UNIVERSITY OF CALIFORNIA
RIVERSIDE

Peptides, Diamines, and Nucleobases: Investigations in Ion Chemistry

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

Doctor of Philosophy

In

Chemistry

by

Aaron Robert Moehlig

December 2011

Dissertation Committee:

Dr. Thomas H. Morton, Chairperson
Dr. Christopher Switzer
Dr. Richard Hooley

The Dissertation of Aaron Robert Moehlig is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENT

Graduate school has been a rewarding and bewildering experience. Chemistry has become infinitely interesting, something I would not have thought possible ten years ago. I want to thank Dr. Cliff Harris for first making me excited about chemistry and for being a great mentor. I wouldn't have the outlook on the world I have now without the guidance of Dr. Thomas Morton. I want to thank Tom for his inspiration and for making me curious about everything around me, not just in a chemical sense. It has been an immense pleasure to be his graduate student and I have had many great opportunities because of it. I want to thank Dr. Rich Kondrat for teaching me about mass spectroscopy and for helping me whenever I struggled with it. I want to thank Dr. Dan Borchardt for help with NMR and Raman. I would like to thank Dr. Jianshuang Wang, Dr. Tony Ly, and Prof. Ryan Julian for all the work they did on the peptides before I took over the project. I want to thank Dr. Audrey Solgadi, Dr. Isabelle Schmitz-Afonso, Dr. Olivier LaPrevote, Prof. Bruno Figadere for making my time in France fun and constructive. I would also like to thank Dr. Luke Daemen and Dr. Jos Oomens for all their work on the INS and IRMPD experiments, respectively. I could not have completed this thesis without their help. I also want to thank Dr. Richard Hooley for the conversations we've had about my research and for his careful reading of this thesis.

I could not have gotten through graduate school without the friends I have made while at UCR. They have always been there to bounce ideas off, talk chemistry, or just wind down with after a long day. I would like to thank all of my current and former lab mates, Dr.Tony Cao, Hou Ung, Jay-Ar Bendo, and especially Dr. Sepideh Yaghmaei for all their help and support over the years. To Conor Pierce, thanks for being a good friend and for all the fun we've had. To Mellissa Padilla, I know I cannot thank you enough for everything you've done for me, but I will try. Without you this thesis would not be finished and I would not be the person I am.

Finally, I would not be here if it was not for my family. My parents Tony and Julie have always been there for me and I am proud to be their son. I want to thank my brothers and sister, Chaesa, Jeremy, Jono, and Megan, for their love and support.

ABSTRACT OF THE DISSERTATION

Peptides, Diamines, and Nucleobases: Investigations in Ion Chemistry

by

Aaron Moehlig

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, December 2011
Dr. Thomas H. Morton, Chairperson

Gas-phase ion chemistry offers a variety of experimental tools that can assist in determining the structure and properties of molecules that are difficult to purify and isolate or available in only small concentrations. This thesis highlights three areas where ion chemistry has been successfully used to give new insight into the structures and properties of various peptides, diamines, and nucleobases.

Peptides

Peptide biosynthesis incorporates only **L**-isomers of the chiral amino acids, however, **D**-isomers can result from post-translational modifications performed by epimerase enzymes or through spontaneous racemization. Sequential mass spectrometry can be used to measure quantitative differences in the fragmentation patterns of peptides containing a single D-amino acid. This thesis describes the differences in the collsionally activated dissociation (CAD) patterns of the diastereomers of the tripeptide Gly-Ser-Lys (GSK) and the tetrapeptide Gly-Leu-Ser-Lys (GLSK). The latter is among the tryptic fragments expected from the digest of alpha-synuclein, a protein found in Lewy bodies implicated in Parkinsonism and other neurological

disorders. The differences between the CAD fragmentation patterns is used to develop a method to quantitate the amount of the GLdSK diastereomer in a mixture of GLSK.

Diamines

Vibrational spectra of the conjugate acid of the linear diamine *N,N,N',N'*-tetramethylputrescine has been recorded in the gas-phase with infrared multiple photon dissociation (IRMPD) and solid-phase by inelastic neutron scattering (INS). A band near 550 cm^{-1} is observed in both spectra and is assigned as the asymmetric stretch of a proton between two nitrogens based on deuterium substitution. The position of this band agrees with the value predicted from a 2-dimensional potential energy surface. The reduced dimensionality of the potential energy surface, which treats the ion as though it has a linear geometry, predicts the zero-point energy level to be slightly above the calculated barrier to proton transit suggesting that this is a low-barrier hydrogen bond and the proton is equally shared between the two nitrogens. The cyclic diamine *N,N,N',N'-tetramethyl-cis-1,5-diaminocyclooctane* was synthesized and its 1-dimensional potential energy surface predicted the zero-point energy to be below the calculated barrier to proton transit suggesting that it is not a low-barrier hydrogen bond. The IRMPD and INS spectra will be recorded to determine if the value of the N-H-N asymmetric stretch matches with theory.

Nucleobases

The vibrational spectrum of proton-bound dimers consisting of 1-methylcytosine, 1,5-dimethylcytosine, and 5-fluoro-1-methylcytosine has been recorded and the structure of the dimers as well as the location of the band corresponding to the N-H-N asymmetric

stretch were determined based upon comparisons with theoretical spectra and isotopic substitution. The proton between the two nucleobases prefers to be situated on one side of the dimer and is not equally shared between the two nitrogens based upon the presence of two IR bands in the spectrum of the proton-bound dimer formed between 1-methylcytosine and 1,5-dimethylcytosine that correspond to the N-H-N asymmetric stretches of the two tautomers of the dimer.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter I:

Peptides

Introduction

Ribosomal biosynthesis incorporates only L-isomers of chiral amino acids into peptides, while non-ribosomal peptide synthetases can make use of a wider range of substrates. D-amino acids can be incorporated into polypeptides synthesized by ribosomes through post-translational modifications by site-specific epimerases or through spontaneous racemization.[1,2] Non-ribosomal biosynthesis can use D-aminio acids directly in their peptide synthesis.[3,4] Mass spectrometry has the capacity to distinguish between peptide diasteromers through enzyme digestion followed by mass spectrometric analysis[5] and via a variety of sequential mass spectrometric (MS/MS) techniques. MS/MS methods include collisionally activated dissociation (CAD)[6] and electron capture dissociation (ECD).[7] Both techniques rely on the isolation of ions of a single mass-to-charge ratio in the first stage of analysis followed by decomposition of those isolated ions by means of collisions with an inert gas (CAD) or neutralization of a positive charge by capture of a free electron (ECD). Both CAD and ECD require a second stage of mass analysis to record the resulting fragmentation pattern. ECD requires the initially isolated ion to have more than one positive charge so that fragmentation after electron capture will produce daughter ions. CAD is not confined by this limitation and can be applied to positive and negative ions regardless of the number charges. CAD is the method of choice for analysis of small peptides as they can often only carry a single charge.

CAD experiments may be performed using a variety of mass spectrometers that can be divided into two categories: tandem mass spectrometers and storage mass spectrometers. Tandem instruments involve multiple mass spectrometers with a fragmentation sector between mass analyzing segments. One example of a tandem instrument is a triple quadrupole mass spectrometer. A triple quadrupole mass spectrometer consists of an ionization source connected to two mass analyzing quadrupoles separated by a radio-frequency only quadrupole that acts as a collision cell. The first triple quadrupole mass spectrometer was developed by Enke and Yost at Michigan State University.[8] The first and third quadrupoles are under a static vacuum and serve as mass selection filters. The middle quadrupole is filled with an inert gas such as argon or helium. Ions are created by an ionization source such as electrospray, accelerated into the first quadrupole where ions with a specific mass-to-charge ratio are selected to enter the second quadrupole. The selected ions collide with the inert gas that fills the middle quadrupole and fragments the parent ions in a single high energy collision. The fragments then enter the third quadrupole mass spectrometer which scans the entire m/z range to reveal the mass-to-charge ratio of the fragments (daughter ions). The triple quadrupole mass spectrometer is useful in determining the structure of parent ions and in elucidating the amino acid sequence of a protein.

Storage instruments usually have only a single mass spectrometer and utilize an ion trap or an ICR in order to store selected ions. Once ions are created

they can be stored in a linear or a three-dimensional ion trap using constant DC voltage and radio-frequency oscillating AC electric fields. Linear ion traps have more ion storage capacity and quicker scan times than three-dimensional traps.[9] A linear trap confines ions radially by a set of quadrupole rods and axially by a static electric potential on-end electrodes.[10] The ions are not motionless in the trap; rather, they experience restoring forces that drive them towards the center of the trap. The motion of the ions in the trap can be described by the Mathieu equation.[9] Sequential mass spectrometric methods take advantage of the motion of the ions by allowing an inert gas to enter the trap after the ions have been isolated. The trapped ions experience many small energy collisions with the gas molecules, which excite the parent ions and cause fragmentation to occur. Once the parent ion has been fragmented the trap is scanned and the masses of the daughter ions are recorded. The types and relative amounts of fragmentation in a linear ion trap can differ drastically from tandem instruments because of the number of excited states the ions can enter due to the multiple collisions they experience in the trap.

One storage instrument, an orbitrap, can be coupled to another storage instrument, such as a linear ion trap, to form a hybrid mass spectrometer where the orbitrap acts as a mass analyzer of fragments ejected from the front-end ion trap. The orbitrap consists of an outer electrode and a coaxial inner electrode that uses an electrostatic field to trap ions. Ions are injected tangentially into the electric field between the two electrodes and trapped due to their attraction to the

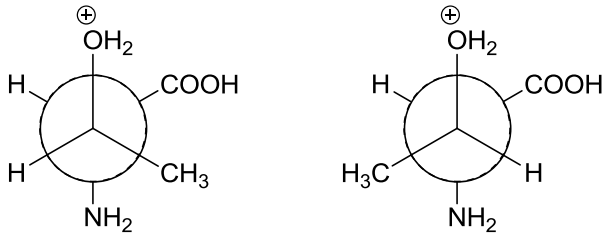
inner electrode and their offsetting centrifugal forces. The ions cycle around the center electrode and move back and forth along its axis. The frequency of the ions' harmonic oscillation is inversely proportional to the square root of the mass-to-charge ratio. Mass spectra are obtained by Fourier transformation of the recorded image currents.[11] Operating with a very low background pressure, the orbitrap can sense the ion oscillation and records the masses of the trapped ions with a high mass accuracy and high resolution.[12] The combination of a linear ion trap containing a mass analyzer with a high resolving power like an orbitrap allows access to a precise record of robust fragmentation patterns.

All three of the mass spectrometers mentioned can be used to isolate an ion of a single mass-to-charge ratio and record the mass-to-charge ratio of the ions that result from its fragmentation. These spectrometers are useful in determining a peptide sequence, location of functional groups, and can be used to distinguish between two molecules with the same molecular formula, among other things. For these reasons mass spectrometry is a great resource for distinguishing between isomers of amino acids.

Protonated amino acids readily expel water in the gas phase. Usually, the water lost comes from the carboxylic group and is often accompanied by the loss of carbon monoxide to produce an iminium ion.[13] Threonine presents one example where water loss from the side chain OH group predominates.[14] Threonine has two asymmetric carbons, and the diastereomers (**L** or **D**)

threonine and (**L** or **D**) *allo***-**threonine can be differentiated using CAD sequential mass spectrometry.[15] Both parent ions (MH^+) of threonine and *allo*-threonine have the same mass-to-charge ratio (m/z 120 for the MH^+) and differ only in the orientation of the OH group on the β-carbon (Figure 1.1).

Comparing the intensity of the peak corresponding to the loss of water (m/z 102) with the peak corresponding to the loss of water followed by the loss of carbon monoxide (m/z 74) in the CAD spectra of threonine and *allo*-threonine allows for the differentiation of the diastereomers. The enantiomers, where the stereochemistry of both carbons is changed, cannot be differentiated using this technique. When the orientation of the α-carbon is constant and the orientation of the β-carbon is varied MSMS can be used to differentiate between threonine and *allo*-threonine. Plotting the ratio of the MH^+-H_2O and MH^+-H_2O+CO (the y-axis in Figure 1.2) versus the lab frame collision energy of the parent ion conveys the difference in daughter ion spectra between the two diastereomers. As can be seen from comparison of the filled and open symbols in Figure 1.2, the y-values of pairs of enantiomers are the same.



Protonated Threonine

Protonated *allo*-Threonine

Figure 1.1. Newman projection of the amino acid L-threonine and its diastereomer L-*allo*-threonine

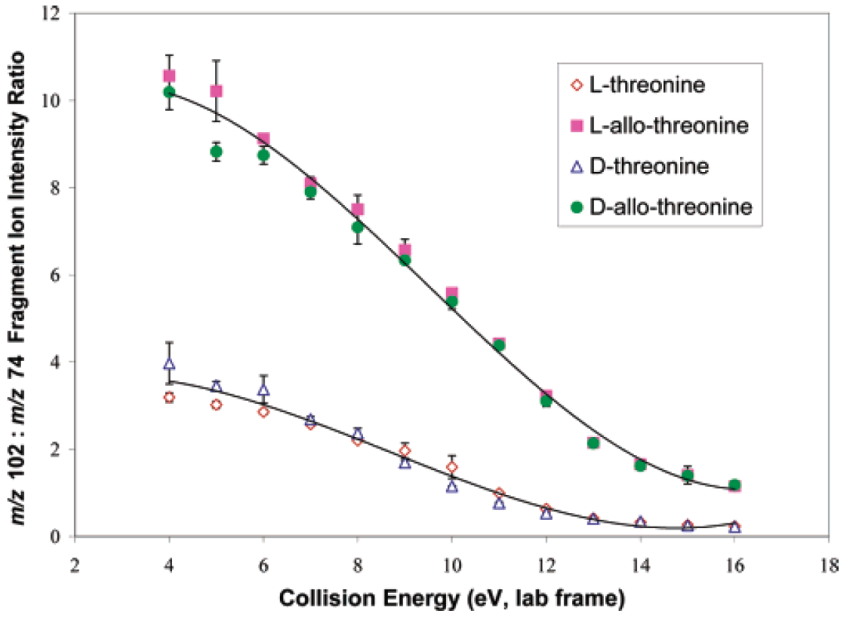


Figure 1.2. Ratios of H_2O loss to H_2O + CO loss from D- and L-threonine (open symbols) and their *allo* diastereomers (solid symbols) as a function of collision energy of *m/z* 120 parent ions.[15]

The ratio of y-values between each diastereomer can be used to demonstrate the magnitude of the difference between each MS/MS spectra. That is, the intensity of the m/z 102 divided by the intensity of the m/z 74 peak (plotted on the y axis of Figure 1.2) for the L-threonine diastereomer at a collision energy

of 4 eV is about 3 times that of the same ratio of the L-*allo*-threonine diastereomer at the same collision energy. This ratio of ratios (r) is a figure of merit used to describe how easily two diastereomers can be differentiated based on the difference between two daughter ion spectra.[16] The larger the r value the more robust the difference between the spectra and the easier it is to differentiate the diastereomers. The r value of L-threonine versus D-threonine (as well as L-*allo*-threonine versus D-*allo*-threonine) is close to 1. This signifies no difference between the relative intensities of the m/z 102 and m/z 74 ions exists in the daughter ion spectra.

The intensity of the m/z 102 ion for L-*allo*-threonine is over ten times as intense as the m/z 74 ion at a collision energy of 4 eV. The m/z 102 peak is about 3.5 times as intense as the m/z 74 peak for L-threonine at the same collision energy. Because the loss of H_2O from the carboxylic group of amino acids is followed quickly by the loss of CO it can be inferred that L-*allo*-threonine more readily loses water from the side chain (which is not coupled with the loss of CO) than does L-threonine.

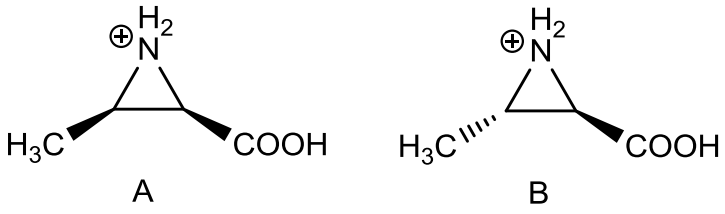


Figure 1.3. Aziridine ring produced by the loss of water from the side chain of threonine (A) and *allo*-threonine (B)

The loss of H_2O from the side chain of threonine can be accomplished by an S_N2 like displacement of water by the amino group, resulting in a stereochemically different 3-membered aziridine ring from each diastereomer (Figure 1.3). The relative ease in which the threonine diastereomers lose water can be attributed to the steric hindrance introduced by placing the methyl and carboxylic groups cis to one another during side chain loss of H_2O from L-threonine.

The amino acid serine, like threonine, has a hydroxyl group on its side chain. Unlike threonine, serine only has a single asymmetric carbon and enantiomers, L and D, of serine would produce identical protonated aziridine rings if they lost H_2O from their side chain in the same manner as threonine. The lack of a second chiral carbon makes it impossible for sequential mass spectrometry to differentiate between L and D-serine. When additional chiral amino acids are coupled with serine, diastereomers are created. In some cases diastereomeric peptides containing a single D-serine residue can be differentiated from the all L peptide using CAD.[6]

Racemization of serine side chains has been known to be accomplished enzymatically, as in the case of the venom glands of the common grass spider[17] and is suspected to be spontaneous within the serine residues that compose β-amyloid, a peptide implicated in Alzheimer's disease.[18] When serine is part of a polypeptide that contains at least one other asymmetric carbon it is possible to

use MS/MS on a tandem mass spectrometer (Q-Tof Ultima-Global, Micromass) to differentiate the diastereomers with r values as high as 3.8 for a single D-serine containing peptide (NH_2-Gly-Leu-Ser-Phe-Ala-OH). A dodecapeptide, H_2N-Leu-Val-Phe-Phe-Ala-Glu-Asp-Val-Gly-Ser-Asn-Lys-OH (LVFFAEDVGSNK) that is a predicted tryptic digest fragment of β-amyloid, also exhibits differences between the CAD MS/MS spectra of the all L peptide and the peptide containing a D-serine residue.[6] Doubly protonated LVFFAEDVGSNK and LVFFAEDVG^{d}SNK did not permit an easy distinction among diastereomeric peptides. A D-serine does induce small changes in the intensities of some fragment ions. Doubly charged peptide-complexed metal ions required large collision energies for fragmentation and no significant differences were observed between the diastereotopic peptides. Triply charged, metal-complexed, protonated dodecapeptide ions did allow for differentiation of diastereomers, however.

Experiments such as the ones mentioned above are done using diastereomeric pure peptides or digests of pure peptides in order to develop the best mass spectrometric conditions to induce differences between the daughter ion spectra of each diastereomer. *In vivo* enzymatic and spontaneous racemization of serine residues does not result in 100% conversion of serine residues from L to D in every peptide. Although the sequential mass spectrometric techniques mentioned above can differentiate between diastereomers, none of the methods quantified a mixture of a D-serine containing

peptide in a mixture of an all L-peptide. This thesis describes the development of a MS/MS method to differentiate between a diastereomeric peptide that is a predicted tryptic digest fragment of α-synuclein, containing a singly substituted D-serine, and quantify the amount of D-serine in a mixture of peptides containing both L and D-serine.

Background

In ancient India "Kampavata", described by Galen in AD 175 as, "a shaking palsy" and finally recognized as a distinct medical disorder after "An Essay on the Shaking Palsy," was published by Dr. James Parkinson in 1817. Motor related symptoms of Parkinson's disease (PD) are characterized by tremors, stiffness of limbs, and impaired balance and movement.[19] Psychotic symptoms exhibit themselves in the form of illusions, a false sense of presence, hallucinations, and delusions.[20] Caused by the degradation of dopaminergic neurons, PD affects over 500,000 people in the United States with 50,000 new cases reported each year.[21] Current treatments revolve around increasing the amount of dopamine in the brain to offset the degeneration of dopamine producing nerve cells. Virtually all dopaminergic medications are known to induce or heighten psychosis in PD by stimulating or inducing hypersensitivity of mesocorticolimbic dopamine receptors.[22] Non- dopaminergic substrates have been developed in order to treat PD induced psychosis and attempt to treat the cause of PD rather than just replace the loss of dopamine.[23] New research strategies must be aware of relationships between among other neurochemical

and neuroanatomical substrates and consider those elements as targets for drug development.

There is no universally accepted theory as to the pathogenesis of this disease, but a cause and effect relationship between the neuronal protein α-synuclein and PD was established in 2000 by Peter Lansbury at the Harvard Medical School.[24] α-Synuclein is an intrinsically disordered 140 amino acid protein found at nerve terminals of the basal ganglia and substantia nigra, making up about 1% of all proteins in the cytosol.[25] Aggregates (also known as fibrils or Lewy Bodies) of α-synuclein have been discovered in the brains of PD at a much higher rate than those without PD. The currently accepted theory in PD progression, “The Pre-Fibril Hypothesis” speculates that: because all animals have α-synuclein, but not PD, the monomeric protein is not a direct cause of nerve damage. Because mice have PD symptoms without fibrils (insoluble, polymeric aggregates of α-synuclein) that were found in the brain of humans with PD, that polymer of the protein cannot be responsible for the nerve cell damage either. Therefore, the theory concludes that there must be a partially folded monomer or a collection of small oligomers that triggers the nerve cell damage.[26] Investigations into the mechanics of α-synuclein folding led to an important discovery about post-translational modifications to α-synuclein in the brain, namely that α-synuclein can be phosphorylated at the serine-129.[27] Later it was discovered that 90% of the fibrillated α-synuclein that accumulated in the brain was phosphorylated at Ser-129 as opposed to only 4% of the soluble, monomeric

α-synuclein.[28] No research has focused on the epimerization of serine-129 or the other three serine residues that α-synuclein contains, however. It stands to reason that α-synuclein may have epimerized serine residues much like β-amyloid, because they are in the same class of neuronal proteins and because Lewy bodies start to form in the central nervous system decades before symptoms emerge. The question becomes whether epimerization of serine residues in α-synuclein takes place and, if so, does the amount of epimerization correlate with the age of a Lewy body or with fibril formation or disease onset and progression?

Results

The observation that other neuropathologically relevant peptides contain appreciable levels of D-serine (notably, β-amyloid from Alzheimer's plaque[29]) prompts the inquiry whether Lewy bodies might also include epimerized serine residues, particularly at sites that do not undergo extensive phosphorylation. Before attempts can be made to quantify the amount of D-serine in the α-synuclein of PD affected brains a method must be developed to distinguish between peptide diastereomers of a digest fragment of the full protein. Gly-Leu-Ser-Lys (GLSK) is among the tryptic fragments expected from α-synuclein, consisting of residues 7-10 of the 140 amino acid protein.

Shorter peptides Ser-Lys (SK), Leu-Lys (LK) Leu-Ser (LS), and Gly-Ser-Lys (GSK) were investigated to determine if the all L-peptides could be

distinguished from their D-serine containing diastereomers and, if so, whether or not the stereochemistry of serine is specifically involved in the different fragmentation pathways of the diastereomers. CAD experiments were performed using an LTQ linear ion trap mass spectrometer in collaboration with the research group of Prof. Ryan Julian.[30] Peptide samples were prepared and purified using LC by Dr. Jianshwang Wang. The greatest difference in the spectra of the CAD spectra of the dipeptides SK and ^{d}SK was in the MH^{+}-$2H_2O$:MH^{+}-H_2O ratio. Protonated SK has a ratio of 0.137 while ^{d}SK has a ratio of 0.112 which gives an r value of 1.2. Both sets of data had a standard deviation of 0.002 which means that the diastereomers do produce significantly different CAD spectra, but the differences are not robust enough to be able to quantify a mixture of diastereomers. The truncation dipeptide from GLSK, LS, and the deletion dipeptide, LK, exhibit even smaller effects.

The deletion tripeptide, GSK, does show a quanitative difference between protonated GSK (m/z 291) and its diastereomer G^{d}SK. The MH^{+}-H_2O (m/z 273) ion from G^{d}SK decomposes less readily than does the same ion from GSK. CAD produces fragments in the same proportions for each diastereomer, except that the MH^{+}-H_2O ion is 1.9 times more intense (relative to any other fragment) in the spectra of G^{d}SK as compared to GSK (Figure 1.4). MS/MS/MS confirms this interpretation: under identical activation conditions, the MH^{+}-H_2O ion from GSK dissociates nearly twice as much as does the same ion from G^{d}SK. An r value of

1.9 is greater than that for the dipeptide diasteromers, but less than the value seen in other serine containing diastereomers referenced in the introduction.

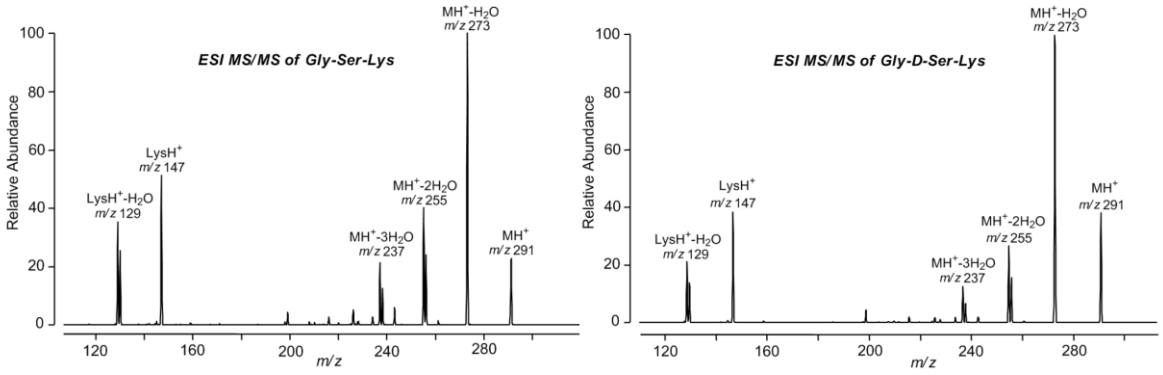


Figure 1.4. ESI MS/MS spectra of protonated GSK and G^{d}SK performed on a LTQ linear ion trap mass spectrometer

GSK and its serine diastereomer were also subjected to MS/MS experiments using a triple quadrupole mass spectrometer.[31] The fragmentation of the parent ion gave the same daughter ions as the spectra recorded on the linear trap instrument, but did not give the same relative intensities (Figure 1.5). The ratio of daughter peaks that gave the largest difference between diastereomers (r=1.94) was the same as in the linear trap spectra (MH^+-H_2O:$LysH^+$). As the lab frame collision energy was increased incrementally from 10 eV to 20 eV the difference between the two diasteromers decreases. Figure 1.6 shows the r values for the MH^+-H_2O:$LysH^+$ ratios for GSK versus G^{d}SK at varying collision energies along with their standard deviations. The increase in collision energy indicates that the parent ion is colliding with the gas molecules in the middle quadrupole with more and more velocity. With enough velocity the collision with the gas molecules is sufficiently violent that the diastereomers no longer show a

significant difference in their fragmentation. Although the r value of 1.94 observed at the lowest collision energy is sufficient to distinguish between the two diastereomers, the error bars associated with the spectra are too large to be able to accurately quantify a mixture of the two diastereomers.

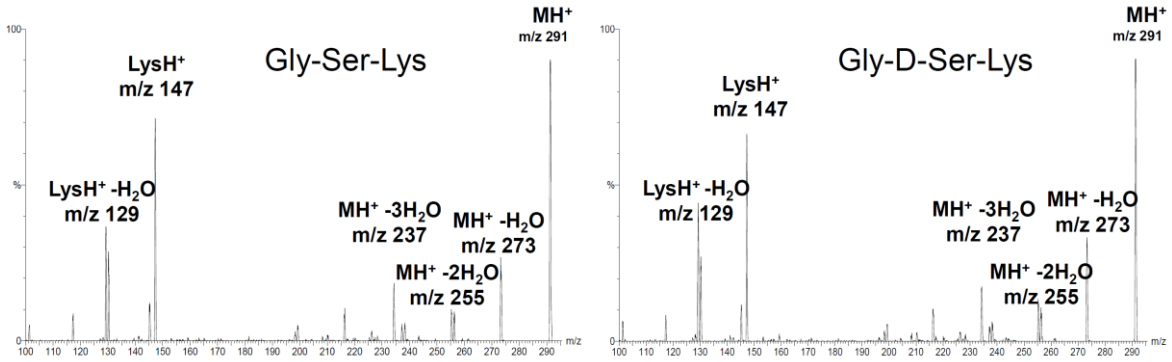


Figure 1.5. ESI MS/MS spectra of GSK and G^{d}SK performed on a triple quadrupole mass spectrometer.

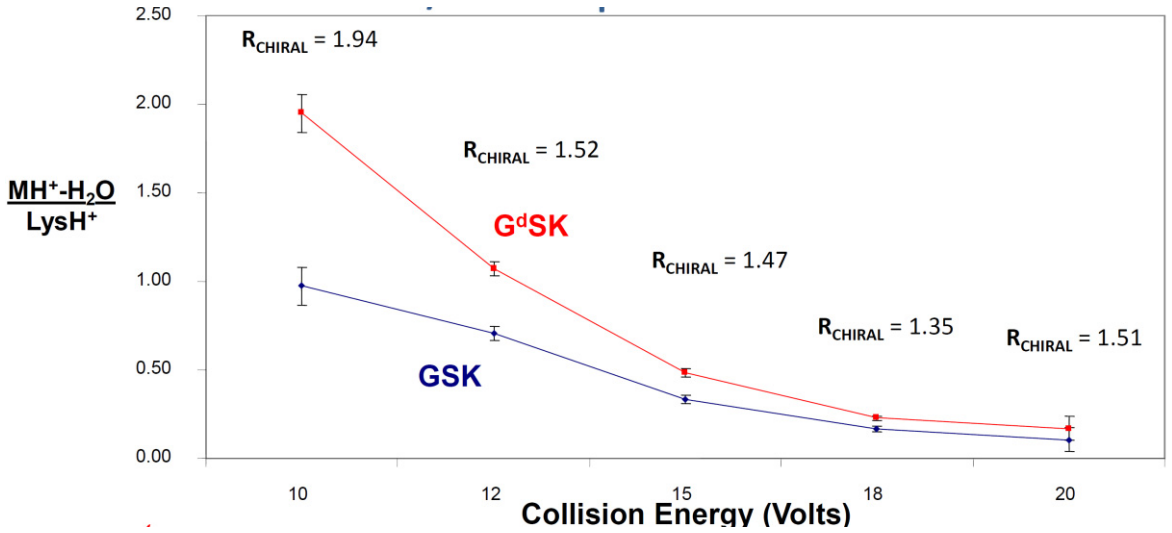


Figure 1.6. Plot of the ratio of the two daughter ions from GSK and G^{d}SK parent ions at various lab frame collision energies of a triple quadrupole mass spectrometer.

The Electrospray Ionization (ESI) of GLSK (m/z 404) and of its diastereomer GL^{d}SK followed by CAD in a linear ion trap mass spectrometer produces the five major fragments indicated in the spectra reproduced in Figure 1.7. The relative intensities of two fragment ions, loss of water (m/z 386) and loss of water and ammonia (m/z 369), do not differ significantly between the two diastereomers. No simple ammonia loss is observed in either spectrum; hence, water loss must precede ammonia loss in the dissociation that produces the m/z 369 ion. Acetylation of GLSK at the N-terminus (Ac-GLSK) suppresses the formation of the MH^+-H_2O-NH_3 ion, further implicating the N-terminal nitrogen of glycine as the origin of the loss of ammonia and not the amino group of lysine.

