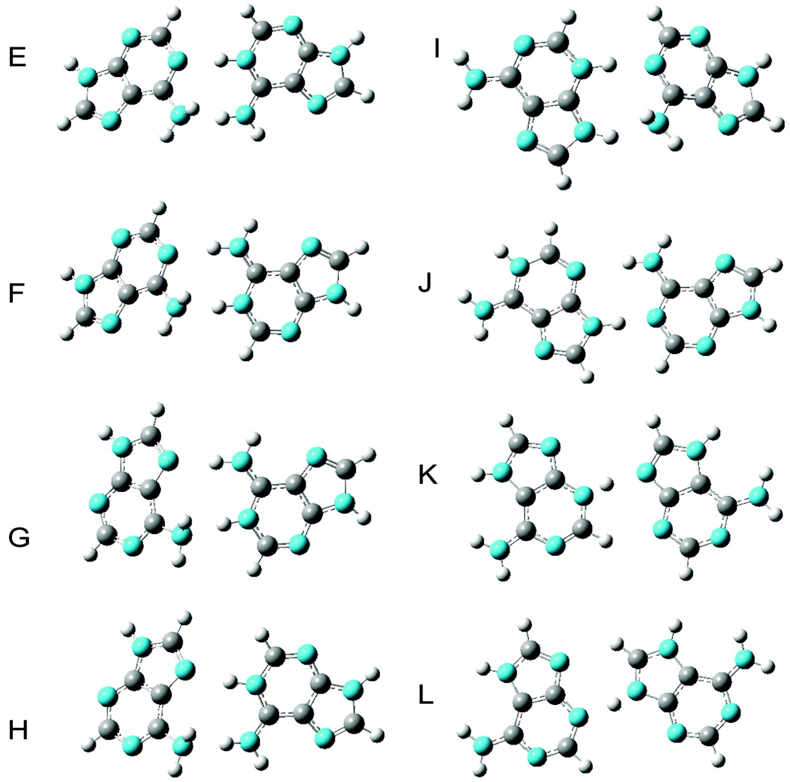


Figure 3.3: B3LYP/6-31+G(d,p) structures of eight high-energy proton-bound adenine dimers

To begin, we validate the B2P3LYP functional for these systems by comparing the predicted binding energies for the proton-bound adenine dimer with the experimental value of ΔH(500 K) = 127 ± 4 kJ mol^{-1} measured by Mautner using high-pressure mass spectrometry experiments. [114] The binding energies are computed relative to the adenine N9H tautomer and the N1-protonated adenine N9H tautomer. [140] These results are summarized in Table 3.1. As has been noted earlier, [132] B3LYP underestimates the interaction energy for the proton-bound dimer by about 20 kJ mol^{-1}. In contrast, three of the B2P3LYP-predicted isomer interaction energies lie within the experimental uncertainty, and the fourth lies only 1 kJ mol^{-1} outside of that range. Both B3LYP and

Table 3.1: Binding Energies (kJ mol^{-1}) of the Four Lowest-Energy Isomers of the Proton-Bound Adenine Dimer Relative to N9H Adenine and N1-Protonated-N9H Adenine (Experimental Value for ΔH(500 K) = 127 ± 4 kJ mol^{-1})

	B3LYP/6-31+G(d,p)		B2P3LYP/cc-pVTZ^{a}	
Isomer	Δ $E_{elec}CP^{b}$	$\Delta H(500\text{ K})^{c}$	Δ $E_{elec}CP^{b}$	$\Delta H(500\text{ K})^{c}$
A	113.8	106.8	132.4	125.4
B	111.3	104.1	129.5	122.3
C	111.0	104.7	131.1	124.8
D	109.4	102.9	129.5	123.1

[a] Using B3LYP/6-31+G(d,p)-optimized geometries.
[b] Counterpoise-corrected electronic interaction energies.
[c] Enthalpies computed using B3LYP/6-31+G(d,p) structures and frequencies.

B2P3LYP suggest that isomer A is the most stable, but the small energy differences between the four structures prevent a definitive prediction of the most stable gas-phase structure. Overall, these results suggest that the B2P3LYP functional performs better than B3LYP for this system. In the remaining sections, we will use B2P3LYP to confirm the B3LYP predictions.

The relative energies of the four lowest-energy proton-bound dimers computed here and energies for these and other structures published previously [96] are summarized in Table 3.2. MP2/6-311++G(2d,p)//B3LYP/6-31+G** energies agree with the relative energetic ordering from Hud and Morton [132] and Liu et al. [96] The relative energies of these species suggest that the higher energy structures, E-L, are unlikely to be present in significant quantities experimentally.

3.3.2 IRMPD Spectroscopy

Upon absorption of infrared radiation from the OPO laser, the adenine proton-bound dimer (m/z 271) dissociates to produce protonated adenine (m/z 136). No other products were observed.

The IRMPD efficiency spectrum in the 3250-3650 cm^{-1} region is shown as the

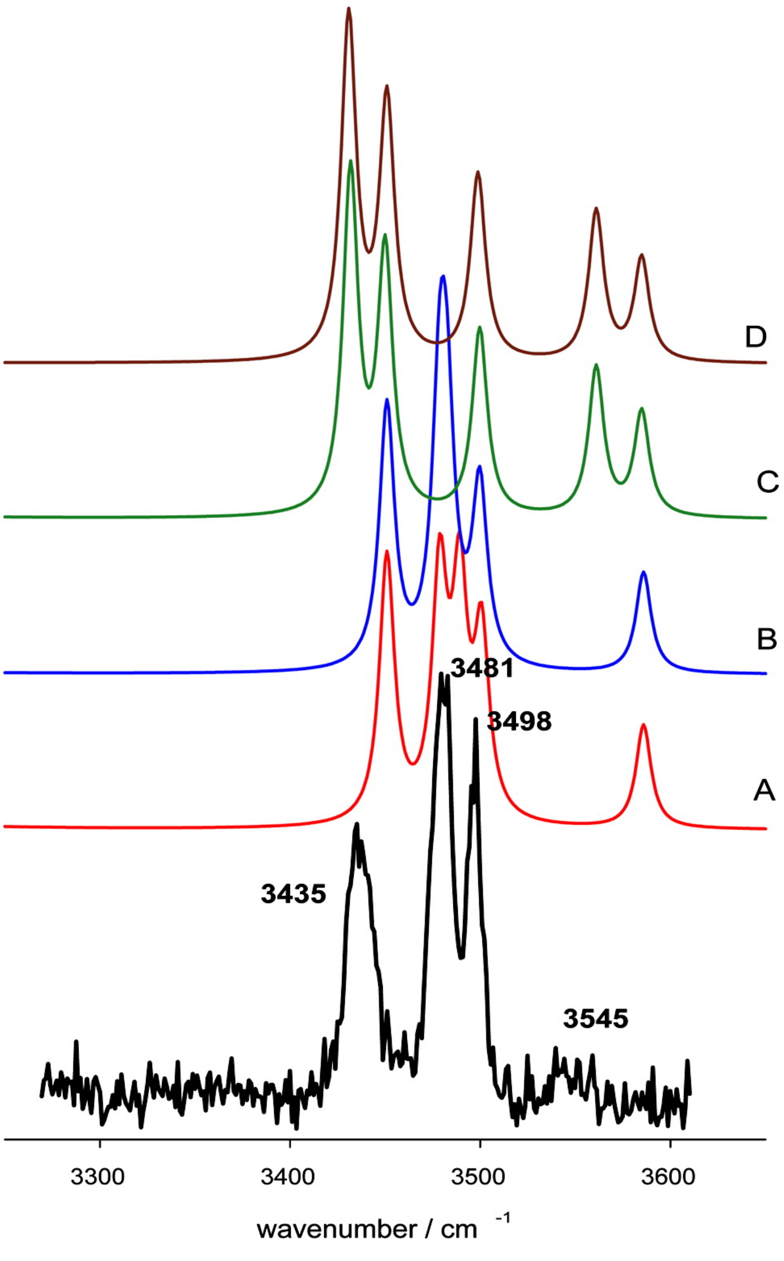


Figure 3.4: Comparison of the experimental IRMPD spectrum of the adenine proton-bound dimers with the B3LYP/6-31+G(d,p) predicted IR spectra for structures A-D in Figure 3.2

Table 3.2: Computed relative enthalpies for various adenine proton-bound dimer structures.
Method 1 = B3LYP/6-31++G(d)//B3LYP/6-31G(d,p)
Method 2 = B3LYP/6-311+G(d) for both energies and geometries
Method 3 = MP2/6-311++G(2d,p)//B3LYP/6-31+G(d,p)
Method 4 = B2P3LYP/cc-pVTZ//B3LYP/6-31+G(d,p)

Structure	Method 1^{ac}	Method 2^{b}	Method 3^{c}	Method 4^{c}
A	0		0.0 (0.0)	0.0 (0.0)
B	2.6		3.1 (1.9)	3.1 (1.9)
C	2.5		0.7 (1.3)	0.5 (1.1)
D	4		2.3 (2.4)	2.2 (2.3)
E	24.3	0	17.7 (21.1)	
F	27.8	3.8	23.1 (26.6)	
G	31.5	13.1	25.5 (30.0)	
H		15.1	24.5 (28.6)	
I		14.1	23.0 (26.1)	
J		15.9	36.2 (35.8)	
K			37.4 (34.5)	
L			46.2 (39.7)	

^{a} [132].

^{b} [96] and relative to structure E.

^{c} 298 K enthalpies and (free energies in parentheses), relative to structure A.

black trace at the bottom of Figure 3.4. The computed spectra in Figure 3.4 will be discussed below. There are three strong features at 3435, 3481, and 3498 cm^{-1} as well as a much weaker absorption at 3545 cm^{-1}. The two flanking bands can be assigned to the NH_2 symmetric stretching (3435 cm^{-1}) and the NH_2 antisymmetric stretching (3545 cm^{-1}) vibrations based on comparison with experiments conducted on neutral adenine in the gas phase (see Figure 3.5) and isolated in cryogenic matrices. [110,141,142] There have been some major differences reported between the experimental and theoretical positions of the NH_2 symmetric and antisymmetric stretching bands for the DNA bases. This is important to justify when comparing the experimental and computed spectra and will be discussed now.

In Figure 3.5 we compare the experimental IRMPD spectrum of the adenine proton-bound dimer and the IR spectrum of neutral adenine in the gas phase, [141,142]

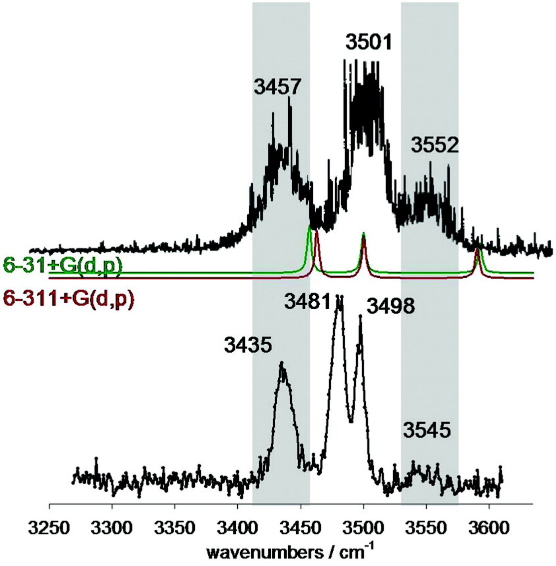


Figure 3.5: Comparison of the experimental IRMPD spectrum of the adenine proton-bound dimer with the gas-phase spectrum of neutral adenine. Also shown are predicted spectra for the N9H tautomer of neutral adenine

It is clear from Figure 3.5 that the NH_2 antisymmetric stretch and symmetric stretches are in very similar positions for the two species. Also in this figure are the spectra predicted by both B3LYP/6-31+G(d,p) and B3LYP/6-311+G(d,p) scaled by 0.957 and 0.9595, respectively. The scaling factor for the B3LYP/6-311+G(d,p) calculations was chosen to match the position of the N-H stretch predicted by the B3LYP/6-31+G(d,p) calculations, the scaling factor of which is a standard one used by our group. [126, 131] It can be seen that the observed NH_2 symmetric and antisymmetric stretching bands are significantly to the red of the predicted band positions in both cases. The same disagreement in the positions of these particular modes, NH_2 symmetric and antisymmetric stretching, has also been observed with B3LYP/6-31G(d,p) calculations of adenine [89] and cytosine. [99] The reason for the disagreement between experiment and theory is beyond the scope of this paper, but it is important to point out that this disagreement exists and that it must be considered when comparing the experimental and observed spectra of DNA bases and complexes containing an -NH_2 group. We are confident in assigning the 3435 and 3545 cm^{-1} features to the symmetric and antisymmetric stretching vibrations of the proton-bound dimer. Finally, very recent anharmonic calculations on neutral adenine predict the asymmetric and symmetric stretching vibrations at 3539 and 3432 cm^{-1}, respectively, in excellent agreement with experimental values. [143] The N-H stretching vibrations predicted by the scaled harmonic calculations and the anharmonic calculations are virtually identical.

The two bands in the experimental proton-bound dimer spectrum observed at 3481 and 3498 cm^{-1} (Figure 3.4) may be assigned to free N-H stretching bands based on their positions. Throughout this paper "free N-H" denotes an N-H group that is not involved in hydrogen bonding, which would strongly red-shift the N-H stretch out of the observable IR region. The band observed at 3501 cm^{-1} (Figure 3.5) for neutral adenine is the N9-H stretch. In the proton-bound dimer the band at 3498 cm^{-1} is in agreement with the theoretically predicted N7-H stretch of the neutral adenine moiety, for all four lowest energy structures, A-D (see Figure 3.2). In the proton-bound dimer spectrum the band centered at 3481 cm^{-1} is not observed in the neutral adenine spectrum and is not predicted for structures C and D. In structures C and D as well as neutral adenine, there is only one free N-H moiety. However, in proton-bound dimers A and B there are two free N-H groups. This closely resembles the positions of the predicted absorptions for the N9-H stretch of the protonated adenine moiety for proton-bound dimer structures A and B (see Table 3.3). At almost the same frequency is the free N-H stretch of the amino group involved in hydrogen bonding. This feature observed in the experimental spectrum does not rule out structures C and D from being present in the gas phase, but it does show that structures A and B are present and perhaps even dominant species based upon the relative intensities.

Also in Figure 3.4 are the B3LYP/6-31+G(d,p)-predicted IR spectra for the

Table 3.3: Table of Assignments for Experimental IRMPD Bands for the Adenine Proton-Bound Dimer and Predicted Bands for Structures A-D

		B3LYP/6-31+G(d,p)/cm^{-1}			
Observed/cm^{-1}	Assignment	A	B	C	D
3435	NH_2 symmetric stretch	3448	3448	3429^a 3447^b	3428^a 3447^b
3481	A(9)-H stretch	3479	3480		
	Free N-H stretch of H-bonded NH_2 group	3490	3483		
3498	A(7)-H stretch	3497	3497	3497	3496
3545	NH_2 antisymmetric stretch	3582	3583	3558^a 3581^b	3557^a 3582^b

[a] Protonated adenine moiety.
[b] Neutral adenine moiety.

four lowest energy proton-bound dimers, A-D. As discussed above, C and D alone cannot account for the experimental spectrum since the second N-H stretches at 3481 cm^{-1} is not predicted for them. Either of the structures A or B could account for the experimental spectrum since all of the features are accounted for in the predicted spectra. According to the predicted gas-phase thermochemistries, all four isomers are essentially isoenergetic, at least to within the computational error bars. However, if structures C and D are prevalent contributors to the experimental spectrum, the predicted spectra suggest that there would be two each of the NH_2 symmetric and antisymmetric stretching bands or that the experimental bands might be significantly broader than observed. While it is difficult to ascertain whether the weak band at 3545 cm^{-1} is split, the strong band centered at 3435 cm^{-1} does not resemble two bands nor is it significantly broadened. Calculations predict a difference of about 20 cm^{-1} between the two NH_2 symmetric stretching vibrations of C and D (see Table 3.3). The two bands at 3481 and 3498 cm^{-1} in the Experimental spectrum are well-resolved, so we would expect, if C and D are present, that there would be two bands around 3435 cm^{-1}.

Table 3.4: Relative enthalpies[a] (relative free energies in parentheses) of solvated and unsolvated adenine proton-bound dimers A, B, C, and D

	Unsolvated	PCM B3LYP/6-31+G**	5 waters	5 waters + PCM
A	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
B	3.1 (1.9)	-0.1 (-0.1)	3.5 (4.2)	-0.2 (0.6)
C	0.5 (1.1)	9.5 (10.4)	21.8 (20.7)	10.6 (11.2)
D	2.2 (2.3)	9.3 (10.5)	22.8 (22.9)	10.2 (11.7)

[a] Unless otherwise stated, all energies are B2P3LYP/cc-pVTZ//B3LYP/6-31+G(d,p), 298 K values.

3.3.3 Aqueous Solvation Effects on Dimer Stabilities

Working under the hypothesis that structures A and B are mainly responsible for the experimental infrared spectrum of the adenine proton-bound dimer, an explanation for why structures C and D can be ruled out, even though the gas-phase thermochemistries predict virtually equal amounts of all four (Table 3.4), is needed. Because these proton-bound dimers are initially formed in solution before being electrosprayed and desolvated, we explore the effects of solvation on the relative stabilities of the different isomers. Solvent effects remain challenging to address quantum mechanically, so two different approaches are used to investigate them. As we will demonstrate, both solvent models predict qualitatively similar results, suggesting that the models are capturing important physical effects.

First, a polarizable continuum model is used to approximate bulk solvation effects. The geometries of all four structures were therefore reoptimized using the PCM with water (dielectric = 78.39) as the solvent at the B3LYP/6-31+G(d,p) level. Frequency calculations were also performed in the presence of the PCM. The thermochemical results of the PCM calculation are shown in Table 3.4. According to the PCM calculations, water preferentially stabilizes structures A and B significantly with respect to C and D.