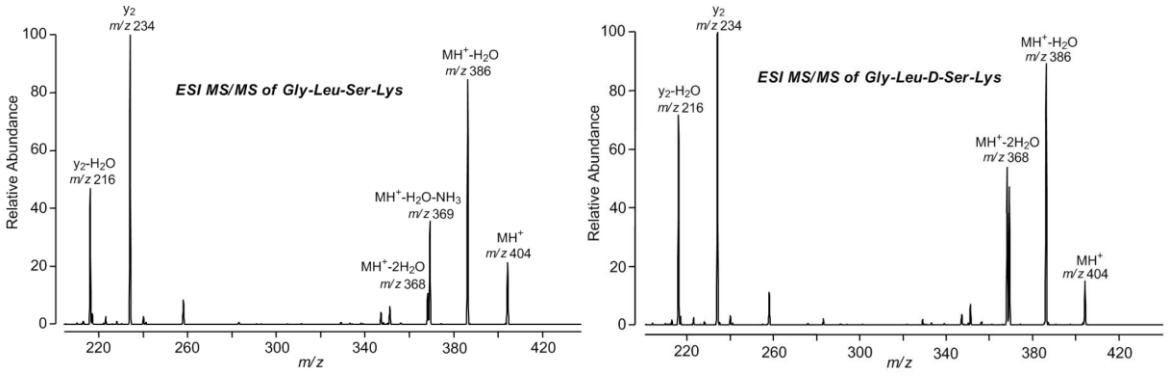


Figure 1.7. ESI MS/MS spectra of protonated GLSK and GL^{d}SK performed on a LTQ linear ion trap mass spectrometer

The greatest r value for the tetrapeptide diastereomers comes from comparing the ratio of the MH^+-$2H_2O$ ion (m/z 368) with the MH^+-H_2O-NH_3 ion (m/z 369). The MH^+-$2H_2O$: MH^+-H_2O-NH_3 ratio for GLSK has a value of 0.31, while the intensity ratio from GL^{d}SK is 1.15, resulting in an r value of 3.7. Adding another chiral amino acid (leucine) to GSK increases the diastereomeric

difference almost two-fold. Other differences between the two spectra are apparent, such as extent of cleavage of the middle peptide bond to give protonated Ser-Lys (a y_2 fragmentation) relative to the MH^+-H_2O, but none give an r value approaching 3.7.

In an effort to ascertain the origin of the lost water molecules, the C-termini of the tetrapeptides were esterified with methanol. The protonated methyl esters lose water, but the loss of CH_3OH or of two water molecules or of water and CH_3OH are suppressed. It is known that the acidic proton of free serine and free threonine transposes to the side chain hydroxyl group and is lost as water.[7] The O-methylated ether of GLSK, abbreviated GLS(OMe)K, was prepared in order to assess the role of the side chain of serine in water loss. Only racemic protected O-methylserine was available and the CAD spectrum for GLS(OMe)K corresponds to a diasteromeric mixture of GLSK ethers. Like the GLSK ester protonated GLS(OMe)K loses water or CH_3OH, or two water molecules, but exhibits a much smaller peak from the loss of water and CH_3OH. This suggests that the initial water loss exhibited by the GLSK diastereomers can originate from multiple sources: the C-terminal carboxylic group, the serine side chain hydroxyl group, and at least one of the peptide backbone amide oxygens. Although the serine hydroxyl group is not exclusively involved in one specific step in the sequential loss of water molecules, its stereochemistry still affects the fragmentation pattern of the parent ion.

CAD spectra recorded using a triple quad mass spectrometer showed no difference between the GLSK and GL^{d}SK diastereomers. This suggests differences arise between the relative intensities of the MH^{+}-$2H_2O$ (m/z 368) and MH^{+}-H_2O-NH_3 (m/z 369) daughter ions for the two diastereomers due to of the multistep excitation that occurs in a storage instrument. When only a single collision of parent ion with an inert gas molecule is responsible for fragmentation the two diastereomers show no significant difference in daughter ion intensities. However, when the parent ions undergo multiple collisions as they oscillate in an ion trap they are able to enter a number of energetically excited states. When each diastereomer is able to sample many excited states the stereochemistry of GLSK allows it to lose ammonia after its first water loss more easily than a second water molecule. The opposite is true for GL^{d}SK. After multiple excitations it more easily loses a second molecule of water than ammonia after initial water loss. The difference in these spectra can be used to distinguish between the two diastereomers.

DFT geometry optimizations of the protonated GSK and G^{d}SK and the lowest energy conformation of the loss of one and two water molecules from the parent ions are shown in Figure 1.8. Protonated G^{d}SK is calculated to be more stable than GSK by 1.86 kcal/mol. Loss of the first water molecule from the serine side chain by the glycine carbonyl oxygen is calculated to be endothermic for both diastereomers by 16.54 kcal/mol for GSK and 20.47 kcal/mol for G^{d}SK. The protonated oxazolidines that result from the initial water loss can dissociate

to lose a second water molecule, originating from the serine carbonyl group, to endothermically form a cyclic iminium ion by 21.33 and 20.43 kcal/mol, respectively. The calculated change in enthalpy for the first and second water loss is essentially the same for G^{d}SK, while the endothermicity differs by about 5 kcal/mol for GSK. Although kinetic control determines CAD fragmentation patterns, thermochemical variations on the order of 1 kcal/mol or more represent the appropriate magnitude needed to account for the observed effects. The thermochemical effects in CAD experiments are more apparent in a storage mass spectrometer. The ions trapped in a storage MS undergo multiple collisions and, therefore, access multiple excitation states. Thermodynamic control plays a larger role in a storage CAD experiment than in a tandem mass spectrometer, where only a single collision is responsible for fragmentation.

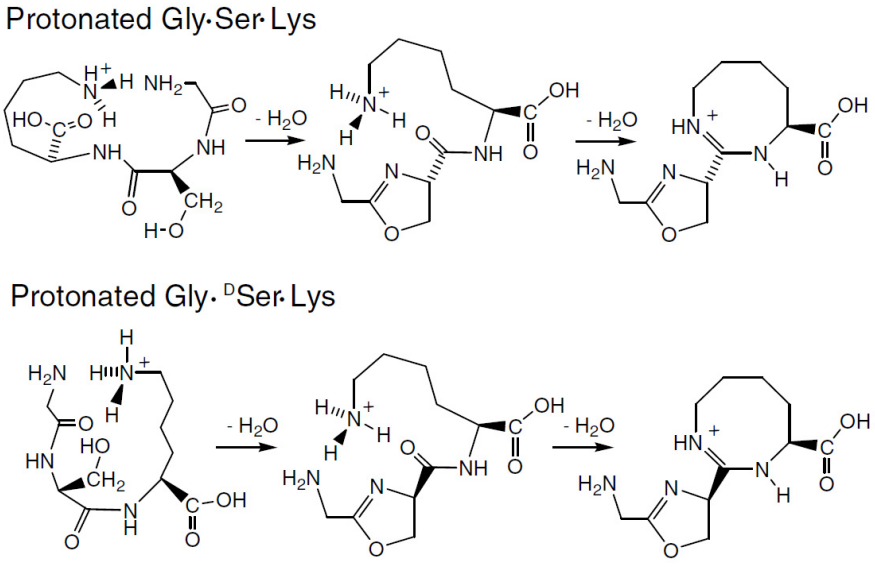


Figure 1.8. Representations of DFT (B3LYP/6-31G) calculated geometries of protonated GSK and GS^{d}K along with their subsequent loss of two water molecules.**

Mixtures of GLSK and GL^{d}SK were prepared and CAD experiments of the mixtures were performed on the linear ion trap instrument on the basis of the discussion above. The plot of the m/z 368:369 intensity ratios versus the mole fraction of GL^{d}SK is shown in Figure 1.9. The plot deviates from linearity because GL^{d}SK dissociates more easily than GLSK. The data points can be fit by the equation in the insert using a factor, *W*, that weights the contribution of GLSK differently from GL^{d}SK. The equation of the data fit can be used to quantify a mixture of the two diastereomers and could allow for a tryptic digest of α-synuclein recovered from a human autopsy to be surveyed to determine how much of the serine at the 9-position has epimerized.

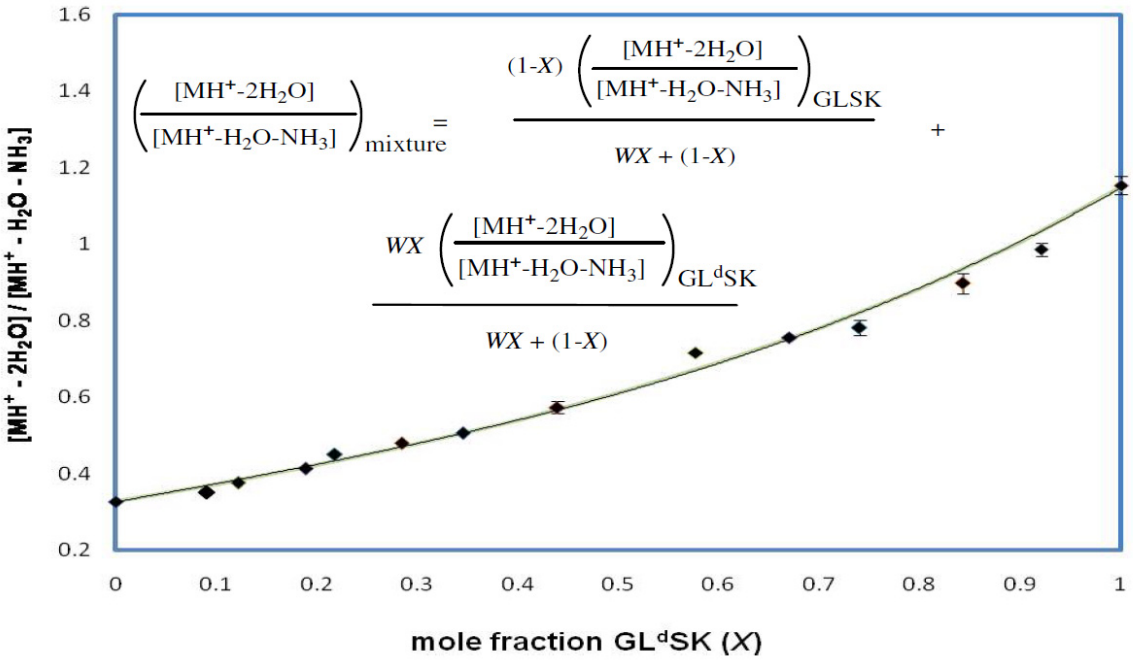


Figure 1.9. m/z 368:369 intensity ratios of various GLSK/GL^{d}SK mixtures performed on a linear ion trap mass spectrometer.

GLSK and GL^{d}SK were also subjected to CAD experiments using a hybrid linear trap/orbitrap mass spectrometer (LTQ Orbitrap™) at the Institute de Chimie des Substances Naturelles (ICSN) at Gif-sur-Yvette outside Paris, with the assistance of Dr. Isabelle Schmitz-Afonso. Mixtures of the two diasteromers were first run through a liquid chromatographic (LC) column prior to electrospray ionization and isolation in the ion trap to insure that the m/z 404 ions correspond to the tetrapeptides only. After optimizing LC conditions the trapped ions were subjected to inert gas collisions at various energies and activation times. Once the parent ions were fragmented in the linear trap they entered the orbitrap sector and the mass-to-charge ratio and intensity of the fragment ions were recorded. The collision energy and activation time had to be optimized prior to each experiment to ensure the parameters of the experiment resulted in greatest difference between the spectra of the diastereomers. Five repetitions of each of seven different mixtures of diastereomers had to be performed in series in order to ensure reproducibility. The spectra of GLSK and GL^{d}SK recorded on the LTQ-Orbitrap mass spectrometer are reproduced in Figure 1.10. The fragment ions and their relative intensities are similar to that recorded by the linear ion trap mass spectrometer mentioned above. The spectra performed on the LTQ-orbitrap shows a more resolved mass spectrum and a more robust difference between the diastereomers than the linear ion trap mass spectrometer. The biggest difference between the two diastereomers is the ratio between the MH^{+}-$2H_2O$:MH^{+}-H_2O-NH_3 ions, just as the linear ion trap spectra. The r value for the

LTQ-orbitrap spectra is 4.3, 16% larger than that observed in the linear ion trap instrument. The plot of the m/z 368:369 intensity ratios versus the mole fraction of GL^{d}SK is shown in Figure 1.11. The plot is linear and does not require a weighting factor to fit the data, unlike the plot of the data recorded on the linear ion trap instrument. The equation of the trend line is related to the percentage of GL^{d}SK in the mixture.

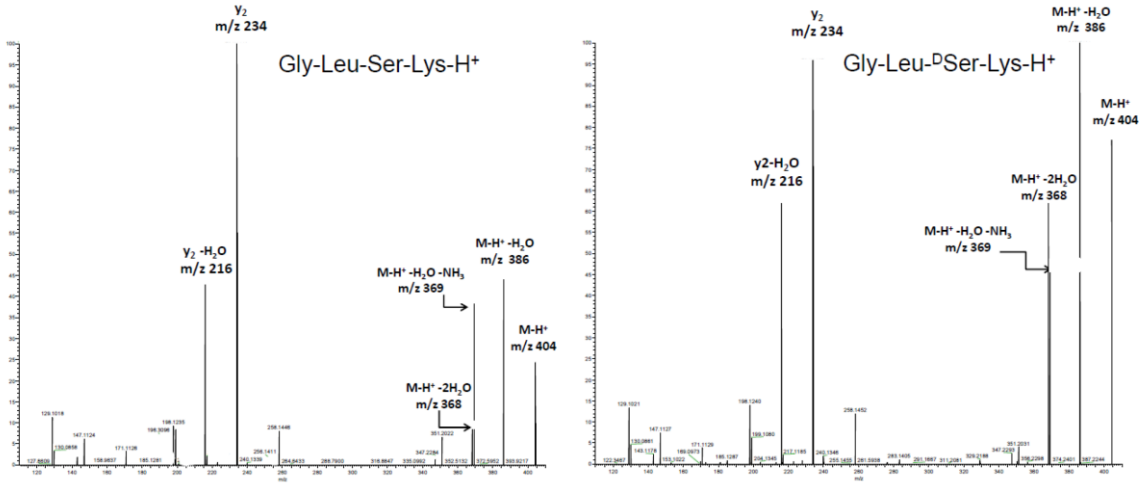


Figure 1.10. ESI MS/MS spectra of protonated GLSK and GL^{d}SK performed on a LTQ Orbitrap mass spectrometer

The equation of the trendline in Figure 1.11 is y= 0.0113x + 0.3249, where the percentage of GL^{d}SK in the mixture is x and the recorded m/z 368:369 intensity ratio is y. Rearrangement of the equation gives %GL^{d}SK= (m/z 368:369 ratio – 0.3249)/0.0113. By recording the average m/z 368:369 intensity ratio of unknown percentages of GL^{d}SK mixtures its mole fraction of the D-serine containing diastereomer can be obtained with the rearranged equation. Measurements of “unknown” mixtures of diastereomers were +/- 4% of their

actual values. The standard deviation of the measurements were the largest for the measurement of pure GLSK, while the other mixtures had standard deviations of under 4%. Proper LC separation of GLSK from tryptic digests of Lewy bodies extracted from autopsied brains could be submitted to the LTQ-orbitrap for CAD analysis. The average m/z 368:369 ratio could be compared to a plot of known mixtures of GLSK and GL^{d}SK prepared under identical conditions and the percentage of D-serine could be measured.

It's possible that he phosphorylation of serine-129 in α-synuclein occurs before or after fibril formation. The serine-129 is located in a more accessible location on the folded protein and the phosphorylation could be a result of the fibril formation and not the cause. If it occurs, epimerization of the serine-9 residue of α-synuclein, could play a role in fibril formation and onset or progression of PD. Even if phosphorylation doesn't play a role in PD, the methods discussed in this chapter can be used to determine the extent of epimerization of serine-9 of α-synuclein. These methods could prove useful in determining the mechanism of fibril formation and could lead to new therapy strategies in treating PD.

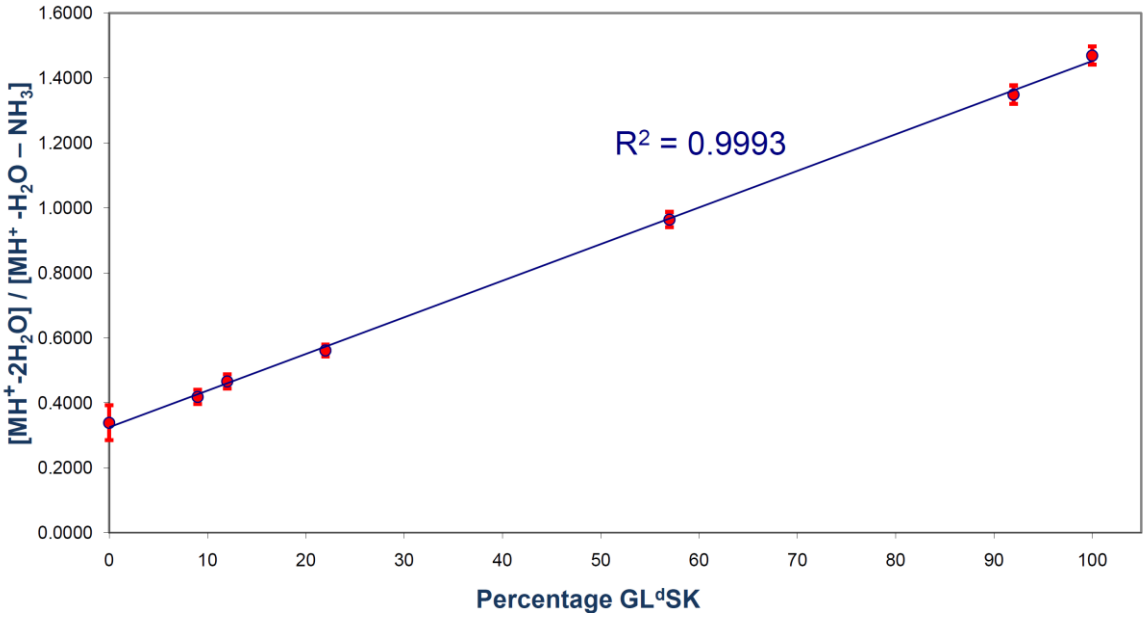


Figure 1.11. m/z 368:369 intensity ratios of various GLSK/GL^{d}SK mixtures performed on a linear ion trap-orbitrap hybrid mass spectrometer.

REFERENCES

1. Murkin, A. S.; Tanner, M. E., Dehydroalanine-based inhibition of a peptide epimerase from spider venom. *J. Org. Chem.* **2002,** *67*, 8389-8394.

2. Kuchel, P. W.; Torres, A. M.; Tsampazi, M.; Kennett, E. C.; Belov, K.; Geraghty, D. P.; Bansal, P. S.; Alewood, P. F., Characterization and isolation of L-to-D-amino-acid- esidue isomerase from platypus venom. *Amino Acids* **2007,** *32*, 63-68.

3. Stephenson, R. C.; Clarke, S., Succinimide Formation from Aspartyl and Asparaginyl Peptides as a Model for the Spontaneous Degradation of Proteins. *J. Bio.l Chem.* **1989,** *264*, 6164-6170.

4. Lowenson, J. D.; Clarke, S.; Roher, A. E., Chemical modifications of deposited amyloid-beta peptides. *Method. Enzymol.* **1999,** *309*, 89-105.

5. Sweedler, J. V.; Ewing, M. A.; Wang, J.; Sheeley, S. A., Detecting D-Amino acid-containing neuropeptides using selective enzymatic digestion. *Anal. Chem.* **2008,** *80*, 2874-2880.

6. Serafin, S. V.; Maranan, R.; Zhang, K. L.; Morton, T. H., Mass spectrometric differentiation of linear peptides composed of L-amino acids from isomers containing one D-amino acid residue. *Anal. Chem.* **2005,** *77*, 5480-5487.

7. Adams, C. M.; Zubarev, R. A., Distinguishing and quantifying peptides and proteins containing D-amino acids by tandem mass spectrometry. *Anal. Chem.* **2005,** *77*, 4571-4580.

8. Yost, R. A.; Enke, C. G., Selected Ion Fragmentation with a Tandem Quadrupole Mass-Spectrometer. *J. Am. Chem. Soc.* **1978,** *100*, 2274-2275.

9. March, R. E., An introduction to quadrupole ion trap mass spectrometry. *J. Mass Spectrom.* **1997,** *32*, 351-369.

10. Douglas, D. J.; Frank, A. J.; Mao, D. M., Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **2005,** *24*, 1-29.

11. Cooks, R. G.; Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005,** *40*, 430-443.

12. Makarov, A.; Denisov, E.; Kholomeev, A.; Baischun, W.; Lange, O.; Strupat, K.; Horning, S., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **2006,** *78*, 2113-2120.

13. Heerma, W.; Kulik, W., Identification of Amino-Acids in the Fast Atom Bombardment Mass-Spectra of Peptides. *Biomed. Environ. Mass.* **1988,** *16*, 155-159.

14. van Dongen, W. D.; van Wijk, J. I. T.; Green, B. N.; Heerma, W.; Haverkamp, J., Comparison between collision induced dissociation of electrosprayed protonated peptides in the up-front source region and in a low-energy collision cell. *Rapid Commun. Mass. Sp.* **1999,** *13*, 1712-1716.

15. Serafin, S. V.; Zhang, K. L.; Aurelio, L.; Hughes, A. B.; Morton, T. H., Decomposition of protonated threonine, its stereoisomers, and its homologues in the gas phase: Evidence for internal backside displacement. *Org. Lett.* **2004,** *6*, 1561-1564.

16. Taphanel, M. H.; Morizur, J. P.; Leblanc, D.; Borchardt, D.; Morton, T. H., Stereochemical analysis of monodeuterated isomers by GC/MS/MS. *Anal. Chem.* **1997,** *69*, 4191-4196.

17. Heck, S. D.; Siok, C. J.; Kelbaugh, P. R.; Thadeio, P. F.; Welch, M. J.; Williams, R. D.; Ganong, A. H.; Kelly, M. E.; Lanzetti, A. J.; Phillips, D.; Ahlijanian, M. K., Functional Consequences of Posttranslational Isomerization of Ser(46) in a Calcium-Channel Toxin *Science* **1994,** *266*, 1464-1464.

18. Kubo, T.; Kumagae, Y.; Miller, C. A.; Kaneko, I., beta-amyloid racemized at the Ser(26) residue in the brains of patients with Alzheimer disease: Implications in the pathogenesis of Alzheimer disease. *J. Neuropath. Exp. Neur.* **2003,** *62*, 248-259.

19. http://www.parkinsons.org/index.html.

20. Ravina, B.; Marder, K.; Fernandez, H. H.; Friedman, J. H.; McDonald, W.; Murphy, D.; Aarsland, D.; Babcock, D.; Cummings, J.; Endicott, J.; Factor, S.; Galpern, W.; Lees, A.; Marsh, L.; Stacy, M.; Gwinn-Hardy, K.; Voon, V.; Goetz, C., Diagnostic criteria for psychosis in Parkinson's disease: Report of an NINDS, NIMH work group. *Movement Disord.* **2007,** *22*, 1061-1068.

21. National Institute of Neurological Disorders and Stroke. www.ninds.nih.gov.

22. Wolters, E. C., PD-related psychosis: pathophysiology with therapeutical strategies. *J. Neura.l Transm-Supp.* **2006,** 71, 31-37.

23. Goldman, J., New Thoughts on Thought Disorders in Parkinson's Disease: Review of Current Research Strategies and Challenges. *Parkinson's Disease* **2011**, 12 pages.

24. Goldberg, M. S.; Lansbury, P. T., Is there a cause-and-effect relationship between alpha-synuclein fibrillization and Parkinson's disease? *Nat. Cell. Biol.* **2000,** *2*, E115-E119.

25. Iwai, A.; Masliah, E.; Yoshimoto, M.; Ge, N. F.; Flanagan, L.; Desilva, H. A. R.; Kittel, A.; Saitoh, T., The Precursor Protein of Non-a-Beta Component of Alzheimers-Disease Amyloid Is a Presynaptic Protein of the Central-Nervous-System. *Neuron.* **1995,** *14*, 467-475.

26. Fink, A. L.; Uversky, V. N.; Li, J., Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **2001,** *276*, 10737-10744.

27. Walter, J.; Schindzielorz, A.; Hartung, B.; Haass, C., Phosphorylation of the beta-amyloid precursor protein at the cell surface by ectocasein kinases 1 and 2. *J. Bio.l Chem.* **2000,** *275*, 23523-23529.

28. Nishie, M.; Mori, F.; Fujiwara, H.; Hasegawa, M.; Yoshimoto, M.; Iwatsubo, T.; Takahashi, H.; Wakabayashi, K., Accumulation of phosphorylated alpha-synuclein in the brain and peripheral ganglia of patients with multiple system atrophy. *Acta. Neuropathol.* **2004,** *107*, 292-298.

29. Roher, A. E.; Lowenson, J. D.; Clarke, S.; Wolkow, C.; Wang, R.; Cotter, R. J.; Reardon, I. M.; Zurcherneely, H. A.; Heinrikson, R. L.; Ball, M. J.; Greenberg, B. D., Structural Alterations in the Peptide Backbone of Beta-Amyloid Core Protein May Account for Its Deposition and Stability in Alzheimers-Disease. *J. Biol. Chem.* **1993,** *268*, 3072-3083.

30. Morton, T. H.; Julian, R. R.; Ly, T.; Finaldi, A. M., Dissociation of a protonated secondary amine in the gas phase via an ion-neutral complex. *International Journal of Mass Spectrometry* **2007,** *265*, 302-307.

31. Karolak, S.; Nefau, T.; Bailly, E.; Solgadi, A.; Levi, Y., Estimation of illicit drugs consumption by wastewater analysis in Paris area (France). *Forensic Sci. Int.* **2010,** *200*, 153-160.

Chapter II

Diamines

Introduction

Medium-sized cycloalkanes can adopt a variety of conformations. Their lowest energy conformations in the crystalline state may differ from their lowest energy conformations in the gas or solution phase. In some cases, a variety of molecular conformations or crystal packing arrangements may be exhibited for a single molecule.[1] Computational methods can, in principle, include the effects of crystal packing in their prediction of a cycloalkane's lowest energy conformation. Crystallography is often the gold standard for determining a molecule's geometry, but the presence of multiple phase transitions and crystal disorder compromises crystallography's effectiveness. Furthermore, the lowest energy conformation for an isolated molecule may not even be present in the crystalline form due to the fact that crystal forces may favor a higher energy conformation. In these cases, isolated molecule computations do not provide accurate predictions of solid-state conformations. Of the cycloalkanes between six and fourteen carbons, only four have known crystal structures (C_6H_{12}, $C_{10}H_{20}$, $C_{12}H_{24}$, and $C_{14}H_{28}$). For the remaining medium sized cycloalkanes, those without a solved crystal structure, another solid state experimental method is necessary to confirm the conformation in the solid-phase.

Vibrational spectroscopy can be used as a method of crystal structure analysis along with comparisons to computed molecular vibrational frequencies for various conformations.[2] Vibrational spectroscopy has the advantage over

NMR or diffraction data of representing disorder as a sum of contributions rather than being dynamically or spatially oriented. Inelastic Neutron Scattering (INS) has been used to establish what conformation is present in molecular crystals of cycloalkanes containing 6, 7, 8, 10, 12, and 14 carbons.[3] INS has a few advantages over IR and Raman spectroscopy stemming from the lack of interaction with the electrons of a molecule. Neutrons interact only with the nucleus of an atom and, therefore, INS has no selection rules. This allows for a more facile quantitative comparison between experiment and theory. The neutron undergoes a change in momentum when it excites a molecular vibration. The incoherent scattering cross section of hydrogen is roughly 80 barns (80 $x10^{-24}$ cm^2) for hydrogen, 2 barns for deuterium, and 10^{-3} for carbon. When investigating the vibrational spectrum of cycloalkanes hydrogen is the only atom contributing to the INS spectrum.

Of the four crystal phases of cyclohexane the phase observed below 186 K is referred to as phase II.[4] In comparing the INS spectrum of cyclohexane, recorded at the pulsed neutron source at the Rutherford Appelton Laboratories, with simulated spectra of the chair and twist-boat conformations calculated using Density Functional Theory (DFT), it was shown that the dominant conformation is the chair (Figure 2.1). When the computed vibrational frequencies of both conformers are compared with the INS spectrum the χ^2 value of the twist-boat frequencies is over three times greater than the chair conformer. DFT calculations (gas-phase) predict that the energy of the twist-boat conformation is

6.3 kcal/mol greater than that of the chair conformation. This is consistent with the interpretation that the chair is the dominant conformation of the cyclohexane crystal.

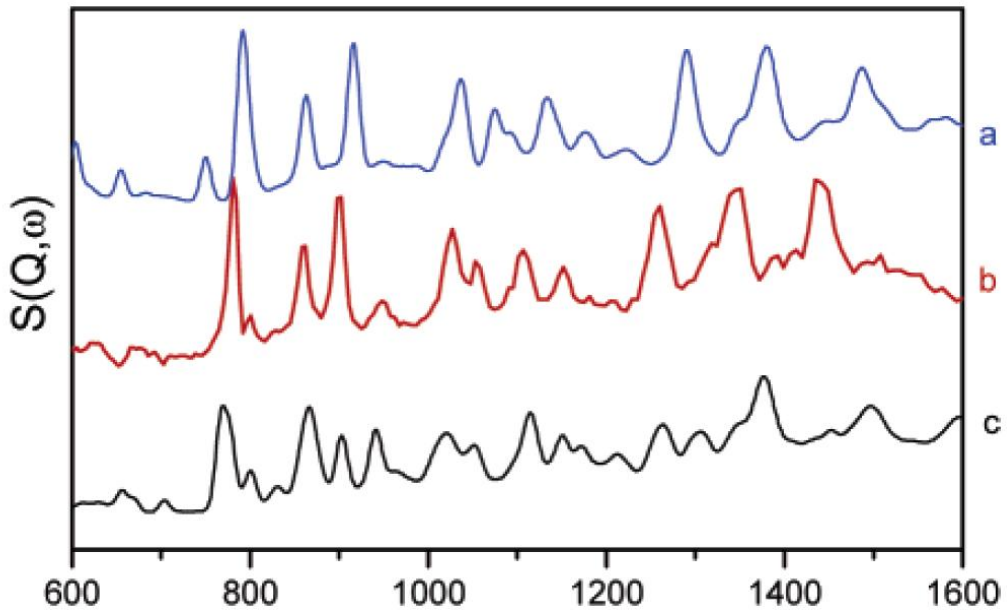


Figure 2.1. Chair (a, blue) and twist-boat (c) cyclohexane spectra calculated with B3LYP/6-311G(d,p) and simulated using aClimax, compared with the experiment cyclohexane spectrum (b, red).[5]

The crystal structure of cycloheptane is unknown. X-ray experiments have been performed on derivatives of cycloheptane in an attempt to determine the most likely conformation adopted by the unsubstituted cycloheptane. The most common conformation of the derivatives of cycloheptane is the twist-chair conformation.[5] The twist-chair conformation is also predicted from isolated molecule calculations, and DFT calculations done at the CCSD and MP2 levels of theory. [6] The INS spectrum of cycloheptane is compared with the four lowest energy conformations (boat, chair, twist-boat, and twist-chair) calculated by DFT

in Figure 2.2. The χ^2 value for the boat, chair, and twist boat conformations are greater than 1.9 times that of the twist-chair conformation, indicating that the twist-chair is the most dominant conformer in the cycloheptane crystal. The peak at 1450 cm^{-1} in the experimental spectrum of cycloheptane is not present in any of the theoretical spectra. The authors do not address this discrepancy, but it may arise due to a number of possibilities. The simplest explanation is that the authors did not properly scale their simulated spectra. It appears if the simulated spectra for the twist-chair conformation were scaled by 0.97 its theoretical peaks would better match the experimental spectrum above 900 cm^{-1}. The experimental spectrum overlaps better with the twist-chair conformation's theoretical spectra below 600 cm^{-1}. However, it is possible that another conformation present in the crystal of cycloheptane, which the authors failed to identify, and that this conformation is responsible for the experimental peaks that do not overlap with any of the peaks in the theoretical spectra.

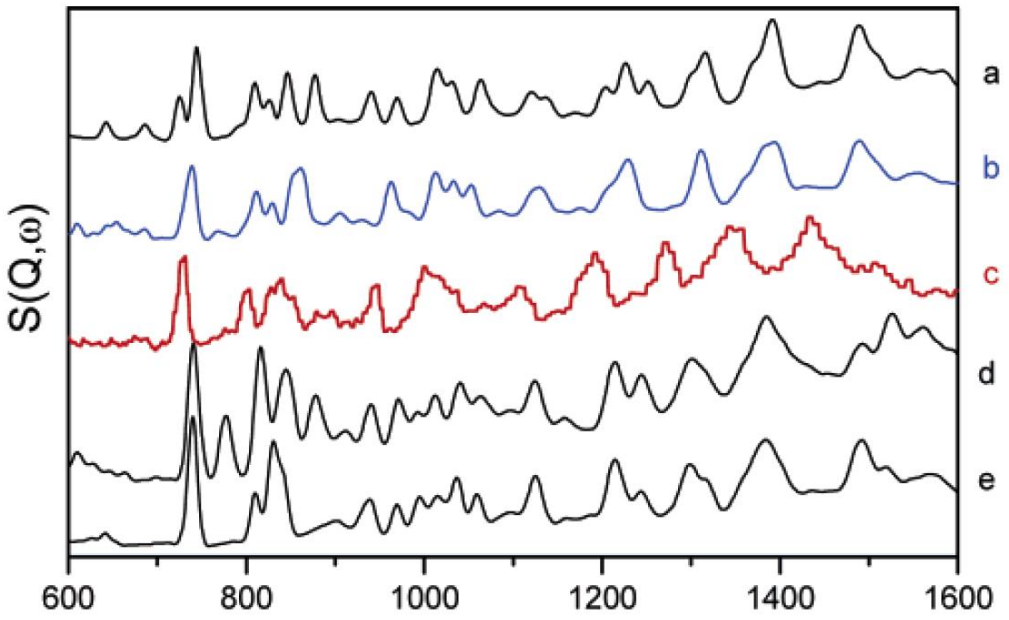


Figure 2.2. Chair (a), twist-chair (b), boat (d), and twist-boat (e) cylcoheptane spectra, calculated with B3LYP/6-311G and simulated using aClimax, compared with the experimental INS cycloheptane spectrum (c). The spectra displayed in the higher frequency range are magnified by a factor of 2 relative to those in the low-frequency panel.[3]**

Cyclooctane has no known crystal structure, although it has been investigated by optical spectroscopy.[7,8] Raman and IR spectroscopy identified three stable phases of solid cyclooctane and a metastable phase. Phase III is stable below 166 K and is the most likely phase present in the INS spectrum. Crystal structures of cyclooctane derivatives at low temperatures tend to adopt a boat-chair conformation.[9] Calculations have supported the conclusion that the boat-chair conformation is the most energetically favorable conformation.[6] The INS spectrum of cyclooctane appears to confirm this, although it appears the frequencies over 900 cm^{-1} in the theoretical spectrum of the boat-chair

conformation should be scaled by 0.97 in order to better overlap with the experimental spectrum (Figure 2.3). The experimental spectrum of cyclooctane has two peaks below 600 cm^{-1} that do not overlap with the theoretical spectrum of the boat-chair conformer. The experimental peak near 225 cm^{-1} overlaps well with a peak in the simulated spectra of the boat-boat conformer, and the experimental peak at 140 cm^{-1} does not overlap with any of the computed spectra. While the boat-chair conformation appears to be the dominant conformer of solid cyclooctane the data suggest at least one other conformer, most likely the boat-boat, exists at these temperatures.

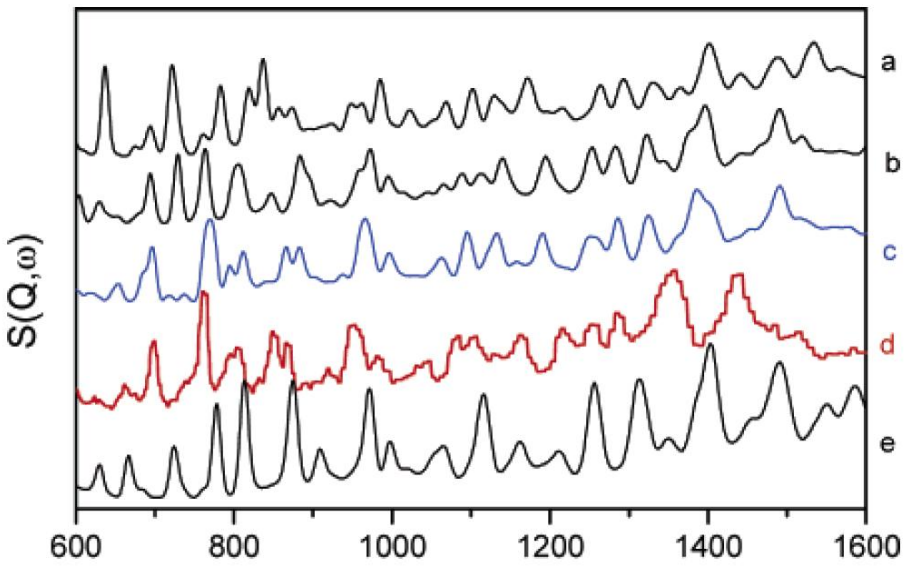


Figure 2.3. Boat-boat (a), twist-boat-chair (b), boat-chair (c), and crown (e) cyclooctane spectra, calculated with B3LYP/6-311G(d,p) and simulated using aClimax, compared with the experimental cyclooctane spectrum (d, red).[5]

The energetics of the conformations of cyclooctane were first calculated over 45 years ago.[10] Wiberg, as well as Hendrickson, Bixon, and Allinger using

subsequent calculations shortly following the initial report, predicted that the boat-chair conformation was the most stable.[11,12,13] The interconversion paths linking the ten lowest energy conformations were calculated by Anet and Krane in 1973.[14] Anet and Krane found the lowest energy path between the boat-chair and the crown family is obtained by driving the ω_1 angle of the twist-boat-chair conformation, while maintaining a C_2 axis of symmetry. Anet and Krane calculated the barrier for the interconversion from the twist-boat-chair conformation to the twist-chair-chair to be 10.3 kcal/mol.

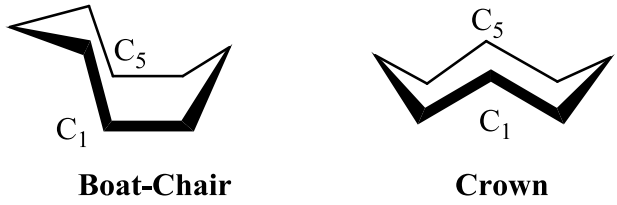


Figure 2.4. Lowest energy conformations of cyclooctane

Both the boat-chair and crown conformations bring the C_1 and C_5 carbons close together and are calculated to be the two lowest energy conformations of cyclooctane (Figure 2.4). The crystal structure of the 1,5-disubstituted molecule 1,5-cyclooctanedione dioxime is found in the boat-chair conformation where the substituents at the one and five positions are close together and the molecule has a mirror plane of symmetry (Figure 2.5).[15] The molecule *cis*-1,5-Diaminocyclooctane takes advantage of this geometry in a way that allows the two amino groups to share a proton between them in the gas-phase. The cooperativity between the two amino groups brought together, in part, by the

preferred geometry of the eight membered ring makes it the most basic primary amine, as determined by the kinetic method, to date.[16]

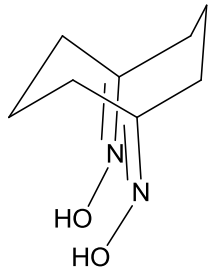


Figure 2.5. Geometry of the crystal structure of 1,5-cyclooctanedione dioxime

Background

Enzymatic catalysis often is capable of removing a proton from a site on a substrate that does not have a high acidity. For example, the N delta 1 of His57 and the beta-carboxyl group of Asp102 of chymotrypsin can cooperate to remove a proton during serine protease action to form a tetrahedral adduct.[17] The tetrahedral adduct has been shown to involve a short, strong hydrogen bond, in which a proton is bridged between two basic atoms having similar electronegativities.[18] It is known that, after formation of an imine bond with the cofactor PLP, the hydrogen on the α-carbon of glycine has a pK_a of 17. The conjugate acid ions of the two most basic amino acids, arginine and lysine, have pK_a values of 12.48 and 10.53 respectively. This implies that, on their own, arginine and lysine are not sufficiently basic to remove the hydrogen on the α-carbon. The Richards group hypothesizes that two basic amino acid side chains

work together in order to overcome their individual lack of basicity to form a strong hydrogen bond between two side chains and remove the hydrogen.[19] Strong hydrogen bonds favor a linear arrangement of the two electronegative atoms and the bridging proton.[20] In the case of two amino groups, when neutral, the nitrogens are positioned at a distance of D_{NN}, upon protonation the nitrogens move towards one another to a distance of d_{NN}, where the strong hydrogen bond stabilizes the conjugate acid (Figure 2.6). The increased basicity of linear diamines (such as 1,4-diaminobutane and 1,5-diaminopentane) is believed to be a result of strong hydrogen bonds that result in a cyclic structure upon protonation.[21] When protonated 1,4-diaminobutane forms a 7-membered ring and is the most basic of the linear diamines.

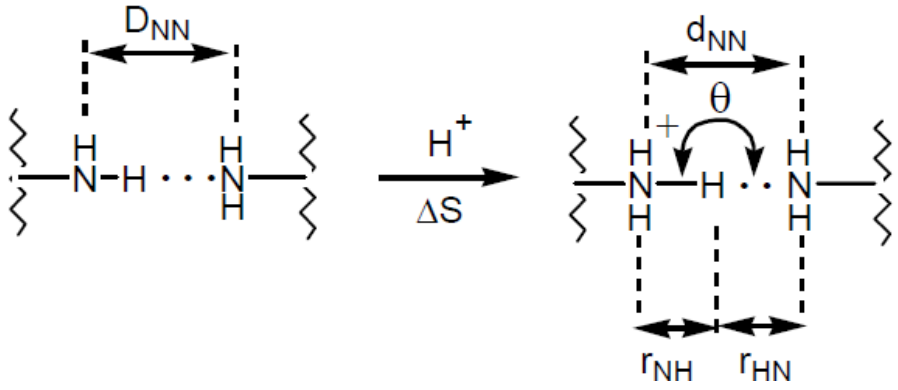


Figure 2.6. Geometric features of neutral and protonated diamines.[16]

Internally hydrogen bonded neutral diamines represent energetic minima, but entropy does not favor the cyclic structures for neutral linear diamines as illustrated in Figure 2.6. Monoprotonated diamines can crystallize in three

general motifs: salt-bridged, clustered, or cyclic. With linear tertiary diamines salt-bridged clusters become less favorable and, depending on the chain length, can form either cyclic or clustered crystals. The monoprotonated salt of *N,N,N',N'*-tetramethyl-1,4-diaminobutane (with a triflate counter ion) forms a cyclic crystal structure, while the triflate salt of *N,N,N',N'*-tetramethyl-1,5-diaminopentane forms a dimeric cluster.[22] The entropy change for making an internal hydrogen bond in $HO(CH_2)_4OCH_3$ in solutions is -16.2 e.u. at 298 K,[23] which is not far from the entropy change for transferring a proton from a monoamine to 1,4-diaminobutane (-14.3 e.u.).[24] This implies that the topicalogical change from a chain to a ring is responsible for most of the entropy change upon protonation of a linear diamine.

Taking advantage of the more rigid, cyclic structure of cyclooctane the molecule *cis*-1,5-diaminocyclooctane has a greater gas-phase basicity than does 1,4-diaminobutane. The greater basicity of *cis*-1,5-diaminocyclooctane is due to the more favorable entropy even though the proton affinity of 1,4-diaminobutane is greater than *cis*-1,5-diaminocyclooctane.[16] Poutsma and Andriole determined the proton affinities for 1,4-diaminobutane, 1,5-diaminopentane, *cis*-1,5-diaminocyclooctane, and canavanine using the extended kinetic method described elsewhere.[25] In this instance, proton-bound heterodimer of canavanine and the diamines were produced by electrospray, mass selected, and dissociated by collisions with background gas. The molecule that retains the proton after dissociation is judged to be the more basic. Dissociation of 1,4-

diaminobutane and 1,5-diaminopentane with canavanine produced much more protonated canavanine than diaminopentane. The dissociation of the *cis*-1,5-diaminocyclooctane/canavanine proton-bound dimer yielded more protonated *cis*-1,5-diaminocyclooctane than protonated canavanine. This means that *cis*-1,5-diaminocyclooctane is more basic than canavanine or either of the linear diamines. The proton affinity of *cis*-1,5-diaminocyclooctane was determined to be between that of 1,4-diaminobutane and 1,5-diaminopentane by a single reference variant of the kinetic method demonstrated by Wenthold and co-workers.[26]

The remainder of this chapter will discuss the determination of the vibrational frequencies of the N-H-N bonds in tertiary derivitives of 1,4-diaminobutane and *cis*-1,5-diaminocyclooctane. INS and Infrared Mulitiple Photon Dissociation (IRMPD) spectra of these two compounds will be compared with DFT vibrational calculations and potential energy curves in order to identify the vibrational frequencies associated with internal hydrogen bonds and to determine if these N-H-N bonds are low-barrier hydrogen bonds (LBHB).

Standard infrared methods that rely on direct absorption of the incident light beam require samples with a relatively high density. When an analyte is difficult to isolate, or impossible to isolate without a counterion or solvent present, infrared spectroscopy can be performed on mass selected ions stored in an ion trap through infrared multiple photon dissociation (IRMPD).[27] Molecular ions can

undergo dissociation when the wavelength of radiation is in resonance with an absorption band. Monitoring the dissociation yield in the mass spectrometer as a function of the wavelength allows for an infrared spectrum to be recorded.[28] Free-electron lasers can be a source of radiation in the infrared and far-infrared spectral regions, tunable by samarium-cobalt undulator magnets.[29] In 2000 it was shown that a free-electron laser can be tuned to induce multiple photon dissociation of covalently bound mass-selected ions stored in an ion trap. The Free-Electron Laser for Infrared eXperiments (FELIX) facility at FOM Rijnhuizen can be used to study trapped molecular ions with infrared photo-dissociation (generated from the free-electron laser) spectroscopy. IRMPD spectroscopy done at FELIX couples a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR) to the free-electron laser scanned over the infrared fingerprint region of 300-1800 cm^{-1} (Figure 2.7).

The FT-ICR consists of a 4.7 T superconducting magnet, a home built vacuum system housing an open-ended ICR cell, and is equipped with an electrospray ion source.[30] Precursor ions from the electrospray source are accumulated in an RF hexapole before being transferred to an ICR cell via a quadrupole ion deflector and a 1 meter RF octapole ion guide. Once in the ICR cell, ions can be mass isolated by a stored waveform inverse Fourier transform (SWIFT) excite pulse that ejects all other masses from the trap. After the desired ions are isolated they are exposed to the focused output of the free-electron laser, which crosses the trapped ion cloud many times as it is reflected off the

inner surfaces of the ICR cell. This sequence is repeated while the wavelength of the free-electron laser is scanned. An IRMPD spectrum is obtained by plotting the fragment ion intensity as a function of infrared wavelength.

One drawback of mass spectrometry is that it can only record the mass-to-charge ratio of ions that result from dissociation of a parent ion. While this information provides accurate atom composition of a dissociation product it does not offer any information about its structure. IRMPD experiments performed with the free-electron laser can provide both a mass-to-charge ratio of a dissociation product, while also providing structural information by way of IR spectroscopy.

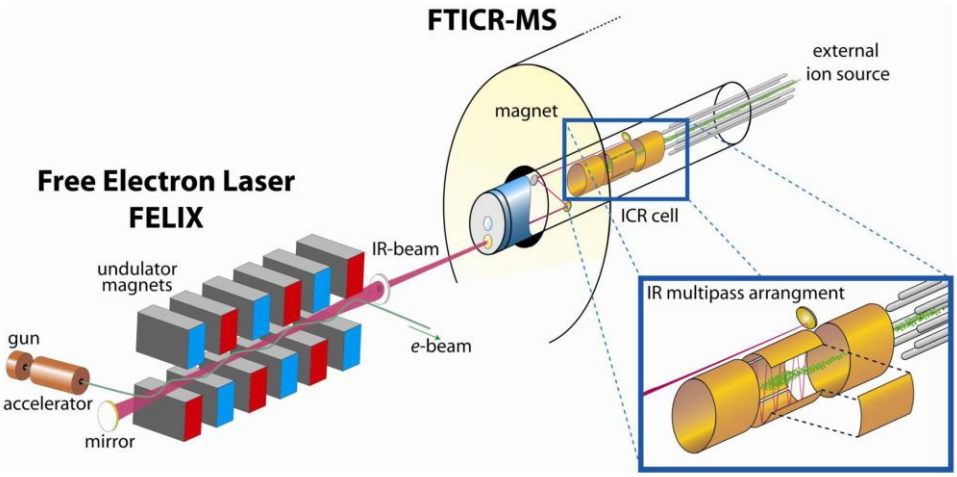


Figure 2.7. Diagram of the experimental apparatus, consisting of the free-electron laser on the left and the FTICR mass spectrometer on the right. The overall length of the free-electron laser is over 20 m, whereas the length of the FTICR magnet is about 1 m.[28]

INS was performed on the Filter Difference Spectrometer (FDS) at the Los Alamos Neutron Scattering Center (LANSCE) (Figure 2.8). Neutrons are

produced by accelerating protons and colliding them with a tungsten target. Neutrons then enter the FDS through a small window, interact with the sample, and pass through low-energy band-pass filters. The scattered neutrons are sorted by their total time of flight from source to detector. Incoming neutrons excite molecular vibrations in molecules by giving up energy upon scattering from the sample. Because the proton source is pulsed, the time that the neutrons enter the FDS is known, and therefore, the time it takes the neutrons to hit the detector is related to the energy lost by the neutron and to the frequency of the molecular vibrations.

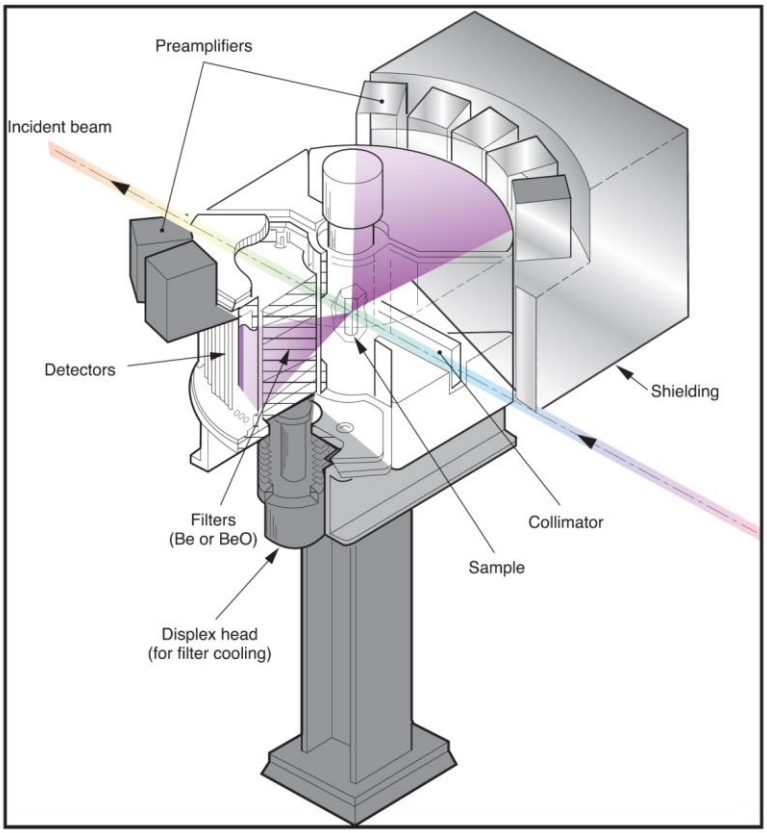


Figure 2.8. Schematic representation of the Filter Difference Spectrometer at LANSCE

The total neutron scattering cross section of hydrogen is more than 40 times that of deuterium, about 160 times that of nitrogen and roughly 14.5 times that of carbon. This makes the motions of any atom other than hydrogen essentially nonexistent in the INS spectrum. After recording the spectrum of a protonated diamine the proton can be substituted with deuterium causing the vibrations associated with that proton to disappear. The motions associated with the bridging proton can be more easily identified than they could be with IR or RAMAN spectroscopy because the motions of the non-hydrogen atoms would only shift based on their reduced mass, but would still produce intense peaks in the spectrum. The lack of "noise" corresponding to the motions of the non-hydrogen atoms in the diamines should allow for the identification of the N-H-N vibrations.

Results

N,N,N',N'-Tetramethyl-1,4-diaminobutane: Inelastic Neutron Scattering

Slightly less than one equivalent of hydriodic acid was added to the commercially available *N,N,N',N'*-tetramethyl-1,4-diaminobutane (also known as tetramethylputrescine) in methanol and the resulting salt was recrystallized from ethanol. The ratio of iodide counterion to diamine was determined to be 1:1 based on gravimetric analysis with silver nitrate. 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.

The monodeuteronated diamine salt was prepared by bubbling gaseous HI (from reacting I_2 with tetralin) into a solution of excess tetramethylputrescine in EtOD, and the resulting crystalling precipitate was subsequently recrystallized three times from ethanol-OD containing a slight excess of neutral tetramethylputrescine to produce the monoiodide of **1b.** An equivalent INS spectrum was then collected with the aim of comparing the two spectra reproduced in Figure 2.9. Subtraction (not shown) eliminated all bands, except for those strongly coupled to displacements of the hydrogen bonded proton. 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.

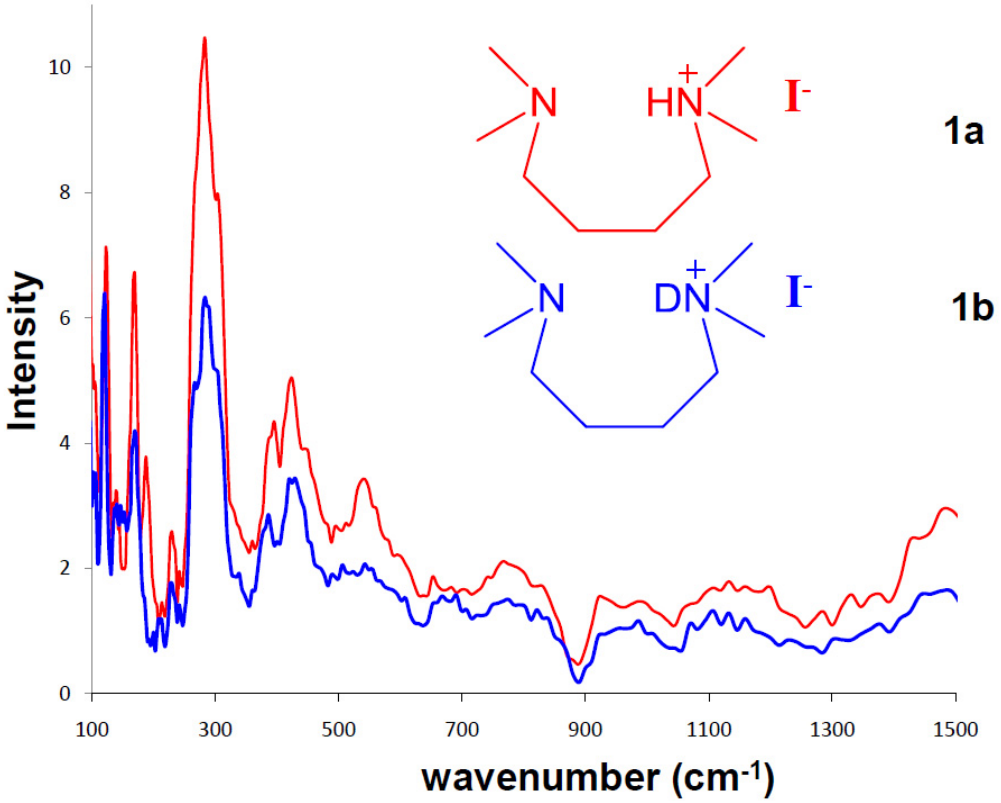


Figure 2.9. Inelastic Neutron Scattering (INS) spectrum of the monoiodide salt of monoprotonated tetra-methylputrescinium (1a, red trace) and of the monoiodide salt of monodeuteronated tetramethylputrescinium (1b, blue trace).

Two bands essentially disappear when the bridging proton of **1a** is replaced by a deuterium 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** (540 cm^{-1}).[31] The band at 525 cm^{-1} also coincides with Dr. Sepideh Yaghmaei's predicted value for the N-H-N asymmetric stretch based on her analysis of a calculated two-dimensional potential energy surface (544 cm^{-1}).[31]

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 N-H-N stretch means that either (1) linear motion of the bridging proton remains unaffected by the presence of the iodide counterion and (2) the Gruneisen parameter for that vibration is small; or else (3) that a variety of effects (changes in the harmonicity of the potential energy curve, interactions between ions within the lattice, coupling among vibrations, etc.) fortuitously cancel. Pressure dependant Raman spectroscopy performed by Dr. Andrew Rice confirmed the assignment of the band observed at 525 cm^{-1} in the INS spectrum as the N-H-N asymmetric stretch.

N,N,N',N'-Tetramethyl-*cis*-1,5-diaminocyclooctane: 1-D Potential Energy Surface

One-dimensional potential energy surfaces for the motion of proton from one basic site to another can be plotted as the electronic energy of the PBD (potential energy, V, in kJ/mol) versus the asymmetric stretch coordinates (S_2= (R-r)/$\sqrt{2}$, in angstroms) as the proton, shared between the two ring nitrogens moves from one side of the dimer to the other (Figure 2.10). Initially, the equilibrium (EQ) and transition state (TS) geometries, as well as their corresponding electronic energies, were calculated. The N-H distance of the shared proton was then varied and electronic energies were calculated (B3LYP/6-311++G**) along the path of proton transit. After the data points are

graphically fit by a polynomial, the energy levels and wave functions for the potential energy surface are calculated by solving the one-dimensional Schrödinger equation for Equation 1 in terms of the asymmetric stretch (S_2) using the Mathematica 7.0 program with a set of 30-40 basis functions.

$$E\ (\Psi) = (-\hbar^2/2m_2)\ (\sigma^2\Psi/\sigma S_2{}^2) + V\Psi \qquad \textbf{(1)}$$

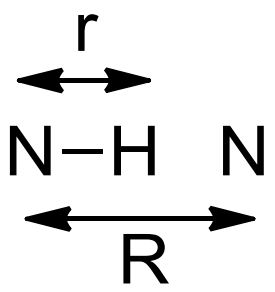


Figure 2. 10. Diagram of the N-H stretch coordinates, S_2, defined as (R-r)/√2

The 1-dimensional potential energy surface (PES), along with its energy levels, for the monoprotonated ion of *N,N,N',N'*-tetramethyl-*cis*-1,5-diaminocyclooctane (Me_4DACO H^+, **2a**) is shown in Figure 2.11. The calculated zero-point energy (ZPE) level lies 0.71 kJ/mol below the transition state based on a proton effective mass of 14/15. Low-barrier Hydrogen bonds (short, strong hydrogen bonds) tend to have ZPE's near or above the barrier top so that the bridging proton has enough energy to move from one side of the hydrogen bond to the other at room temperature. The ZPE predicted from the 1-dimensional PES analysis is close enough to the transition state energy level that is unclear if

2a should contain a LBHB. The PES can be used to predict N-H-N asymmetric stretching frequency by calculating the energy difference between the first energy level below the barrier top (v=0) and the first level above it (v=1).[31] The energy difference between the v=0 and the v=1 levels predicted from the 1-dimensional PES analysis is equal to 372 cm^{-1}. IRMPD and INS spectra of the protonated and deuterated ions of Me_4DACO were recorded in order to determine the accuracy of this prediction.

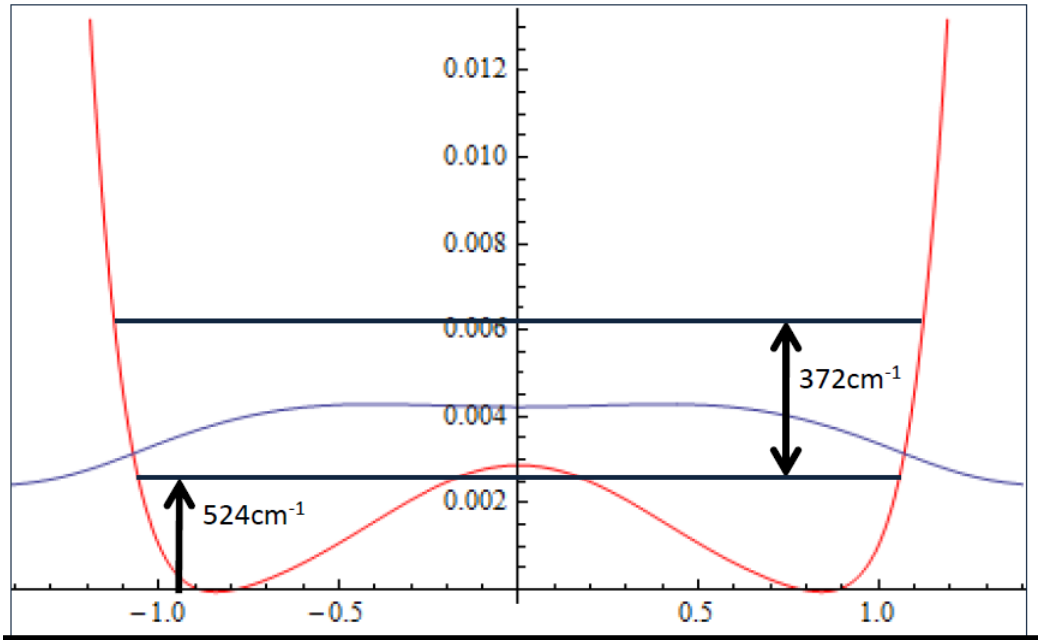


Figure 2.11. One-dimensional potential energy surface for motion of the proton Me_4DACO H^+ along with predicted energetic transitions from the ground state to the first vibrationally excited state. Plotted as N-H coordinates versus energy.

N,N,N',N'-Tetramethyl-*cis*-1,5-diaminocyclooctane: IRMPD

A solution of *N,N,N',N'*-tetramethyl-*cis*-1,5-diaminocyclooctane (Me_4DACO) dissolved in a mixture of water and methanol was electrosprayed, the protonated ion (m/z 199, **2a**) mass-selected in an ICR cell, and the isolated ions subjected to the free electron laser for analysis by IRMPD. Me_4DACO was also dissolved in ethanol-OD to exchange the H-bond proton with deuterium. An equivalent INS spectrum was then collected on the deuterated ion (m/z 200, **2b**) isolated in the ICR cell with the aim of comparing the two spectra reproduced in Figure 2.12.