To look at specific localized solvent interaction, the explicit microsolvation using a handful of water molecules was examined. By use of five water molecules, all hydrogen-bond donors and acceptors on the proton-bound dimers can be saturated with water molecules. These structures, which could reasonably correspond to the innermost solvent coordination layer in bulk water, are presented in Figure 3.6, and their relative B2P3LYP/cc-pVTZ energies are listed in Table 3.4. In contrast to the gas-phase results, and in good agreement with the PCM ones, the A/B microsolvated structures are substantially more stable than the C/D ones. Similarly, calculations employing both the explicit solvent molecules and the PCM model provide energies consistent with the explicit-only and PCM calculations (Table 3.4). Simple equilibrium constant calculations based on either the microsolvated or PCM relative free energies suggest that mixtures of isomers A-D in solution will contain less than 1% each of C and D at 298 K. The ratio of A to B depends on the exact free energies used, but it ranges from almost a 50:50 mixture (PCM) to 84% A (5-water microsolvated). These results are consistent with the interpretation of the IRMPD spectrum discussed above.

Examining these results in more detail, Table 3.5 lists the relative stabilities of the singly solvated A and C dimer structures presented in Figure 3.7. (Solvated structures for B and D are in Figure 3.8.) The four single water molecule solvation sites are the most acidic available sites on each of the proton-bound dimers. Each solvation site donates a proton from the adenine to the water, and two of these sites also have an adjacent proton-acceptor site, which accepts a hydrogen bond from the water molecule. The key difference between the A/B and the C/D dimers is the orientation of the NH_2 group on the right adenine molecule. In A/B, this group is directly involved in the proton-bound dimer formation, while in C/D it is completely accessible to the solvent. A water-adenine hydrogen bond at this site on the side opposite the 5-membered ring is

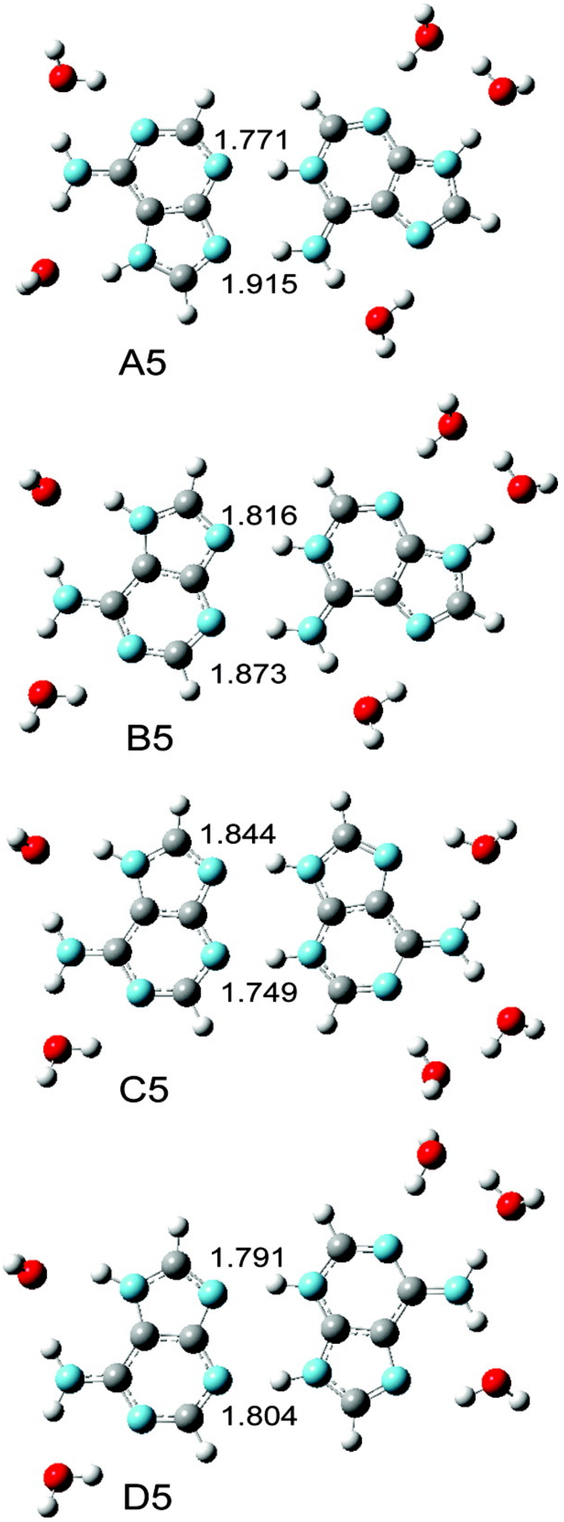


Figure 3.6: B3LYP/6-31+G(d,p) structures of the four lowest-energy adenine proton-bound dimers microsolvated with five water molecules.

Table 3.5: 298 K relative enthalpies and free energies for singly solvated adenine proton-bound dimers as calculated at B2P3LYP/cc-pVTZ//B3LYP/6-31+G** level of theory (see Figure 3.7 for structures)

Structure	Relative H	Relative G
A1i	0.0	0.0
A1ii	12.4	8.3
A1iii	17.0	4.5
A1iv	23.8	20.8
B1i	3.0	3.1
B1ii	15.7	13.4
B1iii	21.5	10.5
B1iv	26.3	24.2
C1i	-0.3	-0.8
C1ii	14.2	11.9
C1iii	24.4	21.2
C1iv	28.6	14.3
D1i	1.4	0.9
D1ii	15.0	12.9
D1iii	25.8	22.9
D1iv	30.3	15.9

the least favored position, as is clear from the relative energies of structures A1iv, C1iii, and C1iv (23.8, 24.4, and 28.6 kJ mol^{-1}, respectively). Because C/D have two such solvent-accessible sites, they are stabilized less in water than are the A/B proton-bound dimers.

Neither the PCM model nor the explicitly microsolvated model truly describes the bulk water solvation effects. However, both models provide the same results: structures C and D are significantly less stable in water than are A and B. Because the dimers are formed in solution prior to electrospray, we propose that the relative stability of these dimers in solution determines which isomers are observed in the IRMPD spectrum. This hypothesis assumes that the dimers A and/or B do not isomerize to C and/or D in the gas phase. This assumption seems reasonable, given that the binding energies of these proton-bound dimers are approximately 120-130 kJ/mol and that substantial disruption of the hydrogen bonding and tautomerization would be required,

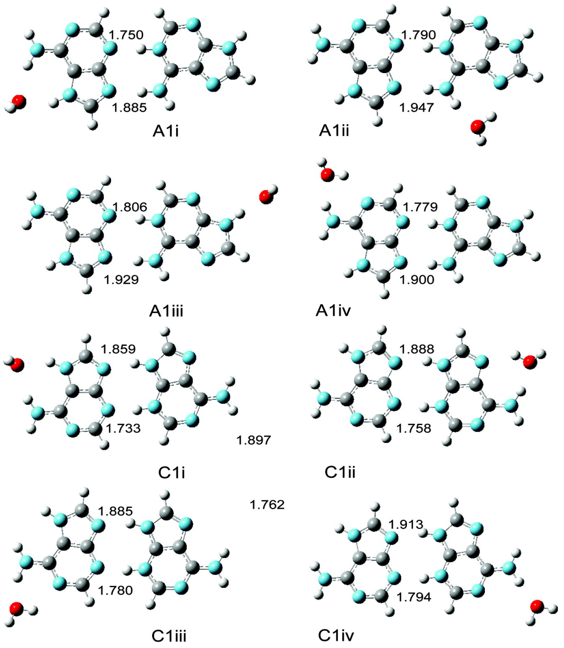


Figure 3.7: The four B3LYP/6-31+G(d,p) singly microsolvated structures each for A and C. Microsolvated structures for B and D are in Figure 3.8

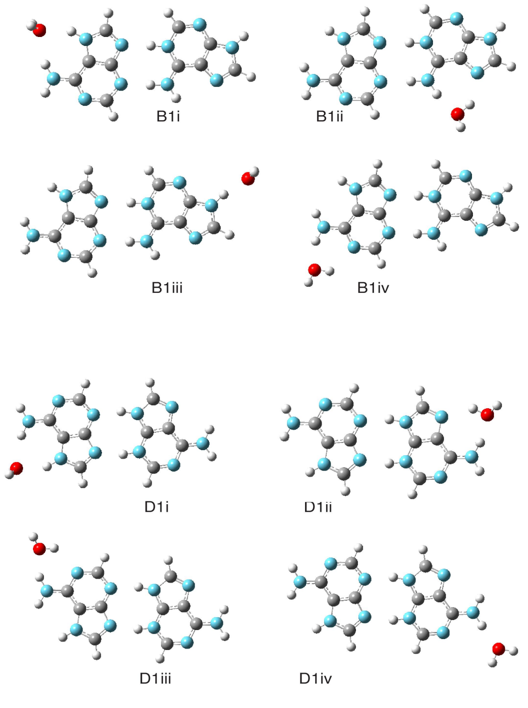


Figure 3.8: Singly microsolvated structures for B and D

resulting in a significant energy barrier for isomerization of A or B to C or D. In other words, isomerization seems unlikely on the time scale of the electrospray/desolvation process. This hypothesis is consistent with the apparent absence of strong signals for isomers C and D in the spectra. In contrast, the gas-phase energies which find all four structures nearly degenerate, cannot explain this observation.

3.4 Comparison of IRMPD Spectrum With Higher-Energy Isomers

In Figure 3.9 the IRMPD spectrum is compared to the computed spectra for structures E-L. Structures I and J have multiple strong bands in the lower-energy portion of the spectrum and therefore can be ruled out on spectroscopic grounds. However, structures E, F, G H, K, and L cannot be equivocally ruled out based solely by comparing of the experimental and computed spectra. These structures alone cannot account for the spectrum, but the absence of predicted bands does not rule them out. Structures E through L can, though, be ruled out based on their computed energies. Even PCM calculations with these structures do not lower their energies with respect to A and B.

3.5 Conclusions

Gas-phase calculations predict four isoenergetic isomers of the proton-bound adenine dimer. In contrast the IRMPD spectrum of electrosprayed adenine proton-bound dimers with the predicted IR spectra reveal that only two of the isomers are present. PCM model calculations and microsolvation calculations with 5 explicit water molecules qualitatively agree that solvation significantly stabilized two of the four lowest energy isomers. These two solvent-stabilized isomers are consistent with the in-

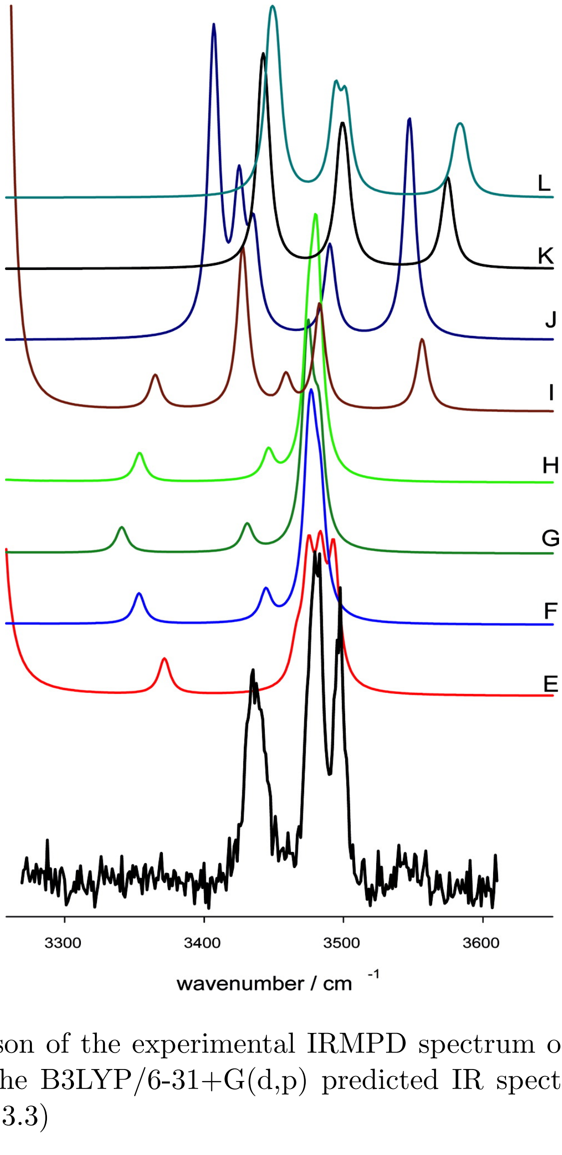


Figure 3.9: Comparison of the experimental IRMPD spectrum of the adenine proton-bound dimers with the B3LYP/6-31+G(d,p) predicted IR spectra for structures E-L (structures in Figure 3.3)

terpretation of the IRMPD spectrum. This work demonstrates that in some cases, using gas-phase calculations to predict the structures of ions born in solution and transferred to the gas-phase via electrospray ionization can give misleading results.

Chapter 4

Constructing a quasi-implicit solvent model through rigorous coarse-graining of explicit solvent in QM/MM simulations

4.0.1 Introduction

Solvation effects play a crucial role in chemistry, but their accurate and practical theoretical modeling has been a long-standing challenge. For example, changing from a protic to aprotic solvent can accelerate the Cope elimination reaction a millionfold. [6–8] Conventional, computationally inexpensive polarizable continuum models which treat the solvent as a featureless dielectric fail to capture these solvent effects. On the other hand, explicitly solvated hybrid quantum solute/classical molecular mechanics solvent (QM/MM) simulations do correctly reproduce the strong solvent effects in this system. [9]

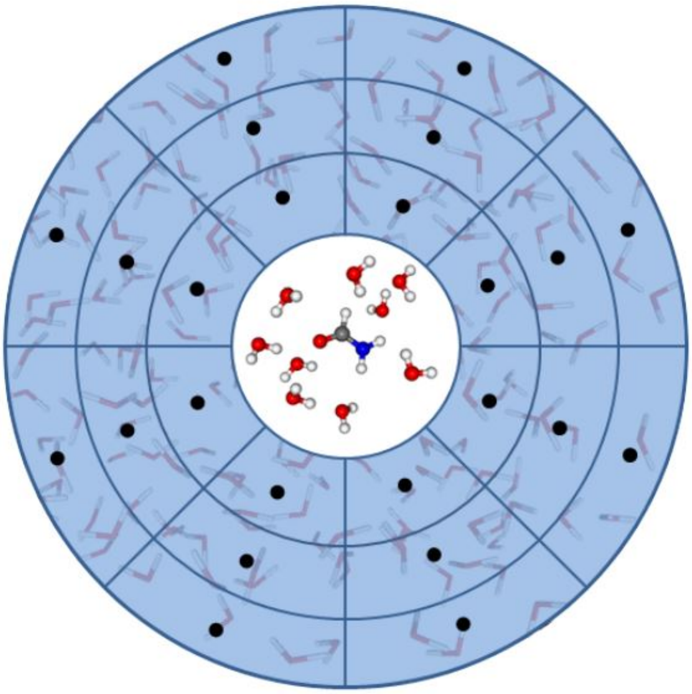


Figure 4.1: Schematic of the coarse-grained solvent model. The inner solute region (solute plus nearby solvent molecules) are treated explicitly, while the outer solvent region is coarse-grained into cells. The effective force field parameters in each cell are assigned to a grid point in the center of that cell and averaged over many solvent configurations.

Unfortunately, modeling explicit solvent requires large numbers of solvent molecules and extensive configuration sampling. While the use of MM solvent in QM/MM simulations reduces some of the configuration sampling costs, the expensive solute QM calculations still need to be performed for each solvent configuration. At the same time, the many successful applications of implicit solvent models [144, 145] suggest that the sampling of the solvent degrees of freedom can be decoupled from the solute degrees of freedom to some extent, thereby greatly reducing the number of times the solute QM calculations need to be performed. On the other hand, most implicit models lose information about the structure of the solvent imparted by the solute.

This work explores a new strategy for capturing important solvent effects, including solvent structure, without incurring the high computational cost of full QM/MM configuration sampling. In this model, the solute and inner-most solvent molecules are treated explicitly with QM/MM, while the longer-range solvent degrees of freedom are decoupled and sampled separately. As illustrated in Figure 4.1, the solvent region is then divided into spatial cells, and the force field electrostatic, polarization, and van der Waals dispersion parameters of the molecules in each cell are rigorously mapped onto a point at the center the cell. The properties in each cell are then averaged over many solvent configurations, and the effective solute-solvent interactions are calculated in a single QM/MM calculation using these averaged properties.

The idea of averaging over solvent degrees of freedom for a QM calculation is not new. For example, mean-field models such as the averaged solvent electrostatic potential (ASEP) and related methods embed the system in point charges obtained by averaging over many solvent configurations. [10–14] Alternatively, one can pre-compute the response of the solute to the solvent and use that in place of QM calculations during the solvent sampling. This has traditionally been done by computing point charges or higher-order multipoles for the solute, [146–151] though more recent efforts use elaborate generalized multipole/polarizability expansions for the solute [152, 153] or continuous solute charge densities. [154] Integral equation theories like the reference interaction site model (RISM) also provide a useful means of obtaining information about the statistical structure and properties of the solvent. [155]

The approach used here is similar to the mean-field ASEP model in that it obtains a set of averaged solvent properties for use in a single quantum calculation. However, it uses a mathematically rigorous coarse-graining that does not rely on statistical data fitting, and it includes electrostatics, polarization and dispersion effects, all of which prove important. As we demonstrate below in a proof-of-principle application to a formamide molecule in water, a single QM/MM calculation with coarse-grained and averaged solvent reproduces the average over many explicit-solvent QM/MM calculations to within kJ/mol accuracy at much lower cost. Furthermore, the polarizable MM model used here is very faithful to QM, meaning that the interaction energies obtained in this fashion should be very close to what one would obtain in a set of fully QM calculations.

4.1 Overview of the coarse graining method

The proposed model divides the system into solute and solvent regions. The solute region contains a solute molecule plus some number of first-solvent-shell solvent molecules. This solute region will be treated explicitly with our fragment-based QM/MM hybrid many-body interaction (HMBI) approach [15–17] to capture important, specific solute-solvent interactions. HMBI treats individual molecules and their short-range (SR) pairwise intermolecular interactions using QM, while the longer-range (LR) pairwise and many-body interactions are approximated with an anisotropic polarizable force field whose parameters are computed on the fly from density functional theory.

$$E_{HMBI} = E^{QM}_{1-body} + E^{QM}_{SR\ 2-body} + E^{MM}_{LR\ 2-body} + E^{MM}_{many-body} \tag{4.1}$$

Because the HMBI QM/MM partitioning is based on the interaction type, all molecules in the solute region exhibit both QM and MM interactions. The *ab initio* force field (AIFF) used for the MM interactions in HMBI is parameterized using distributed multipoles (up to hexadecapole), distributed polarizabilities (up to quadrupole-quadrupole), and distributed dispersion (anisotropic terms up to R^{-8}). [16, 18] The HMBI fragment model provides substantial computational savings compared to a full QM treatment while introducing only small errors, and it has proved very effective in modeling molecular clusters [156] and crystals. [16, 157, 158]

The solvent region is treated purely with MM using the same AIFF. To coarse-grain the solvent region, the region is partitioned into spatial "cells" defined in terms of spherical polar angles and distances from the solute center. The coarse-grained properties reside on a grid point located at the center of each cell. The coarse-graining procedure seeks to determine effective multipoles, polarizabilities, and dispersion coefficients for each cell resulting from the molecules present in that cell during the solvent configuration sampling. To do so, we exploit the fact that a multipole moment Q_{lm} (using spherical polar notation) at a given point in space $\mathbf{r}$ can be exactly represented by a separate, infinitely long multipole expansion at another point in space $\mathbf{r}'$.

$$Q_{lm}(\mathbf{r}) = \sum_{l' \geq l} \sum_{m' \geq m} c_{lm,l'm'} Q_{l'm'}(\mathbf{r}') \tag{4.2}$$

where the coefficients $c_{lm,l'm'}$ involve spherical harmonics and numerical coefficients. [159]

In other words, the multipole moments on a given atom can be translated to a nearby coarse-graining (CG) grid point, at the expense of introducing higher-order multipole moments into the expansion. In this manner, the force field electrostatic parameters for all molecules in a given spatial cell can be translated and combined in the center of that cell.

$$Q_{lm}^{CG} \leftarrow \sum_{A}^{atoms} \sum_{l'm'} Q_{l'm'}^{A} \tag{4.3}$$

The same idea of combining multipoles is used in the fast multipole method to accelerate the evaluation of long-range electrostatics. [160]

One can go a step further and coarse-grain the atom-centered polarizabilities and frequency-dependent polarizabilities by recognizing that a polarizability can be expressed in terms of multipole operators $\hat{Q}_{lm}$, [159]

$$\alpha_{lm,l'm'} = \sum_{n}{}' \frac{\langle 0|\hat{Q}_{lm}|n\rangle\langle n|\hat{Q}_{l'm'}|0\rangle + \langle 0|\hat{Q}_{l'm'}|n\rangle\langle n|\hat{Q}_{lm}|0\rangle}{W_n - W_0} \tag{4.4}$$

Therefore, the polarizabilities can be translated to the cell center by inserting expressions for multipole translations for the $\hat{Q}_{lm}$ and $\hat{Q}_{l'm'}$ operators. See the Supporting Information for more details of the coarse-graining. The coarse-grained multipoles and polarizabilities at the grid point in each cell are subsequently averaged over many solvent configurations.

If infinite-order multipole expansions were used, this procedure for translating the multipoles and polarizabilities in individual solvent configurations would be formally exact (though subject to the often poor convergence of the multipole series). In practice, the expansions are truncated at relatively low order, which limits the size of the coarse-graining cells that can be used, since the weights of the new higher-order multipole terms grow with translation distance.

The interactions between the solute region and the coarse-grained solvent cells are evaluated readily in the HMBI framework. Each coarse-grained cell behaves like an additional monomer consisting of a single "atom" in the HMBI model, albeit with no QM contributions. The total energy is given by,

$$\begin{aligned} E_{total} &= E^{QM}_{1\text{-}body}(solute) + E^{QM}_{SR\ 2\text{-}body}(solute\text{-}solute) + E^{MM}_{LR\ 2\text{-}body}(solute\text{-}solute) \\ &+ E^{MM}_{2\text{-}body}(solute\text{-}CG) + E^{MM}_{2\text{-}body}(CG\text{-}CG) + E^{ind}_{many\text{-}body}(solute, CG) \end{aligned} \tag{4.5}$$

The first three terms correspond to the standard HMBI energy within the solute region excluding the many-body induction. The fourth and fifth terms describe the MM pairwise interactions between each solute molecule and coarse-grained cell or between two coarse-grained cells, respectively. The final term describes the many-body induction for the entire system computed self-consistently among all molecules in the solute region and the coarse-grained cells. Because we are interested here in the interaction between the solute and the solvent, no intramolecular (1-body) solvent region contributions are included.

The solute-solvent two-body electrostatics and polarization are computed from the multipolar interactions between multipoles/polarizabilities on each solute molecule and each coarse-grained cell. Solute-solvent two-body dispersion interactions are computed via Casimir-Polder integration over the solute and coarse-grained cell anisotropic frequency-dependent polarizabilities. The MM interactions between two coarse-grained cells are computed analogously. In principle, three-body dispersion could be handled similarly, though those effects are omitted here. Instead, we provide a proof-of-principle demonstration that this approach reproduces the full HMBI QM/MM results at much lower computational cost.

4.2 Theoretical details of coarse graining

4.2.1 Translation equations for multipole moments, Rank 0 - Rank 4

One of the keys to the success of this method is the mathematically rigorous algorithm for translating multipoles and polarizabilities. As stated earlier, the translation equations are exact if the multipole expansion is carried out to infinite rank, subject to convergence of the multipole expansion. In our implementation, we include up to hexadecapole moments (rank 4) in the model. Higher order terms might prove to be necessary if the translation distances are increased substantially, however for our sample system rank 4 proves to be adequate.

Using the standard addition theorem for spherical harmonics, one can show that the multipole components at a new position $\mathbf{r}'$ are determined as a linear combination of the multipoles at the original position $\mathbf{r}$. [159]

$$Q'_{lk}(\mathbf{r}') = \sum_{l'=0}^{l} \sum_{k'=-l'}^{l'} \left[\binom{l+k}{l'+k'} \binom{l-k}{l'-k'} \right]^{1/2} Q_{l'k'}(\mathbf{r}) R_{l-l',k-k'}(\mathbf{r}-\mathbf{r}') \tag{4.6}$$

Thus, $\mathbf{r}$ is the point about which the original multipole $Q^O_{l'k'}$ is expanded, and $\mathbf{r}'$ is the point about which the new multipole Q^C_{lk} is then expanded. The translation vector is $\{x, y, z\}$ in Bohr radii. The subscripts l and k refer to the spherical harmonic indices of the translated multipole, and l' and k' refer to the subscripts of the original multipole before translation. The binomial coefficients are expanded, and the original multipole is multiplied by the corresponding spherical harmonic, real or complex. The terms are then summed and converted into real functions. A few representative multipole translations are worked out by hand, to assist the reader, in Appendix B.

Charge (Rank 0)

$\mathbf{Q_{00}}(\mathbf{r}') = Q_{00}(\mathbf{r})$

Dipole (Rank 1)

$\mathbf{Q_{11c}}(\mathbf{r}') = Q_{11c}(\mathbf{r}) - xQ_{00}(\mathbf{r})$

$\mathbf{Q_{11s}}(\mathbf{r}') = Q_{11s}(\mathbf{r}) - yQ_{00}(\mathbf{r})$

$\mathbf{Q_{10}}(\mathbf{r}') = Q_{10}(\mathbf{r}) - zQ_{00}(\mathbf{r})$

Quadrupole (Rank 2)

$$\mathbf{Q_{20}}(\mathbf{r}') = Q_{20}(\mathbf{r}) + \tfrac{1}{2}(2xQ_{11c}(\mathbf{r}) + 2yQ_{11s}(\mathbf{r}) - 4zQ_{10}(\mathbf{r}) + 2z^2Q_{00}(\mathbf{r}) - x^2Q_{00}(\mathbf{r}) - y^2Q_{00}(\mathbf{r}))$$

$$\mathbf{Q_{21c}}(\mathbf{r}') = Q_{21c}(\mathbf{r}) - \sqrt{3}zQ_{11c}(\mathbf{r}) - \sqrt{3}xQ_{10}(\mathbf{r}) + \sqrt{3}xzQ_{00}(\mathbf{r})$$

$$\mathbf{Q_{21s}}(\mathbf{r}') = Q_{21s}(\mathbf{r}) - \sqrt{3}zQ_{11s}(\mathbf{r}) - \sqrt{3}yQ_{10}(\mathbf{r}) + \sqrt{3}yzQ_{00}(\mathbf{r})$$

$$\mathbf{Q_{22c}}(\mathbf{r}') = Q_{22c}(\mathbf{r}) + \tfrac{1}{2}(2\sqrt{3}yQ_{11s}(\mathbf{r}) - 2\sqrt{3}xQ_{11c}(\mathbf{r}) + \sqrt{3}x^2Q_{00}(\mathbf{r}) - \sqrt{3}y^2Q_{00}(\mathbf{r}))$$

$$\mathbf{Q_{22s}}(\mathbf{r}') = Q_{22s}(\mathbf{r}) - \sqrt{3}xQ_{11s}(\mathbf{r}) - \sqrt{3}yQ_{11c}(\mathbf{r}) + \sqrt{3}xyQ_{00}(\mathbf{r})$$

Octupole (Rank 3)

$$\mathbf{Q_{30}}(\mathbf{r}') = Q_{30}(\mathbf{r}) + \tfrac{1}{2}(3x^2zQ_{00}(\mathbf{r}) + 3y^2zQ_{00}(\mathbf{r}) - 2z^3Q_{00}(\mathbf{r}) - 3x^2Q_{10}(\mathbf{r}) - 3y^2Q_{10}(\mathbf{r}) + 6z^2Q_{10}(\mathbf{r}) - 6zQ_{20}(\mathbf{r}) - 6xzQ_{11c}(\mathbf{r}) - 6yzQ_{11s}(\mathbf{r}) + 2\sqrt{3}xQ_{21c}(\mathbf{r}) + 2\sqrt{3}yQ_{21s}(\mathbf{r}))$$

$$\mathbf{Q_{31c}}(\mathbf{r}') = Q_{31c}(\mathbf{r}) + \tfrac{1}{4}(\sqrt{6}x^3Q_{00}(\mathbf{r}) + \sqrt{6}xy^2Q_{00}(\mathbf{r}) - 4\sqrt{6}xz^2Q_{00}(\mathbf{r}) + 8\sqrt{6}xzQ_{10}(\mathbf{r}) - 4\sqrt{6}xQ_{20}(\mathbf{r}) - 3\sqrt{6}x^2Q_{11c}(\mathbf{r}) - \sqrt{6}y^2Q_{11c}(\mathbf{r}) + 4\sqrt{6}z^2Q_{11c}(\mathbf{r}) - 2\sqrt{6}xyQ_{11s}(\mathbf{r}) - 8\sqrt{2}zQ_{21c}(\mathbf{r}) + 2\sqrt{2}xQ_{22c}(\mathbf{r}) + 2\sqrt{2}yQ_{22s}(\mathbf{r}))$$

$$\mathbf{Q_{31s}}(\mathbf{r}') = Q_{31s}(\mathbf{r}) + \tfrac{1}{4}(\sqrt{6}x^2yQ_{00}(\mathbf{r}) + \sqrt{6}y^3Q_{00}(\mathbf{r}) - 4\sqrt{6}yz^2Q_{00}(\mathbf{r}) + 8\sqrt{6}yzQ_{10}(\mathbf{r}) - 4\sqrt{6}yQ_{20}(\mathbf{r}) - 2\sqrt{6}xyQ_{11c}(\mathbf{r}) - \sqrt{6}x^2Q_{11s}(\mathbf{r}) - 3\sqrt{6}y^2Q_{11s}(\mathbf{r}) + 4\sqrt{6}z^2Q_{11s}(\mathbf{r}) - 8\sqrt{2}zQ_{21s}(\mathbf{r}) - 2\sqrt{2}yQ_{22c}(\mathbf{r}) + 2\sqrt{2}xQ_{22s}(\mathbf{r}))$$

$$\mathbf{Q_{32c}}(\mathbf{r}') = Q_{32c}(\mathbf{r}) + \tfrac{1}{2}(-\sqrt{15}x^2zQ_{00}(\mathbf{r}) + \sqrt{15}y^2zQ_{00}(\mathbf{r}) + \sqrt{15}x^2Q_{10}(\mathbf{r}) - \sqrt{15}y^2Q_{10}(\mathbf{r}) + 2\sqrt{15}xzQ_{11c}(\mathbf{r}) - 2\sqrt{15}yzQ_{11s}(\mathbf{r}) - 2\sqrt{5}xQ_{21c}(\mathbf{r}) + 2\sqrt{5}yQ_{21s}(\mathbf{r}) - 2\sqrt{5}zQ_{22c}(\mathbf{r}))$$

$$\mathbf{Q_{32s}}(\mathbf{r}') = Q_{32s}(\mathbf{r}) - \sqrt{15}xyzQ_{00}(\mathbf{r}) + \sqrt{15}xyQ_{10}(\mathbf{r}) + \sqrt{15}yzQ_{11c}(\mathbf{r}) + \sqrt{15}xzQ_{11s}(\mathbf{r}) - \sqrt{5}yQ_{21c}(\mathbf{r}) - \sqrt{5}xQ_{21s}(\mathbf{r}) - \sqrt{5}zQ_{22s}(\mathbf{r})$$

$$\mathbf{Q_{33c}}(\mathbf{r}') = Q_{33c}(\mathbf{r}) + \tfrac{1}{4}(-\sqrt{10}x^3Q_{00}(\mathbf{r}) + 3\sqrt{10}xy^2Q_{00}(\mathbf{r}) + 3\sqrt{10}x^2Q_{11c}(\mathbf{r}) - 3\sqrt{10}y^2Q_{11c}(\mathbf{r}) - 6\sqrt{10}xyQ_{11s}(\mathbf{r}) - 2\sqrt{30}xQ_{22c}(\mathbf{r}) + 2\sqrt{30}yQ_{22s}(\mathbf{r}))$$

$$\mathbf{Q_{33s}}(\mathbf{r}') = Q_{33s}(\mathbf{r}) + \tfrac{1}{4}(-3\sqrt{10}x^2yQ_{00}(\mathbf{r}) + \sqrt{10}y^3Q_{00}(\mathbf{r}) + 6\sqrt{10}xyQ_{11c}(\mathbf{r}) + 3\sqrt{10}x^2Q_{11s}(\mathbf{r}) - 3\sqrt{10}y^2Q_{11s}(\mathbf{r}) - 2\sqrt{30}yQ_{22c}(\mathbf{r}) - 2\sqrt{30}xQ_{22s}(\mathbf{r}))$$

Hexadecapole (Rank 4)

$\mathbf{Q_{40}}(\mathbf{r}') = Q_{40}(\mathbf{r}) + \frac{1}{8}(3x^4Q_{00}(\mathbf{r}) + 6x^2y^2Q_{00}(\mathbf{r}) + 3y^4Q_{00}(\mathbf{r}) - 24x^2z^2Q_{00}(\mathbf{r}) - 24y^2z^2Q_{00}(\mathbf{r}) +$
$8z^4Q_{00}(\mathbf{r}) + 48x^2zQ_{10}(\mathbf{r}) + 48y^2zQ_{10}(\mathbf{r}) - 32z^3Q_{10}(\mathbf{r}) - 24x^2Q_{20}(\mathbf{r}) - 24y^2Q_{20}(\mathbf{r}) +$
$48z^2Q_{20}(\mathbf{r}) - 32zQ_{30}(\mathbf{r}) - 12x^3Q_{11c}(\mathbf{r}) - 12xy^2Q_{11c}(\mathbf{r}) + 48xz^2Q_{11c}(\mathbf{r}) - 12x^2yQ_{11s}(\mathbf{r}) -$
$12y^3Q_{11s}(\mathbf{r}) + 48yz^2Q_{11s}(\mathbf{r}) - 32\sqrt{3}xzQ_{21c}(\mathbf{r}) - 32\sqrt{3}yzQ_{21s}(\mathbf{r}) + 4\sqrt{3}x^2Q_{22c}(\mathbf{r}) - 4\sqrt{3}y^2Q_{22c}(\mathbf{r}) +$
$8\sqrt{3}xyQ_{22s}(\mathbf{r}) + 8\sqrt{6}xQ_{31c}(\mathbf{r}) + 8\sqrt{6}yQ_{31s}(\mathbf{r}))$

$\mathbf{Q_{41c}}(\mathbf{r}') = Q_{41c}(\mathbf{r}) + \frac{1}{4}(-3\sqrt{10}x^3zQ_{00}(\mathbf{r}) - 3\sqrt{10}xy^2zQ_{00}(\mathbf{r}) + 4\sqrt{10}xz^3Q_{00}(\mathbf{r}) + 3\sqrt{10}x^3Q_{10}(\mathbf{r}) +$
$3\sqrt{10}xy^2Q_{10}(\mathbf{r}) - 12\sqrt{10}xz^2Q_{10}(\mathbf{r}) + 12\sqrt{10}xzQ_{20}(\mathbf{r}) - 4\sqrt{10}xQ_{30}(\mathbf{r}) + 9\sqrt{10}x^2zQ_{11c}(\mathbf{r}) +$
$3\sqrt{10}y^2zQ_{11c}(\mathbf{r}) - 4\sqrt{10}z^3Q_{11c}(\mathbf{r}) + 6\sqrt{10}xyzQ_{11s}(\mathbf{r}) - 3\sqrt{30}x^2Q_{21c}(\mathbf{r}) - \sqrt{30}y^2Q_{21c}(\mathbf{r}) +$
$4\sqrt{30}z^2Q_{21c}(\mathbf{r}) - 2\sqrt{30}xyQ_{21s}(\mathbf{r}) - 2\sqrt{30}xzQ_{22c}(\mathbf{r}) - 2\sqrt{30}yzQ_{22s}(\mathbf{r}) - 4\sqrt{15}zQ_{31c}(\mathbf{r}) +$
$2\sqrt{6}xQ_{32c}(\mathbf{r}) + 2\sqrt{6}yQ_{32s}(\mathbf{r}))$