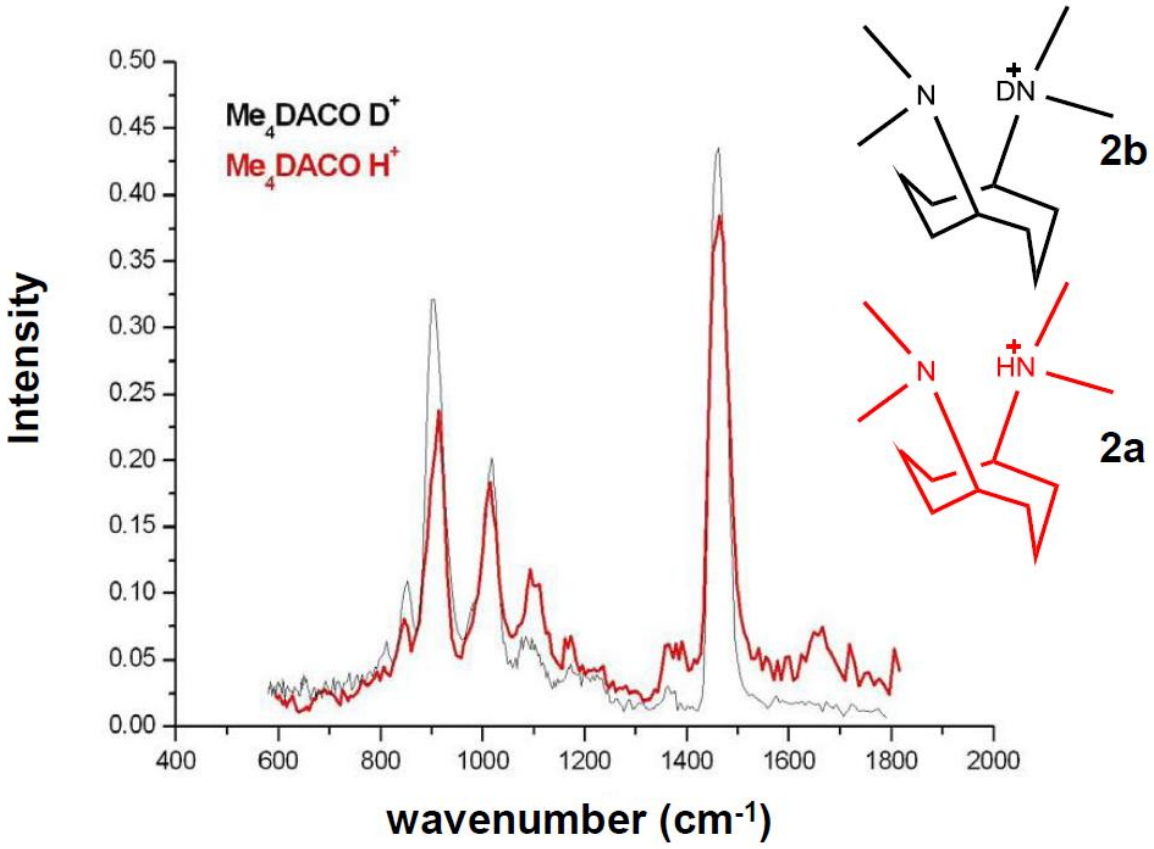


Figure 2.12. Infrared Multiple Photon Dissociation (IRMPD) spectrum of the monoprotonated tetramethyldiaminocyclooctane (2a, red trace) and of the monodeuteronated tetramethyldiaminocyclooctane (2b, black trace).

Unfortunately the IRMPD experiment was not able to investigate the region below 600 cm^{-1}, where analysis of the 1-dimensional PES predicts the N-H-N asymmetric stretch to be observed. The band at 1670 cm^{-1} band essentially disappear when the bridging proton of **2a** is replaced by a deuteron in **2b.** The only difference between the two IRMPD spectra lies in the bridging atom therefore, the band that disappeared must correspond to a motion of that atom.

The band at 1670 cm^{-1} can be assigned to one of the N-H-N bending motions. The validity of various approaches to theoretical predictions will be discussed more thoroughly in Chapter III.

N,N,N',N'-Tetramethyl-*cis*-1,5-diaminocyclooctane: INS

N-methyl torsions are predicted to occur near the region where the N-H-N asymmetric stretch is expected to be observed. Due to the low level of incoherent scattering by deuterium (compared to hydrogen) the twelve methyl hydrogens were replaced with deuterium in order to decrease the noise level around that region. The mono-protonated *N,N,N',N'*-Tetramethyl-*cis*-1,5-diaminocyclooctane-d_{12} (**2c**) monoiodide salt was loaded in an aluminum sample container under helium atmosphere for data collection at 10 K. The monoprotonated d_{12}-diamine salt was subsequently recrystallized three times from ethanol-OD to exchange the H-bond proton with deuterium, yielding the monodeuteronated (**2d**) monoiodide salt. An equivalent INS spectrum was then collected with the aim of comparing the two spectra reproduced in Figure 2.13. Because the amount of sample was limited, stoichiometry was not assayed by the silver iodide precipitations method (as it had been for the salt of tetramethylputrescine).

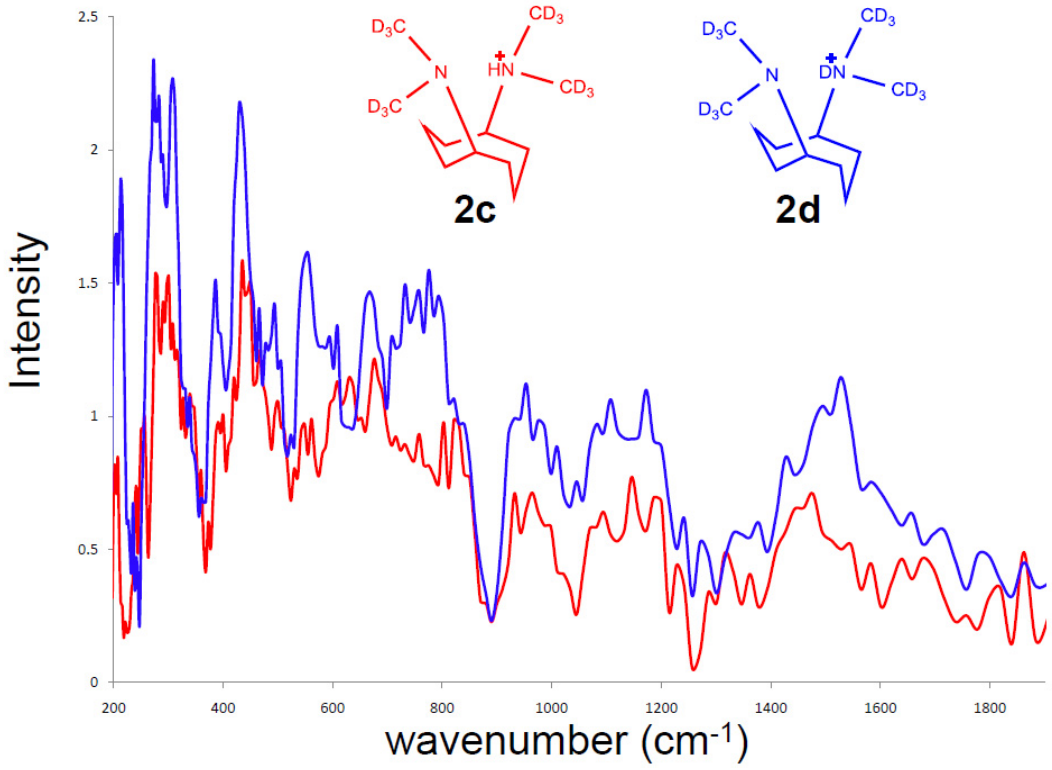


Figure 2.13. Inelastic Neutron Scattering (INS) spectrum of the monoiodide salt of monoprotonated tetramethyldiaminocylooctane (2c, red trace) and of the monoiodide salt of monodeuteronated tetramethyldiaminocylooctane (2d, blue trace).

The INS spectra does not provide a clear picture as much of the monodeuterated spectrum has a greater intensity than the monoprotonated spectrum. One very small band at 1750 cm^{-1} in the spectrum of **2c** disappears in the **2d** spectrum. No peaks under 500 cm^{-1} disappear after the bridging proton is replaced by deuterium, and analysis of the INS spectra does result in a clear conclusion. Two plausible explanations could be responsible for the lack of conclusive data. It is possible that the bridging proton did not exchange with deuterium during the recrystalization from ethanol-OD. If this were true the

fluctuation in the intensity between the two samples could be attributed to the difference in the duration spent in the neutron beam and/or the intensity of the neutron beam during data collection. The other possibility is that what is assumed to be the monoprotonated diamine might actually have been diprotonated.

Not all proton bridges necessarily form low-barrier hydrogen bonds, but the structural features of salts of monoprotonated tetramethylputrescine, the INS described here, along with gas-phase vibrational spectroscopy of isolated ions, and Raman spectroscopy done by others, 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 vibrational frequency of the N-H-N stretch is identified at 525 cm^{-1} for the monoprotonated tetramethylputrescine monoiodide salt and supports the prediction by Dr. Sepideh Yaghmaei based upon her analysis of its two-dimensional potential energy surface.

Synthesis and Characterization

General

Formic acid and formaldehyde were purchased from Fisher. Formic acid-d_2 and paraformaldehyde were purchased from Cambridge Isotope Laboratories. 1M 9-Borabicyclo[3.3.1]nonane in THF and hydroxylamine-O-sulfonic acid were purchased from Acros. 300 MHz 1H NMR spectra were recorded using an Inova 300 (2 channel triple axis gradient spectrometer). Exact masses were recorded on a Waters GCT Premier ToF instrument using direct injection of samples in methanol/water into the electrospray source.

A. N,N,N',N'-Tetramethyl-*cis*-1,5-diaminocyclooctane

cis-1,5-Diaminocyclooctane was prepared according to procedures reported elsewhere. [31] To 0.60 g (11.4 mmol) of 90% formic acid in a 25 ml round bottom flask, equipped with a stir bar, is added slowly 200 mg (1.14 mmol) of *cis*-1,5-diaminocyclooctane. To the resulting clear solution is added 412 mg (6.84 mmol) of formaldehyde (37 % in methanol/water). The flask is equipped with a reflux condenser and placed in an oil bath heated to 90^{o} C. Evolution of carbon dioxide begins after 2-3 minutes, at which time the flask is removed from the bath until the gas evolution subsides; then it is returned to the bath and heated at 100^{o} C for 8 hours.

After the solutions has been cooled, 15 ml of 3M hydrochloric acid is added and the solution is evaporated to dryness under reduced pressure. The yellow syrup residue is dissolved in 10 ml of water and the organic base is liberated by the addition of 5 ml of saturated sodium hydroxide solution. The upper (organic) phase is separated and the lower (aqueous) phase is extracted with two 5 ml portions of THF. The combined organic layers are dried over 1 g of anhydrous sodium sulfate. The sodium sulfate is removed by vacuum filtration, the THF removed under reduced pressure, and the product distilled. The yield of pale yellow *N,N,N',N'*-Tetramethyl-*cis*-1,5-diaminocyclooctane boiling at 95-97^{o} / 0.1 mm. is 60 mg (27 % yield).

^{1}H NMR (300MHz, $CHCl_3$): δ 2.47 (2H m) δ 2.2 (12H s) 1.3-1.9 (12H m)

APCI-ESI MS: MH^+: 199.2087 (calc. 199.2169)

B. N,N,N',N'-Tetramethyl-*cis*-1,5-diaminocyclooctane-d_{12}

To 0.930 g (19.4 mmol) of 95% formic acid-d_2 in a 25 ml round bottom flask, equipped with a stir bar, is added slowly 275 mg (1.94 mmol) of *cis*-1,5-diaminocyclooctane. To the resulting clear solution is added 372 mg (11.6 mmol) of solid paraformaldehyde-d_2. The flask is equipped with a reflux condenser and is placed in an oil bath heated to 90^{o} C. Evolution of carbon dioxide begins after

2-3 minutes, at which time the flask is removed from the bath until the gas evolution subsides; then it is returned to the bath and heated at 100° C for 8 hours.

The reaction is worked up in the same manner as described above. The yield of pale yellow *N,N,N',N'*-Tetramethyl-*cis*-1,5-diaminocyclooctane-d_{12} boiling at 96-98° / 0.1 mm. is 90 mg (22 % yield).

^{1}H NMR (300MHz, $CHCl_3$): δ 2.47 (2H m) δ 1.3-1.9 (12H m)

APCI-ESI MS: MH^+:211.2904 (calc. 211.2922)

C. cis-1,5-Diaminocyclooctane

A new synthesis of *cis*-1,5-diaminocyclooctane was developed based on H.C. Brown's hydroboration-amination reaction.[32] To a 100 ml round bottom flask containing 25 ml (25 mmol) of a 1M 9-Borabicyclo[3.3.1]nonane (9-BBN) solution in THF is added 7.07 g (63 mmol) of hydroxylamine-*O*-sulfonic acid. The reaction was heated under reflux for 3 hours.

After the solution has been cooled, 25 ml of 3M hydrochloric acid is added and the solution is evaporated to dryness under reduced pressure. The yellow syrup residue is dissolved in 10 ml of water and the organic base is liberated by the addition of 5 ml of saturated sodium hydroxide solution. The upper (organic)

phase is separated and the lower (aqueous) phase is extracted with two 5 ml portions of THF. The combined organic layers are dried over 1 g of anhydrous sodium sulfate. The sodium sulfate is removed by vacuum filtration, the THF is removed under reduced pressure, and the product distilled. The yield of a clear oil boiling at 62-64^{o}/ 0.5 mm is 0.294 g (8.1 % yield).

^{1}H NMR (300MHz, $CHCl_3$): δ 2.92 (2H m) δ 1.2-2.0 (12H m)

APCI-ESI MS: MH^+: 143.1527 (calc. 143.1543)

References

1. Howard, J. A. K.; Bilton, C.; Madhavi, N. N. L.; Nangia, A.; Desiraju, G. R.; Allen, F. H.; Wilson, C. C., When is a polymorph not a polymorph? Helical trimeric O-H center dot center dot center dot O synthons in trans-1,4-diethynylcyclohexane-1,4-diol. *Chem. Commun.* **1999,** 1675-1676.

2. Decius, J. C., *Molecular Vibrations in Crystals*. McGraw-Hill: New York, 1977.

3. Hudson, B. S.; Verdal, N.; Wilke, J. J., Conformational distinguishability of medium cycloalkanes in crystals via inelastic neutron scattering. *J. Phys. Chem. A* **2006,** *110*, 2639-2646.

4. Kahn, R.; Fourme, R.; Andre, D.; Renaud, M., Crystal Structures of Cyclohexane I and Ii. *Acta. Crystallogr. B* **1973,** *B 29*, 131-138.

5. Thomas, S. A.; Ajibola, V. O., Molecular conformation of the cycloheptane ring in the solid state. *B. Chem. Soc. Ethiopia* **1999,** *13*, 71-82.

6. Wiberg, K. B., The C7-C10 cycloalkanes revisited. *J. Org. Chem.* **2003,** *68*, 9322-9329.

7. Huang, Y. N.; Gilson, D. F. R.; Butler, I. S., Vibrational Spectroscopic Studies of Phase-Transitions in Solid Cyclooctane. *J. Phys. Chem.-US.* **1991,** *95*, 5051-5054.

8. Huang, Y. N.; Butler, I. S.; Gilson, D. F. R., A Vibrational Spectroscopic Investigation of Solid Cyclooctanone. *J. Phys. Chem.-US.* **1992,** *96*, 1151-1156.

9. Srinivas.R; Srikrish.T, Studies in Molecular Structure, Symmetry and Conformation .4. Conformation of Cyclooctane Ring System from X-Ray Studies. *Tetrahedron.* **1971,** *27*, 1009.

10. Wiberg, K. B., A Scheme for Strain Energy Minimization - Application to Cycloalkanes. *J. Am. Chem. Soc.* **1965,** *87*, 1070.

11. Bixon, M.; Lifson, S., Potential Functions and Conformations in Cycloalkanes. *Tetrahedron.* **1967,** *23*, 769.

12. Hendrick.J.B., Molecular Geometry .V. Evaluation of Functions and Conformations of Medium Rings. *J. Am. Chem. Soc.* **1967,** *89*, 7036

13. Allinger, N. L.; Hirsch, J. A.; Miller, M. A.; Tyminski, I. J.; Vancatle.Fa, Conformational Analysis .60. Improved Calculations of Structures and Energies of Hydrocarbons by Westheimer Method. *J. Am. Chem. Soc.* **1968,** *90*, 1199.

14. Anet, F. A. L.; Krane, J., Strain Energy Calculation of Conformations and Conformational-Changes in Cyclooctane. *Tetrahedron Lett.* **1973,** 50, 5029-5032.

15. Winkler, F. K.; Seiler, P.; Chesick, J. P.; Dunitz, J. D., Medium Ring Compounds .29. Cyclooctane-1,5-Dione-Dioxime. *Helv. Chim. Acta.* **1976,** *59*, 1417-1423.

16. Poutsma, J. C.; Andriole, E. J.; Sissung, T.; Morton, T. H., cis-1,5-diaminocyclooctane: the most basic gaseous primary amine? *Chem. Commun.* **2003,** 16, 2040-2041.

17. Frey, P. A.; Whitt, S. A.; Tobin, J. B., A Low-Barrier Hydrogen-Bond in the Catalytic Triad of Serine Proteases. *Science.* **1994,** *264*, 1927-1930.

18. Houk, R. J. T.; Monzingo, A.; Anslyn, E. V., Electrophilic coordination catalysis: A summary of previous thought and a new angle of analysis. *Accounts. Chem. Res.* **2008,** *41*, 401-410.

19. Richard, J. P.; Toth, K., Covalent catalysis by pyridoxal: Evaluation of the effect of the cofactor on the carbon acidity of glycine. *J. Am. Chem. Soc.* **2007,** *129*, 3013-3021.

20. Morton, T. H.; Beaucham.J. L., "Chemical Consequences of Strong Hydrogen Bonding in the Reactions of Organic Ions in the Gas Phase. Interaction of Remote Functional Groups," *J. Am. Chem. Soc.* **1972**, *94*, 3671-3672.

21. Aue, D. H.; Webb, H. M.; Bowers, M. T., Quantitative Evaluation of Intramolecular Strong Hydrogen-Bonding in Gas-Phase. *J. Am. Chem. Soc.* **1973,** *95*, 2699-2701.

22. Morton, T. H.; Yaghmaei, S.; Khodagholian, S.; Kaiser, J. M.; Tham, F. S.; Mueller, L. J., Chelation of a proton by an aliphatic tertiary diamine. *J. Am. Chem. Soc.* **2008,** *130*, 7836.

23. Belhekar, A. A.; Jose, C. I., Spectral Study of Hydrogen-Bonding in Monopotassium Salts of Cycloalkane 1-1 Dicarboxylic-Acids. *J. Chem. Soc. Farad. T2.* **1982,** *78*, 201-204.

24. Meot-Ner, M.; Hamlet, P.; Hunter, E. P.; Field, F. H., Internal and External Solvation of Polyfunctional Ions. *J. Am. Chem. Soc.* **1980,** *102*, 6393-6399.

25. Cooks, R. G.; Patrick, J. S.; Kotiaho, T.; Mcluckey, S. A., Thermochemical Determinations by the Kinetic Method. *Mass Spectrom. Rev.* **1994,** *13*, 287-339.

26. Wenthold, P. G.; Lardin, H. A.; Squires, R. R., Determination of the electron affinities of alpha- and beta-naphthyl radicals using the kinetic method with full entropy analysis. The C-H bond dissociation energies of naphthalene. *J. Mass Spectrom.* **2001,** *36*, 607-615.

27. Bagratashvilli, V. N.; Letokhoc, V. S.; Markov, A. A.; Ryabov, E. A. *Mulitple Photon Infrared Photophyscis and Photochemistry;* Harwood Academic Publishers: Chur, Switzerland, 1985.

28. Polfer, N. C.; Oomens, J. Reaction Products in Mass Spectrometry Elucidated with Infrared Spectroscopy. *Phys. Chem. Chem. Phys.* **2007,** *9,* 3804-3817.

29. Oepts, D.; van der Meer, A. F. G.; van Amersfoort, P. W. The Free-Electron Laser User Facility. *Infrared Phys. Technol.* **1995,** *36*, 297-308.

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

31. Yaghmaei, S. *In Search of a Low Barrier Hydrogen Bond in Proton Bridged Diamines;* University of California, Riverside, 2008.

32. Brown, H. C.; Heydkamp, W.R.; Breuer, E.; Murphy, W. S. The Reaction **of** Organoboranes with Chloramine and with Hydroxylamine-0-sulfonic Acid. **A** Convenient Synthesis of Amines from Olefins *via* Hydroboration. *J. Am. Chem. Soc.* **1964,** 86, 3565-3566.

Chapter III

Nucleobases

INTRODUCTION

Hydrogen bonding has been the basis for molecules having peculiar physical properties for almost 100 years, ever since Moore and Winmill first discovered that trimethylammonium hydroxide is a weaker base than its tetramethyl equivalent in 1912. [1] Since then, hydrogen bonding, the attractive force between a covalently bound hydrogen and more electronegative atoms, has been used to describe everything from the anomalous properties of water to the complex secondary structure of macromolecules. Even after 100 years there is much to be studied about hydrogen bonds. As more research is performed in every realm of science, hydrogen bonds continue to shape how the world, and specifically chemistry, is understood.

The DNA double helix was first suggested in 1953. [2] It was discovered that hydrogen bonding is responsible fro the connectivity of complementary bases on anti-parallel strands (Figure 3.1), allowing Watson and Crick to rationalize the previously undiscovered three dimensional structure of DNA. Hydrogen bonding (and specifically hydrogen bonding between DNA base pairs) continues to be an area of considerable research.

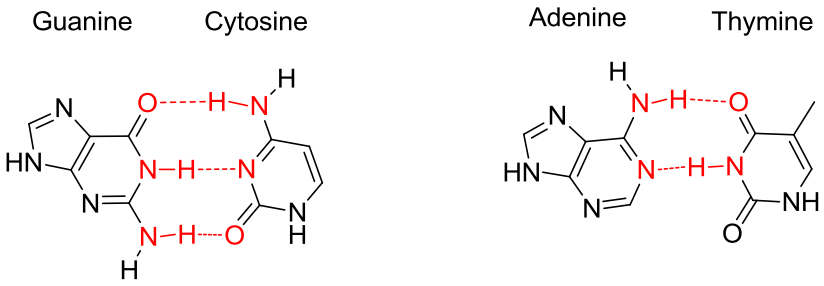


Figure 3.1. Hydrogen bonding (red) between complimentary DNA base pairs

The neutral hydrogen bonds that are shown in Figure 3.1 between hydrogen and either nitrogen or oxygen provide a stabilization energy that is worth approximately 10 kJ/mol per hydrogen bond in aqueous solution. [3] This stabilization energy allows the three hydrogen bonds that occur between guanine (G) and cytosine (C) to provide for a more stable complex than the two hydrogen bonds between adenine (A) and thymine (T). Double stranded DNA domains that are rich in G-C base pairs are more stabilized than those rich in A-T pairs. The hydrogen bonds that hold the two strands of DNA together are not covalent bonds and, therefore, can be pulled apart relatively easily, which allows DNA replication to occur. [4]

There are cases, however, where hydrogen bonding between non-complementary base pairs exists. One of the most studied non-complementary DNA base pair hydrogen bonding motifs observed in nature is found in G-rich repeats in single-stranded overhangs at the 3' end of chromosomes known as telomeres. Single stranded poly dG form four stranded helical structures with G-tetrads (Figure 3.2) stacked on top of one another, held together by hydrogen

bonding from the Watson-Crick face to the Hoogsteen face between two guanines. Formation of these tetrameric structures is possible at the ends of chromosomes because telomeres are single stranded. [5] Telomeres impart stability to linear eukaryotic chromosomes, preventing two chromosomes from fusing together or being degraded, and distinguishing normal ends from ends produced by chromosome breakage events. [6] This G-rich region of non-coding DNA forms a G-quadruplex consisting of 4 guanine residues acting as hydrogen bond donors on the Watson-Crick face and hydrogen bond acceptors on the Hoogsteen face surrounding a metal ion (Figure 3.2). [7] Telomeres protect the ends of chromosomes from degradation by exonucleases [8] and from the shortening that occurs at the end of each replication. [9] Telomeres get shorter and shorter in each successive replication resulting in the loss of many of their telomeric, G-rich repeats. Ultimately chromosomes lose coding portions of DNA. The stabilization provided by the G-quadruplex prevents the loss of coding DNA regions. The preservation of these complexes allow the chromosomes to go on replicating indefinitely. [10] The G-quadruplex prevents telomerase from synthesizing guanine repeats to add to the 3' overhang. [6]

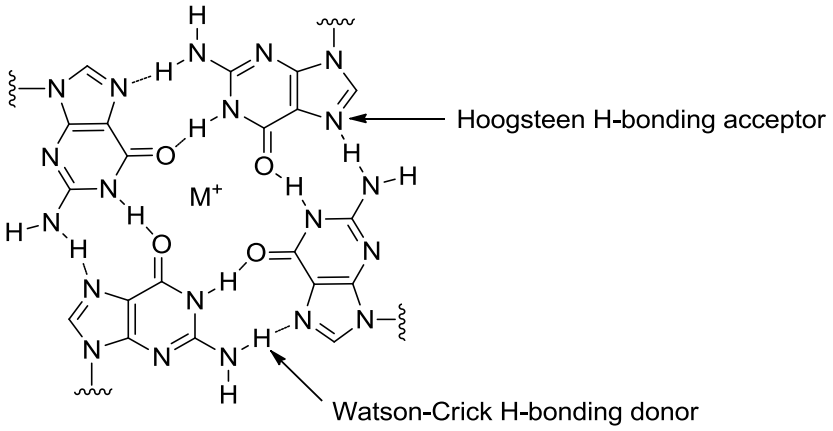


Figure 3.2. G-quadruplex: four guanine residues hydrogen bonded together, surrounding a metal ion.

More recently there has been evidence that G-quadruplexes occur in double stranded G-rich regions of DNA (usually within oncogene promoters) as well as at the single stranded overhang at the end of chromosomes. [11, 12] Oncogenes are genes that, when mutated or expressed at abnormally high levels, assist in transforming a normal cell into a cancer cell. Activated oncogenes can cause abnormal cells to avoid apoptosis, allowing them to survive and flourish. Six cellular and microenvironmental processes are regulated during oncogenic transformation and malignancy. [13] When these six processes (self sufficiency for growth signals, insensitivity to anti-growth signals, evasion of apoptosis, sustained angiogenesis, limitless replicative potential, and tissue invasion and metastasis) are examined, G-quadruplexes are found with critical proteins in the core or proximal promoter. [13] Promoters are regions near a gene in DNA that facilitates the expression of that particular gene. Eight human genes with G-quadruplexes in the double-stranded region highlight the

microenvironmental processes that lead to malignant tumors. PDGF, c-Kit, KRAS, and VEGF-A all code for growth factor proteins. Specifically, PDGF codes for a platelet-derived growth factor, [14] c-Kit codes for a cytokine receptor, [15] KRAS codes for a GTPase enzyme, [16] and VEGF-A codes for a vascular endothelial growth factor protein. [17] Bcl-2, pRb, c-Myc, and hTERT all code for regulator proteins. Specifically, Bcl-2 codes for a protein that regulates apoptosis of cells in immune responses, [18] pRb codes for a protein that prevents damaged DNA from being replicated, [19] c-Myc codes for a transcription factor regulating expression of almost fifteen percent of all human genes, [20] and hTERT codes for human telomerase reverse transcriptor. [21]

G-quadruplexes in promoter regions near these genes have numerous folding patterns and loop lengths, making them desirable targets for drugs. [22] The greatest variability among these G-rich secondary structures in the promoters is in their loop length and constituent bases. Four classes of quadruplexes (I, II, III, IV) have been identified differing in the number of G-quadruplexes formed and their relative position within the promoter (Figure 3.3). [23] The classes are separated based on the number of G-quadruplexes, the loops that connect them, and the orientation of the DNA strands. Class I has three G-quadruplexes stacked on top of one another connected by three small DNA strand loops. Class I consists of four genes having loops using either four or five residues and the 3'-(1:2:1)-5' loop size predominates. An example of Class I comes from the c-Myc gene whose DNA sequence, **GGG**X**GGG**XX**GGG**X**GGG** (X's represent the DNA

bases that make up the loops connecting the guanine residues), orients the sequential guanine residues that form the vertices of each of the stacked G-quadruplexes parallel to one another. Class II has two sets of stacked G-quadruplexes, with both small and medium sized DNA loops connecting the G-quadruplexes. There is only one example of a Class II G-quadruplex (c-Kit), in which there are two stacked G-quadruplexes separated by three turns (35 bases) of DNA. The downstream-stacked G-quadruplexes have a 3'-(1:5:1)-5' loop arrangement (**GGG**X**GGG**XXXXX**GGG**X**GGG**), while the upstream-stacked G-quadruplexes have a 3'-(4:4:1)-5', loop arrangement (**GG**XXX**GGG**X**G**XX**GGG**X**GGG)**. Both stacked G-quadruplexes orient the guanine vertices parallel to each other. Class III also has two sets of stacked G-quadruplexes, but with different loop sizes and strand orientation than Class II. One example of the Class III quadruplex is found in the hTERT gene which has two sets of G-quadruplexes separated by 7 DNA bases. The downstream stacked quadruplexes have a loop orientation of 3'-(1:4:1)-5' (**GGG**X**GGG**XXXX**GGG**X**GGG**), while the upstream stacked quadruplexes have a loop orientation of 3'-(1:26:3)-5' (**GGG**X**GGG**X_{26}**GGG**XXX**GGG**). The 3' G-quadruplexes have two small loops and one very large loop connecting the vertices of the G-quadruplexes. The large loop connecting the second and third sequential guanine residues results in one of the vertices being antiparallel to the other three. The 5' G-quadruplexes resemble the Class I orientation with small loop sizes and parallel guanine vertices. Class IV, like Class I, has only a single

set of stacked G-quadruplexes. The loop sizes in Class IV are both small and medium in size and the medium sized loop results in one set of sequential guanine vertices to be anti-parallel to the other three. The Bcl-2 gene's stacked G-quadruplexes have a 3'-(1:7:3)-5' loop orientation (**GGG**X**GGG**XXXXXXX**GGG**XXX**GGG**) where the third set of guanine residues run anti-parallel to the other three.