$\mathbf{Q_{41s}}(\mathbf{r}') = Q_{41s}(\mathbf{r}) + \frac{1}{4}(-3\sqrt{10}x^2yzQ_{00}(\mathbf{r}) - 3\sqrt{10}y^3zQ_{00}(\mathbf{r}) + 4\sqrt{10}yz^3Q_{00}(\mathbf{r}) + 3\sqrt{10}x^2yQ_{10}(\mathbf{r}) +$
$3\sqrt{10}y^3Q_{10}(\mathbf{r}) - 12\sqrt{10}yz^2Q_{10}(\mathbf{r}) + 12\sqrt{10}yzQ_{20}(\mathbf{r}) - 4\sqrt{10}yQ_{30}(\mathbf{r}) + 6\sqrt{10}xyzQ_{11c}(\mathbf{r}) +$
$3\sqrt{10}x^2zQ_{11s}(\mathbf{r}) + 9\sqrt{10}y^2zQ_{11s}(\mathbf{r}) - 4\sqrt{10}z^3Q_{11s}(\mathbf{r}) - 2\sqrt{30}xyQ_{21c}(\mathbf{r}) - \sqrt{30}x^2Q_{21s}(\mathbf{r}) -$
$3\sqrt{30}y^2Q_{21s}(\mathbf{r}) + 4\sqrt{30}z^2Q_{21s}(\mathbf{r}) + 2\sqrt{30}yzQ_{22c}(\mathbf{r}) - 2\sqrt{30}xzQ_{22s}(\mathbf{r}) - 4\sqrt{15}zQ_{31s}(\mathbf{r}) -$
$2\sqrt{6}yQ_{32c}(\mathbf{r}) + 2\sqrt{6}xQ_{32s}(\mathbf{r}))$

$\mathbf{Q_{42c}}(\mathbf{r}') = Q_{42c}(\mathbf{r}) + \frac{1}{4}(-\sqrt{5}x^4Q_{00}(\mathbf{r}) + \sqrt{5}y^4Q_{00}(\mathbf{r}) + 6\sqrt{5}x^2z^2Q_{00}(\mathbf{r}) - 6\sqrt{5}y^2z^2Q_{00}(\mathbf{r}) -$
$12\sqrt{5}x^2zQ_{10}(\mathbf{r}) + 12\sqrt{5}y^2zQ_{10}(\mathbf{r}) + 6\sqrt{5}x^2Q_{20}(\mathbf{r}) - 6\sqrt{5}y^2Q_{20}(\mathbf{r}) + 4\sqrt{5}x^3Q_{11c}(\mathbf{r}) - 12\sqrt{5}xz^2Q_{11c}(\mathbf{r}) -$
$4\sqrt{5}y^3Q_{11s}(\mathbf{r}) + 12\sqrt{5}yz^2Q_{11s}(\mathbf{r}) + 8\sqrt{15}xzQ_{21c}(\mathbf{r}) - 8\sqrt{15}yzQ_{21s}(\mathbf{r}) - 2\sqrt{15}x^2Q_{22c}(\mathbf{r}) -$
$2\sqrt{15}y^2Q_{22c}(\mathbf{r}) + 4\sqrt{15}z^2Q_{22c}(\mathbf{r}) - 2\sqrt{30}xQ_{31c}(\mathbf{r}) + 2\sqrt{30}yQ_{31s}(\mathbf{r}) - 8\sqrt{3}zQ_{32c}(\mathbf{r}) + 2\sqrt{2}xQ_{33c}(\mathbf{r}) +$
$2\sqrt{2}yQ_{33s}(\mathbf{r}))$

$\mathbf{Q_{42s}}(\mathbf{r}') = Q_{42s}(\mathbf{r}) + \frac{1}{2}(-\sqrt{5}x^3yQ_{00}(\mathbf{r}) - \sqrt{5}xy^3Q_{00}(\mathbf{r}) + 6\sqrt{5}xyz^2Q_{00}(\mathbf{r}) - 12\sqrt{5}xyzQ_{10}(\mathbf{r}) +$
$6\sqrt{5}xyQ_{20}(\mathbf{r}) + 3\sqrt{5}x^2yQ_{11c}(\mathbf{r}) + \sqrt{5}y^3Q_{11c}(\mathbf{r}) - 6\sqrt{5}yz^2Q_{11c}(\mathbf{r}) + \sqrt{5}x^3Q_{11s}(\mathbf{r}) + 3\sqrt{5}xy^2Q_{11s}(\mathbf{r}) -$
$6\sqrt{5}xz^2Q_{11s}(\mathbf{r}) + 4\sqrt{15}yzQ_{21c}(\mathbf{r}) + 4\sqrt{15}xzQ_{21s}(\mathbf{r}) - \sqrt{15}x^2Q_{22s}(\mathbf{r}) - \sqrt{15}y^2Q_{22s}(\mathbf{r}) +$

$2\sqrt{15}z^2Q_{22s}(\mathbf{r})-\sqrt{30}yQ_{31c}(\mathbf{r})-\sqrt{30}xQ_{31s}(\mathbf{r})-4\sqrt{3}zQ_{32s}(\mathbf{r})-\sqrt{2}yQ_{33c}(\mathbf{r})+\sqrt{2}xQ_{33s}(\mathbf{r}))$

$\mathbf{Q_{43c}}(\mathbf{r}') = Q_{43c}(\mathbf{r})+\frac{1}{4}(\sqrt{70}x^3zQ_{00}(\mathbf{r})-3\sqrt{70}xy^2zQ_{00}(\mathbf{r})-\sqrt{70}x^3Q_{10}(\mathbf{r})+3\sqrt{70}xy^2Q_{10}(\mathbf{r})-$

$3\sqrt{70}x^2zQ_{11c}(\mathbf{r})+3\sqrt{70}y^2zQ_{11c}(\mathbf{r})+6\sqrt{70}xyzQ_{11s}(\mathbf{r})+\sqrt{210}x^2Q_{21c}(\mathbf{r})-\sqrt{210}y^2Q_{21c}(\mathbf{r})-$

$2\sqrt{210}xyQ_{21s}(\mathbf{r})+2\sqrt{210}xzQ_{22c}(\mathbf{r})-2\sqrt{210}yzQ_{22s}(\mathbf{r})-2\sqrt{42}xQ_{32c}(\mathbf{r})+2\sqrt{42}yQ_{32s}(\mathbf{r})-$

$4\sqrt{7}zQ_{33c}(\mathbf{r}))$

$\mathbf{Q_{43s}}(\mathbf{r}') = Q_{43s}(\mathbf{r})+\frac{1}{4}(3\sqrt{70}x^2yzQ_{00}(\mathbf{r})-\sqrt{70}y^3zQ_{00}(\mathbf{r})-3\sqrt{70}x^2yQ_{10}(\mathbf{r})+\sqrt{70}y^3Q_{10}(\mathbf{r})-$

$6\sqrt{70}xyzQ_{11c}(\mathbf{r})-3\sqrt{70}x^2zQ_{11s}(\mathbf{r})+3\sqrt{70}y^2zQ_{11s}(\mathbf{r})+2\sqrt{210}xyQ_{21c}(\mathbf{r})+\sqrt{210}x^2Q_{21s}(\mathbf{r})-$

$\sqrt{210}y^2Q_{21s}(\mathbf{r})+2\sqrt{210}yzQ_{22c}(\mathbf{r})+2\sqrt{210}xzQ_{22s}(\mathbf{r})-2\sqrt{42}yQ_{32c}(\mathbf{r})-2\sqrt{42}xQ_{32s}(\mathbf{r})-$

$4\sqrt{7}zQ_{33s}(\mathbf{r}))$

$\mathbf{Q_{44c}}(\mathbf{r}') = Q_{44c}(\mathbf{r})+\frac{1}{8}(\sqrt{35}x^4Q_{00}(\mathbf{r})-6\sqrt{35}x^2y^2Q_{00}(\mathbf{r})+\sqrt{35}y^4Q_{00}(\mathbf{r})-4\sqrt{35}x^3Q_{11c}(\mathbf{r})+$

$12\sqrt{35}xy^2Q_{11c}(\mathbf{r})+12\sqrt{35}x^2yQ_{11s}(\mathbf{r})-4\sqrt{35}y^3Q_{11s}(\mathbf{r})+4\sqrt{105}x^2Q_{22c}(\mathbf{r})-4\sqrt{105}y^2Q_{22c}(\mathbf{r})-$

$8\sqrt{105}xyQ_{22s}(\mathbf{r}) - 8\sqrt{14}xQ_{33c}(\mathbf{r}) + 8\sqrt{14}yQ_{33s}(\mathbf{r}))$

$\mathbf{Q_{44s}}(\mathbf{r}') = Q_{44s}(\mathbf{r})+\frac{1}{2}(\sqrt{35}x^3yQ_{00}(\mathbf{r})-\sqrt{35}xy^3Q_{00}(\mathbf{r})-3\sqrt{35}x^2yQ_{11c}(\mathbf{r})+\sqrt{35}y^3Q_{11c}(\mathbf{r})-$

$\sqrt{35}x^3Q_{11s}(\mathbf{r})+3\sqrt{35}xy^2Q_{11s}(\mathbf{r})+2\sqrt{105}xyQ_{22c}(\mathbf{r})+\sqrt{105}x^2Q_{22s}(\mathbf{r})-\sqrt{105}y^2Q_{22s}(\mathbf{r})-$

$2\sqrt{14}yQ_{33c}(\mathbf{r}) - 2\sqrt{14}xQ_{33s}(\mathbf{r}))$

4.2.2 Translation of polarizabilities

Polarizabilities, which describe how the electron density changes in a non-uniform electric field, can also be expressed in terms of multipole moment operators using the standard sum-over-states form from Rayleigh-Schrodinger perturbation theory. [159]

$$\alpha_{l\kappa,l'\kappa'} = \sum_n{}' \frac{\langle 0|\,\hat{Q}_{l\kappa}\,|n\rangle\,\langle n|\,\hat{Q}_{l'\kappa'}\,|0\rangle + \langle 0|\,\hat{Q}_{l'\kappa'}\,|n\rangle\,\langle n|\,\hat{Q}_{l\kappa}\,|0\rangle}{W_n - W_0} \tag{4.7}$$

Here, the polarizability, $\alpha_{l\kappa,l'\kappa'}$, is a function of the multipole moments, $Q_{l\kappa}$ and $Q_{l'\kappa'}$. The translated polarizability, $\alpha'_{l\kappa,l'\kappa'}$, then, is a function of the translated multipoles $Q'_{l\kappa}, Q'_{l'\kappa'}$. First, translated multipoles are expressed in terms of the original multipoles, using the equations above. Then, the terms are grouped so as to be expressions of the original polarizabilities, resulting in the polarizability translation in terms of the original polarizability.

For example, the translations for $Q_{10}(\mathbf{r'})$ and $Q_{11c}(\mathbf{r'})$ are given below:

$Q_{10}(\mathbf{r'}) = Q_{10}(\mathbf{r}) - zQ_{00}(\mathbf{r})$

$Q_{11c}(\mathbf{r'}) = Q_{11c}(\mathbf{r}) - xQ_{00}(\mathbf{r})$

We then substitute these into the equation for polarizability, such that $\alpha'_{10,11c} =$

$\sum_{n}{}' \frac{<0|(\hat{Q}_{10}(\mathbf{r})-z\hat{Q}_{00}(\mathbf{r}))|n><n|(\hat{Q}_{11c}(\mathbf{r})-x\hat{Q}_{00}(\mathbf{r}))|0>+<0|(\hat{Q}_{11c}(\mathbf{r})-x\hat{Q}_{00}(\mathbf{r}))|n><n|(\hat{Q}_{10}(\mathbf{r})-z\hat{Q}_{00}(\mathbf{r}))|0>}{W_n-W_0}$

This can be expanded, and because the charge Q_{00} is a constant and the states $|0>$ and $|n>$ are orthogonal, the matrix elements involving Q_{00} are 0. The terms are rearranged, so that the new polarizability is expressed in terms of original polarizabilities.

$\alpha'_{10,11c} = \sum_{n}{}' \frac{<0|\hat{Q}_{10}(\mathbf{r})|n><n|\hat{Q}_{11c}(\mathbf{r})|0>+<0|\hat{Q}_{11c}(\mathbf{r})(\mathbf{r})|n><n|\hat{Q}_{10}(\mathbf{r})|0>}{W_n-W_0} = \alpha_{10,11c}$

It can be seen that this dipole-dipole polarizability, the response of the z-component of the dipole moment (Q_{10}) due to the x-component of a dipole (Q_{11c}) has no change upon translation. The actual equations to translate multipoles of rank 0 - rank 2 are provided in Appendix C. The same method is used to translate the frequency dependent polarizabilities, except instead of just one polarizability, they must be translated for each value of the frequency.

4.3 Results

To test the performance of this approach, a molecular dynamics (MD) simulation was performed on one formamide solvated by 213 water molecules in an 18.7 Å box using the OPLS-AA (formamide) [161] and TIP3P (water) [162] force fields. The formamide and 17 water molecules lying within 4.5 Å of the formamide in a single time step were chosen as the solute region. Earlier work indicates that the AIFF is very faithful to QM for water around 4–5 Å (shorter distances require exchange contributions, which are not currently included in the AIFF). [17] Then the solvent was sampled for 2 nanoseconds around the fixed solute region using MD. Fifty well-separated solvent configurations were taken from this second simulation and used for the subsequent calculations. Periodic boundary conditions were employed during the MD configuration sampling, but for simplicity, they are not utilized in the QM/MM coarse-graining tests described below. Modification of the method to include periodic boundary conditions would be straightforward. [16]

Full HMBI calculations (RI-MP2/aug-cc-pVDZ [163] QM and AIFF MM parameters computed using the Sadlej triple-zeta basis [164]) were performed on each sampled configuration, and the interaction energy $E_{int} = E_{full} - E_{water} - E_{formamide}$ between the formamide and water was computed for each one. The average interaction over all 50 configurations was found to be -65.9 kJ/mol. Detailed structures and energetics are provided in Supporting Information.

The proposed solvation model introduces errors by (1) treating the solvent region with pure MM instead of HMBI QM/MM, (2) coarse-graining each individual solvent configuration, and (3) averaging the solvent properties before interacting them with the solute, instead of the traditional “interact then average” approach. The good

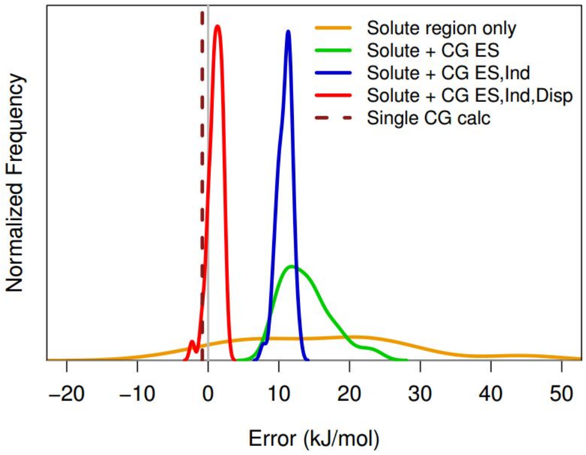


Figure 4.2: Error distributions for the coarse-grained model formamide-water interaction energy relative to conventional HMBI for each of the 50 configurations. The dotted line corresponds to the error in a single composite coarse-grained calculation.

Table 4.1: Comparison of the single coarse-grained and averaged formamide-water interaction energy calculation with the average of 50 full HMBI calculations, in kJ/mol. The data show the improvements as solvent-region electrostatics (ES), induction (Ind), and dispersion (Disp) are added to the model.

	Avg. of 50 CG calcs		1 composite CG calc	
	Contrib.	E_{int}	Contrib.	E_{int}
Solute only		-49.8		-49.8
Solute & CG ES	-2.2	-52.0	-2.2	-52.0
Solute & CG ES, Ind	-3.0	-55.0	-4.9	-56.9
Solute & CG ES, Ind, Disp	-9.8	-64.8	-9.8	-66.7
Avg. of 50 Standard HMBI calcs			-65.9	
Avg. of 50 HMBI solute & MM solvent calcs, no CG			-67.1	

quality of the AIFF makes the first source of error small. Switching to a purely AIFF MM explicit treatment of the solvent region (no coarse-graining) leads to an average energy of -67.1 kJ/mol, or an error of -1.2 kJ/mol.

To gauge the second type of errors, the coarse-graining procedure was applied separately to each of the 50 solvent configurations and compared against the full HMBI results. The coarse-graining used cells in radial increments of 1.0 Å and angular increments of $\Delta\theta = 5°$ and $\Delta\phi = 5°$. 4.2 shows how the distribution of errors changes as more physical terms are added to the model, and 4.1 summarizes the resulting average interaction energies. Complete energetics are provided as Supporting Information.

If the solvent region is neglected entirely, the error distribution is extremely broad, with a mean error and standard deviation of 16.1 ± 13.0 kJ/mol. Despite lying 4.5 Å or more from the formamide, the solvent region molecules account for 25% of the total interaction energy between formamide in water on average, and they contribute much more in many individual configurations. Including the coarse-grained solvent electrostatics for each configuration reduces the error to 13.9 ± 3.6 kJ/mol. In other words, the solvent-region electrostatics have a modest impact on the mean error, but they substantially reduce the width of the error distribution. Next, accounting for induction throughout the solute and coarse-grained solvent improves both the mean error and standard deviation to 10.8 ± 1.0 kJ/mol. This combination of solvent electrostatics and induction provides interaction energies with good precision but a large systematic error due to the neglect of dispersion.

Including the dispersion energies eliminates most of this residual error and leads to a final error of 1.0 ± 1.0 kJ/mol compared to full HMBI. The similar shape of the error distributions with and without dispersion suggests that the dispersion energy contributions are fairly uniform across the different solvent configurations. Overall, the electrostatics, induction, and dispersion terms are each important for describing the interaction between the formamide solute and the solvent, and performing the coarse-graining procedure on individual configurations introduces only small errors into the interaction energies.

Finally, we examine the degree to which one can obtain the same results by averaging the coarse-grained solvent properties over all 50 configurations and then performing a single, composite coarse-grained QM/MM calculation. Doing so gives an interaction energy of -66.7 kJ/mol (4.1), which differs from the explicit HMBI average over 50 configurations by only 0.8 kJ/mol (1%), and it can be obtained with orders of magnitude less computational time. Furthermore, previous work on molecular crystals has demonstrated that HMBI can predict molecular crystal lattice energies to within a couple kJ/mol of full QM results or experiment. [16] Therefore, the coarse-grained results here are also likely to be close to what one would obtain by performing 50 full MP2 calculations on this entire 214-molecule system.

Further insights are obtained by comparing the average of 50 coarse-grained calculations and the single composite calculation in 4.1. Because their contributions are pairwise additive, the average contribution due to the solvent electrostatics and dispersion are mathematically identical regardless of whether one "interacts then averages" or "averages then interacts." In contrast, the self-consistent induction contribution to the average interaction energy differs by 2 kJ/mol depending on the order of operations, and this accounts for the difference between the average over 50 calculations and the single composite calculation. So while the overall 0.8 kJ/mol accuracy clearly benefits from error cancellation among the three approximations described above, each of the errors involved is still only 1–2 kJ/mol.

Next, we examine the coarse-graining and averaging behavior in further detail. The individual cells used here are relatively small, particularly those closest to the solute region. A typical cell contains either zero or one water molecule(s) in a single solvent configuration. Considering 196 solvent-region molecules in 50 distinct configurations leads to 9,800 water molecules total. These molecules are distributed across 893 coarse-grained cells, with an average of 11 water molecules per cell across the 50 configurations. 4.3 plots a histogram of the number of water molecules in each coarse-grain cell. In this example, about half the cells exhibit total occupations of five or fewer, but the other half are occupied more frequently, including several with occupations of 50. Cells further

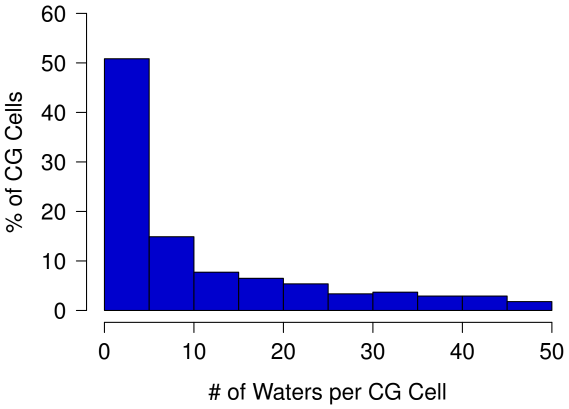


Figure 4.3: Histogram of the number of water molecules per cell in the averaged 50-configuration coarse-grained calculation.

from the solute are larger and therefore tend to have slightly higher occupations, though this gets tempered quickly as the cells reach the edges of the finite water cluster used here.

Of course, the distribution of cell occupations depends strongly on the coarse-graining cell size. Simple preliminary tests varying the size of the coarse-graining cells by integer multiples in r, θ, and ϕ produce fluctuations in the overall errors on the order of a couple kJ/mol, consistent with the magnitudes of the errors discussed above. Therefore, one can probably use larger coarse-grained cells as well, which would skew the distribution in 4.3 toward higher occupancies. Future work will examine the trade-offs between reducing the number of coarse-grained degrees of freedom using larger cells with the increased errors introduced by translating the multipoles and polarizabilities further.

A key benefit to coarse-graining explicit solvent is that captures the underlying solvent structure. From the radial distribution function, one expects that waters in the inner solvent shells will be more ordered than those further away. Indeed, the

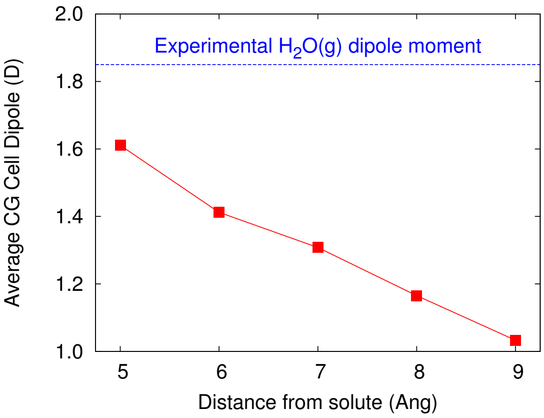


Figure 4.4: Averaged permanent cell dipole moments (in Debeye) as a function of the cell distance from the formamide solute.

average permanent cell dipole moment (i.e. not considering the induction contribution) in the inner-most shell of coarse-grained cells is 1.6 D, which is relatively close to the 1.85 D dipole of an isolated water molecule (4.4). However, the average permanent cell dipole moment decreases 40% by the time one is 9 Å away from the formamide solute molecule. The solvent shell structure restricts the orientational freedom of the inner water molecules, leading to larger average cell dipole moments, but more facile rotational averaging in the cells further from the solute leads to smaller average cell dipole moments.

4.4 Conclusions and future work

In conclusion, we have demonstrated a new approach for constructing an implicit solvent model in QM/MM calculations through coarse-graining and averaging over many solvent configurations. This coarse-graining procedure, which is based on the fact that a given multipole/polarizability can be expanded in terms of a set of different multipoles/polarizabilities centered on a new point in space, introduces only small errors in the solute-solvent interaction energies for both individual and averaged solvent configurations. By decoupling the MM solvent sampling from QM/MM solute region, it drastically reduces the computational costs associated with sampling the solvent degrees of freedom.

Many issues must be addressed to transform the model proposed here into a practical procedure. For example, to perform the configuration averaging, one might use a dual-sampling strategy in which a variety of configurations of the solute region are first sampled. Then, for each solute-region configuration, one would freeze the solute region and sample the solvent. The AIFF force field parameters are relatively expensive to generate, but they can be pre-computed, tabulated, and interpolated easily. Future work will investigate how the results depend on the size of the coarse-graining grid cells, the solute region size, and to what extent a simpler isotropic treatment of dispersion can be used instead of the anisotropic one used here. Overall, the high-quality results obtained here suggest that this sort of approach could prove fruitful for accurately modeling solution-phase chemistry with much lower computational costs.

Chapter 5

Conclusions

Our modeling of LBHBs demonstrates how complimentary theoretical calculations can be to experimental work. Our calculations helped assign the key stretching modes in IRMPD spectra, and predicted whether the ZPE was above the proton-transfer barrier, as expected for a low-barrier hydrogen bond. Also, in another practical collaboration with experimentalists, we showed how solvation affected IRMPD experiments on proton-bound dimers of adenine. Low levels of theory gave good results in the simple system of interest, and showed that surprisingly, the solvent was determining the structure of the proton-bound dimer investigated in the gas phase electrospray experiment.

Most importantly, we developed a new, coarse-grained model that effectively and affordably captures solvent structure and solute-solvent energetics. This model was demonstrated on a small molecule system, solvated in explicit water. The results were exceptional, thanks to a coarse graining scheme that uses a rigorous mathematical algorithm to translate and average the properties of each solvent atom, including multipoles, polarizabilities, and dispersion. Since we explore the configuration space with inexpensive molecular dynamics, then do a single high-level calculation that includes the explicit solute region and the coarse grained solvent, this method is also relatively affordable.

Bibliography

[1] Gregory J. O. Beran, Eric L. Chronister, Luke L. Daemen, Aaron R. Moehlig, Leonard J. Mueller, Jos Oomens, Andrew Rice, David R. Santiago-Dieppa, Fook S. Tham, Kelly Theel, Sepideh Yaghmaei, and Thomas Hellman Morton. Vibrations of a chelated proton in a protonated tertiary diamine. *Phys. Chem. Chem. Phys.*, 13:20380–20392, 2011.

[2] Khadijeh Rajabi, Kelly Theel, Elizabeth A. L. Gillis, Gregory Beran, and Travis D. Fridgen. The structure of the protonated adenine dimer by infrared multiple photon dissociation spectroscopy and electronic structure calculations. *J. Phys. Chem. A*, 113(28):8099–8107, 2009.

[3] W. Wallace Cleland, Perry A. Frey, and John A. Gerlt. The low barrier hydrogen bond in enzymatic catalysis. *J. Biol. Chem.*, 273(40):25529–25532, 1998.

[4] Aaron E Engelhart, Thomas Hellman Morton, and Nicholas V Hud. Evidence of strong hydrogen bonding by 8-aminoguanine. *J. Chem. Soc., Chem. Commun.*, (6):647–649, 2009.

[5] Frank Jensen. *Introduction to Computational Chemistry.* Wiley, West Sussex, 2007.

[6] D. J. Cram and J. E. McCarty. Studies in sterochemistry. XXIV. The preparation and determination of configuration of the isomers of 2-amino-3-phenylbutane, and the steric course of the amine oxide pyrolysis reaction in this system. *J. Am. Chem. Soc.*, 76:5740–5745, 1954.

[7] D. J. Cram and M. R. V. Sahyun. Room temperature Wolff-Kishner reduction and Cope elimination reactions. *J. Am. Chem. Soc.*, 84:1734–1735, 1962.

[8] M. R. V. Sahyun and D. J. Cram. Studies in stereochemistry. XXXV. Mechanism of E_i reaction of amine oxides. *J. Am. Chem. Soc.*, 85:1263–1268, 1963.

[9] Orlando Acevedo and William L Jorgensen. Cope elimination: elucidation of solvent effects from QM/MM simulations. *J. Am. Chem. Soc.*, 128(18):6141–6, May 2006.

[10] M L Sanchez, M A Aguilar, and F J Olivares. Study of solvent effects by means of averaged solvent electrostatic dynamics data. *J. Comp. Chem.*, 18(3):313–322, 1997.

[11] M. L. Sanchez, M. E. Martin, M. A. Aguilar, and F. J. Olivares del Valle. Solvent effects by means of averaged solvent electrostatic potentials: Coupled method. *J. Comp. Chem.*, 21(9):705–715, July 2000.

[12] M E Mart, M A Aguilar, and S Chalmet. A comparative study of two QM/MM methods testing the validity of the mean-field approximation. *Chem. Phys. Lett.*, 344:107–112, 2001.

[13] Zongrong Xu and Spiridoula Matsika. Combined multireference configuration interaction/molecular dynamics approach for calculating solvatochromic shifts: Application to the $n_O \rightarrow \pi^*$ electronic transition of formaldehyde. *J. Phys. Chem. A*, 110:12035–12043, 2006.

[14] K. Coutinho, H.C. Georg, T.L. Fonseca, V. Ludwig, and S. Canuto. An efficient statistically converged average configuration for solvent effects. *Chem. Phys. Lett.*, 437(1-3):148–152, March 2007.

[15] G. J. O. Beran. Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. *J. Chem. Phys.*, 130:164115, 2009.

[16] S. Wen and G. J. O. Beran. Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly *ab initio* force-field parameterization. *J. Chem. Theory Comput.*, 7:3733–3742, 2011.

[17] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.*, 14:7578–7590, 2012.

[18] A. Sebetci and G. J. O. Beran. Spatially homogeneous QM/MM for systems of interacting molecules with on-the-fly *ab initio* force-field parameterization. *J. Chem. Theory Comput.*, 6:155–167, 2010.

[19] Eugen Merzbacher. *Quantum Mechanics.* Wiley, New York, 1961.

[20] Christer E. Nordman and William N. Lipscomb. Isotope effect and thermal expansion coefficients of crystals. *J. Chem. Phys.*, 21(11):2077–2077, 1953.

[21] R. L. Somorjai and D. F. Hornig. Double-minimum potentials in hydrogen-bonded solids. *J. Chem. Phys.*, 36(8):1980–1987, 1962.

[22] N. D. Sokolov and V. A. Savel'ev. Dynamics of the hydrogen bond: two-dimensional model and isotope effects. *Chem. Phys.*, 22(3):383 – 399, 1977.

[23] Maurice M. Kreevoy and Tai Ming Liang. Structures and isotopic fractionation factors of complexes, $A_1HA_2^{-1}$. *J. Am. Chem. Soc.*, 102(10):3315–3322, 1980.

[24] Sepideh Yaghmaei, Sevana Khodagholian, J. Michael Kaiser, Fook S. Tham, Leonard J. Mueller, and Thomas Hellman Morton. Chelation of a proton by an aliphatic tertiary diamine. *J. Am. Chem. Soc.*, 130(25):7836–7838, 2008. The calculated NN distance derived from the B2-P3LYP/cc-pVTZ//B3LYP/6-31G(d, p) 2-dimensional surface described in this communication is R = 2.655 Å.

[25] J. Klimkiewicz, L. Stefaniak, B. Brzeziski, E. Grech, S. Kuroki, I. Ando, and G.A. Webb. A ^{13}C and ^{15}N NMR investigation of two proton sponges in solution and in the solid state. *J. Mol. Struct.*, 323:193–195, 1994.

[26] Yevhen Horbatenko and Sergei F. Vyboishchikov. Hydrogen motion in proton sponge cations: A theoretical study. *Chem. Phys. Chem.*, 12(6):1118–1129, 2011.

[27] Z. Tian, A. Fattahi, L. Lis, and S. R. Kass. Neutral intramolecular hydrogen-bonded bases. *Croat. Chem. Acta*, 82(1), 2009.

[28] Michael D. Pluth, Robert G. Bergman, and Kenneth N. Raymond. Proton-mediated chemistry and catalysis in a self-assembled supramolecular host. *Acc. Chem. Res.*, 42(10):1650–1659, 2009.

[29] Tobias Benighaus, Robert A. DiStasio, Rohini C. Lochan, Jeng-Da Chai, and Martin Head-Gordon. Semiempirical double-hybrid density functional with improved description of long-range correlation. *J. Phys. Chem. A*, 112(12):2702–2712, 2008.

[30] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J.Chem. Phys.*, 90(2):1007, 1989.

[31] Sándor Kristyán and Péter Pulay. Can (semi)local density functional theory account for the London dispersion forces? *Chem. Phys. Lett.*, 229(3):175–180, 1994.

[32] Tobias Schwabe and Stefan Grimme. Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability. *Phys. Chem. Chem. Phys.*, 9(26):3397–3406, 2007.

[33] Yihan Shao, Laszlo Fusti Molnar, Yousung Jung, Jorg Kussmann, Christian Ochsenfeld, Shawn T. Brown, Andrew T.B. Gilbert, Lyudmila V. Slipchenko, Sergey V. Levchenko, Darragh P. O'Neill, Robert A. DiStasio Jr, Rohini C. Lochan, Tao Wang, Gregory J.O. Beran, Nicholas A. Besley, John M. Herbert, Ching Yeh Lin, Troy Van Voorhis, Siu Hung Chien, Alex Sodt, Ryan P. Steele, Vitaly A. Rassolov, Paul E. Maslen, Prakashan P. Korambath, Ross D. Adamson, Brian Austin, Jon Baker, Edward F. C. Byrd, Holger Dachsel, Robert J. Doerksen, Andreas Dreuw, Barry D. Dunietz, Anthony D. Dutoi, Thomas R. Furlani, Steven R. Gwaltney, Andreas Heyden, So Hirata, Chao-Ping Hsu, Gary Kedziora, Rustam Z. Khalliulin, Phil Klunzinger, Aaron M. Lee, Michael S. Lee, WanZhen Liang, Itay Lotan, Nikhil Nair, Baron Peters, Emil I. Proynov, Piotr A. Pieniazek, Young Min Rhee, Jim Ritchie, Edina Rosta, C. David Sherrill, Andrew C. Simmonett, Joseph E. Subotnik, H. Lee Woodcock III, Weimin Zhang, Alexis T. Bell, Arup K. Chakraborty, Daniel M. Chipman, Frerich J. Keil, Arieh Warshel, Warren J. Hehre, Henry F. Schaefer III, Jing Kong, Anna I. Krylov, Peter M. W. Gill, and Martin Head-Gordon. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.*, 8:3172–3191, 2006.

[34] Karin Eichkorn, Oliver Treutler, Holger Öhm, Marco Häser, and Reinhart Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 240(4):283–290, 1995.

[35] Florian Weigend, Andreas Kohn, and Christof Hattig. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.*, 116(8):3175–3183, 2002.

[36] Ryan P. Steele, Robert A. DiStasio, Jr., Yihan Shao, Jing Kong, and Martin Head-Gordon. Dual-basis second-order Møller-Plesset perturbation theory: A reduced-cost reference for correlation calculations. *J. Chem. Phys.*, 125(7):074108, 2006.

[37] D. I. Bower and W. F. Maddams. *Vibrational Spectroscopy of Polymers.* Cambridge University Press, Cambridge, 1992.

[38] Vincenzo Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.*, 122(1), 2005.

[39] Vernon M. Ingram. The identification of peptide end-groups as dimethylamino acids. *J. Biol. Chem.*, 202(1):193–201, 1953.

[40] Edward C. Lis, Rebecca J. Salomon, Michal Sabat, William H. Myers, and W. Dean Harman. Synthesis of 1-oxadecalins from anisole promoted by tungsten. *J. Am. Chem. Soc.*, 130(37):12472–12476, 2008.

[41] Jos Oomens, Elfi Kraka, Michael K. Nguyen, and Thomas Hellman Morton. Structure, vibrational spectra, and unimolecular dissociation of gaseous 1-fluoro-1-phenethyl cations. *J. Phys. Chem. A*, 112(43):10774–10783, 2008.

[42] J. Lemaire, P. Boissel, M. Heninger, G. Mauclaire, G. Bellec, H. Mestdagh, A. Simon, S. Le Caer, J. M. Ortega, F. Glotin, and P. Maitre. Gas phase infrared spectroscopy of selectively prepared ions. *Phys. Rev. Lett.*, 89:273002, 2002.

[43] Barbara Chiavarino, Maria E. Crestoni, Simonetta Fornarini, Joel Lemaire, Luke Mac Aleese, and Philippe Matre. Infrared absorption features of gaseous isopropyl carbocations. *Chem. Phys. Chem.*, 5(11):1679–1685, 2004.