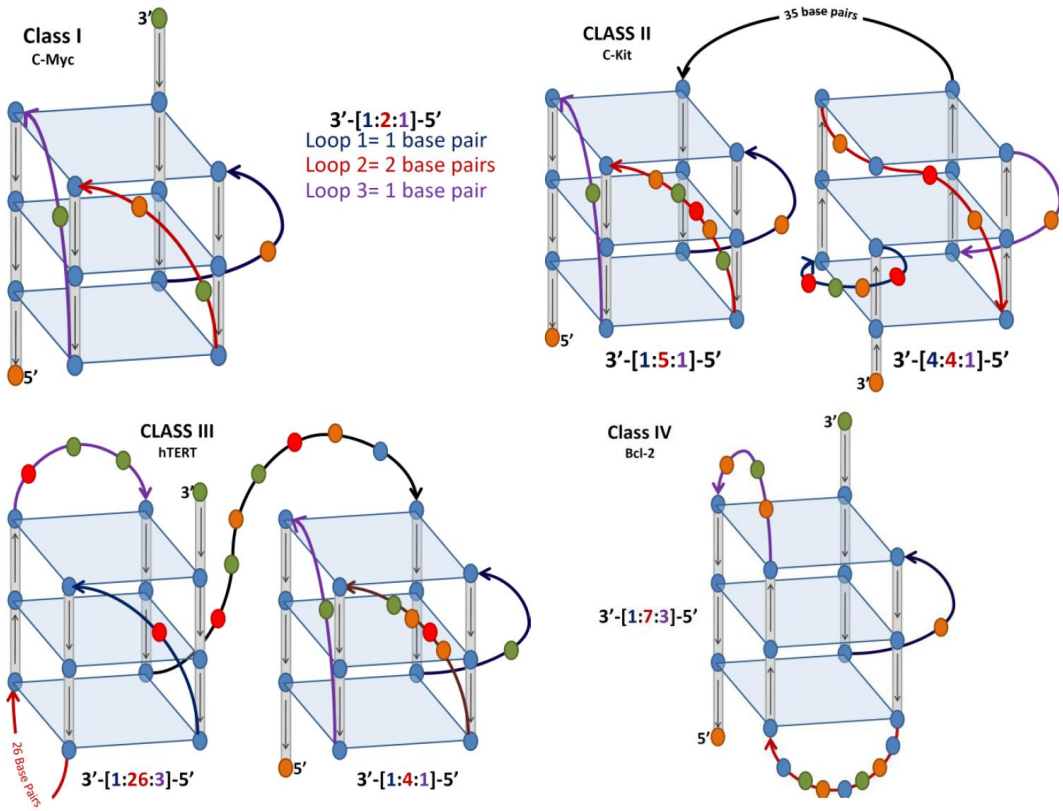


Figure 3.3. Proposed classes of unimolecular G-quadruplexes found in eukaryotic promoter elements. Blue parallelograms represent G-quadruplexes (as shown in Figure 1.2) and the filled ovals represent DNA base pairs (blue being guanine). Class I (A) is represented by the single G-quadruplex found in the c-Myc promoter element. Class II (B) contains a pair of different G-quadruplexes separated by about three turns of DNA. Class III (C) is represented by the tandem G-quadruplexes from the hTERT promoter. Class IV (D) represents multiple overlapping G-quadruplexes. The example shown is from the Bcl-2 promoter and the G-quadruplex shown (MidG4) is the most stable of the three structures.

BACKGROUND

Cytosine Proton-Bound Dimers in DNA

Opposite the G-rich region of double stranded DNA there is a C-rich region. After separation of the G-rich tract to form a quadruplex, hydrogen bonding no longer connects the two anti-parallel strands as complimentary base pairs, and the C-rich region is now left to self-associate into proton-bound dimers (PBDs).

Cytosine-like PBDs were first observed in crystals of acetyl cytosine. [24] After repeated recrystallization from hot water, cytosine-5-acetic acid formed rectangular plates. The resulting X-ray study revealed the proton transfer from the carboxyl oxygen of the acetic acid side chain, to the nitrogen atom of the pyrimidine ring in one-half of the molecules of the crystal. Consequently, a single proton is shared between two pyrimidine ring nitrogens on neighboring molecules, leaving half the molecules in the crystal uncharged and half as zwitterions. This allows the zwitterionic pyrimidine nitrogen to act as a strong hydrogen bond donor and the uncharged pyrimidine ring nitrogen as an acceptor, binding together the two molecules of cytosine-5-acetic acid through a shared proton. The three-dimensional structure shows the molecules are held together by hydrogen bonds involving all donor protons, aligning the cytosine pairs about a point of symmetry between the two pyrimidine ring nitrogens.

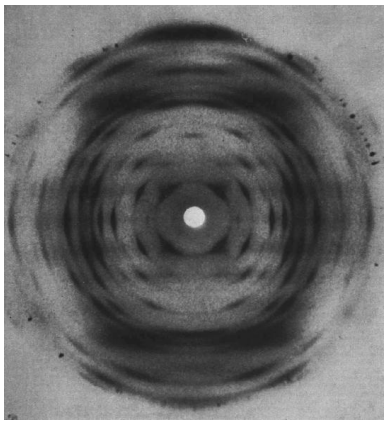


Figure 3.4. X-ray pattern of polycytidylic acid. [25]

The X-ray of polycytidylic acid, a polymer of cytidine linked through phosphodiester bonds, at a slightly alkaline or neutral pH is amorphous, containing no internal order or sharp reflections. Upon lowering the pH to 5.5 a highly ordered diffraction pattern appears (Figure 3.4). [25] With over fifty individual reflections the X-ray diffraction pattern taken under acidic conditions is unusual among synthetic polynucleotides and is surpassed in order only by highly oriented DNA preparations. [26] This highly ordered pattern does not change when humidity is varied, establishing that these molecules fit together very closely with very few water molecules in the lattice. The structure determined by X-ray crystallographic analysis shows that the molecule exists in a helical form, but does not reveal whether or not the helix consists of one or more strands, nor if the strands are parallel. Analysis of the X-ray diffraction of polycytidylic acid at pH 4 produced a slightly less ordered diffraction pattern. The structure determined by X-ray crystallographic analysis suggests the bases are at the

center of the molecule and that the packing has been disturbed at the lower pH. The increase in the concentration of protons destroys the two-fold symmetry of the molecule, which indicates that the highly ordered X-ray pattern is a result of two parallel strands related by a two-fold rotation axis. Analysis of the X-ray diffraction patterns at various pH determined that the hydrogen bonding scheme between cytosine base pairs in polycytidylic acid would have a charged hydrogen bond flanked by two ordinary hydrogen bonds (Figure 3.5).

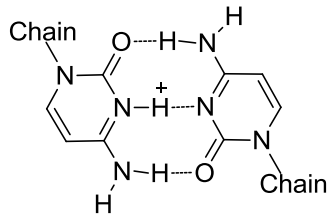


Figure 3.5. Proposed H-bonding between hemiprotonated cytosines in polycytidylic acid

The hydrogen bond scheme between cytosine bases in the helical form of polycytidylic acid was confirmed by titration studies, infrared and ultraviolet spectroscopy. [27] Figure 3.6 shows the titration of HCl into poly C demonstrating a two step process. The first step occurs at pH 5.7 where there is a fast uptake of protons into the molecule without any change in the pH of the solution. The rapid uptake of hydrogen ions is characteristic of cooperativity throughout the polymer, resulting in a structural change where many residues accept a proton. At pH 3 there is another transition in that protons are rapidly taken up by the polymer that becomes fully protonated at a pH just under 3. Comparing the titration curve of

poly C with that of cytidine (that shows no cooperativity between residues) leads to the conclusion that the polymer forms an ordered three dimensional structure from pH 5.7-3.

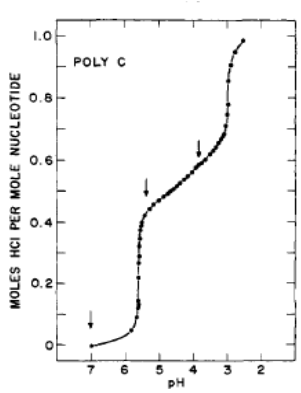


Figure 3.6. The titration curve of poly C shows that upon adding acid, protons are taken on abruptly at two points with pK values of 5.7 and 3.0. Near pH 3, the solution of poly C becomes turbid, suggesting that the molecule becomes insoluble when it is neutral. Almost half the bases are protonated just below pH 5.7. [27]

Infrared spectra of cytidine and poly C in D_2O show changes in the 1450-1800 cm^{-1} region as the pD is lowered. These bands were not assigned at the time, however. Lowering the pH from 6.4 to 3 shifts the UV absorption maximum from 270 to 278 nm, highlighted by an abrupt red shift at pH 5.7. This red shift was accompanied by an increase in the intensity of the band. This suggested that poly C showed an ordered structural change at pH 5.7, at which point half of the cytosine bases are protonated. Unlike the neutral hydrogen bonds seen between complementary base-pairs, the cytosine dimers must be held together by a strong hydrogen bond, i.e. a short hydrogen bond where a proton is held

between two basic sites having similar proton affinities flanked by two ordinary hydrogen bonds. Without the charge in the middle of the two monomers the cytosines would be unable to orient themselves to share three hydrogen bonds. If there were no proton between the two ring nitrogens, the lone pairs on those nitrogens would repel each other and the highly ordered helix would not form. The neutral cytosine-cytosine dimer has been shown to have two hydrogen bonds as shown by UV-UV and IR-UV double resonance laser spectroscopy. [28] Nir and colleagues also showed that this is true for the cytosine: 1-methylcytosine neutral heterodimer, as it has the same hydrogen binding motif as the neutral cytosine dimer (Figure 3.7). The hydrogen bonding motif of the cytosine proton-bound dimer (Figure 3.5) is vastly different from that in the neutral cytosine dimer or the G-quadruplex, which contains only neutral hydrogen bonds to hold the base pairs together. The strong hydrogen bond flanked by the two ordinary hydrogen bonds in the cytosine PBD stabilizes the two base pairs by approximately 140-150 kJ/mol [29] compared to the 10 kJ/mol stabilization energy per neutral hydrogen bond in the G-quadruplex. This is a larger energy stabilization than that observed for the difference in proton affinity for many monoamines and their corresponding molecular pairs (Table 3.1). [30]

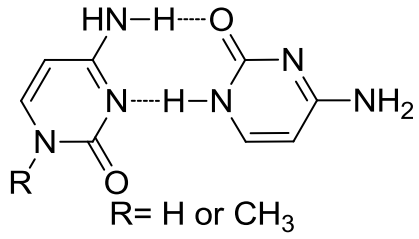


Figure 3.7. Hydrogen bonding motif for the neutral cytosine-cytosine and 1-methylcytosine-cytosine dimer.

More recently, many structures determined by X-ray crystallographic analysis have provided atomic resolution to the PBD of free cytosine or 1-methylcytosine salts. An X-ray crystal structure of two molecules of 1-methylcytosine, with iodide as a counterion, is shown to be bound together by a strong, asymmetric, hydrogen bond flanked by two normal hydrogen bonds (Figure 3.8). [31]

Table 3.1. Gas-Phase Proton Affinities (PA) and Molecular-Pair Proton Affinities (MPPA) of amines at 25 ^{0}C.

Amine	PA (kJ/mol)	MPPA (Kj/mol)	ΔPA (kJ/mol)
NH_3	857.7	961.5	103.8
$MeNH_2$	895.8	986.6	90.8
Me_2NH	922.6	1009.6	87.0
Me_3N	938.5	1032.6	94.1
Allyl NH_2	905.8	993.3	87.4
$EtNH_2$	908.3	995.8	87.4
n-$PrNH_2$	914.2	1001.2	87.0
i-$PrNH_2$	918.0	1002.5	84.5
n-$BuNH_2$	916.3	1004.6	88.3
i-$BuNH_2$	918.4	1004.6	86.2
s-$BuNH_2$	922.6	1007.5	84.9
Pyridine	922.2	1012.5	90.4
c-hexyl NH_2	925.9	1014.6	88.7
Azetidine	931.8	1023.4	91.6
MeNHEt	931.4	1017.1	85.8
Pyrollidine	938.5	1028.4	90.0

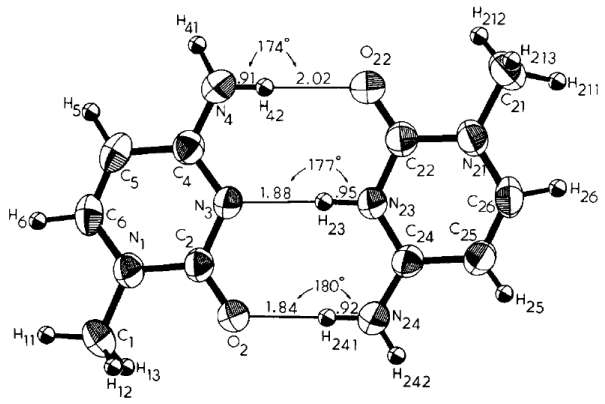


Figure 3.8. Hemiprotonated 1-Methylcytosine, 1-Methylcytosine complex. [31]

X-ray crystallographic analysis cannot identify the proton that resides in between the two monomers, but asymmetry is inferred from the different distances between the hydrogen bonded amino nitrogen and the corresponding carbonyl oxygen on either side of the strong hydrogen bond between the two ring nitrogens sharing the proton. The hydrogen bond formed between the amino hydrogen on the charged 1-methylcytosine and the carbonyl oxygen of the neutral 1-methylcytosine is shorter than the hydrogen bond formed by their respective counterparts on the other side of the PBD by about 0.2 Å. Upon further investigation of these PBDs, Kistenmacher and colleagues found that these PBDs were associated into stacked dimers with a crystallographic center of symmetry. The stacked bases aligned so that protonated 1-methylcytosine is sandwiched between two neutral molecules of 1-methylcytosine and vice-versa (see Figure 3.9). The analysis of the structures determined by X-ray analysis suggest that the asymmetry in the hydrogen bonded structures is stimulated by

these base stacking factors. The authors suggest that these base stacking interactions would also govern the structure of hemiprotonated, helical polydeoxyribocytidylic nucleotide. That is, polyC DNA stretches would form highly stacked pairs of asymmetrically hydrogen bonded bases as proposed in Figure 3.10.

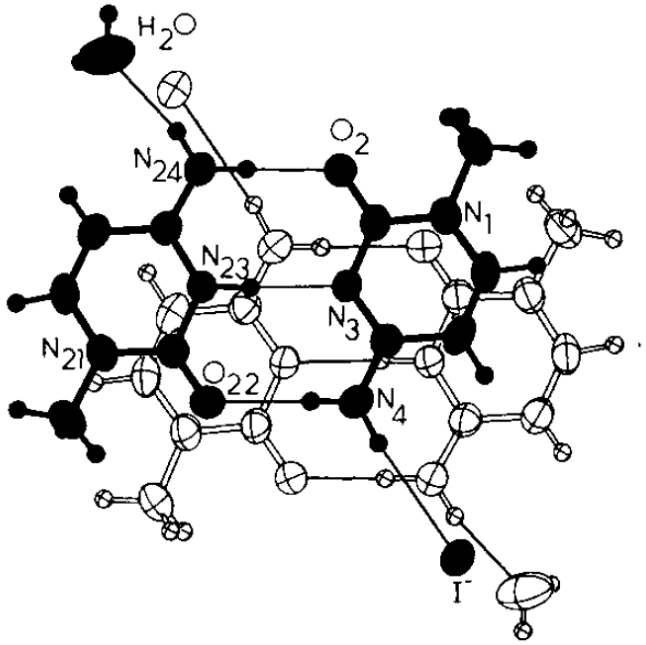


Figure 3.9. X-ray structure showing the base stacking of 1-methylcytosine proton-bound dimer. [31]

The infrared absorption spectrum of a crystalline salt of the proton-bound dimer of the nucleoside cytidine exhibits four bands in the 1600-1800 cm^{-1} domain. [32] These bands correspond to bands that are present in the spectra of both charged and neutral monomers of cytidine. This result implies that the two nucleobases do not share the bridging proton equally in the equilibrium geometry. Kistenmacher and colleagues concluded from the crystallographic

data and an IR band of the polycrystalline sample (nujol mull) at 1890 cm^{-1} (presumed to be indicative of an asymmetric hydrogen bond observed in cytidinium-cytidine complex) [32] that the structure is asymmetric as indicated in Figure 3.8. The IR band at 1890 cm^{-1} was assigned as it was not present in the IR spectra of either the protonated or neutral monomer of cytosine. Although this may not be sufficient evidence, that region warrants further scrutiny (vide infra).

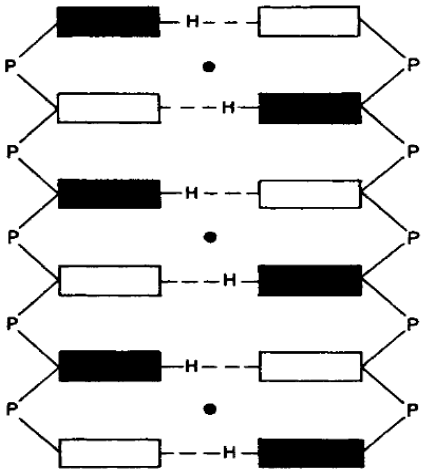


Figure 3.10. Proposed model of helical polyribocytidylic acid. Rectangles represent cytosine bases and the solid circles show centers of symmetry between base-pairs. Black boxes represent the charged 1-methylcytosine monomer. White boxes represent the neutral 1-methylcytosine monomer. [31]

Asymmetry could be caused by a bent complex but, it is more likely that it is due to the off-center location of the proton shared between the two ring nitrogens. Theoretical calculations support this conclusion, predicting amino nitrogen-to-carbonyl oxygen distances that differ by about ten percent for the ordinary hydrogen bonds that flank the proton. [33] The calculations indicate that a

high barrier to proton transit is not the basis for asymmetry, as the computed barrier to proton transit from one nitrogen to the other (which renders the two amino nitrogen carbonyl oxygen hydrogen bonds equivalent) is only about 7 kJ/mol after zero-point corrections. [33]

The stacked helical structure of polydeoxyribocytidylic acid proposed by Kistenmacher and colleagues was not found in the NMR structure of the DNA oligomer 5-d(TCCCCC). [34] The ^{1}H NMR spectrum of d(TC_5) in D_2O shows five resonances between 15 and 16 ppm that integrate to one proton per pair of cytidines. These resonances suggest strong hydrogen bonding associated with the ring nitrogens. If the imino proton were not involved in hydrogen bonding it would be accessible to the solvent and thus exchangeable. The absence of peak broadening signifies that the PBD is linked at the ring nitrogen, as shown in Figure 3.5. Leroy and co-workers hypothesized that poly C-rich strands of DNA form a four stranded complex, where two parallel-stranded duplexes have their base pairs completely intercalated, an orientation later termed the *i*-motif (Figure 3.11) [34]. NMR studies on the oligomer d(TCC) confirmed the occurrence of the *i*-motif, as it formed a tetrameric structure with intercalated base pairs between two parallel strands (Figure 3.12). [35] The symmetrical, hemiprotonated cytosine base pairs are characterized by cytidine imino protons at 15.13 and 15.81 ppm along with two pairs of amino protons whose proton shifts are between those of neutral and protonated cytosine at pH 4.2, -5 ^{o}C. This indicates the strands associate into a multimer, and the number of peaks demonstrates that all the strands are

equivalent on the NMR timescale. Accordingly, the proton jumps quickly between the two ring nitrogens. Analysis of the NMR spectrum also shows that some peaks increase in intensity upon dilution and are assigned to the single stranded species. The dissociation constant for $[d(TCC)]_4$ at pH 4 and 6 ^oC was determined to be 3.6 x 10^{-3} M^3. The NOESY spectrum in H_2O contains cross peaks in the amino proton region that are indicators of intercalation, as they appear in regions that are devoid of peaks in the NOESY spectra of Watson-Crick duplexes.

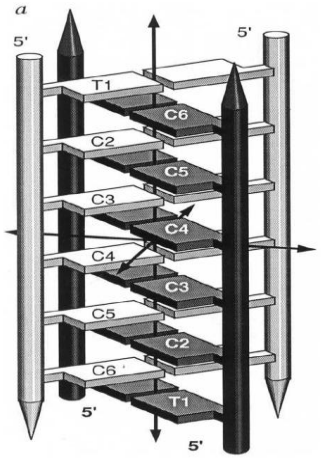


Figure 3.11. Model of d(TCCCCC) with intercalated base pairs of two parallel duplexes. [34]

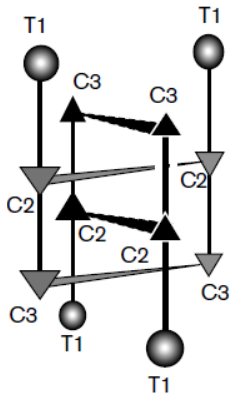


Figure 3.12. Diagram of $d(TCC)_4$ showing the intercalated cytosines held together by hydrogen bonds. [35]

Recently, synchrotron radiation circular dichroism (SRCD) experiments have been performed on various lengths of single stranded poly deoxyC DNA $(dC)_n$ at variable pH. [36] Two well known SRCD markers of hemi-protonated cytosine, a positive band at 288 and a negative band at 266 nm [37] were used to characterize the structure. At pH 1, all $(dC)_n$ (n=1-11) had a positive band at 288 nm, but no negative band at 266 nm was observed. Since at pH 1 all cytosines bases are protonated, the positive band at 288 nm must correspond to the protonated base. The presence of a negative band at 266 nm would correspond to interbase Hoogsteen hydrogen bonding. Increasing the pH to 3 does not drastically change the SRCD spectra for $(dC)_n$ for $n \leq 5$, but for $n \geq 6$ a negative band at 266 nm does appear. This means that only the strands with greater than five cytosine residues exhibit interactions between neighboring nucleobases. Principal component analysis (PCA) shows that for $n \geq 6$ $(dC)_n$ is

hemiprotonated. This is ascribed to be the onset of the *i*-motif, which concur with the discussion above. None of the results for any strand length at pH 7 resemble that of an all-protonated single strand. The data did indicate that at pH 7 the structure of all strand lengths is a mixture of the *i*-motif and something else, most likely an all neutral single strand. Both bands at 288 and 266 nm are absent at pH 9 for all strand lengths, indicating that there is no electronic interaction between base pairs and that the structure resembles stacked single strands with all neutral bases. Unlike the X-ray studies of polycytidylic acid discussed above, these results suggest that the *i*-motif can occur at physiological pH.

Kang and colleagues have solved the crystal structure of $d(ACCCT)_4$ and confirmed the intercalation of the cytosine base pairs, as diagramed in Figure 3.13. [38] Although Kistenmacher and colleagues did not correctly predict the base pair intercalation seen in polydeoxy C rich strands of nucleic acids, they were correct in predicting that the PBDs would be asymmetric, but only in certain instances. The distances between the electronegative heteroatoms involved in hydrogen bonding between cytosines in the crystal of $d(ACCCT)_4$ are listed in Table 3.2. The PBDs at the top and bottom of the crystal of $d(ACCCT)_4$, such as the one between C2 and C12, have amino nitrogen to carbonyl oxygen distances that differ from one another by over one- third of an angstrom. This is much larger than in any of the published crystal structures of PBDs between cytosine or 1-methylcytosine discussed below. Again, the hydrogen atoms cannot be determined by X-ray crystallographic analysis, but the large difference between

flanking hydrogen bonds signify the middle proton is located on one side of the dimer. As one advances inward from the ends of the stack to the middle, the differences between the flanking hydrogen bonds decrease until they become almost equal, as in the case of the PBD between C3 and C13. The slight difference in the flanking N---O distances (~0.02 Å) and the shortest distance observed between the ring nitrogens (2.58 Å) in the crystal suggest that the proton is equally shared between the two ring nitrogens.

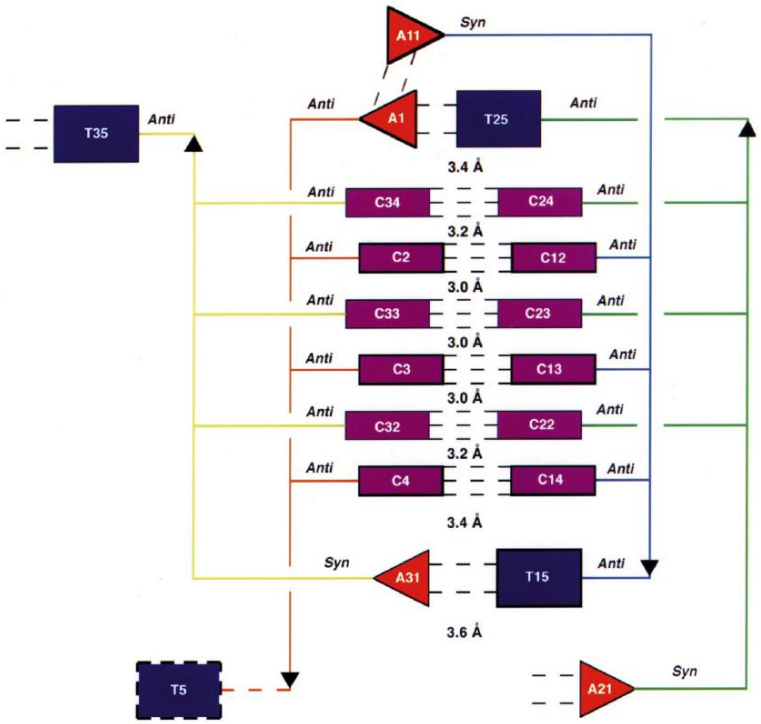


Figure 3.13. Diagram of the crystal structure of d(ACCCT)$_4$, cytosines are represented by purple rectangles. [38]

Table 3.2. Distances between heteroatoms involved in hydrogen bonding between hemiprotonated cytosine(C) bases in Figure 3.13.

C2 -- C12		
	Distance (Å)	
Amino N	2.9830	Carbonyl O
Ring N	2.7151	Ring N
Carbonyl O	2.6217	Amino N

C34 -- C24		
	Distance (Å)	
Amino N	2.8205	Carbonyl O
Ring N	2.6835	Ring N
Carbonyl O	2.6571	Amino N

C3 -- C13		
	Distance (A)	
Amino N	2.6251	Carbonyl O
Ring N	2.5822	Ring N
Carbonyl O	2.6467	Amino N

C33 -- C23		
	Distance (Å)	
Amino N	2.7670	Carbonyl O
Ring N	2.7330	Ring N
Carbonyl O	2.7968	Amino N

C4 -- C14		
	Distance (Å)	
Amino N	2.9029	Carbonyl O
Ring N	2.6957	Ring N
Carbonyl O	2.6040	Amino N

C32 -- C22		
	Distance (Å)	
Amino N	2.7904	Carbonyl O
Ring N	2.8058	Ring N
Carbonyl O	2.9714	Amino N

i-Motifs found in nature opposite the G-rich regions in oncogene promoters are divided into two classes; Class I consisting of small loops and Class II consisting of large loops (Figure 3.14). [23] The driving force for *i*-motif formation arises from maximizing the number of hemiprotonated cytosine-cytosine base pairs. [39] Only c-Myc relies on negative supercoiling, rather than acidic pH, to drive *i*-motif formation. [40] Class I *i*-motifs have loop sizes of 5'-(2:3/4:2)-3', while Class

II have loop sizes of 5'-(6/8:2/5:6/7)-3', the promoter region of the oncogene Bcl-2 having the largest observed cumulative loop size of twenty base pairs, 5'-(8:5:7)-3'. [23]

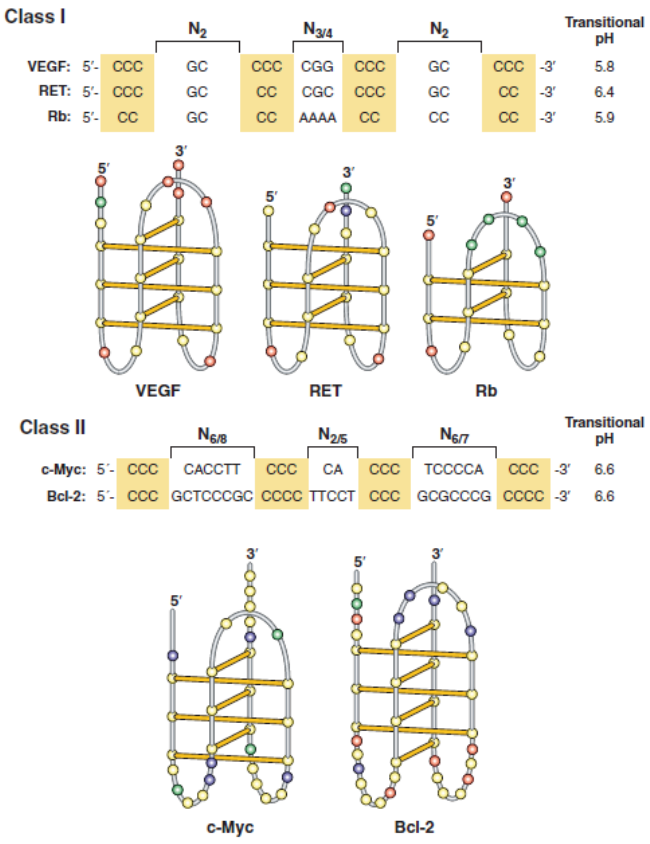


Figure 3.14. Sequences and folding patterns of i-motifs in the two proposed classes of i-motifs found in eukaryotic promoter elements. Class I, having small loop sizes, is found in the VEGF, RET and Rb promoter elements, and Class II, having larger loop sizes, is found in the c-Myc and Bcl-2 promoter elements. [23]

Two questions arise when considering the biological role of secondary DNA structures such as those seen in G-quadruplexes and the *i*-motif: (a) how can these structures develop from double stranded DNA; and (b) once formed,

how are they disbanded? Supercoiling, which has been a known factor in eukaryotic transcription, has been shown to inspire G-quadruplex and *i*-motif formation. [22] It has also been shown that binding proteins (such as nucleolin) can facilitate G-quadruplex formation in c-Myc, and therefore, allow the opposite C-rich strand to form an *i*-motif. [41] This evidence suggests that these secondary DNA structures in promoter regions of oncogenes are important to eukaryotic transcriptional regulation. No definitive proof for their existence in cells is available, however.

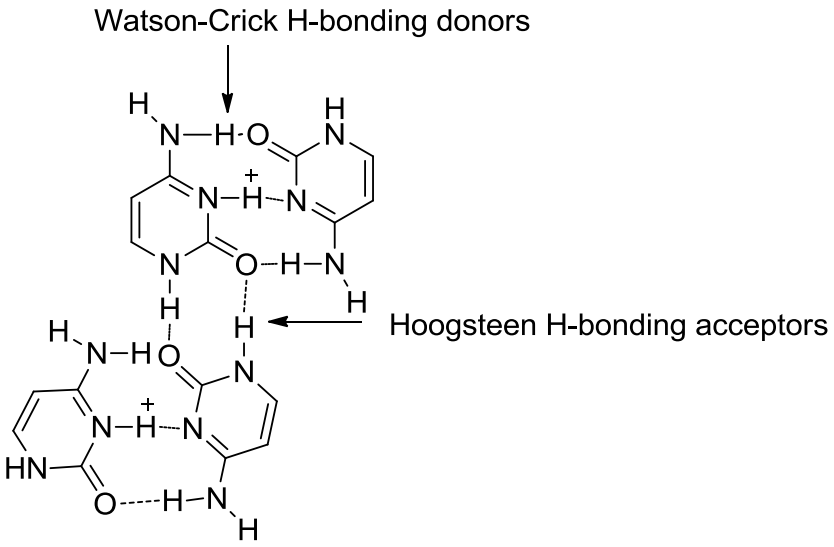


Figure 3.15. Hydrogen bonding motif between protonated cytosines.