[44] Jose J. Valle, John R. Eyler, Jos Oomens, David T. Moore, A. F. G. van der Meer, Gert von Helden, Gerard Meijer, Christopher L. Hendrickson, Alan G. Marshall, and Gregory T. Blakney. Free electron laser-Fourier transform ion cyclotron resonance mass spectrometry facility for obtaining infrared multiphoton dissociation spectra of gaseous ions. *Review of Scientific Instruments*, 76(2):023103, 2005.

[45] Nick C. Polfer and Jos Oomens. Reaction products in mass spectrometry elucidated with infrared spectroscopy. *Phys. Chem. Chem. Phys.*, 9:3804–3817, 2007.

[46] Jos Oomens, Jeffrey D. Steill, and Britta Redlich. Gas-phase IR spectroscopy of deprotonated amino acids. *J. Am. Chem. Soc.*, 131(12):4310–4319, 2009.

[47] A.D. Taylor, E.J. Wood, J.A. Goldstone, and J. Eckert. Lineshape analysis and filter difference method for a high intensity time-of-flight inelastic neutron scattering spectrometer. *Nucl. Instrum. Methods*, 221(2):408–418, 1984.

[48] Charles L. Perrin and Brian K. Ohta. Symmetry of N-H-N hydrogen bonds in 1,8-bis(dimethylamino)naphthalene·H^+ and 2,7-dimethoxy-1,8-bis(dimethylamino)naphthalene·H^+. *J. Am. Chem. Soc.*, 123(27):6520–6526, 2001.

[49] Charles L. Perrin. Are short, low-barrier hydrogen bonds unusually strong? *Acc. Chem. Res.*, 43(12):1550–1557, 2010.

[50] Irena Majerz and Ivar Olovsson. Comparison of the proton-transfer path in hydrogen bonds from theoretical potential-energy surfaces and the concept of conservation of bond order. II. $(\mathrm{N—H}\cdots\mathrm{N})^+$ hydrogen bonds. *Acta Crystallogr., Sect. B*, 63(4):650–662, 2007.

[51] Irena Majerz and Ivar Olovsson. The shape of the potential energy curves for NHN^+ hydrogen bonds and the influence of non-linearity. *Phys. Chem. Chem. Phys.*, 10:3043–3051, 2008.

[52] Michael Meot-Ner. Ion thermochemistry of low-volatility compounds in the gas phase. 2. Intrinsic basicities and hydrogen-bonded dimers of nitrogen heterocyclics and nucleic bases. *J. Am. Chem. Soc.*, 101(9):2396–2403, 1979.

[53] Michael Meot-Ner (Mautner). The ionic hydrogen bond. *Chem. Rev.*, 105(1):213–284, 2005.

[54] Khadijeh Rajabi, Kelly Theel, Elizabeth A. L. Gillis, Gregory Beran, and Travis D. Fridgen. The structure of the protonated adenine dimer by infrared multiple photon dissociation spectroscopy and electronic structure calculations. *J. Phys. Chem. A*, 113(28):8099–8107, 2009.

[55] Nicholas V. Hud and Thomas Hellman Morton. DFT energy surfaces for aminopurine homodimers and their conjugate acid ions. *J. Phys. Chem. A*, 111(17):3369–3377, 2007.

[56] S. Yaghmaei. *Ph.D. thesis*, University of California, Riverside, 2008.

[57] S. Bratoz and D. Hadzi. Infrared spectra of molecules with hydrogen bonds. *J. Chem. Phys.*, 27(5):991–997, 1957.

[58] Proceedings of the chemical society, vol. 30, no. 436. *Proc. Chem. Soc., London*, 30:301–310, 1963.

[59] Friedrich Klages, John E. Gordon, and Herbert A. Jung. NMR- und IR-spektren von oxoniumsalzen. *Chem. Ber.*, 98(11):3748–3756, 1965.

[60] M. Calleri and J. C. Speakman. Crystal structure of bis-(1-methyl-2-quinolone) hydrogen hexafluoroarsenate(v): a structure of pseudo-A-type involving a very short hydrogen bond. *J. Chem. Soc. A*, 1969.

[61] L. Angeloni, M.P. Marzocchi, S. Detoni, D. Hadi, B. Orel, and G. Sbrana. Infrared and Raman spectra of potassium hydrogen malonate single crystals. *Spectrochim. Acta A*, 34(3):253–261, 1978.

[62] Roger W. Alder, Richard E. Moss, and Richard B. Sessions. Intrabridgehead hydrogen-bonded ions: spectroscopic characteristics and the question of single vs. double minimum potentials. *J. Chem. Soc., Chem. Commun.*, 0:1000–1002, 1983.

[63] Evgenii S. Stoyanov, Irina V. Stoyanova, and Christopher A. Reed. IR spectroscopic properties of $\mathrm{H(MeOH)}_n^+$ clusters in the liquid phase: Evidence for a proton wire. *Chem. Eur. J.*, 14(12):3596–3604.

[64] Mikhail V. Vener, Alexey V. Manaev, Dusan Hadzi, and Vladimir G. Tsirelson. DFT study of proton dynamics in the potassium hydrogen maleate crystal: the infrared versus the inelastic neutron scattering spectra. *Hoppe-Seyler's Z. Physiol. Chem.*, 222:1349–1358.

[65] Jean Louis Leroy, Kalle Gehring, Abdelali Kettani, and Maurice Gueron. Acid multimers of oligodeoxycytidine strands: Stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry*, 32(23):6019–6031, 1993.

[66] Christof-G. Hoelger and Hans-Heinrich Limbach. Localization of hydrons in hydrogen bonds using dipolar solid state NMR spectroscopy. *J. Phys. Chem.*, 98(45):11803–11810, 1994.

[67] Hans Benedict, Hans-Heinrich Limbach, Martin Wehlan, Wolf-Peter Fehlhammer, Nicolai S. Golubev, and Rudolf Janoschek. Solid state ^{15}N NMR and theoretical studies of primary and secondary geometric H/D isotope effects on low-barrier nhn-hydrogen bonds. *J. Am. Chem. Soc.*, 120(12):2939–2950, 1998.

[68] Concepción López, Philippe Lorente, Rosa María Claramunt, Javier Marín, Concepcion Foces-Foces, Antonio L. Llamas-Saiz, Jose Elguero, and Hans-Heinrich Limbach. Localization of hydrogen bond deuterons in proton sponges by dipolar solid state ^{15}N NMR spectroscopy. *Ber. Bunsen-Ges. Phys. Chem.*, 102(3):414–418, 1998.

[69] V. Rani Parvathy, Sukesh R. Bhaumik, Kandala V. R. Chary, Girjesh Govil, Keliang Liu, Frank B. Howard, and H. Todd Miles. NMR structure of a parallel-stranded DNA duplex at atomic resolution. *Nucleic Acids Res.*, 30(7):1500–1511, 2002.

[70] Aaron E. Engelhart, Thomas Hellman Morton, and Nicholas V. Hud. Evidence of strong hydrogen bonding by 8-aminoguanine. *Chem. Commun.*, 0:647–649, 2009.

[71] David T. Moore, Jos Oomens, Lex van der Meer, Gert von Helden, Gerard Meijer, Jose Valle, Alan G. Marshall, and John R. Eyler. Probing the vibrations of shared, OH^+O-bound protons in the gas phase. *Chem. Phys. Chem.*, 5(5):740–743, 2004.

[72] Jeffrey M. Headrick, Eric G. Diken, Richard S. Walters, Nathan I. Hammer, Richard A. Christie, Jun Cui, Evgeniy M. Myshakin, Michael A. Duncan, Mark A. Johnson, and Kenneth D. Jordan. Spectral signatures of hydrated proton vibrations in water clusters. *Science*, 308(5729):1765–1769, 2005.

[73] Nathan I. Hammer, Eric G. Diken, Joseph R. Roscioli, Mark A. Johnson, Evgeniy M. Myshakin, Kenneth D. Jordan, Anne B. McCoy, Xinchuan Huang, Joel M. Bowman, and Stuart Carter. The vibrational predissociation spectra of the $H_5O_2^+\cdot RG_n$ (RG = Ar, Ne) clusters: Correlation of the solvent perturbations in the free OH and shared proton transitions of the Zundel ion. *J. Chem. Phys.*, 122(24):244301, 2005.

[74] Nicola Solcà and Otto Dopfer. Hydrogen-bonded networks in ethanol proton wires: IR spectra of $(EtOH)_qH^+$-L_n clusters ($L = Ar/N_2$, $q \leq 4$, $n \leq 5$). *J. Phys. Chem. A*, 109(28):6174–6186, 2005.

[75] Knut R. Asmis, Yonggang Yang, Gabriele Santambrogio, Mathias Brümmer, J. Robert Roscioli, Laura R. McCunn, Mark A. Johnson, and Oliver Kühn. Gas-phase infrared spectroscopy and multidimensional quantum calculations of the protonated ammonia dimer $N_2H_7^+$. *Angew. Chem. Int. Ed.*, 46(45):8691–8694, 2007.

[76] Laura R. McCunn, Joseph R. Roscioli, Mark A. Johnson, and Anne B. McCoy. An H/D isotopic substitution study of the $H_5O_2^+\cdot$Ar vibrational predissociation spectra: Exploring the putative role of Fermi resonances in the bridging proton fundamentals. *J. Phys. Chem. B*, 112(2):321–327, 2008.

[77] Y. Yang, O. Kuhn, G. Santambrogio, D. J. Goebbert, and K. R. Asmis. Vibrational signatures of hydrogen bonding in the protonated ammonia clusters $NH_4^+(NH_3)_{1-4}$. *J. Chem. Phys.*, 129(22):224302, 2008.

[78] P. Garcia-Fernandez, L. Garcia-Canales, J. M. Garcia-Lastra, J. Junquera, M. Moreno, and J. A. Aramburu. Pseudo-Jahn-Teller origin of the low barrier hydrogen bond in $N_2H_7^+$. *J. Chem. Phys.*, 129(12), 2008.

[79] Yonggang Yang and Oliver K˙ H/D isotope effects on the geometry and infrared spectrum of the protonated ammonia dimer. *Chem. Phys. Lett.*, 505(1-3):1–4, 2011.

[80] Juan Miguel Lopez, Ferdinand Mannle, Iwona Wawer, Gerd Buntkowsky, and Hans-Heinrich Limbach. NMR studies of double proton transfer in hydrogen bonded cyclic N,N'-diarylformamidine dimers: conformational control, kinetic HH/HD/DD isotope effects and tunneling. *Phys. Chem. Chem. Phys.*, 9:4498–4513, 2007.

[81] Oriol Vendrell and Hans-Dieter Meyer. A proton between two waters: insight from full-dimensional quantum-dynamics simulations of the H_2O-H-OH_2^+ cluster. *Phys. Chem. Chem. Phys.*, 10:4692–4703, 2008.

[82] Oriol Vendrell, Fabien Gatti, and Hans-Dieter Meyer. Strong isotope effects in the infrared spectrum of the Zundel cation. *Angew. Chem. Int. Ed.*, 48(2):352–355, 2009.

[83] Jerry L. Atwood, Leonard J. Barbour, and Agoston Jerga. Supramolecular stabilization of $N_2H_7^+$. *J. Am. Chem. Soc.*, 124(10):2122–2123, 2002.

[84] Jos Oomens, Aaron R. Moehlig, and Thomas Hellman Morton. Infrared multiple photon dissociation (IRMPD) spectroscopy of the proton-bound dimer of 1-methylcytosine in the gas phase. *J. Phys. Chem. Lett.*, 1(19):2891–2897, 2010.

[85] Oscar Hur, Catherine Leja, and Michael F. Dunn. Evidence of a low-barrier hydrogen bond in the tryptophan synthase catalytic mechanism. *Biochemistry*, 35(23):7378–7386, 1996.

[86] Gudrun R. Stranzl, Karl Gruber, Georg Steinkellner, Klaus Zangger, Helmut Schwab, and Christoph Kratky. Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from Hevea brasiliensis explains a large pKa shift of the catalytic base induced by the reaction intermediate. *J. Biol. Chem.*, 279(5):3699–3707, 2004.

[87] Sathyapriya Rajagopal and Saraswathi Vishveshwara. Short hydrogen bonds in proteins. *FEBS Journal*, 272(8):1819–1832, 2005.

[88] Nadrian C. Seeman. DNA in a material world. *Nature*, 421:427–431, 2003.

[89] Chengde Mao, Thomas H. LaBean, John H. Reif, and Nadrian C. Seeman. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 407:493–496, 2000.

[90] Ya-Jun Cheng, Zhe-Ming Wang, Chun-Sheng Liaoa, and Chun-Hua Yan. Novel sheet-like supramolecular architectures constructed from infinite hydrogen-bonded, protonated adeninewaterhalide and polyiodide ribbons. *New. J. Chem.*, 26:1360–1364, 2002.

[91] Jiande Gu and Jerzy Leszczynski. A DFT study of the water-assisted intramolecular proton transfer in the tautomers of adenine. *J. Phys. Chem. A*, 103:2744–2750, 1999.

[92] Oleg S. Sukhanov, Oleg V. Shishkin, Leonid Gorb, Yeugen Podolyan, and Jerzy Leszczynski. Molecular structure and hydrogen bonding in polyhydrated complexes of adenine: a DFT study. *J. Phys. Chem. B*, 107:2846, 2003.

[93] Janet E. Del Bene. Molecular orbital study of the protonation of DNA bases. *J. Phys. Chem.*, 87:367–371, 1983.

[94] Asit K.Chandra, Minh Tho Nguyen, Tadafumi Uchimaru, and Thrse Zeegers-Huyskens. Protonation and deprotonation enthalpies of guanine and adenine and implications for the structure and energy of their complexes with water: comparison with uracil, thymine, and cytosine. *J. Phys. Chem. A*, 103:8853–8860, 1999.

[95] Benedetta Mennucci, Alessandro Toniolo, and Jacopo Tomasi. Theoretical study of the photophysics of adenine in solution: tautomerism, deactivation mechanisms, and comparison with the 2-aminopurine fluorescent isomer. *J. Phys. Chem. A*, 105:4749–4757, 2001.

[96] Bing Ping Liu, Yang Jun Ding, and Xiang Ai Yuan. First-principle applications on protonated adenine dimers with NH···N intermolecular coupling. *THEOCHEM*, 848:47–55, 2008.

[97] Hye Sun Park, Sang Hwan Nam, Jae Kyu Song, Seung Min Park, and Seol Ryu. Density functional study of intradimer proton transfers in hydrated adenine dimer ions, $A^{2+}(H_2O)_n$ (n = 02). *J. Phys. Chem. A*, 112:9023–9030, 2008.

[98] Vassil B. Delchev, Ivan G. Shterev, Hans Mikosch, and Nikolay T. Kochev. Investigation of the intermolecular proton transfer in the supersystems adenine-methanol/ethanol/i-propanol: MP2 and DFT levels study. *J. Mol. Model*, 13:1001–1008, 2007.

[99] Feng Wang, Matthew T. Downton, and Nader Kidwani. Adenine tautomer electronic structural signatures studied using dual space analysis. *J. Theor. Comput. Chem.*, 4:247, 2005.

[100] Wiktor Zierkiewicz, Ludwik Komorowski, and Danuta Michalska. The amino group in adenine: MP2 and CCSD(T) complete basis set limit calculations of the planarization barrier and DFT/B3LYP study of the anharmonic frequencies of adenine. *J. Phys. Chem. B.*, 112:16734–16740, 2008.

[101] Bárbara Herrera and Alejandro Toro-Labbé. The role of reaction force and chemical potential in characterizing the mechanism of double proton transfer in the adenineuracil complex. *J. Phys. Chem. A*, 111:5921–5926, 2007.

[102] Jan Florián and Jerzy Leszczyński. Spontaneous DNA mutations induced by proton transfer in the guaninecytosine base pairs: an energetic perspective. *J. Am. Chem. Soc.*, 118:3010–3017, 1996.

[103] Per-Olov Löwdin. Quantum genetics and the aperiodic solid: Some aspects on the biological problems of heredity, mutations, aging, and tumors in view of the quantum theory of the DNA molecule. *Adv. Quantum Chem.*, 2:213–360, 1965.

[104] Michael D. Topal and Jacques R. Fresco. Complementary base pairing and the origin of substitution mutations. *Nature*, 263:285–289, 1976.

[105] Michal Hanus, Martin Kabeláč, Jaroslav Rejnek, Filip Ryjáček, and Pavel Hobza. Correlated ab initio study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment, and in aqueous solution. Part 3. Adenine. *J. Phys. Chem. B*, 108:2087–2097, 2004.

[106] C. Fonseca Guerra, F. M. Bickelhaupt, S. Saha, and F. Wang. Adenine tautomers: Relative stabilities, ionization energies, and mismatch with cytosine. *J. Phys. Chem. A*, 110(11):4012–4020, 2006.

[107] František Turecček and Xiaohong Chen. Protonated adenine: Tautomers, solvated clusters, and dissociation mechanisms. *J. Am. Soc. Mass Spectrom.*, 16:1713–1726, 2005.

[108] R. L. Bell. *The Tunnel Effect in Chemistry.* Chapman and Hall, New York, 1980.

[109] Larry A. Curtiss, David L. Drapcho, and John A. Pople. An ab initio molecular orbital study of the potential energy surface of the N_2H N_2 + H system. *Chem. Phys. Lett.*, 103(6):437–442, 1984.

[110] Maciej J. Nowak, Leszek Lapinski, Jzef S. Kwiatkowski, and Jerzy Leszczyski. Molecular structure and infrared spectra of adenine. Experimental matrix isolation and density functional theory study of adenine 15N isotopomers. *J. Phys. Chem.*, 100(9):3527–3534, 1996.

[111] R. Bonaccorsi, A. Pullman, E. Scrocco, and J. Tomasi. The molecular electrostatic potentials for the nucleic acid bases: Adenine, thymine, and cytosine. *Theor. Chim. Acta*, 24:51–60, 1972.