Free Cytosine and 1-MethylCytosine Proton-Bound Dimers

Synthetic proton-bound between cytosine or 1-methylcytosine nucleobases that are not incorporated into DNA strands have been investigated to gain insight into the binding motif assumed to occur in biology. X-ray

crystallogrphic analysis of these PBDs has consistently shown structures with a well-defined geometry independent of the counterion used. [42] A summary of data from the published crystal structures of cytosine and 1-methylcytosine PBDs is shown in Table 3.3. It is important to note that the PBDs between two molecules of cytosine have one additional hydrogen bond donor than their 1-methylcytosine analogue. The ring N-H is allowed to hydrogen bond with a neighboring PBD carbonyl oxygen, as shown in Figure 3.15. The asymmetric proton-bound homodimers of 1-methylcytosine form even in the presence of a coordinating species, such as $AuCl_3$. [42] When the crystallization is performed from calcium perchlorate the N---O distances of the flanking hydrogen bonds are equivalent (Figure 3.16). The position of the proton cannot be assigned to either of the ring nitrogens, indicating that it is equally shared between them. [43] The x-ray analysis of this crystal structure is reported in a journal (JASMS) that does not require the authors to deposit the X-ray data in the Cambridge Crystallographic Data Centre (CCDC), making conformation structure impossible. There are two other examples of symmetric PBDs in the literature: one of 1-methylcytosine using iodide as counterion, and the other of cytosine with trichloroacetate as a counterion. [44, 45] These studies directly dispute the previously published crystal structure of Kistenmacher and colleagues, who used iodide as a counter ion. The absence of water in the two symmetric PBD X-rays (and its presence in the Kistenmacher X-ray) may be responsible for its asymmetry, but the symmetric PBD with iodine as a counterion proposed by Kruger was disputed a year after it

was published by Muller and Freisinger. [46] The newer crystal structure of [(1-methylcytosine)$_2$H] I is asymmetric, with the N---O distances differing by 0.14 Å. Muller and Freisinger attribute the conflict in the data between two identically prepared crystal structures to temperature-dependent change from disorder to order. Muller and Freisinger conducted their X-ray study at 162 K while Kruger had performed the analysis at 293 K. The lower temperature resulted in a lack of disorder of the hydrogen atom in the central hydrogen bond, which allowed for a more accurate assignment of the crystal structure. Disorder of the hydrogen atom shared between the two ring nitrogens is also suggested as the cause of the incorrect assignment of the PBD as symmetric in the case of the PBD of cytosine with a trichloroacetate counterion.

Table 3.3. Crystallographic data on published cytosine and 1-methycytosine proton-bound homodimers.

Proton-Bound Dimer	Counter Ion	H_2O present?	N-O bond length difference
Cytosine [43]	Tetrachlorozinc	no	0.17 Å
Cytosine [44]	PF_6	no	0.12Å
Cytosine [44]	Et_2TCNQ	no	0.10 Å
Cytosine [47]	F_4TCNQ-OMe	no	N/A, but asym
Cytosine [45]	Trichloroacetate	no	0Å
1-MeCyt [31]	Iodine	yes	0.17 Å
1-MeCyt [46]	Iodine	no	0.14Å
1-MeCyt [42]	$Au(CN)_4$	yes	0.11Å
1-MeCyt [44]	Iodine	no	0Å
1-MeCyt [43]	ClO_4	no	0Å

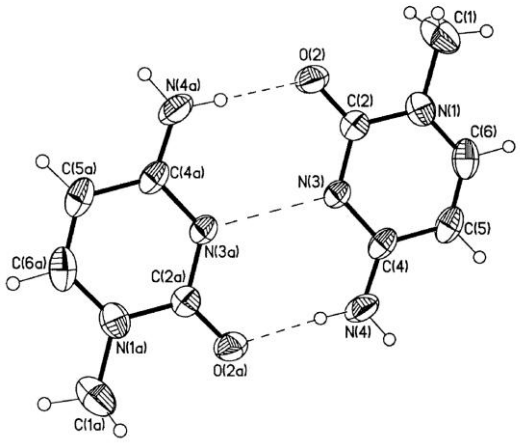


Figure 3.16. X-ray structure of the symmetric proton-bound dimer of 1-methylcytosine with a potassium perchlorate counter ion. [43]

It appears that both environment and temperature play a major role in the structure of cytosine and cytosine-like PBDs. The X-ray structure of a cytosine PBD has been studied as a function of temperature, and the results show that a phase change occurs in this particular cytosine PBD at 180K. [48] The co-crystal of decavanadate with cytosine and water was synthesized forming a $Na_3[V_{10}O_6](C_4N_3OH_5)_3(C_4N_3OH_6)_3 \bullet 10H_2O$ complex. [48] Three different molecules of cytosine (A, B, and C) make up the PBDs seen in the X-ray unit cell. Two molecules of cytosinium A are paired together and each molecule is protonated at the ring nitrogen by 0.5 hydrogen atoms, giving each ring a charge of +0.5 and a crystallographic center of symmetry. For the PBD between two molecules of cytosinium A the two flanking amino nitrogen and carbonyl oxygens (N4-O2) have identical distances (2.841 Å) and dihedral (N4-H3---O2) bond angles (177.12^{o}). The distance between the two ring nitrogens is 2.838 Å with an N3-H2-

N3 angle of 167.72°. The hydrogen atom between the two ring nitrogens is disordered over two equally occupied positions (Figure 3.17).

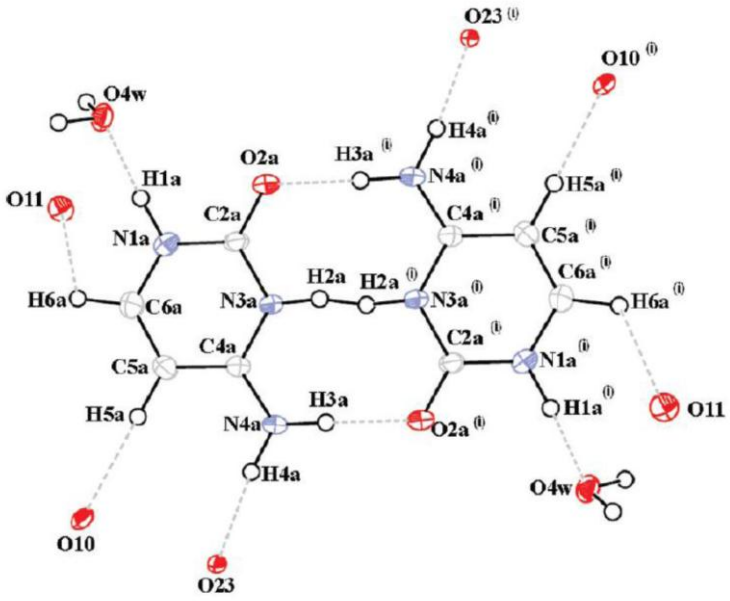


Figure 3.17. Structure of cytosine-cytosinium interactions between two molecules of cytosinium A determined by X-ray crystallographic analysis. [48]

Cytosine C is neutral and is paired with cytosinium B, which is protonated at the ring nitrogen position giving cytosinium B a charge of +1. Some of the interatomic distances in cytosine C, namely the carbonyl (C=O) and amino (C-N) bonds, differ from the corresponding distances in cytosinium A and B by over 10% (Figure 3.18). Protonation also induces changes between the neutral cytosine C and the charged cytosinium B. The distances between the ring carbons and the ring nitrogen in cytosinium B is longer than the corresponding ring carbon ring nitrogen (that acts as the hydrogen bond acceptor) bond in neutral cytosine C. The dihedral angles of the ring carbon-nitrogen-carbon bonds

also differ between cytosine C and cytosinium B. Cytosine C and cytosinium B are connected in the same manner as the cytosinium A PBD with a distance of 2.816Å between the two ring nitrogens (N3c-N3b) and N3c-H2b-N3b dihedral angle of 173.39°. The flanking hydrogen bonds between the amino nitrogens and the carbonyl oxygens (N4b-O2c and N4c-O2b) differ by 0.12 Å, confirming the asymmetry of the PBD. There also exist differences in the distances and the arrangements of the hydrogen bonds between the cytosine molecules and the surrounding water molecules, with cytosinium C having the longest (weakest) hydrogen bonds with the external water molecules.

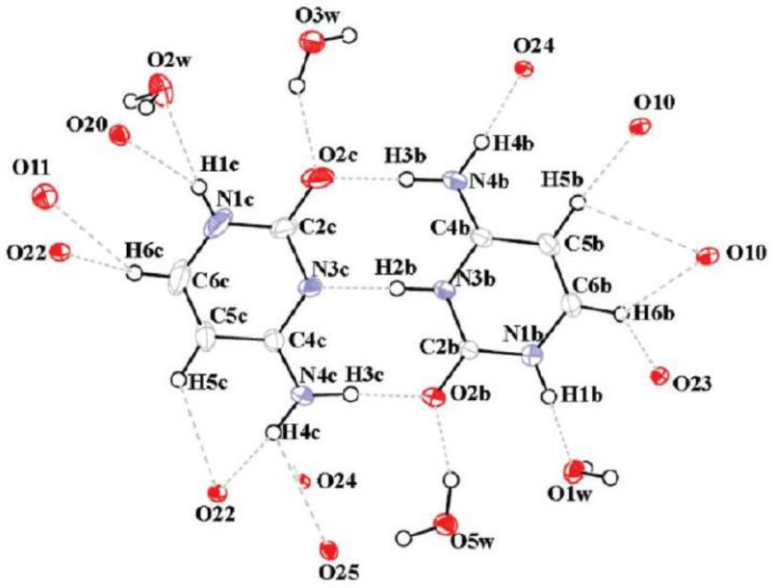


Figure 3.18. Srtucture of cytosine-cytosinium interactions between a molecule of cytosine C and a molecule of cytosinium B determined by X-ray crystallographic anlaysis. [48]

At 210 and 300 K the unit cell of the crystal contains three cytosine PBDs, three sodium atoms, and one decavanadate molecule, and exhibits a centrosymmetric $P\bar{1}$ space group where the triclinic unit cell has an inversion

point of symmetry. At these temperatures the proton is statistically equally shared between two molecules of cytosinium A, as the analysis of the X-ray diffraction patter indicates than only half a proton is present on each ring nitrogen of the PBD. The proton can jump back and forth between ring nitrogens at temperatures above 180K, a dynamic that leads to a point of symmetry between the two ring nitrogens. Once the temperature is lowered at 10 K/hr to 100 K this dynamic is no longer observed. At the lower temperature the unit cell, solved in the P1 space group, has expanded to contain six PBDs, six molecules of sodium and two molecules of decavanadate, and the unit cell itself is asymmetric. The hydrogen bonding of the carbonyl oxygen of each of the cytosine molecules can be used to discriminate between the two types of PBDs (A-A and B-C). The carbonyl oxygen in both cytosine C and cytosinium B is hydrogen-bonded to both the amino hydrogen of the opposite cytosine molecule and to a hydrogen atom belonging to one of the external water molecules. The carbonyl oxygen of cytosinium A is hydrogen-bonded to the amino hydrogen of the opposite molecule, akin to cytosine C and cytosinium B, but it is also coordinated to a sodium atom. The assignment was confirmed by the ring carbon-nitrogen-carbon dihedral angles. Of the twelve PBDs in the unit cell at 100 K, two of the dimers are pairs of cytosine A, but the other four dimers are of cytosine C paired with cytosinium B.

Using the electron density coordinates taken from the X-ray data, the topological analysis of the cytosine PBDs were investigated using Atoms in

Molecules (AIM) theory. As expected the bond critical points (BCP) of the heteronuclear carbon-nitrogen and carbon-oxygen bonds are located closer to the less electronegative carbon atoms than the nitrogen or oxygen atoms. The BCP between the ring nitrogen and the proton in the asymmetric dimer is located 0.158 Å from the proton and 0.856 Å from the nitrogen, the electron density at this point is 1.49 e/Å^3. The bond critical points between other nitrogen-hydrogen bonds are further away from the proton and have electron densities between 1.9 and 3.0 e/Å^3. The amine N-H that is hydrogen-bonded to the opposite carbonyl has the lowest electron density of any of the neutral N-H BCPs, but still has 0.42 e/Å^3 more electron density at the BCP than the BCP between the ring nitrogen and the proton in the asymmetric dimer. The Laplacian values at the reported BCPs are all negative indicating that all the bonds are covalent. Table 3.4 shows the electron density (Rho) of the AIM calculations performed on the B3LYp/6-311++G** optimized equilibrium geometry of the cytosine proton-bound dimer (cytosinium B-cytosine C), the protonated cytosinium monomer, and the neutral cytosine monomer, compared with the experimental values. The calculated dimer has an N-H BCP with an electron density 32% greater than the experimental value and is located 0.837 Å from the nitrogen and 0.251 Å from the proton. The N-H BCP in the calculated dimer is closer to the nitrogen and further from the proton than the X-ray structure of the cytosine asymmetric dimer. The BCP for the amino N-H that is hydrogen-bonded to the opposite carbonyl on the protonated side of the dimer is calculated to have eight percent electron density

less than that of the corresponding BCP in the crystal structure. The nonhydrogen-bonded amino N-H BCP on the protonated side of the dimer has a calculated value near the experimental value, while the opposite amino N-H bond critical points are calculated to have about 17% more electron density than the equivalent experimental BCP. The protonated and neutral monomers predicted BCP with electron densities within 2.7% of the densities of the calculated dimer. All but one of the calculated monomeric electron densities is greater than the corresponding BCP of the calculated dimer due to the lack of a hydrogen bonding partner. In every case, except for the N4b-H3b BCP the calculated values overestimate the electron density of the BCPs. The solvated water molecules that surround and hydrogen bond to the cytosine PBD in the crystal structure most likely remove electron density from the cytosine dimer itself. The disagreement between the experimental and calculated electron densities is due to the lack of consideration of solvent molecules or counterions in DFT-calculations.

Table 3.4. Calculated electron densities (Rho) of bond critical points of the various N-H bonds for the neutral cytosine monomer (C), the protonated cytosine monomer, and cytosine proton-bound dimer between B and C (B3LYP/6-311+G) compared with the experimental electron densities for the cytosine proton-bound dimer shown in Figure 3.18.**

Structure	N4b-H3b Rho (e/A^3)	N4b-H4b Rho (e/A^3)	N3b-H2b Rho (e/A^3)	N4c-H3c Rho (e/A^3)	N4c-H4c Rho (e/A^3)
Neutral Monomer (C)	N/A	N/A	N/A	2.29	2.31
Protonated Monomer (B)	2.08	2.28	2.02	N/A	N/A
PBD (B-C)	2.03	2.29	1.97	2.23	2.3
Experimental PBD	2.21	2.24	1.49	1.91	1.97

The remainder of this chapter will discuss the comparison of gas-phase experiments (IRMPD and mass spectrometry) performed on substituted 1-methylcytosines with *ab inito* studies, as well as the solid phase data discussed above.

IRMPD (as discussed in Chapter II) is especially useful in identifying the structure of a dissociation product where numerous fragment isomers are possible by comparing the IR spectra of the fragment ion with DFT-calculated vibrational spectra for the possible isomers. It is also possible to plot the depletion of the parent ion as function of infrared wavelength to determine the structure of a parent ion that may adopt multiple isomers.

The disorder of the hydrogen atoms involved in the strong hydrogen bond lead to disagreement about whether the crystal structure of these PBDs are symmetric or not (Table 3.3). The variable distances between cytosine PBDs seen in the X-ray of $d(ACCCT)_4$ suggest that the environment in which these PBDs come together influence its ultimate structure. Gas phase infrared studies of isolated PBD ions would remove the medium from the investigation and allow for the determination of the preferred position of the shared proton. Much of the remainder of this chapter will focus on IRMPD experiments done at the FELIX facility to determine the preferred position of the shared proton (and its N-H-N stretching frequency) of cytosine-like proton-bound dimers in the gas-phase by mass selecting for the unsolvated PBD and recording its IRMPD spectrum.

References

1. Moore, T. S.; Winmill, T. F. The State of Amines in Aqueous Solution. *J. Chem. Soc.* **1912,** 1635-1676.

2. Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids. *Nature* **1953,** 735-738.

3. Hashimoto, H.; Nelson, M. G.; Switzer, C. Zwitterionic DNA. *J. Am. Chem. Soc.* **1993,** *115,* 7128-7134.

4. Clausen-Schaumann, H.; Rief, M.; Tolksdorf, C.; Gaub, H. Mechanical Stability of single DNA Molecules. *Biophys. J.* **2000,** *78,* 1997-2007.

5. Burge, S.; Parkingson, G.; Hazel, P.; Todd, A.; Neidle, S. Quadruplex DNA: Sequence topology and structure. *Nucleic Acid Research* **2006,** *34* (19), 5402-5415.

6. Gottschling, D. E.; Stoddard, B. Structure of a chromosome aglet. *Current Biology* **1999,** *9,* R164-R167.

7. Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric G-tetraplex. *Structure* **1993,** *1,* 263-282.

8. Blackburn, E. H. Telomeres: No End in Sight. *Cell* **1994,** *92,* 621-623.

9. Harley, C. B.; Futcher, A. B.; Greider, C. W. Telomeres Shorten During Aging of Human Fibroblasts. *Nature* **1990,** *219,* 130-136.

10. Rhodes, D.; Giraldo, R. Telomere Structure and Function. *Curr. Opin. Struct. Biol.* **1995,** *5,* 311-322.

11. Simonsson, T.; Pecinka, P.; Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Research* **1998,** *26*, 1167-1172.

12. Siddiqul-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *PNAS* **2002,** *99*, 11593-11598.

13. Hanahan, D.; Weinberg, R. A. The hallmarks of cancer. *Cell* **2000,** *100,* 57-70.

14. Hannink, M.; Donoghue, D. J. Structure and function of platelet-derived growth factor

(PDGF) and related proteins. *Biochim. Biophys. Acta* **1989,** *989,* 1-10.

15. Edling, C. E.; Hallberg, B. c-Kit- a hemtopoietic cell essential receptortyrosine kinase. *Int. J. Biochem. Cell Biol.* **2007,** *39,* 1995-1998.

16. Kranenburg, O. The KRAS oncogene: past, present, and, future. *Biochim. Biophys. Acta* **2005,** *1756,* 81-82.

17. Mattei, M. G.; Borg, J. P.; Rosnet, O.; Marme, D.; Birnbaum, D. Assignment of vascular endothelial growth factor (VEGF) and placenta growth factor (PLGF) genes to human chromosome 6p12-p21 and 14q24-q31 regions, respectively. *Genomics* **1996,** *32,* 168-169.

18. Tsujimoto, Y.; Finger, L. R.; Yunis, J.; Nowell, P. C.; Croce, C. M. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* **1984,** *226,* 1097-9099.

19. Das, S. K.; Hashimote, T.; Shimizu, K. Fucoxanthin induces cell cycle arrest at G0/G1 phase in human colon carcinoma cells through up-regulation of p21WAF1/Cip1. *Biochim. Biophys. Acta* **2005,** *1726,* 328-335.

20. Cotterman, R.; Jin, V. X.; Krig, S. R.; Lemen, J. M.; Wey, A.; Farnham, P. J.; Knoepfler, P. S. N-Myc regulates a widespread euchromatic program in the human genome partially independant of its role as a classical transcription factor. *Cancer Res.* **2008,** *68,* 9654-9662.

21. Kirkpatric, K. L.; Mokbel, K. The significance of human telomerase reverse transcriptase (hTERT) in cancer. *Eur. J. Surg. Oncol.* **2001,** *27,* 754-760.

22. Brooks, T. A.; Hurley, L. H. The role of supercoilingin transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **2009,** *9,* 849-861.

23. Brooks, T. A.; Kendrick, S.; Hurley, L. Making sense od G-quadruplex and i-motif functions in oncogene promoters. *FEBS Journal* **2010,** *277,* 3459-3469.

24. Marsh, R. E.; Bierstedt, R.; Eichhorn, E. L. The crystal structure of cytosine-5-acetic acid. *Acta. Crystallogr.* **1962,** *15,* 310-316.

25. Langridge, R.; Rich, A. Molecular structure of helical polycytidylic acid. *Nature* **1963,** *198,* 725-728.

26. Langridge, R.; Wilson, H. R.; Hooper, C. W.; Wilkins, M. H. F.; Hamilton, L. D. The

molecular configuration of deoxyribonucleic acid. *J. Mol. Biol.* **1963,** *2,* 19-37.

27. Hartman, K. A.; Rich, A. The Tautomeric Form of Helical Polyribocytidylic Acid. *J. Am. Chem. Soc.* **1965,** *87,* 2033-2039.

28. Nit, E.; Hunig, I.; Kleinermanns, K.; de Vries, M. S. The nucelobase cytosine dimer by double resonance laser spectroscopy and ab initio calculations. *Phys. Chem. Chem. Phys.* **2003,** *5,* 4780-4785.

29. Neilson, P. V.; Bowers, M. T.; Chau, M.; Davidson, W. R.; Aue, D. H. Energy Transfer in Excited Ionic Species, Rates and Mechanisms of Dimerization of Protonated Amines with Their Neutral Bases. *J. Am. Chem. Soc.* **1978,** *100*, 3649-3658.

30. Aue, D. H.; Bowers, M. T. Stabilites of positive ions from equilibrium gas-phase basicity measurements. In *Gas Phase Ion Chemistry;* Bowers, M. T., Ed.; Academic Press In.: London, 1979; Vol. 2, pp2-47.

31. Kistenmacher, T. J.; Rossi, M.; Marzilli, L. G. A Model for Interrelationship Between Asymmetric Interbase Hydrogen Bonding and Base-Base Stacking in Hemiprotonated Polyribocytidylic Acid: Crystal Structure of 1-Methylcytosine Hemihydroiodide Hemihydrate. *Biopolymers* **1979,** *17,* 2581-2585.

32. Borah, B.; Wood, J. L. The cytidinium—cytidine complex: Infrared and Raman Spectroscopy Studies. *J. Mol. Struct.* **1976,** *30,* 13-30.

33. Han, S. Y.; Oh, H. B. Theoretical study on the ionic hydrogen bond in the isolated proton-bound dimer of cytosine. *Chem. Phys. Lett.* **2006,** *432,* 269-274.

34. Gehrig, K.; Leroy, J. L.; Gueron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **1993,** *363,* 561-563.

35. Leroy, J.-L.; Gueron, M. Solution structures of the i-motif tetramers of d(TCC), d(5-methylCCT) and d(T5-methylCC): novel NOE connections between amino protons and sugar protons. *Structure* **1995,** *3*, 101-120.

36. Holm, A. I. S.; Nielson, L. M.; Kohler, B.; Hoffman, S. V.; Nielson, S. B. Electronic Coupling between Cytosine bases in DNA single strands and i-motifs revealed from synchrotron radiation circular dichroism experiments. *Phys. Chem. Chem. Phys.* **2010,** *12,* 3420-3426.

37. Inman, R. B. Transistions of DNA homopolymers. *J. Mol. Biol.* **1964,** *9,* 624-637.

38. Weil, J.; Min, T.; Yang, C.; Wang, S.; Sutherland, C.; Sinha, N.; Kang, C. Stabilization of the i-motif by intramolecular adenine-adenine-thymine base pair triple in the structure of d(ACCCT). *Acta Cryst.* **1999,** *D55,* 422-429.

39. Jin, K. S.; Shin, S. R.; Ahn, B.; Rho, Y.; Kim, S. J.; Ree, M. pH-dependent structures of an i-motif DNA in solution. *J. Phys. Chem. B* **2009,** *113,* 1852-1856.

40. Sun, D.; Hurley, L. H. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* **2009,** *52,* 2863-2874.

41. Qin, Y.; Fortin, J. S.; Tye, D.; Gleason-Guzman, M.; Brooks, T. A.; Hurley, L. H. Molecular cloning of the human platelet-derived growth factor receptor b(PDGFR-b) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-b expression. *Biochemistry* **2010,** *49,* 4208-4219.

42. Schimanski, A.; Freisinger, E.; Erxleben, A.; Lippert, B. Interactions between [AuX4]− (X = Cl, CN) and cytosine and guanine model nucleobases: salt formation with (hemi-) protonated bases, coordination, and oxidative degradation of guanine. *Chim. Acta.* **1998,** *283,* 223-232.

43. Armentano, D.; De Munno, G.; Di Donna, L.; Sindona, G. Self-Assembling of Cytosine Nucleoside Into Triply-Bound Dimers in Acid Media. *J.Am. Soc. Mass Spectrom.* **2004,** *15,* 268-279.

44. Kruger, T.; Bruhn, C.; Steinborn, D. Synthesis of [(MeCyt)2H]I-Structure and Stability of a Dimeric Three fold Hydrogen-Bonded 1-Methylcytosinium1-Methylcytosine Cation. *Org. Biomol. Chem.* **2004,** *2,* 2513-2516.

45. Gdaniec, M.; Brycki, B.; Szafran, M. Crystal and Molecular Structure of Cytosine Hemitrichloroacetate. *J. Chem. Soc. Perkin. Trans. II* **1988,** 1775-1779.

46. Muller, J.; Freisinger, E. [(1-Methylcytosine)2H] I, an Asymmetric Base Pair. *Acta Crystallogr. Sect. E* **2005,** *E61,* o320-o322.

47. Murata, T.; Saito, G. Properties of Reaction Products between Cytosine and F4TCNQ in MeOH: Two Hemiprotonated Cytosine Salts with F4TCNQ Radical Anion and Methoxy Adduct Anion. *Chem. Lett.* **2006,** *35,* 1342-1343.

48. Bosnjakovic, N.; Spasojevic-de Brie, A. Cytosine-cytosinium dimer behavior in a cocrystal with decavanadate anion as a function of temperature. *J. Phys. Chem A*

2010, *114,* 10664-10675.

49. Bagratashvilli, V. N.; Letokhoc, V. S.; Markov, A. A.; Ryabov, E. A. *Mulitple Photon Infrared Photophyscis and Photochemistry;* Harwood Academic Publishers: Chur, Switzerland, 1985.

50. Polfer, N. C.; Oomens, J. Reaction Products in Mass Spectrometry Elucidated with Infrared Spectroscopy. *Phys. Chem. Chem. Phys.* **2007,** *9,* 3804-3817.

51. Oepts, D.; van der Meer, A. F. G.; van Amersfoort, P. W. The Free-Electron Laser User Facility. *Infrared Phys. Technol.* **1995,** *36*, 297-308.

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

53. Rajabi, K.; Theel, K.; Gillis, E. A.; Beran, G.; Fridgen, T. D. The structure of the protonated adenine dimer by infrared multiphoton dissociation spectroscopy and electronic structure calculations. *J. Phys. Chem. A* **2009,** *113,* 8099-8107.

54. Yaghmaei, S. *In Search of a Low Barrier Hydrogen Bond in Proton Bridged Diamines;* University of California, Riverside, 2008.

55. Pine, A. S.; Laffery, W. J.; Howard, B. J. Vibrational predissociation, tunneling, and rotational saturation in the HF and DF dimers. *J. Chem. Phys.* **1984,** *81*, 2939-2950.

56. Salpin, J.-Y.; Guilaumont, S.; Tortajada, J.; MacAleese, L.; Lemaire, J.; Maitre, P. Infrared spectra of protonated uracil, thymine and cytosine. *Chem. Phys. Phys. Chem.* **2007,** *8,* 2235-2244.

57. Liu, M.; Li, T.; Cardoso, D. S.; Fu, Y.; Lee, J. K. Gas phase thermochemical properties of pyrimidine nucelobases. *J. Org. Chem.* **2008,** *73,* 9283-9291.

58. Purrello, R.; Molina, M.; Wang, Y.; Smulevich, G.; Fresco, J. R.; Spiro, T. G.; Fossela, J. Keto-Iminol Tautomerism of Protonated Cytidine Monophosphate Characterized by Ultraviolet Resonance Raman Spectroscopy: Implications of C+ Iminol Tautomer. *J. Am. Chem. Soc* **1993,** *115,* 760-767.

59. Szczesniak, M.; Leszczynski, J.; Person, W. B. Identification of the Imino-Oxo Form of 1-Methylcytosine. *J. Am. Chem. Soc.* **1992,** *114,* 2731-2733.

60. Bowen, R. D. Ion-Neutral Complexes. *Acc. Chem. Res.* **1991,** *24,* 364-371.

61. Sivaramakrishnan, R.; Michael, J. V.; Klippenstein, S. J. Direct Observation of Roaming Radicals in the Thermal Decomposition of Acetaldehyde. *J. Phys. Chem. A* **2010,** *114,* 755-765.

62. Harding, L. B.; Georgievskii, Y.; Klippenstein, S. J. Roaming Radical Kinetics in the Decomposition of Acetaldehyde. *J. Phys. Chem. A* **2010,** *114,* 765-777.

63. Fujinami, F.; Ogawa, K.; Arakawa, Y.; Shirotake, S.; Fujii, S.; Tomita, K. The Structure of the Complex Cytosinium Hemitetrachlorozincate-Cytosine. *Acta. Crystallogr., Sect. B* **1979,** *35,* 968-970.

64. Murata, T.; Enomoto, Y.; Saito, G. Exploration of Charge-Transfer Complexes of a Nucleobase: Crystal Structure and Properties of Cytosine-Et2TCNQ Salt. *Solid State Sci.* **2008,** *10,* 1364-1368.

65. Morton, T. H. The Reorientation Criterion and Positive Ion-Neutral Complexes. *Org. Mass Spectrom* **1992,** *27,* 353-368.

66. Longevialle, P. Ion-Neutral Complexes in the Unimolecular Reactivity of Organic Cations in the Gas Phase. *Mass Spectrom. Rev* **1992,** *11,* 157-192.

67. McAdoo, D. J.; Morton, T. H. Gas Phase Analogues of Cage Effects. *Acc. Chem. Res.* **1993,** *26,* 295-302.

68. Morton, T. H. Collisional Activation and Dissociation: Decomposition via Ion-Neutral Complexes. In *The Encyclopedia of Mass Spectrometry;* Armentrout, P. B., Ed.; Elservier: London, 2003; Vol. 1, pp 467-479.

69. Morton, T. H. Theoretical Models for Ion-Neutral Complexes in Unimolecular Ion Decompositions. In *The Encyclopedia of Mass Spectrometry;* Nibbering, N. M. M., Ed.; Elservier: London, 2004; Vol. 4, pp 165-173.

70. Julian, R. R.; Ly, T.; Finaldi, A. M.; Morton, T. H. Dissociation of a Protonated Secondary Amine in the Gas Phase via an Ion-Neutral Complex. *Int. J. Mass Spectrom.* **2007,** *265,* 302-307.

Density Functional Theory and Potential Energy Surface Calculations

Unlike quadruplexes, which require cooperation among several nucleotides, proton-bound dimers between nucleobases form readily under electrospray conditions.[1] The vibrational spectrum of the hemiprotonated pairs of three homo and three hetero proton-bound dimers (Figure 3.19), along with the spectra where the five exchangeable hydrogens have been replaced with deuterium, have been recorded in the 300-1800 cm^{-1} domain using action spectroscopy on mass-selected electrosprayed positive ions isolated in an ICR cell using a free-electron laser tuned to scan the IR fingerprint region.

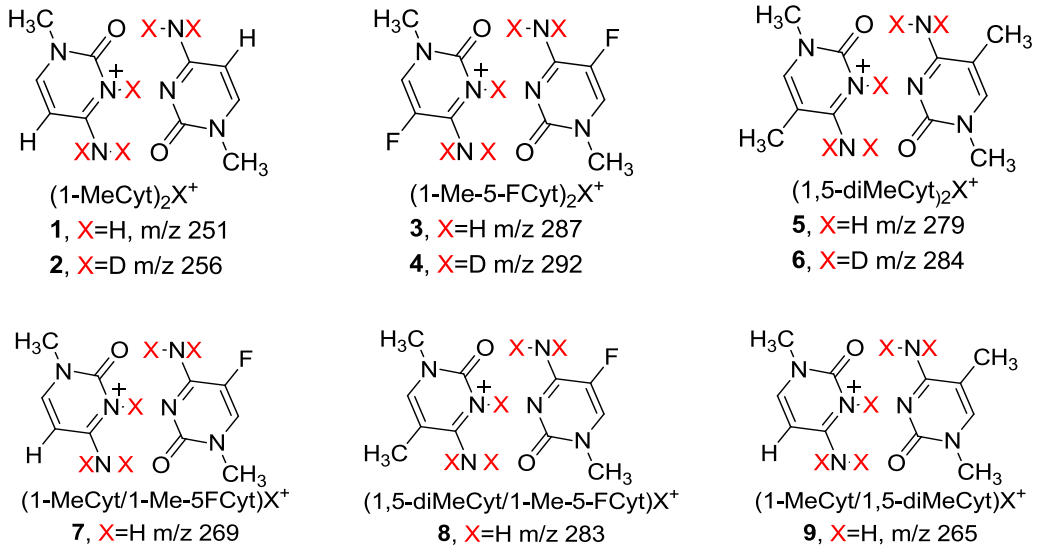


Figure 3.19. Proton-bound homo (1-6) and heterodimers (7-9) investigated by IRMPD.

Before IRMPD experiments were carried out, DFT-calculations were performed on the undeuterated PBD structures **1-9** in order to predict the energetic barrier to proton transit and to estimate where the N-H-N asymmetric

stretch would appear in the IR spectrum. Electronic energies and geometry optimizations were performed using the B3LYP/6-311++G** basis set. Harmonic and anharmonic modes were predicted at both the B3LYP/6-31G** (small basis set or SBS) and B3LYP/6-311++G** (large basis set or LBS) levels of theory. Theoretical one and two-dimensional potential energy surfaces can be used to predict N-H-N asymmetric stretches.[2] The one-dimensional potential energy surfaces for **1, 3, 7,** and **9** were plotted as the electronic energy of the PBD (potential energy, V, in kJ/mol) versus the asymmetric stretch coordinates (S_2= (R-r)/$\sqrt{2}$, in angstroms) as the proton, shared between the two ring nitrogens, moves from one side of the dimer to the other as discussed in Chapter II.

HOMODIMERS

1-Methylcytosine PBD (1)

Table 3.5 shows the unscaled and scaled by 0.97 harmonic and unscaled anharmonic DFT-calculated frequencies performed at the small and large basis sets for the N-H-N asymmetric stretch for **1**. The N-H bond strength in this PBD is weaker than a usual N-H bond, as the opposite ring nitrogen has pulled the proton 0.044 Å away from where it is calculated to be located on the N-protonated 1-methylcytosine monomer. The calculated frequencies for the N-H-N asymmetric stretch are lower than those of usual N-H bonds, typically observed at 3400 cm^{-1}.

Table 3.5. Predicted N-H-N asymmetric stretch frequencies for structure 1.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2673 cm^{-1}	2593 cm^{-1}	2201 cm^{-1}
B3LYP/6-311++G**	2750 cm^{-1}	2668 cm^{-1}	2330 cm^{-1}

The equilibrium geometry of **1**, calculated using LBS, has an N-H distance of 1.0596 Å, an N-N distance of 2.8483 Å and an electronic energy of -869.1936407 a.u. The transition state where the proton is directly between the two ring nitrogens has an N-H distance of 1.3215 Å, an N-N distance of 2.6431 Å, and an electronic energy 18.5 kJ/mol higher in energy than that of the equilibrium geometry (Table 3.6).

Atoms in Molecules 2000 was used to calculate bond critical points (BCP) between the ring nitrogens and the bridging proton. AIM has fairly accurately predicted the location and amount of electron density at BCPs for the solid-phase cytosine-cytosine PBD. Because the PBDs under investigation are in the gas phase, AIM should provide more accurate estimations of the BCPs for these dimers. The N-H BCP for the 1-methylcytosine PBD has an electron density of 1.978 $e/Å^3$ and is located 0.808 Å from the nitrogen and 0.251 Å from the proton. The Laplacian for the BCP is negative, indicative of a covalent bond. The BCP between the hydrogen-bonded nitrogen has an electron density of 0.304 $e/Å^3$ and a positive Laplacian, indicative of a non-covalent hydrogen bond.

Table 3.6. Predicted bond lengths and electronic energies as the proton moves from one side of the 1-methylcytosine PBD to the other (B3LYP/6-311++G).**

N-H Dist.(Å)	N-N Dist.(Å)	Energy (a.u.)	E_{rel} (kJ/mol)
0.98	2.8885	-869.1900005	9.5
1.0596 (EQ)	2.8483	-869.1936407	0.0
1.13	2.7960	-869.1920388	4.2
1.19	2.7405	-869.1894731	10.9
1.25	2.6877	-869.1874567	16.2
1.3215 (TS)	2.6431	-869.1865839	18.5

The potential energy surface (PES), along with its energy levels, are shown in Figure 3.20. The calculated zero-point energy (ZPE) level lies 12.75 kJ/mol below the transition state based on a proton effective mass of 14/15, suggesting that this is not a low-barrier hydrogen bond (LBHB), as LBHBs tend to have ZPE's near or above the barrier top. The height of the barrier to proton transit suggests that, at room temperature, the PBD of 1-methylcytosine does not have the necessary energy for the proton to traverse from one ring nitrogen to the other

The PES results in two energy levels below the transition state, differing by 1.1 kJ/mol (91 cm^{-1}). Structures with non-rigid hydrogen bonds can exhibit spectral doubling due to tunneling between indistinguishable conformers involving the exchange of the hydrogen bond.[3] These two energy levels are two different sum and difference ground states that have resulted from the tunneling of the proton between the two identical tautomers. The calculated E_0 to E_1 energetic transition (or, in this case, the E_0 to E_2 transition), where E_1 or E_2 is

above the barrier to proton transit, has been used to predict the frequency of the N-H-N asymmetric stretch observed in IRMPD spectra for other systems where a proton is shared between two nitrogens.[2] Analysis of the one-dimensional potential energy surface predicts this transition to be 1159 cm^{-1} and, if the calculation is correct, this region should display an experimental band that corresponds to the N-H-N asymmetric stretch, unless tunneling is operating (Table 3.7). The large amplitude of the motions that accompany proton transit argues against tunneling, however.

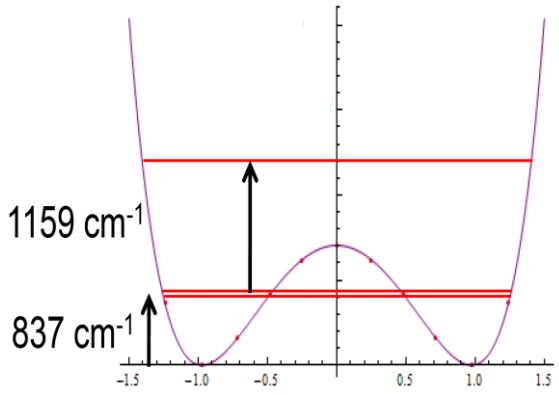


Figure 3.20. One-dimensional potential energy surface for motion of the proton in the 1-methylcytosine PBD along with predicted energetic transitions from the ground state to the first vibrationally excited state. Plotted as N-H coordinates versus energy.

Table 3.7 Predicted energy levels for the 1-methylcytosine PBD.

Energy Level	Height (kJ)	Height (cm^{-1})
E_o	9.98	837
E_1	11.07	928
TS	18.53	1554
E_2	23.80	1996

5-Fluoro-1-methylcytosine PBD (3)

Table 3.8 shows the unscaled and scaled by 0.97 harmonic and unscaled anharmonic DFT-calculated frequencies at the small and large basis sets for the N-H-N asymmetric stretch of **3**. The calculated N-H bond distance of this PBD is 0.046 Å longer than the N-H bond in the N-protonated 5-fluoro-1-methylcytosine monomer. The longer and, therefore weaker, N-H bond lowers the predicted frequencies for the asymmetric stretch compared to a typical N-H stretch as shown in Table 3.7. The predicted values for this proton bound dimer are lower than the frequencies predicted for the 1-methylcytosine PBD, suggesting that the N-H bond is weaker in this dimer. The electron withdrawing nature of the fluorine atom at the 5-position pulls electron density away from the ring nitrogen, weakening the N-H bond, and shifting the predicted N-H stretch to a lower frequency.

Table 3.8. Predicted N-H-N asymmetric stretch frequencies for structure 3.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2620 cm^{-1}	2541 cm^{-1}	2099 cm^{-1}
B3LYP/6-311++G**	2711 cm^{-1}	2630 cm^{-1}	2278 cm^{-1}

The LBS calculated equilibrium geometry of **3** predicts an N-H distance of 1.0613 Å, an N-N distance of 2.8284 Å and an electronic energy of -1067.7001087 a.u. The transition state is calculated to have an N-H distance of 1.3167 Å, an N-N distance of 2.6334 Å, and an electronic energy 16.8 kJ/mol

higher in energy than that for the equilibrium geometry (Table 3.9). The N-H distance in the equilibrium geometry is longer than that for the 1-methylcytosine PBD. The N-H distance is shorter for the 5-fluoro-1-methylcytosine transition state geometry, while the N-N distances are shorter in both geometries. The replacement of a hydrogen with a much more electronegative fluorine atom at the 5-position has pulled electron density out of the cytosine ring and away from the nitrogen that is bound to the bridging proton, resulting in the longer N-H distance. The longer N-H distance also results in a weaker N-H bond; the vibrational predictions for the N-H-N asymmetric stretch are lower than those for the 1-methylcytosine PBD. The nitrogens and the proton involved in transit have less distance to travel from the equilibrium geometry to the transition state and, therefore, the energy required to overcome the barrier to proton transit is less than that for the 1-methylcytosine PBD.

AIM analysis places the N-H BCP 0.810 Å from the nitrogen and 0.251 Å from the proton, with an electron density of 1.963 e/Å^3. The BCP for the N-H bond in the 5-fluoro-1-methylcytosine PBD is almost exactly the same distance away from the nitrogen and the proton as the BCP for the 1-methylcytosine PBD. The electron density for the 5-fluoro-1-methylcytosine PBD is less than that for the 1-methylcytosine PBD, however. This corroborates the claim that the fluorine at the 5-position of the cytosine ring pulls electron density from the ring nitrogen resulting in a weaker and longer N-H bond. The BCP between the hydrogen-bonded nitrogen and the proton in the 5-fluoro-1-methylcytosine PBD has an

electron density of 0.321 e/Å^{-3} which is greater than the corresponding BCP in the 1-methylcytosine PBD. This suggests that the strength of the bond between the nitrogen and the proton is inversely proportional to the distance of the hydrogen bond between the opposite ring nitrogen and the proton. The lack of electron density (weaker bond) between the bound proton and the ring nitrogen in the 5-fluoro-1-methylcytosine PBD means hydrogen between the proton and the opposite ring nitrogen is shorter than the corresponding hydrogen bond in the 1-methylcytosine PBD.

Table 3.9. Predicted bond lengths and electronic energies as the proton moves from one side of the 5-fluoro-1-methylcytosine PBD to the other (B3LYP/6-311++G).**

N-H Dist.(Å)	N-N Dist.(Å)	Energy (a.u.)	E_{rel} (kJ/mol)
0.99	2.8651	-1067.6973217	7.3
1.0613 (EQ)	2.8284	-1067.7001087	0.0
1.13	2.7760	-1067.6986524	3.8
1.19	2.7226	-1067.6963226	9.9
1.25	2.6715	-1067.6944567	14.8
1.3167 (TS)	2.6334	-1067.6937200	16.8

The calculated potential energy surface for **3**, along with its energy levels, are shown in Figure 3.21. The calculations suggest that this PBD will also have tunneling splitting between identical tautomers as it predicts two energy levels below the barrier, differing by 122 cm^{-1}. The calculated ground state (E_o) level lies 7.2 kJ/mol below the transition state, suggesting that this is not a low-barrier hydrogen bond (Table 3.10). The ground state energy level is 9.6 kJ/mol above the equilibrium geometry and has a E_o to E_2 transition only 4 cm^{-1} greater than

the 1-methylcytosine PBD. If the calculation is correct, the region around 1150 cm^{-1} should display experimental peaks that correspond to the N-H-N asymmetric stretch.

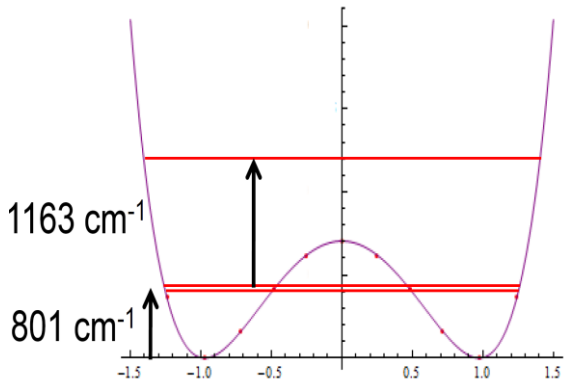


Figure 3.21. One-dimensional potential energy surface for motion of the proton in the 5-fluoro-1-methylcytosine PBD along with predicted energetic transitions from the ground state to the first vibrationally excited state. Plotted as N-H coordinates versus energy.

Table 3.10. Predicted energy levels for the 5-fluoro-1-methylcytosine PBD.

Energy Level	Height (kJ)	cm^{-1}
E_o	9.55	801
E_1	11.01	923
E_2	23.42	1964
TS	16.77	1406

1,5-Dimethylcytosine PBD (5)

DFT-calculated frequencies at the small and large basis sets for the N-H-N asymmetric stretch for **5** are displayed in Table 3.11. Much like the other proton-bound homodimers DFT calculations predict the N-H-N asymmetric stretch to occur in a region much lower than where normal N-H stretches are observed. Unfortunately, the LBS anharmonic vibrations could not be calculated

because the Gaussian 09 program suffered from an error that has not been corrected. The predicted N-H-N asymmetric stretches for the 1,5-dimethylcytosine PBD are higher than those predicted for either of the previously discussed proton-bound homodimers. This suggests that the N-H bond is stronger than either of the other two homodimers. With a methyl group (weak electron donor) at the 5-position of the cytosine ring the nitrogen has more electron density and a slightly stronger bond with the proton than the 1-methylcytosine PBD, and a much stronger bond than the 5-fluoro-1-methylcytosine PBD. Calculations predict the N-H-N asymmetric stretch to be between 10 and 20 cm^{-1} blue-shifted from the 1-methylcytosine PBD and 50-70 cm^{-1} blue-shifted from the 5-fluoro-1-methylcytosine PBD.

Table 3.11. Predicted N-H-N asymmetric stretch frequencies for structure 5.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2693 cm^{-1}	2612 cm^{-1}	2226 cm^{-1}
B3LYP/6-311++G**	2761 cm^{-1}	2678 cm^{-1}	N/A

The equilibrium geometry of **5**, calculated using the LBS, predicts an N-H distance of 1.0588 Å, an N-N distance of 2.8491 Å, and an electronic energy of -947.8544996 a.u. (Table 3.12) The transition state is calculated to have an N-H distance of 1.3234 Å, an N-N distance of 2.6468 Å, and an electronic energy 25.2 kJ/mol higher in energy than that of the equilibrium geometry. The N-H distance for the equilibrium geometry is calculated to be slightly shorter than the N-H bond

in the 1-methylcytosine dimer. The presence of a weak electron donor at the 5-position strengthens the N-H bond, therefore, making it shorter than when a hydrogen is in the 5-position. The calculated relative energy of the transition state geometry is higher than any of the other PBDs being investigated. This is due to the strength of the N-H bond in the equilibrium geometry. The short N-H distance in the equilibrium geometry results in the largest N-N bond length difference between an equilibrium and a transition state geometry, consequently **5** has largest calculated electronic energy difference between the two geometries.

AIM analysis places the BCP between the N-H bond 0.807 Å from the nitrogen and 0.251 Å from the proton, neither of which differs greatly from the BCPs for the other proton-bound homodimers. The electron density at the BCP for the 1,5-dimethylcytosine PBD, 1.983 e/Å^3, is greater than either of the other proton-bound homodimers. The BCP between the hydrogen-bonded nitrogen and the proton has an electron density of 0.301 e/Å^3 which is less than either of the other proton-bound homodimers. This furthers the suggestion that the length and strength of the ring nitrogen bound to the proton influences the length of the opposite N-H hydrogen bond. A shorter N-H bond results in a longer hydrogen bond between the proton and the opposite ring nitrogen; the longer the N-H hydrogen bond the less electrondensity at its BCP. A potential energy surface was not calculated for this homodimer, but previous results suggest that it would not be dissimilar for the other two homodimers.

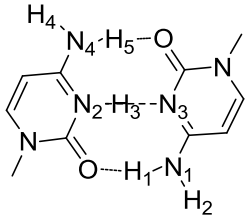
Table 3.12. Predicted bond lengths and electronic energies as the proton moves from one side of the 1,5-dimethylcytosine PBD to the other (B3LYP/6-311++G).**

N-H Dist.(Å)	N-N Dist.(Å)	Energy (a.u.)	E_{rel} (kJ/mol)
1.0588 (EQ)	2.8491	-947.8544996	0.0
1.3234 (TS)	2.6468	-947.8449057	25.2

Table 3.13 displays the electron densities for all the N-H bond critical points for each of the homodimers. The values for the three homodimers are fairly similair, but trends do appear. Namely, the more electron donating the R group at the 5-position of the protonated cytosine ring, the more electron density there is between N2 and the proton (H3). The relative electron donating nature of the substituent at the 5-position on the protonated cytosine ring also governs the electron density between the unprotonated ring nitrogen (N3) and the proton. That is, when the most electron withdrawing substituent, fluorine, is at the 5-position of the protonated ring (and the unprotonated ring) the electron density between the N3-H3 hydrogen bond is greater than the other two homodimers. This is due to the lack of electron density and longer length of the N2-H2 bond. The bond's lack of strength and longer distance allows the unprotonated ring to get closer and form a stronger hydrogen bond with the proton than in the cases of the other proton-bound homodimers.

Table 3.13. Calculated electron densities (Rho) of bond critical points of the various N-H bonds for the three homodimers formed between 1-methylcytosine, 5-fluoro-1-methylcytosine, and 1,5-dimethylcytosine.

Dimer	N4-H5 Rho (e/A^3)	N4-H4 Rho (e/A^3)	N2-H3 Rho (e/A^3)	N3-H3 Rho (e/A^3)	N1-H1 Rho (e/A^3)	N1-H2 Rho (e/A^3)
1-MeCyt PBD	2.03	2.29	1.979	0.304	2.23	2.30
5-F-1-MeCyt PBD	2.02	2.28	1.963	0.321	2.22	2.29
1,5-diMeCyt PBD	2.02	2.29	1.983	0.301	2.19	2.31



Heterodimers

1-Methylcytosine: 5-Fluoro-1-methylcytosine (7)

The tautomer of **7** where the proton rests on the 1-methylcytosine side of the dimer was calculated to be 8.8 kJ/mol more stable than the opposite tautomer and, therefore, only the vibrations for that tautomer were calculated. Table 3.14 shows the DFT-calculated frequencies performed at the small and large basis sets for the N-H-N asymmetric stretch for the more stable tautomer of **7**. All of the predicted vibrational frequencies are higher than the corresponding calculated frequencies for either the 1-methylcytosine or 5-fluoro-1-methylcytosine homodimers. One would expect the predicted vibration to be near the calculated N-H-N vibration for the 1-methylcytosine homodimer because the proton rests on that side of the dimer. The coordinating nitrogen on the 5-fluoro-

1-methylcytosine ring has less electron density on it due to the electron withdrawing fluorine at the 5-position, therefore, the hydrogen-bonded nitrogen has less pull on the proton, resulting in a stronger N-H bond and a higher predicted frequency for the N-H-N asymmetric stretch.

Table 3.14. Predicted N-H-N asymmetric stretch frequencies for structure 7.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2727 cm^{-1}	2645 cm^{-1}	2277 cm^{-1}
B3LYP/6-311++G**	2804 cm^{-1}	2720 cm^{-1}	2390 cm^{-1}

The equilibrium geometry for the more stable tautomer has an N-H distance of 1.0560 Å, an N-N distance of 2.8509 Å, and an electronic energy of -968.4486577 a.u. The less stable tautomer has a longer N-H distance and a shorter N-N distance than the more stable tautomer, suggesting that more electron density on the ring nitrogen and, therefore a shorter N-H bond results in a more stable dimer (Table 3.15). The transition state has a 1-methylcytosine N-H distance of 1.3419 Å and a 5-fluoro-1-methylcytosine N-H distance of 1.2971 Å. The N-N distance is 2.6390 Å and the electronic energy is 22.3 kJ/mol higher than the more stable tautomer's equilibrium energy. This barrier is the highest of the calculated barriers thus far and its magnitude is due to the distance the ring nitrogens have to travel from the equilibrium geometry to the transition state.

Table 3.15. Predicted bond lengths and electronic energies as the proton moves from one side of the 1-methylcytosine/5-fluoro-1-methylcytosine PBD to the other (B3LYP/6-311++G).**

5-F-1-MeCyt N-H Dist.(Å)	1-MeCyt N-H Dist.(Å)	N-N Dist.(Å)	Energy (a.u.)	E_{rel} (kJ/mol)
0.98	1.9067	2.8865	-968.4452241	9.0
1.0560 (EQ)	1.7950	2.8509	-968.4486577	0.00
1.325	1.3196	2.6446	-968.4402085	22.2
1.33	1.3123	2.6424	-968.4401867	22.2
1.3372	1.3	2.6372	-968.4401677	22.3
1.3419 (TS)	1.2971	2.6390	-968.4401623	22.3
1.3448	1.295	2.6398	-968.4401675	22.3
1.3531	1.29	2.6431	-968.4401749	22.3
1.5588	1.175	2.7338	-968.4425677	16.0
1.6440	1.13	2.7740	-968.4440987	12.0
1.7588	1.0656 (EQ)	2.8244	-968.4453007	8.8
1.8906	0.98	2.8706	-968.4413242	19.3

The potential energy surface, along with its energy levels, is shown in Figure 3.22. There are no longer two close energy levels below the barrier top, because the distinguishability of the two tautomers is sufficient that no tunneling splitting occurs. The calculated ground state (E_0) level for the more stable tautomer lies 11.0 kJ/mol below the transition state, and the less stable tautomer's ground state (E_1) is 3.9 kJ/mol below the barrier top (Table 3.16). The ground state energy level (E_0) is 11.3 kJ/mol above the calculated energy for the equilibrium geometry, the ground state energy level for the less stable tautomer is located 9.62 kJ/mol above its equilibrium geometry. The difference between ground state energy levels (7.2 kJ/mol) suggests that only the more stable tautomer should be present at room temperature. If the calculations are correct,

the E_0 to E_2 transition of 1394 cm^{-1} should provide a estimation of the N-H-N asymmetric stretch.

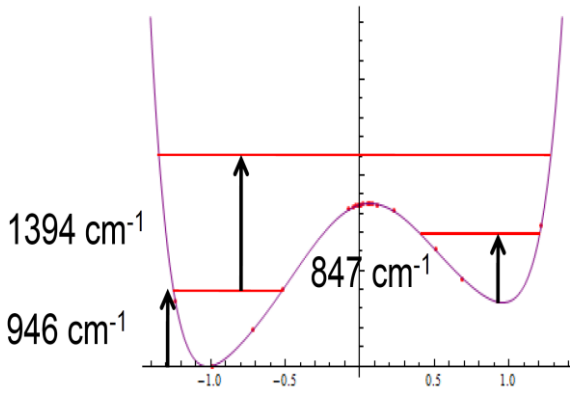


Figure 3.22. One-dimensional potential energy surface for motion of the proton in the 1-methylcytosine/5-fluoro-1-methylcytosine PBD along with predicted energetic transitions from the ground state to the first vibrationally excited state. Plotted as N-H coordinates versus energy.

Table 3.16. Predicted energy levels for the 1-methylcytosine/5-fluoro-1-methylcytosine PBD.

Energy Level	Height (kJ)	Height (cm^{-1})
E_o	11.28	946
E_1	18.44	1546
E_2	27.91	2340
TS	22.30	1870

5-Fluoro-1-methylcytosine: 1,5-Dimethylcytosine (8)

The tautomer of **8** where the proton rests on the 1,5-dimethylcytosine side of the dimer was calculated to be 11.2 kJ/mol more stable than the opposite tautomer and, therefore, only the vibrations for that tautomer were calculated.

Table 3.17 shows the unscaled and scaled by 0.97 harmonic and unscaled anharmonic DFT-calculated frequencies done at the small and large basis sets for the N-H-N asymmetric stretch for the more stable tautomer of **8.** The predicted N-H-N frequencies for this PBD are slightly higher than those of the other 1,5-dimethylcytosine containing dimers. This is due to the lack of electron density on the 5-fluoro-1-methylcytosine ring nitrogen. The hydrogen-bonded nitrogen has less pull on the proton and, therefore, strengthens the N-H bond and increases the predicted IR frequency because of the electron withdrawing nature of fluorine,.

Table 3.17. Predicted N-H-N asymmetric stretch frequencies for structure 8.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2748 cm^{-1}	2666 cm^{-1}	2279 cm^{-1}
B3LYP/6-311++G**	2825 cm^{-1}	2740 cm^{-1}	N/A

.

The lack of electron density on the hydrogen-bonded nitrogen also shortens the N-H bond length in the equilibrium geometry of the most stable tautomer of **8** (Table 3.18). The equilibrium geometry for the less stable tautomer has a longer N-H bond than any of the 5-fluoro-1-methylcytosine-containing dimers due to the electron donating methyl group on the opposite cytosine ring. The methyl group places more electron density on the hydrogen-bonded nitrogen, pulling the proton away from the 5-fluoro-1-methylcytosine ring. The difference in energy between the two tautomers is also related to the nature of

the substituent at the five position. The energy difference between these two tautomers is greater than any of the other PBDs investigated owing to the relative basicity of the ring nitrogen. This PBD also has the largest relative transition state owing to the distance the two ring nitrogens must travel from equilibrium to transition state geometry. The more electron density the ring nitrogen has the more basic it is and the better it can stabilize the positive charge on the proton. The tautomer with the proton bound to 1,5-dimethylcytosine stabilizes the proton much better than the opposite tautomer, leading to a large electronic energy difference between the tautomers and a large electronic energy difference at the transition state.

Table 3.18. Predicted bond lengths and electronic energies as the proton moves from one side of the 1,5-dimethylcytosine/5-fluoro-1-methylcytosine PBD to the other (B3LYP/6-311++G).**

1,5-diMeCyt N-H Dist.(Å)	5-F-1-MeCyt N-H Dist.(Å)	N-N Dist.(Å)	Energy (a.u.)	E_{rel} (kJ/mol)
1.0545 (EQ)	1.8024	2.8567	-1007.7785610	0.00
1.3588	1.2931	2.6517	-1007.7694232	24.0
1.7570	1.0665 (EQ)	2.8235	-1007.7742920	11.2

The potential energy surface for this PBD was not calculated, but the energy difference between the two tautomers likely would have caused a larger E_0 to E_2 transition than the others previously predicted.

1,5-dimethylcytosine: 1-methylcytosine (9)

The tautomer where the proton is situated on the 1,5-dimethylcytosine side of the dimer is calculated to be only 2.44 kJ/mol more stable than the other tautomer. The N-H-N asymmetric stretch was calculated for each of the two tautomers (Tables 3.19 and 3.20). As expected, the predicted frequencies are higher for the tautomer where the proton resides on the 1,5-dimethylcytosine side, as it has a stronger N-H bond, due to the electron donating methyl group. The predicted frequencies for the N-H-N stretch for the more stable tautomer are between the values calculated for 1,5-dimethylcytosine homodimer and the 1,5-dimethylcytosine: 5-fluoro-1-methylcytosine heterodimer due to the relative basicities of the hydrogen-bonded cytosine partner. 1,5-Dimethylcytosine is more basic than 5-fluoro-1-methylcytosine and, therefore, has more of a pull on the proton, thus weakening the N-H bond. The same trends are true for the less stable tautomer. The calculated N-H-N asymmetric stretch frequencies are between those calculated for the 1-methylcytosine homodimer and the 1-methylcytosine: 5-fluoro-1-methylcytosine heterodimer. Fluorine removes electron density from the hydrogen-bonded nitrogen, providing less pull on the proton and strengthening the N-H bond. When a methyl group takes the place of the fluorine, it gives the hydrogen-bonded nitrogen more electron density and more pull on the proton, lengthening the N-H bond.

Table 3.19. Predicted N-H-N asymmetric stretches for the more stable tautomer of structure 9.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2705 cm^{-1}	2612 cm^{-1}	2248 cm^{-1}
B3LYP/6-311++G**	2774 cm^{-1}	2678 cm^{-1}	2357 cm^{-1}

Table 3.20. Predicted N-H-N asymmetric stretches for the less stable tautomer of structure 9.

Basis Set	Harmonic	Scaled Harmonic	Anharmonic
B3LYP/6-31G**	2666 cm^{-1}	2612 cm^{-1}	2193 cm^{-1}
B3LYP/6-311++G**	2737 cm^{-1}	2678 cm^{-1}	2312 cm^{-1}

As with the N-H bond strength, the N-H and N-N distances for each tautomer's equilibrium geometry fall between the values for other proton-bound dimers containing 1-methylcytosine or 1,5-dimethylcytosine (Table 3.21). The electronic energy difference between the equilibrium geometry of the two tautomers (2.4 kJ/mol) suggests that they both may be present at room temperature. The transition state electronic energy is calculated to be 20.2 kJ/mol above the equilibrium geometry of the more stable tautomer. This is lower than for the other proton-bound heterodimers, but still large enough that it should not have enough energy at room temperature to overcome the barrier and oscillate the proton between the two tautomers on the IR time scale.

Table 3.21. Predicted bond lengths and electronic energies as the proton moves from one side of the 1,5-dimethylcytosine/1-methylcytosine PBD to the other (B3LYP/6-311++G).**

1,5-diMeCyt N-H Dist. (Å)	1-MeCyt N-H Dist. (Å)	N-N Dist. (Å)	Energy (a.u.)	E_{rel} (kJ/mol)
1	1.8837	2.8836	-908.5216151	4.80
1.0578 (EQ)	1.7956	2.8535	-908.5234451	0.00
1.1	1.7242	2.8241	-908.5227746	1.76
1.15	1.6340	2.7840	-908.5208576	6.79
1.25	1.4434	2.6934	-908.5168334	17.36
1.32 (TS)	1.3287	2.6487	-908.5157543	20.19
1.35	1.2883	2.6383	-908.5158093	20.05
1.4	1.2333	2.6333	-908.5164134	18.46
1.6281	1.15	2.7781	-908.5201586	8.63
1.7201	1.1	2.8201	-908.5219357	3.96
1.7887	1.0602 (EQ)	2.8489	-908.5225159	2.44
1.8819	1	2.8818	-908.5205868	7.50

The one-dimensional potential energy surface for the motion of the proton from one side of the proton-bound heterodimer to the other is shown in Figure 3.23. The calculated ground state (E_0) level for the more stable tautomer lies 8.51 kJ/mol below the transition state, and the less stable tautomer's ground state (E_1) is 7.3 kJ/mol below the barrier top (Table 3.22). The ground state energy level (E_0) is 11.7 kJ/mol above the calculated energy for the equilibrium geometry, the ground state energy level for the less stable tautomer is located 10.4 kJ/mol above its equilibrium geometry. The first excited state above the barrier to proton transit (E_2) is calculated to be 25.8 kJ/mol above the equilibrium geometries electronic energy and 14.1 kJ/mol above the ground state energy for the more stable tautomer (E_0). This results in a predicted E_0 to E_2 transition at

1183 cm^{-1}, which is very close to the values predicted for the 1-methylcytosine and 5-fluoro-1-methylcytosine proton-bound homodimers, but much lower than the heterodimer between those two.

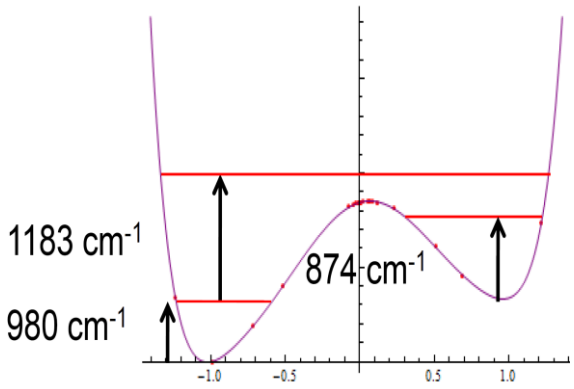


Figure 3.23. One-dimensional potential energy surface for motion of the proton in the 1,5-dimethylcytosine/1-methylcytosine PBD along with predicted energetic transitions from the ground state to the first vibrationally excited state. Plotted as N-H coordinates versus energy.