[112] Nino Russo, Marirosa Toscano, Andr Grand, and Franck Jolibois. Protonation of thymine, cytosine, adenine, and guanine DNA nucleic acid bases: Theoretical investigation into the framework of density functional theory. *J. Comp. Chem.*, 19:989–1000, 1998.

[113] Yevgeniy Podolyan, Leonid Gorb, and Jerzy Leszczynski. Protonation of nucleic acid bases. A comprehensive post-Hartree Fock study of the energetics and proton affinities. *J. Phys. Chem. A*, 104(31):7346–7352, 2000.

[114] Michael Meot-Ner. Ion thermochemistry of low-volatility compounds in the gas phase. 2. Intrinsic basicities and hydrogen-bonded dimers of nitrogen heterocyclics and nucleic bases. *J. Am. Chem. Soc.*, 101(9):2396–2403, 1979.

[115] Jill K. Wolken and František Tureček. Proton affinity of uracil. A computational study of protonation sites. *J. Am. Soc. Mass Spectrom.*, 11(12):1065–1071, 2000.

[116] Francesco Greco, Angelo Liguori, Giovanni Sindona, and Nicola Uccella. Gas-phase proton affinity of deoxyribonucleosides and related nucleobases by fast atom bombardment tandem mass spectrometry. *J. Am. Chem. Soc.*, 112(25):9092–9096, 1990.

[117] Gopinatha Suresh Kumar, Suman Das, Kakali Bhadra, and Motilal Maiti. Protonated forms of poly[d(G-C)] and poly(dG)·poly(dC) and their interaction with berberine. *Bioorg. Med. Chem.*, 11(23):4861–4870, 2003.

[118] Fu Ming Chen. Base protonation facilitates B-Z interconversions of poly(dG-dC).cntdot.poly(dG-dC). *Biochemistry*, 23(25):6159–6165, 1984.

[119] Jens Volker and Horst H. Klump. Electrostatic effects in DNA triple helixes. *Biochemistry*, 33(45):13502–13508, 1994.

[120] W. David Wilson, Farial A. Tanious, Shaikh Mizan, Shijie Yao, Alexander S. Kiselyov, Gerald Zon, and Lucjan Strekowski. DNA triple-helix specific intercalators as antigene enhancers: Unfused aromatic cations. *Biochemistry*, 32(40):10614–10621, 1993.

[121] M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, and M. E. Hogan. Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Science*, 241(4864):456–459, 1988.

[122] Christel Marian, Dirk Nolting, and Rainer Weinkauf. The electronic spectrum of protonated adenine: Theory and experiment. *Phys. Chem. Chem. Phys.*, 7:3306–3316, 2005.

[123] Travis D. Fridgen. Infrared consequence spectroscopy of gaseous protonated and metal ion cationized complexes. *Mass Spectrom. Rev.*, 28(4):586–607, 2009.

[124] Nick C. Polfer and Jos Oomens. Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom. Rev.*, 28(3):468–494, 2009.

[125] John R. Eyler. Infrared multiple photon dissociation spectroscopy of ions in penning traps. *Mass Spectrom. Rev.*, 28(3):448–467, 2009.

[126] Chad G. Atkins, Khadijeh Rajabi, Elizabeth A. L. Gillis, and Travis D. Fridgen. Infrared multiple photon dissociation spectra of proton- and sodium ion-bound glycine dimers in the N-H and O-H stretching region. *J. Phys. Chem. A*, 112(41):10220–10225, 2008.

[127] Khadijeh Rajabi and Travis D. Fridgen. Structures of aliphatic amino acid proton-bound dimers by infrared multiple photon dissociation spectroscopy in the 700-2000 cm^{-1} region. *J. Phys. Chem. A*, 112(1):23–30, 2008.

[128] Sung Hwan Yoon, Julia Chamot-Rooke, Brittany R. Perkins, Amy E. Hilderbrand, John C. Poutsma, and Vicki H. Wysocki. IRMPD spectroscopy shows that AGG forms an oxazolone $b_2{}^+$ ion. *J. Am. Chem. Soc.*, 130(52):17644–17645, 2008.

[129] Jean-Yves Salpin, Sébastien Guillaumont, Jeanine Tortajada, Luke MacAleese, Joël Lemaire, and Philippe Maitre. Infrared spectra of protonated uracil, thymine and cytosine. *Chem. Phys. Chem.*, 8(15):2235–2244, 2007.

[130] Khadijeh Rajabi, Michael L. Easterling, and Travis D. Fridgen. Solvation of electrosprayed ions in the accumulation/collision hexapole of a hybrid Q-FTMS. *J. Am. Soc. Mass Spectrom.*, 20(3):411–418, 2009.

[131] Elizabeth A. L. Gillis, Khadijeh Rajabi, and Travis D. Fridgen. Structures of hydrated Li$^+$-thymine and Li$^+$-uracil complexes by IRMPD spectroscopy in the N-H/O-H stretching region. *J. Phys. Chem. A*, 113(5):824–832, 2009.

[132] Nicholas V. Hud and Thomas Hellman Morton. DFT energy surfaces for aminopurine homodimers and their conjugate acid ions. *J. Phys. Chem. A*, 111(17):3369–3377, 2007.

[133] Joost M. Bakker, Thierry Besson, Joël Lemaire, Debora Scuderi, and Philippe Maître. Gas-phase structure of a π-allyl-palladium complex: Efficient infrared spectroscopy in a 7 T fourier transform mass spectrometer. *J. Phys. Chem. A*, 111(51):13415–13424, 2007.

[134] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. *Gaussian 03, revision B.04*. Gaussian Inc., Wallingford, CT, 2004.

[135] Stefan Grimme. Semiempirical hybrid density functional with perturbative second-order correlation. *J.Chem. Phys.*, 124(3):034108, 2006.

[136] S. F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, 19(4):553–566, 1970.

[137] B. I. Dunlap. Fitting the Coulomb potential variationally in X alpha molecular calculations. *J. Chem. Phys.*, 78(6):3140–3142, 1983.

[138] Martin Feyereisen, George Fitzgerald, and Andrew Komornicki. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.*, 208(56):359–363, 1993.

[139] Maurizio Cossi, Nadia Rega, Giovanni Scalmani, and Vincenzo Barone. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comp. Chem.*, 24(6):669–681, 2003.

[140] The stability of the N3-protonated adenine N7H tautomer is computed to be nearly the same as the N1-protonated adenine N9H tautomer, as they differ by only a fraction of a kJ/mol. Which species is more stable depends on exactly which functional and basis set is used, but for B2P3LYP/cc-pVTZ, the N1H/9H species is 0.4 kJ/mol more stable. These relative stabilities of the two tautomers agrees well with the experimental results from ref 35.

[141] C. Plutzer and K. Kleinermanns. Tautomers and electronic states of jet-cooled adenine investigated by double resonance spectroscopy. *Phys. Chem. Chem. Phys.*, 4:4877–4882, 2002.

[142] Pina Colarusso, KeQing Zhang, Bujin Guo, and Peter F. Bernath. The infrared spectra of uracil, thymine, and adenine in the gas phase. *Chem. Phys. Lett.*, 269:39–48, 1997.

[143] Wiktor Zierkiewicz, Ludwik Komorowski, Danuta Michalska, Jiri Cerny, and Pavel Hobza. The amino group in adenine: MP2 and CCSD(T) complete basis set limit calculations of the planarization barrier and DFT/B3LYP study of the anharmonic frequencies of adenine. *J. Phys. Chem. B*, 112(51):16734–16740, 2008.

[144] C. J. Cramer and D. G. Truhlar. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99:2161–2200, 1999.

[145] J. Tomasi, B. Mennucci, and R. Cammi. Quantum mechanical continuum solvation models. *Chem. Rev.*, 105:2999–3093, 2005.

[146] Jayaraman Chandrasekhar, Scott F Smith, and William L Jorgensen. Theoretical examination of the S_N2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.*, 107:154–163, 1985.

[147] Robert V Stanton, Mikael Pera, Dirk Bakowies, and Peter A Kollman. Combined ab initio and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution. *J. Am. Chem. Soc.*, 7863:3448–3457, 1998.

[148] Yingkai Zhang, Haiyan Liu, and Weitao Yang. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.*, 112:3483–3492, 2000.

[149] Nicolas Ferre and Janos Angyan. Approximate electrostatic interaction operator for QM/MM calculations. *Chem. Phys. Lett.*, 356:331–339, 2002.

[150] Thomas H Rod and Ulf Ryde. Accurate QM/MM free energy calculations of enzyme reactions: Methylation by catechol O-methyltransferase. *J. Chem. Theory Comput.*, 1:1240–1251, 2005.

[151] Johannes Kastner, Hans Martin Senn, Stephan Thiel, Nikolaj Otte, and Walter Thiel. QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: Application to an enzymatic reaction. *J. Chem. Theory Comput.*, 23:452–461, 2006.

[152] Peter Pulay and Tomasz Janowski. Efficient calculation of the energy of a molecule in an arbitrary electric field. *Int. J. Quant. Chem.*, 109:2113–2120, 2009.

[153] Tomasz Janowski, Krzysztof Wolinski, and Peter Pulay. Ultrafast quantum mechanics/molecular mechanics Monte Carlo simulations using generalized multipole polarizabilities. *Chem. Phys. Lett.*, 530:1–9, March 2012.

[154] Hiroshi Nakano and Takeshi Yamamoto. Accurate and efficient treatment of continuous solute charge density in the mean-field QM/MM free energy calculation. *J. Chem. Theory Comput.*, 9(1):188–203, January 2013.

[155] Hirofumi Sato. A modern solvation theory: quantum chemistry and statistical chemistry. *Phys. Chem. Chem. Phys.*, 15(20):7450–65, May 2013.

[156] E. A. L. Gillis, M. Demireva, K. Nanda, G. J. O. Beran, E. R. Williams, and T. D. Fridgen. Structures and energetics of electrosprayed uracil$_n$Ca^{2+} clusters (n=14-4) in the gas phase. *Phys. Chem. Chem. Phys.*, 14:3304–3315, 2012.

[157] S. Wen and G. J. O. Beran. Accidental degeneracy in crystalline aspirin: New insights from high-level ab initio calculations. *Cryst. Growth Des.*, 12:2169–2172, 2012.

[158] S. Wen and G. J. O. Beran. Crystal polymorphism in oxalyl dihydrazide: Is empirical DFT-D accurate enough? *J. Chem. Theory Comput.*, 8:2698–2705, 2012.

[159] A. J. Stone. *The Theory of Intermolecular Forces.* Clarendon Press, Oxford, 2002.

[160] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comp. Phys.*, 73:325–348, 1987.

[161] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[162] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[163] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007–1023, 1989.

[164] A. J. Sadlej. Medium-size polarized basis sets for high-level correlated calculations of molecular electronic properties. *Collect. Czech Chem. Commun.*, 53:1995–2016, 1988.

Appendix A

Multipole Moments in Spherical Tensor Formalism

Please note that the dipole moments are in a non-standard order. This list corresponds to the ordering found in the HMBI C+ code, with Q_{11c} first, Q_{11s} second, and then Q_{10}. The more standard order places Q_{10} first, then Q_{11c}, and lastly Q_{11s}. The indexing (starting with zero) maps to the multipole class multipole vector indexing. (reference Stone)

Point Charges

0. $Q_{00} = q$

Dipoles

1. $Q_{11c} = \mu_x$
2. $Q_{11s} = \mu_y$
3. $Q_{10} = \mu_z$

Quadrupoles

4. $Q_{20} = \Theta_{zz}$

5. $Q_{21c} = \frac{2}{\sqrt{3}}\Theta_{xz}$

6. $Q_{21s} = \frac{2}{\sqrt{3}}\Theta_{yz}$

7. $Q_{22c} = \frac{1}{\sqrt{3}}(\Theta_{xx} - \Theta_{yy})$

8. $Q_{22s} = \frac{2}{\sqrt{3}}\Theta_{xy}$

Octopoles

9. $Q_{30} = \Omega_{zzz}$

10. $Q_{31c} = \sqrt{\frac{3}{2}}\Omega_{xzz}$

11. $Q_{31s} = \sqrt{\frac{3}{2}}\Omega_{yzz}$

12. $Q_{32c} = \sqrt{\frac{3}{5}}(\Omega_{xxz} - \Omega_{yyz})$

13. $Q_{32s} = 2\sqrt{\frac{3}{2}}\Omega_{xyz}$

14. $Q_{33c} = \sqrt{\frac{1}{10}}(\Omega_{xxx} - 3\Omega_{yyy})$

15. $Q_{33s} = \sqrt{\frac{1}{10}}(3\Omega_{xxy} - \Omega_{yyy})$

Hexadecapoles

16. $Q_{40} = \Phi_{zzzz}$

17. $Q_{41c} = \sqrt{\frac{8}{5}}\Phi_{xzzz}$

18. $Q_{41s} = \sqrt{\frac{8}{5}}\Phi_{yzzz}$

19. $Q_{42c} = 2\sqrt{\frac{1}{5}}(\Phi_{xxzz} - \Phi_{yyzz})$

20. $Q_{42s} = 4\sqrt{\frac{1}{5}}\Phi_{xyzz}$

21. $Q_{43c} = 2\sqrt{\frac{2}{35}}(\Phi_{xxxz} - 3\Phi_{xyyz})$

22. $Q_{43s} = 2\sqrt{\frac{2}{35}}(3\Phi_{xxyz} - \Phi_{yyyz})$

23. $Q_{44c} = \sqrt{\frac{1}{35}}(\Phi_{xxxx} - 6\Phi_{xxyy} + \Phi_{yyyy})$

24. $Q_{44s} = 4\sqrt{\frac{1}{35}}(\Phi_{xxxy} - \Phi_{xyyy})$

Appendix B

Sample Translations of Multipoles

Using the following formula from section (place section number here):

$Q^C_{lk} = \sum\limits_{l'=0}^{l} \sum\limits_{k'=-l'}^{l'} [\binom{l+k}{l'+k'}\binom{l-k}{l'-k'}]^{\frac{1}{2}} Q^O_{l'k'} R_{l-l',k-k'}(-\mathbf{c})$

Where the original multipole expansion $Q^O_{l'k'}$, is about the origin, and Q^C_{lk} is the new multipole expansion, about point $\mathbf{c}$, at $\{x, y, z\}$. The spherical harmonics are a function of $-\mathbf{c}$ or $\{-x, -y, -z\}$.

Let $W_{lk,l'k'} = [\binom{l+k}{l'+k'}\binom{l-k}{l'-k'}]^{\frac{1}{2}} R_{l-l',k-k'}(-\mathbf{c})$, such that $Q^C_{lk} = \sum\limits_{l'=0}^{l} \sum\limits_{k'=-l'}^{l'} Q^O_{l'k'} W_{lk,l'k'}(-\mathbf{c})$.

If k is not equal to zero, the resulting spherical harmonic (R) and the corresponding multipole moment (Q) will be complex. In that case, the following formula is used to covert Q or R into linear combinations of regular harmonics, which results in a real function for the translation. While the formula is given in terms of R, Q may be substituted throughout for multipole moments.

$R_{lm} = (R_{lmc} + iR_{lms})/2b_m$

Where if $m > 0$, $b_m = (-1)^m\sqrt{\frac{1}{2}}$; if $m < 0$, $b_m = \sqrt{\frac{1}{2}}$. Also, if m is negative, R_{lms}

becomes $-R_{l|m|s}$, where if m is positive, $R_{lms} = R_{l|m|s}$. All $R_{lmc} = R_{l|m|c}$.

By definition spherical harmonics with l values smaller than the absolute value of the m term are equal to zero.

The following examples are designed to elucidate this theory.

For Rank 0, there is only one term:

$Q^C_{00} = Q^O_{00} W_{00,00}(-\mathbf{c})$

l = 0; k = 0; l' = 0; k' = 0;

$Q^O_{00} W_{00,10}(-\mathbf{c}) = Q^O_{00}[\binom{0}{0}\binom{0}{0}]^{\frac{1}{2}} R_{00}(-\mathbf{c})$

$R_{00}(-\mathbf{c}) = 1$, so:

$Q^C_{00} = Q^O_{00}$

This is consistent with the fact that point charges do not change upon translation.

For Rank 1, we will examine Q^C_{10}.

$Q^C_{10} = Q^O_{00} W_{00,10}(-\mathbf{c}) + Q^O_{10} W_{10,10}(-\mathbf{c}) + Q^O_{11} W_{11,10}(-\mathbf{c}) + Q^O_{1,-1} W_{1,-1,10}(-\mathbf{c})$

Taken term by term:

First term

l = 1; k = 0; l' = 0; k' = 0;

$Q^O_{00} W_{00,10}(-\mathbf{c}) = Q^O_{00}[\binom{1}{0}\binom{1}{0}]^{\frac{1}{2}} R_{10}(-\mathbf{c})$

$R_{10}(-\mathbf{c}) = -z$, so:

$Q^O_{00} W_{00,10}(-\mathbf{c}) = -z Q^O_{00}$

Second term

l = 1; k = 0; l' = 1; k' = 0;

$Q^O_{10} W_{10,10}(-\mathbf{c}) = Q^O_{10}[\binom{1}{1}\binom{1}{1}]^{\frac{1}{2}} R_{00}(-\mathbf{c})$

$R_{00}(-\mathbf{c}) = 1$, so:

$Q^O_{10} W_{10,10}(-\mathbf{c}) = Q^O_{10}$

Third term

l = 1; k = 0; l' = 1; k' = 1;

$Q^O_{11}W_{11,10}(-\mathbf{c}) = Q^O_{11}[\binom{1}{2}\binom{1}{0}]^{\frac{1}{2}}R_{0,1}(-\mathbf{c})$

Since $|m| > l$, the spherical harmonic is equal to zero, so the whole term is equal to zero.

Fourth term

l = 1; k = 0; l' = 1; k' = -1;

$Q^O_{1,-1}W_{1,-1,10}(-\mathbf{c}) = Q^O_{1,-1}[\binom{1}{0}\binom{1}{2}]^{\frac{1}{2}}R_{0,-1}(-\mathbf{c})$

Again, the spherical harmonic is equal to zero, and so the whole term goes to zero.

Summing all the terms:

$Q^C_{10} = Q^O_{10} - zQ^O_{00}$

which is indeed the correct translation for the z diole moment.

We will also examine one Rank 2 multipole, Q^C_{20}, which will cover the complex spherical harmonic case.

$Q^C_{20} = Q^O_{00}W_{00,20} + Q^O_{10}W_{10,20} + Q^O_{11}W_{11,20} + Q^O_{1,-1}W_{1,-1,20} + Q^O_{20}W_{20,20} + Q^O_{21}W_{21,20} +$
$Q^O_{2,-1}W_{2,-1,20} + Q^O_{22}W_{22,20} + Q^O_{2,-2}W_{2,-2,20}$

Again, taking this expression term by term:

First term

l = 2; k = 0; l' = 0; k' = 0;

$Q^O_{00}W_{00,20}(-\mathbf{c}) = Q^O_{00}[\binom{2}{0}\binom{2}{0}]^{\frac{1}{2}}R_{20}(-\mathbf{c})$

$R_{20}(-\mathbf{c}) = \frac{1}{2}(3z^2 - r^2) = \frac{1}{2}(2z^2 - x^2 - y^2)$, so

$Q^O_{00}W_{00,20}(-\mathbf{c}) = \frac{1}{2}(2z^2 - x^2 - y^2)Q^O_{00}$

Second Term

l = 2; k = 0; l' = 1; k' = 0;

$Q^O_{10}W_{10,20} = Q^O_{10}[\binom{2}{1}\binom{2}{1}]^{\frac{1}{2}}R_{10}(-\mathbf{c})$

$R_{10}(-\mathbf{c}) = -z$, so,

$Q^O_{10}W_{10,20} = -2zQ^O_{10}$

Third term

l = 2; k = 0; l' = 1; k' = 1;

$Q^O_{11}W_{11,20}(-\mathbf{c}) = Q^O_{11}[\binom{2}{2}\binom{2}{0}]^{\frac{1}{2}}R_{1,-1}(-\mathbf{c})$

The spherical harmonic is complex.

Using the rules for changing complex spherical harmonics to linear combinations of regular spherical harmonics:

$R_{1,-1} = (R_{11c} - iR_{11s})/(2\sqrt{\frac{1}{2}}) = \sqrt{\frac{1}{2}}(R_{11c} - iR_{11s})$

Using the same rules for the complex multipole:

$Q_{11} = (Q_{11c} + iQ_{11s})/(2(-1)^1\sqrt{\frac{1}{2}}) = -\sqrt{\frac{1}{2}}(Q_{11c} + iQ_{11s})$

Combining these two:

$Q^O_{11}R_{1,-1}(-\mathbf{c}) = -\sqrt{\frac{1}{2}}(Q_{11c} + iQ_{11s})\sqrt{\frac{1}{2}}(R_{11c}(-\mathbf{c}) - iR_{11s}(-\mathbf{c}))$

Simplified,

$Q^O_{11}R_{1,-1}(-\mathbf{c}) = -\frac{1}{2}(Q_{11c}R_{11c}(-\mathbf{c}) + iQ_{11s}R_{11c}(-\mathbf{c}) - iQ_{11c}R_{11s}(-\mathbf{c}) + Q_{11s}R_{11s}(-\mathbf{c}))$

While this term is still complex, we can expect cancellation of terms that will leave the entire translation a real function.

Fourth term

l = 2; k = 0; l' = 1; k' = -1;

$Q^O_{1,-1}W_{1,-1,20}(-\mathbf{c}) = Q^O_{1,-1}[\binom{2}{0}\binom{2}{2}]^{\frac{1}{2}}R_{1,1}(-\mathbf{c}) = Q^O_{1,-1}R_{1,1}(-\mathbf{c})$

Once again, the multipole and spherical harmonic are complex.

$R_{11} = (R_{11c} + iR_{11s})/(2(-1)^1\sqrt{\frac{1}{2}}) = -\sqrt{\frac{1}{2}}(R_{11c} + iR_{11s})$

$Q_{1,-1} = (Q_{11c} - iQ_{11s})/(2\sqrt{\frac{1}{2}}) = \sqrt{\frac{1}{2}}(Q_{11c} - iQ_{11s})$

So the entire term is:

$Q^O_{1,-1}R_{1,1}(-\mathbf{c}) = -\frac{1}{2}(Q_{11c}R_{11c}(-\mathbf{c}) + iQ_{11c}R_{11s}(-\mathbf{c}) - iQ_{11s}R_{11c}(-\mathbf{c}) + Q_{11s}R_{11s}(-\mathbf{c}))$

Fifth term

l = 2; k = 0; l' = 2; k' = 0;

$Q^O_{20} W_{020,20} = Q^O_{20} [\binom{2}{2}\binom{2}{2}]^{\frac{1}{2}} R_{20}(-\mathbf{c}) = Q^O_{20} R_{00}(-\mathbf{c})$

Since $R_{00} = 1$,

$Q^O_{20} W_{020,20} = Q^O_{20}$

We leave it to the reader to show that terms six through nine have R_{lm} with $|m| > l$ and so are equal to zero.

Combining the first, second, third, fourth and fifth terms,

$Q^C_{20} = \frac{1}{2}(2z^2 - x^2 - y^2) Q^O_{00} - 2z Q^O_{10} - \frac{1}{2}(Q_{11c} R_{11c}(-\mathbf{c}) + i Q_{11s} R_{11c}(-\mathbf{c}) - i Q_{11c} R_{11s}(-\mathbf{c}) + Q_{11s} R_{11s}(-\mathbf{c})) - \frac{1}{2}(Q_{11c} R_{11c}(-\mathbf{c}) + i Q_{11c} R_{11s}(-\mathbf{c}) - i Q_{11s} R_{11c}(-\mathbf{c}) + Q_{11s} R_{11s}(-\mathbf{c})) + Q^O_{20}$

Simplified:

$Q^C_{20} = \frac{1}{2}(2z^2 - x^2 - y^2) Q^O_{00} - 2z Q^O_{10} - Q^O_{11c} R_{11c}(-\mathbf{c}) - Q^O_{11s} R_{11s}(-\mathbf{c}) + Q^O_{20}$

Since $R_{11c}(-\mathbf{c}) = -x$ and $R_{11s}(-\mathbf{c}) = -y$

$Q^C_{20} = \frac{1}{2}(2z^2 - x^2 - y^2) Q^O_{00} - 2z Q^O_{10} + x Q^O_{11c} + y Q^O_{11s} + Q^O_{20}$

This is equivalent to the translation given in Chapter 4.

Appendix C

Translations of polarizabilities

The following are the equations used to translate polarizabilites, where $\alpha_{lm,l'm'}$ is an original polarizability, before translation, and $\alpha'_{lm,l'm'}$ is a polarizability that has already been translated to the new coordinates. Note that $\alpha'_{lm,l'm'} = \alpha'_{l'm',lm}$, which reduces the number of equations that are in terms of the original polarizabilities.

Charge has no polarizability, and dipole-dipole polarizabilities do not change upon translation, so equations for translations start with dipole-quadrupole.

dipole - quadrupole

$\alpha'_{11c,20} = \alpha_{11c,20} + \alpha_{11c,11c}x + \alpha_{11c,11s}y - 2\alpha_{11c,10}z$

$\alpha'_{20,11c} = \alpha'_{11c,20}$

$\alpha'_{11c,21c} = \alpha_{11c,21c} - \sqrt{3}\alpha_{11c,10}x - \sqrt{3}\alpha_{11c,11c}z$

$\alpha'_{21c,11c} = \alpha'_{11c,21c}$

$\alpha'_{11c,21s} = \alpha_{11c,21s} - \sqrt{3}\alpha_{11c,10}y - \sqrt{3}\alpha_{11c,11s}z$

$\alpha'_{21s,11c} = \alpha'_{11c,21s}$

$\alpha'_{11c,22c} = \alpha_{11c,22c} - \sqrt{3}\alpha_{11c,11c}x + \sqrt{3}\alpha_{11c,11s}y$

$$\alpha'_{22c,11c} = \alpha'_{11c,22c}$$

$$\alpha'_{11c,22s} = \alpha_{11c,22s} - \sqrt{3}\alpha_{11c,11s}x - \sqrt{3}\alpha_{11c,11c}y$$

$$\alpha'_{22s,11c} = \alpha'_{11c,22s}$$

$$\alpha'_{11s,20} = \alpha_{11s,20} + \alpha_{11c,11s}x + \alpha_{11s,11s}y - 2\alpha_{11s,10}z$$

$$\alpha'_{20,11s} = \alpha'_{11s,20}$$

$$\alpha'_{11s,21c} = \alpha_{11s,21c} - \sqrt{3}\alpha_{11s,10}x - \sqrt{3}\alpha_{11c,11s}z$$

$$\alpha'_{21c,11s} = \alpha'_{11s,21c}$$

$$\alpha'_{11s,21s} = \alpha_{11s,21s} - \sqrt{3}\alpha_{11s,10}y - \sqrt{3}\alpha_{11s,11s}z$$

$$\alpha'_{21s,11s} = \alpha'_{11s,21s}$$

$$\alpha'_{11s,22c} = \alpha_{11s,22c} - \sqrt{3}\alpha_{11c,11s}x + \sqrt{3}\alpha_{11s,11s}y$$

$$\alpha'_{22c,11s} = \alpha'_{11s,22c}$$

$$\alpha'_{11s,22s} = \alpha_{11s,22s} - \sqrt{3}\alpha_{11s,11s}x - \sqrt{3}\alpha_{11c,11s}y$$

$$\alpha'_{22s,11s} = \alpha'_{11s,22s}$$

$$\alpha'_{10,20} = \alpha_{10,20} + \alpha_{11c,10}x + \alpha_{11s,10}y - 2\alpha_{10,10}z$$

$$\alpha'_{20,10} = \alpha'_{10,20}$$

$$\alpha'_{10,21c} = \alpha_{10,21c} - \sqrt{3}\alpha_{10,10}x - \sqrt{3}\alpha_{11c,10}z$$

$$\alpha'_{21c,10} = \alpha'_{10,21c}$$

$$\alpha'_{10,21s} = \alpha_{10,21s} - \sqrt{3}\alpha_{10,10}y - \sqrt{3}\alpha_{11s,10}z$$

$$\alpha'_{21s,10} = \alpha'_{10,21s}$$

$$\alpha'_{10,22c} = \alpha_{10,22c} - \sqrt{3}\alpha_{11c,10}x + \sqrt{3}\alpha_{11s,10}y$$

$$\alpha_{22c,10} = \alpha_{10,22c}$$

$$\alpha_{10,22s} = \alpha_{10,22s} - \sqrt{3}\alpha_{11s,10}x - \sqrt{3}\alpha_{11c,10}y$$

$$\alpha_{22s,10} = \alpha'_{10,22s}$$

quadrupole-quadrupole

$$\alpha_{20,20} = \alpha_{20,20} + 2\alpha_{11c,20}x + \alpha_{11c,11c}xx + 2\alpha_{11s,20}y + 2\alpha_{11c,11s}xy + \alpha_{11s,11s}yy - 4\alpha_{10,20}z - 4\alpha_{11c,10}xz - 4\alpha_{11s,10}yz + 4\alpha_{10,10}zz$$

$$\alpha_{20,21c} = \alpha_{20,21c} - \sqrt{3}\alpha_{10,20}x + \alpha_{11c,21c}x - \sqrt{3}\alpha_{11c,10}xx + \alpha_{11s,21c}y - \sqrt{3}\alpha_{11s,10}xy - \sqrt{3}\alpha_{11c,20}z - 2\alpha_{10,21c}z - \sqrt{3}\alpha_{11c,11c}xz + 2\sqrt{3}\alpha_{10,10}xz - \sqrt{3}\alpha_{11c,11s}yz + 2\sqrt{3}\alpha_{11c,10}zz$$

$$\alpha_{21c,20} = \alpha'_{20,21c}$$

$$\alpha_{20,21s} = \alpha_{20,21s} + \alpha_{11c,21s}x - \sqrt{3}\alpha_{10,20}y + \alpha_{11s,21s}y - \sqrt{3}\alpha_{11c,10}xy - \sqrt{3}\alpha_{11s,10}yy - \sqrt{3}\alpha_{11s,20}z - 2\alpha_{10,21s}z - \sqrt{3}\alpha_{11c,11s}xz - \sqrt{3}\alpha_{11s,11s}yz + 2\sqrt{3}\alpha_{10,10}yz + 2\sqrt{3}\alpha_{11s,10}zz$$

$$\alpha_{21s,20} = \alpha'_{20,21s}$$

$$\alpha_{20,22c} = \alpha_{20,22c} - \sqrt{3}\alpha_{11c,20}x + \alpha_{11c,22c}x - \sqrt{3}\alpha_{11c,11c}xx + \sqrt{3}\alpha_{11s,20}y + \alpha_{11s,22c}y + \sqrt{3}\alpha_{11s,11s}yy - 2\alpha_{10,22c}z + 2\sqrt{3}\alpha_{11c,10}xz - 2\sqrt{3}\alpha_{11s,10}yz$$

$$\alpha_{22c,20} = \alpha'_{20,22c}$$

$$\alpha_{20,22s} = \alpha_{20,22s} - \sqrt{3}\alpha_{11s,20}x + \alpha_{11c,22s}x - \sqrt{3}\alpha_{11c,11s}xx - \sqrt{3}\alpha_{11c,20}y + \alpha_{11s,22s}y - \sqrt{3}\alpha_{11c,11c}xy - \sqrt{3}\alpha_{11s,11s}xy - \sqrt{3}\alpha_{11c,11s}yy - 2\alpha_{10,22s}z + 2\sqrt{3}\alpha_{11s,10}xz + 2\sqrt{3}\alpha_{11c,10}yz$$

$$\alpha_{22s,20} = \alpha'_{20,22s}$$

$$\alpha_{21c,21c} = \alpha_{21c,21c} - 2\sqrt{3}\alpha_{11c,21c}z - 2\sqrt{3}\alpha_{10,21c}x + 3\alpha_{10,10}xx + 3\alpha_{11c,11c}zz + 6\alpha_{11c,10}xz$$

$$\alpha_{21c,21s} = \alpha_{21c,21s} - \sqrt{3}\alpha_{11s,21c}z - \sqrt{3}\alpha_{10,21c}y - \sqrt{3}\alpha_{11c,21s}z + 3\alpha_{11c,11s}zz + 3\alpha_{11c,10}zy -$$

$\sqrt{3}\alpha_{10,21s}x + 3\alpha_{11s,10}xz + 3\alpha_{10,10}xy$

$\alpha_{21s,21c} = \alpha'_{21c,21s}$

$\alpha_{21c,22c} = \alpha_{21c,22c} + \sqrt{3}\alpha_{11s,21c}y - \sqrt{3}\alpha_{11c,21c}x - \sqrt{3}\alpha_{11c,22c}z - 3\alpha_{11c,11s}yz + 3\alpha_{11c,11c}xz -$

$\sqrt{3}\alpha_{10,22c}x - 3\alpha_{11s,10}xy + 3\alpha_{11c,10}xx$

$\alpha_{22c,21c} = \alpha'_{21c,22c}$

$\alpha_{21c,22s} = \alpha_{21c,22s} - \sqrt{3}\alpha_{11s,21c}x - \sqrt{3}\alpha_{11c,21c}y - \sqrt{3}\alpha_{11c,22s}z + 3\alpha_{11c,11s}xz + 3\alpha_{11c,11c}yz -$

$\sqrt{3}\alpha_{10,22s}x + 3\alpha_{11s,10}xx + 3\alpha_{11c,10}xy$

$\alpha_{22s,21c} = \alpha'_{21c,22s}$

$\alpha_{21s,21s} = \alpha_{21s,21s} - 2\sqrt{3}\alpha_{10,21s}y + 3\alpha_{10,10}yy - 2\sqrt{3}\alpha_{11s,21s}z + 6\alpha_{11s,10}yz + 3\alpha_{11s,11s}zz$

$\alpha_{21s,22c} = \alpha_{21s,22c} - \sqrt{3}\alpha_{11c,21s}x + \sqrt{3}\alpha_{11s,21s}y - \sqrt{3}\alpha_{10,22c}y + 3\alpha_{11c,10}xy - 3\alpha_{11s,10}yy -$

$\sqrt{3}\alpha_{11s,22c}z + 3\alpha_{11c,11s}xz - 3\alpha_{11s,11s}yz$

$\alpha_{22c,21s} = \alpha'_{21s,22c}$

$\alpha_{21s,22s} = \alpha_{21s,22s} - \sqrt{3}\alpha_{11s,21s}x - \sqrt{3}\alpha_{11c,21s}y - \sqrt{3}\alpha_{10,22s}y + 3\alpha_{11s,10}xy + 3\alpha_{11c,10}yy -$

$\sqrt{3}\alpha_{11s,22s}z + 3\alpha_{11s,11s}xz + 3\alpha_{11c,11s}yz$

$\alpha_{22s,21s} = \alpha'_{21s,22s}$

$\alpha_{22c,22c} = \alpha_{22c,22c} - 2\sqrt{3}\alpha_{11c,22c}x + 3\alpha_{11c,11c}xx + 2\sqrt{3}\alpha_{11s,22c}y - 6\alpha_{11c,11s}xy + 3\alpha_{11s,11s}yy$

$\alpha_{22c,22s} = \alpha_{22c,22s} - \sqrt{3}\alpha_{11s,22c}x - \sqrt{3}\alpha_{11c,22s}x + 3\alpha_{11c,11s}xx - \sqrt{3}\alpha_{11c,22c}y + \sqrt{3}\alpha_{11s,22s}y +$

$3\alpha_{11c,11c}xy - 3\alpha_{11s,11s}xy - 3\alpha_{11c,11s}yy$

$\alpha_{22s,22c} = \alpha'_{22c,22s}$

$$\alpha_{22s,22s} = \alpha_{22s,22s} - 2\sqrt{3}\alpha_{11s,22s}x + 3\alpha_{11s,11s}xx - 2\sqrt{3}\alpha_{11c,22s}y + 6\alpha_{11c,11s}xy + 3\alpha_{11c,11c}yy$$