Table 3.22. Predicted energy levels for the 1,5-dimethylcytosine/1-methylcytosine PBD.

Energy Level	Height (kJ)	Height (cm^{-1})
E_o	11.68	980
E_1	12.86	1078
E_2	25.80	2163
E_3	35.76	2999
TS	20.19	1693

The basicity of the nitrogen bonded to the proton in these PBDs is directly related to the substituent at the 5-position of the cytosine ring. The more electron donating the atom the more basic the nitrogen. Conversely, the more electron withdrawing the substituent the less basic the nitrogen. The basicity of the nitrogen bound to the proton plays a direct role in all of the calculations discussed above. The basicity is directly proportional to the strength of the N-H

bond and, therefore, the frequency of the predicted N-H-N asymmetric stretch, but inversely related to the N-H distance. The opposite is true of the nitrogen hydrogen-bonded to the proton. The more basic the hydrogen-bonded nitrogen, the more pull it has on the proton and the weaker (longer) the N-H bond is, resulting in a lower predicted vibration for the N-H-N asymmetric stretch. The stronger the N-H bond (and weaker the N---H hydrogen bond) the further the nitrogens must travel from the equilibrium geometry to the transition state. The relative energy of the transition state is directly related to the distance the nitrogens travel and, therefore, to the strength of the N-H bond.

IRMPD Experiments

The DFT-calculated N-H-N asymmetric stretches are between 2100-2800 cm^{-1} is out of range for the free-electron laser at FELIX. The free-electron laser used in the IRMPD experiments can be tuned to investigate the finger print region of the IR spectrum, from 300-1800 cm^{-1}. If no N-H-N asymmetric stretch is observed in the IRMPD spectra it may be that one of the DFT-calculations has accurately predicted its value, but experiment cannot yet test that claim The ground state to excited state transitions predicted for some of the PBDs are between 1100 and 1400 cm^{-1}, therefore, if these predictions are accurate, the N-H-N asymmetric stretch should be visualized by IRMPD.

1-Methylcytosine (1)

A solution of 1-methylcytosine in a mixture of water and methanol was electrosprayed and the ion corresponding to its PBD (m/z 251) was mass-selected in an ICR cell, and the isolated ions subjected to the free electron laser tuned to explore the IR fingerprint region of 300-1800 cm^{-1}. Initially, the structure of the PBD between two molecules of 1-methylcytosine was determined by comparing the IRMPD spectra to scaled harmonic DFT (B3LYP/6-31G**) frequencies of the possible conformations that the proton-bound dimer between two molecules of 1-methylcytosine (m/z 251) might adopt (Figure 3.24).

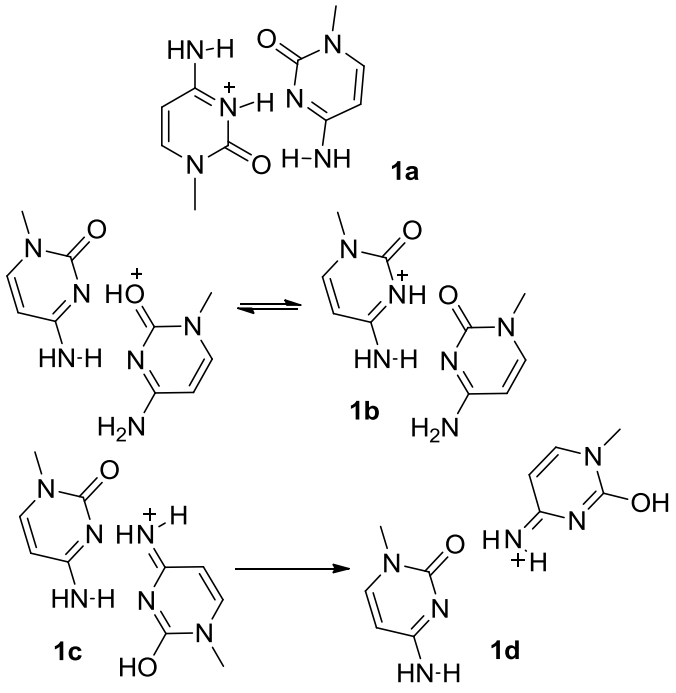


Figure 3.24. Lowest energy conformations of the proton-bound dimer between two molecules of 1-methylcytosine.

Figure 3.25 shows the IRMPD spectra of the m/z 251 ion along with the scaled DFT predicted IR profiles for the four lowest energy geometries a PBD formed between two molecules of 1-methylcytosine can adopt, calculated at the B3LYP/6-31G** level of theory. For dimer **1b** DFT calculations show that the bridging proton moves from the carbonyl oxygen to the nitrogen on the opposite ring without a barrier and prefers to be on the nitrogen on the left hand ring. Dimer **1b** has only two hydrogen bonds while **1a** has three, which is partially responsible for the fact that **1a** is calculated to be 35 kJ/mol more stable. **1b** does have the advantage of possibly being incorporated in two anti-parallel strands of duplex DNA without reversal. DFT predicts the carbonyl stretch of the protonated ring in **1b** to be at 1798 cm^{-1} and, upon deuteration of the five exchangeable hydrogens, this stretch should be red-shifted by 11 cm^{-1}. A 30 cm^{-1} red-shift is observed when the five exchangeable hydrogens have been replaced with deuterium (discussed below) instead. The 30 cm^{-1} red-shift of the carbonyl stretch better agrees with the DFT predicted spectra of **1a** and its d_5 analogue (**2)**. Dimer **1c,** like **1b**, enjoys only two hydrogen bonds between the individual monomers and is almost 130 kJ/mol less favorable than dimer **1a**. After a barrier-free one-hundred and eighty degree twist about the carbonyl bond, **1d** has only a single hydrogen bond and is 79 kJ/mol less favorable than **1a**. The predicted spectrum for structure **1b** provides no overlap for five bands observed in the IRMPD spectrum of the m/z 251 ion (1758 cm^{-1}, 1567 cm^{-1}, 1256 cm^{-1}, 1040 cm^{-1}, and 892 cm^{-1}). Structure **1c** also leaves five experimental bands

without any overlap with theory (1758 cm^{-1}, 1567 cm^{-1}, 1425 cm^{-1}, 892 cm^{-1}, and 455 cm^{-1}). DFT-calculated vibrations of structure **1d** do a better job approximating the experimental spectrum than structure **1c**, but four experimental peaks still remain unmatched by the theoretical spectrum (1758 cm^{-1}, 1567 cm^{-1}, 706 cm^{-1}, and 455 cm^{-1}) . The theoretical predictions for structure **1a** give the best overlap with the experimental spectrum, leaving only the intense band at 1567 cm^{-1} unfit. The relative stabilities of the dimers coupled with the fact that all but one of the observed vibrations can be assigned by DFT scaled harmonic calculations to the vibrations of **1a,** suggest that the structure of the isolated m/z 251 ion will have the proton situated between the two ring nitrogens in the gaseous ion just as in the crystalline samples discussed above. This binding motif is assumed for all of the subsequent PBDs to be discussed.

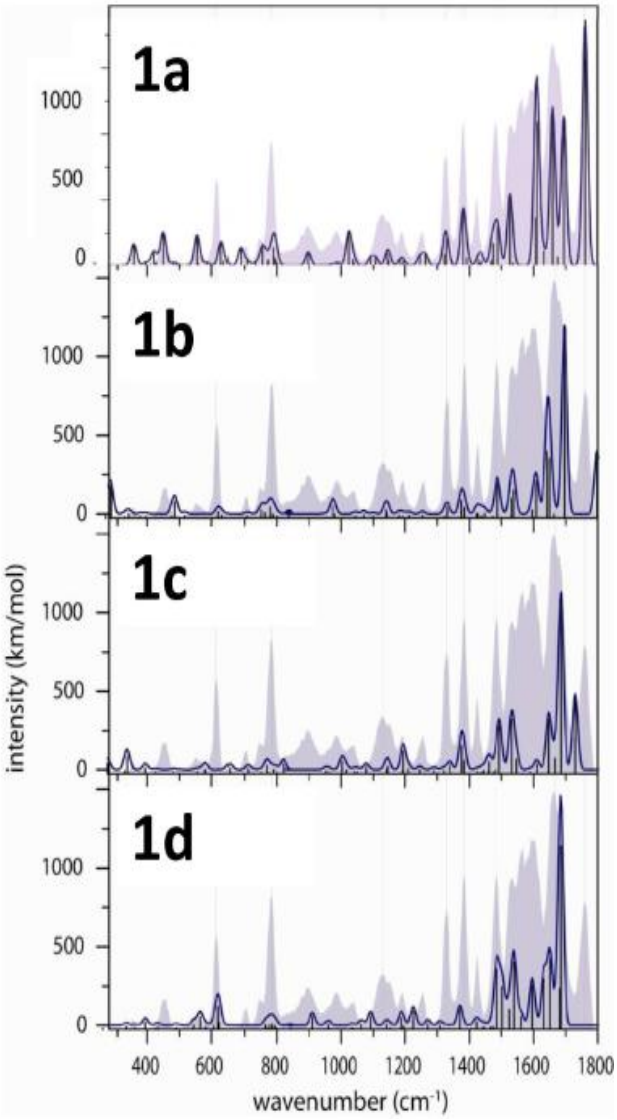


Figure 3.25. Comparison of IR absorption profiles calculated at B3LYP/6-31G for proton-bound dimer structures 1a-1d (20 cm^{-1} Gaussian broadening; frequencies >850 cm^{-1} scaled by 0.97) with the experimental IRMPD spectrum (silhouette).**

Upon IR-induced dissociation of the PBD formed between two molecules of 1-methylcytosine (**1a**) one neutral molecule of 1-methylcytosine would be expelled to leave one of four possible monoprotonated 1-methylcytosinium ions (Figure 3.26). In monitoring the dissociation of **1a** into the protonated monomer,

ten well defined bands are displayed between 1300 and 1800 cm^{-1}, nine of which match theory. At the power level of the free electron laser, multiple photon absorptions causes extensive decomposition of **1a** at resonant frequencies, as the y-axis indicates, with a fractional dissociation of greater than fifty percent for the most intense band.

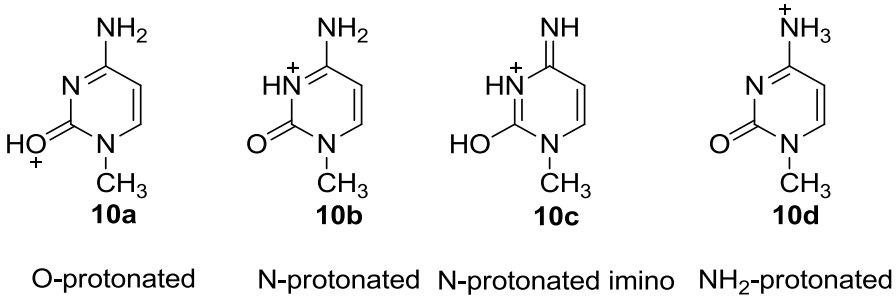


Figure 3.26. Lowest energy conjugate acid ions of 1-methylcytosine.

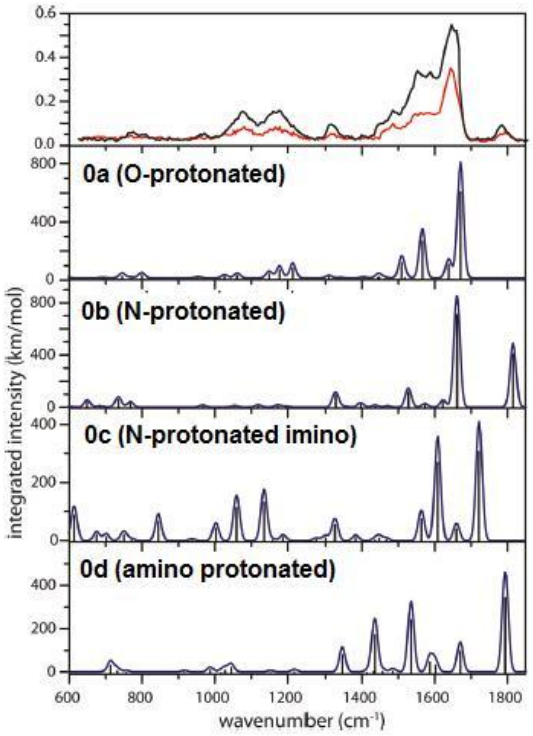


Figure 3.27. Experimental IRMPD spectrum of the protonated monomer of 1-methylcytosine (m/z 126) (top panel: black trace from dissociation of the proton-bound dimer by CO_2 laser photolysis; red trace from electrospray ionization) compared with normal modes calculated at B3LYP/6-31G (for which frequencies >850 cm^{-1} have been scaled by 0.97) with 20 cm^{-1} Gaussian broadening for the four possible tautomers, 0a-0d.**

Salpin, et al. have studied the IRMPD spectrum of the protonated cytosine monomer. Their published theoretical IR absorption profiles for the two most stable tautomeric cations resemble those predicted for isomers **10a** and **10b**.[4] The IRMPD spectrum of induced dissociation of **1a** to the monomeric cation is shown in Figure 3.27, along with the theoretically predicted vibrations of the most stable protonated cytosine monomers **10a-d**. Comparing the previously reported IR spectrum of protonated cytosine with theory shows that O-protonated cytosine predominates under electrospray conditions. The published IRMPD spectrum of

the cytosinium ion shows the most intense band at 1641 cm^{-1}, corresponding to the carbonyl stretch of the O-protonated cytosine. The occurrence of a small band at 1804 cm^{-1} suggests that a small amount of N-protonated cytosine is also present. IR induced dissociation of dimer **1a** to form the protonated 1-methylcytosine monomer shows the most intense band at 1648 cm^{-1} (top panel of Figure 3.27), which overlaps with the in-phase stretch of the carbon-carbon double bond and both carbon-nitrogen double bonds in the theoretical spectrum of ion **10a**. If the band at 1804 cm^{-1} in the IRMPD spectrum from electrosprayed cytosine[4] does come from a small percentage of N-protonated tautomer, then the band at 1779 cm^{-1} observed for protonated 1-methylcytosine, as seen in the top panel of Figure 3.27, should correspond to the carbonyl stretch from ion **10b.** This value is close to the 30 cm^{-1} shift predicted for the carbonyl stretch after methylating cytosine. DFT calculations of O-protonated 1-methylcytosine (**10a**) and N-protonated 1-methylcytosine (**10b**) show the two tautomers have similar basicities, consistent with results from prior investigations.[5]

Geometry optimizations done at similar levels of theory on protonated cytosines (without the attached methyl group) give the same results. Higher level computations (i.e., CCSD(T)/aug-cc-pVTZ) predict that the O-protonated tautomer is about 6 kJ/mol more favorable than the N-protonated tautomer. Applying these results to 1-methylcytosine, which has a gas phase basicity about 8-10 kJ/mol higher than that of cytosine, implies that the IR induced dissociation of **1a** yields predominantly **10a** as the favored tautomer. The relative intensities

of Salpin et al.'s IRMPD bands agree with the 94:6 O-protonated/N-protonated equilibrium proportions that are predicted at 300 K. This suggests that tautomers **10a** and **10b** have comparable dissociation efficiencies under IRMPD conditions. That is, the most stable tautomers of protonated cytosine, and by extension **10a** and **10b**, dissociate into nearly the same proportions as predicted at 300 K. From the IRMPD results it can be concluded that the major tautomer of IR induced dissociation is **10a** with a small amount of **10b**, which DFT (B3LYP/6-31G**) predicts to be 41 kJ mol^{-1} less favorable than **10a**.

The other possible candidates for the minor tautomer of protonated 1-methylcytosine are not as plausible as **10b**. Considerable discussion has focused on the imino form of cytosine[6, 7] whose conjugate acid ion has the proton on the ring nitrogen and whose *N*-methyl homologue (**10c**) has a calculated heat of formation (at 0 K) 125 kJ/mol greater than that of **10a**. If **10c** was present as a minor tautomer the IRMPD results should exhibit a band at 1723 cm^{-1} which, according to computations, corresponds to the in-phase stretching of the carbon-carbon double bond coupled with the exocyclic carbon-nitrogen double bond. This band is not present. Although **10d** cannot be completely eliminated as a possible minor tautomer in the IRMPD spectrum it is calculated to have a heat of formation (at 0 K) 136 kJ mol^{-1} higher than **10a**, which makes it unlikely to be present.

The evidence supports the assignment of **1a** (where the proton is shared between two nitrogens) as the structure of PBD between two molecules of 1-methylcytosine in the gas phase and **10a** (where the proton is located on the carbonyl oxygen) as the major tautomer that results from the IR induced dissociation of the PBD. The proton can move from the ring nitrogen to the carbonyl oxygen in the process of IR induced dissociation by one of two possible mechanisms.

One hypothesis supposes kinetic control: the vibrationally excited dimer passes through a sequence of hydrogen-bonded structures in the course of decomposition, and the resulting monomer ion tends to retain memory of the last structure (e.g. **1d**) that intervenes prior to final dissociation. An alternate interpretation supposes thermodynamic control: the dissociation initially forms an ion-neutral complex, in which a proton passes back and forth between the partners, sampling all of the accessible possibilities. Such noncovalent complexes have many precedents in unimolecular ion decompositions [8-16] and uncharged analogues have recently been inferred in homolyses of gaseous neutrals via “roaming radicals”.[9,10] Assuming that the partners in the complex partition vibrational energy statistically in the course of reversible proton transfer, this picture implies that the products reflect equilibrium at a temperature corresponding to the internal energy shared by the monomeric products. The intensity of the 1779 cm^{-1} band (from **0a**) relative to the 1648 cm^{-1} band (from **0b**) is nearly the same for both the red (electrospray ionization) and the black traces

(IR induced dissociation) in the top panel of Figure 3.27. This suggests that the effective temperature of dissociating dimers does not differ greatly from the temperature of the ions produced by an electrospray source.

The 5-fluoro-1-methylcytosine protonated monomer ion (m/z 144) was isolated in an ICR, the mass selected ions subjected to a CO_2 laser, and the mass spectrum of the IR induced decomposition was recorded. Five fragments were observed from the IR excited 5-fluoro-1-methylcytosinium ion (m/z = 127, 101, 99, 81, and 74). The m/z 127 ion corresponds to parent ion's loss of ammonia, m/z 101 to loss of isocyanic acid (HN=C=O), m/z 99 to the loss of the elements of formamide, m/z 81 to the loss of isocyanic acid followed by the loss of hydrogen fluoride, and m/z 74 to the loss of isocyanic acid followed by the loss of hydrogen isocyanide (Figure 3.28).

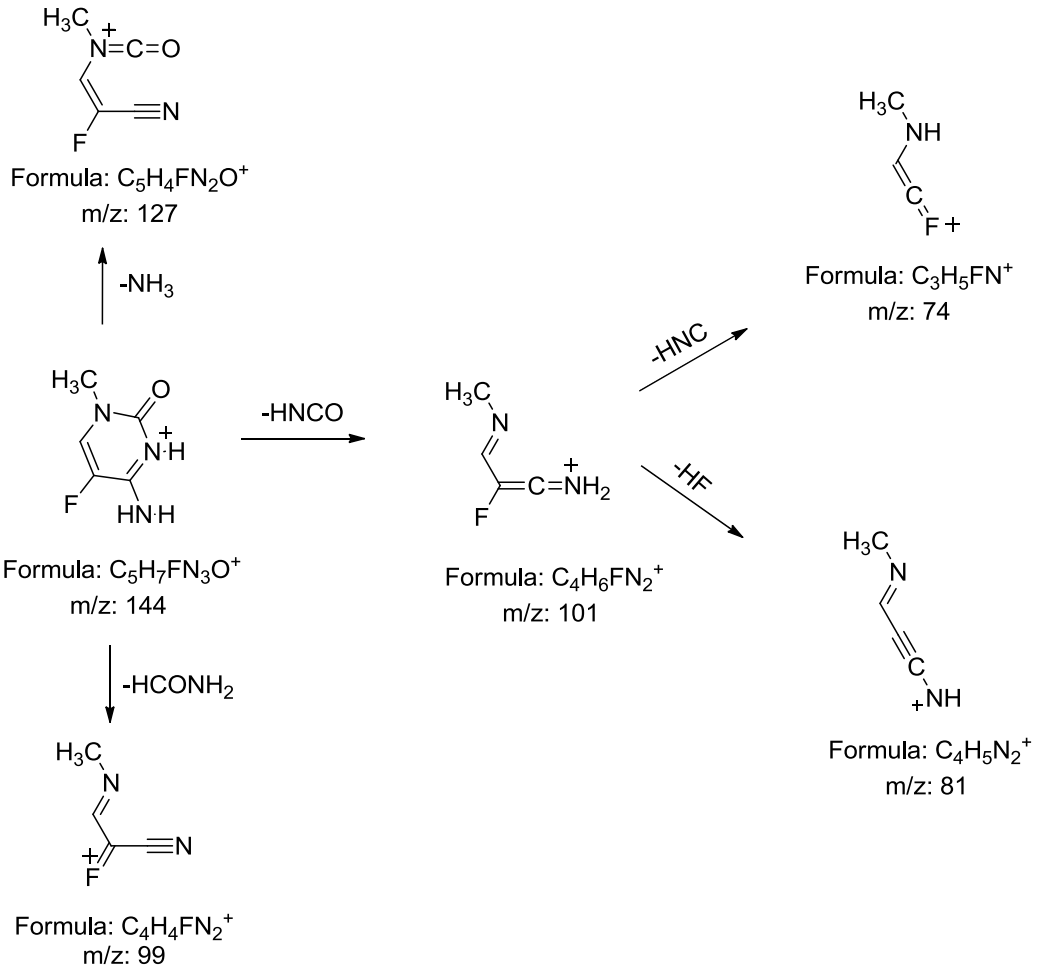


Figure 3.28. Fragmentation pattern inferred for the protonated monomer of 5-fluoro-1-methylcytosine (m/z 144) as deduced from collisionally activated and IR induced dissociation.

Collisionally activated dissociation (CAD) experiments were performed in a triple-quadrupole mass spectrometer with argon as a collision gas on the 5-fluoro-1-methylcytosinium ion, in order to compare its fragmentation pattern with the IRMPD mass spectrum. A center-of-mass-collision energy of 3.2 eV in the triple quadrupole mass spectrometer (15V lab frame) results in a fragmentation pattern similar to that of the IR multiple photon-induced

dissociation. Increasing the center-of-mass collision energy to 4.3 eV (20V lab frame) results in a more robust fragmentation pattern that includes fragments at m/z 113, 103, 72, 54, and 42 as well as the five fragments observed in the IRMPD spectrum.

CAD experiments were also performed on the 5-fluoro-1-methylcytosinium-d_3 ion (m/z 147) in which the 3 exchangeable hydrogens were replaced by deuterium by ESI in D_2O. Five principal fragment ions are produced from CAD experiments with a center-of-mass collision energy of 3.2 eV (m/z = 127, 103, 99, 82, and 75). These fragments correspond to the deuterated analogues of the fragments observed in the IR induced dissociation discussed above. Increasing the collision energy to 4.3 eV results in fragment ions that correspond the deuterated analogues of the higher energy CAD fragment ions mentioned above, with two exceptions. Fragment ions from the d_3 parent ion observed at m/z 76 and 83 have much lower intensity than the m/z 75 and 82 ions, but are still detectable. This result suggests that the hydrogen that is lost by expulsion of neutral hydrogen isocyanide or neutral hydrogen fluoride from the m/z 101 ion produced in the undeuterated CAD experiment can incorporate either exchangeable or non-exchangeable hydrogens.

Figure 3.29 shows the IRMPD spectra of **1a** along with its unscaled harmonic, scaled harmonic, and unscaled anharmonic IR absorption profiles predicted by DFT at small and large basis sets. Neither the large or small basis

set unscaled DFT predictions do an adequate job fitting the experimental spectrum. Two theoretical peaks differ from any observed vibrations by a large margin. Both basis sets predict a carbonyl stretch coupled with the in-plane motion of the bridging proton to be at a higher frequency than any peak that is observed (1818 cm^{-1} for the SBS, and 1793 cm^{-1} for the LBS). They each also predict a vibration slightly above 1700 cm^{-1} that is not observed. The SBS vibration in above 1700 cm^{-1} corresponds to a carbon-carbon double bond and the carbon-nitrogen amino stretches coupled with NH_2 scissoring motions (1713 cm^{-1}), while the LBS prediction corresponds to NH_2 scissoring coupled with both carbonyl stretches (1704 cm^{-1}). This suggests that unscaled harmonic DFT vibrations are not an appropriate theoretical tool for predicting the vibrations for these types of molecules.

Scaling the DFT harmonic vibrations above 800 cm^{-1} by 0.97 gives better agreement with the observed spectrum for both basis sets. Surprisingly, the small basis set produces a theoretical spectrum that better matches what is observed than does the LBS. The LBS scaled harmonic predictions are slightly red shifted from many of the observed bands, while the SBS overlaps with all but one of the observed bands. Although the LBS doesn't completely overlap with the experimental spectrum, both basis sets provide assignments for every experimental band, save for one, when the harmonic vibrations are scaled by 0.97. The two poorly fit SBS unscaled harmonic vibrations discussed above now appear at 1763 cm^{-1} and 1662 cm^{-1} and differ from the experimental peak centers

by 5 cm^{-1} and 1 cm^{-1} respectively. Scaling the poorly fit harmonic LBS vibrations discussed above results in bands at 1739 cm^{-1} and 1653 cm^{-1} which are both red-shifted from the corresponding experimental bands by 19 cm^{-1} and 29 cm^{-1} respectively.

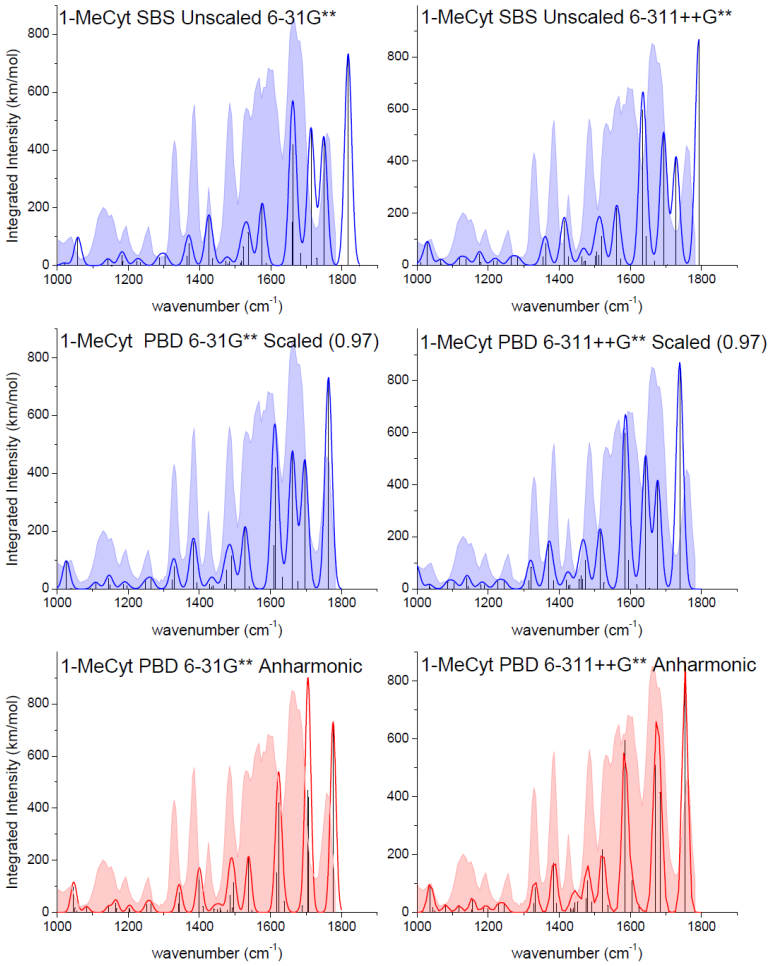


Figure 3.29. Experimental IRMPD spectrum of the proton-bound dimer of 1-methylcytosine (in silhouette) compared with normal modes of structure 1a calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

Table 3.23. Comparison of experimental and theoretical band positions (cm^{-1}) for the 1-methylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	LBS Anharmonic	Assignment
1758	1763	1776	1739	1754	C=O stretch & in-plane NH bend
1682	1697	1707	1677	1683	NH in-plane bends
	1677	1690	1653	1665	C=O stretch
1661	1662	1704	1643	1671	NH in-plane bends
	1634	1639	1618	1624	C=C &C-NH_2 stretches
1606	1613 1610	1624 1618	1595 1585	1604 1584	NH in-plane bends & CO stretch
1567					
1531	1539	1548	1526	1536	C=N stretch & HCH bends
	1529	1538	1515	1521	C=C stretch & HCH bends
1485	1490 1477	1497 1486	1473 1466	1479 1491	HCH & NH in-plane bends C-NH_2 stretch
1426	1430	1459	1420	1444	methyl HCH bends
1386	1394 1383	1410 1399	1384 1371	1393 1385	CH_3 umbrella, in-plane ring CH bends
1329	1330 1324	1344 1341	1321 1314	1332 1329	ring CH in-plane bends
1257	1265 1248	1264 1251	1245 1227	1246 1229	N-CO stretches
1194	1197	1205	1189	1199	ring CH in-plane bends
	1187	1203	1179	1188	
1152	1149 1146	1168 1164	1143 1140	1155 1156	HC=CH in-plane bends
	1119	1134	1114	1129	CH_3 umbrella, HNH
1128	1118 1108	1132 1145	1113 1104	1123 1119	in-plane rocking
	1042	1055	1037	1050	
1040	1038	1050	1034	1044	N-H-N out-of-plane bend
	1026	1046	999	1036	CH_3 rocking

The theoretical anharmonic spectra offer an improvement over their corresponding basis set's unscaled harmonic predictions, but neither provide as much overlap with the experimental spectrum as does the SBS scaled harmonic predictions. The two largest SBS anharmonic peaks (1776 cm^{-1} and 1707 cm^{-1}) both differ from their corresponding experimental bands by over 15 cm^{-1} while the larger basis set was within 7 cm^{-1} for those peaks (1754 cm^{-1} and 1683 cm^{-1}). The large basis set also provides better overlap for the lower frequency peaks as well. Each of the six predicted spectra can assign all but one of the observed bands, but the SBS scaled harmonic and large basis set anharmonic best match the observed spectrum and their assignments are shown in Table 3.23.

In both the scaled harmonic and anharmonic theoretical spectra the only unassigned peak is seen at 1567 cm^{-1}, while the unscaled harmonic predictions do overlap this peak, nor do they fit the peak slightly above 1600 cm^{-1}. None of the other possible dimer structures (**1b-1d**) provide a fit for this intense peak, and it may correspond to the movement of the bridging proton from one ring nitrogen to the other. Neither harmonic unscaled (2673/2750 cm^{-1}), scaled (2593/2668 cm^{-1}) nor anharmonic (2201/2330 cm^{-1}) Gaussian calculations done at either the small or large basis set predict the N-H-N stretch to be in the region of 1570 cm^{-1}. In an attempt to confirm the assignment of the experimental band at 1567 cm^{-1} as the N-H-N asymmetric stretch the IRMPD spectrum of the 1-methylcytosine-d_5 dimer (**2**) was recorded.

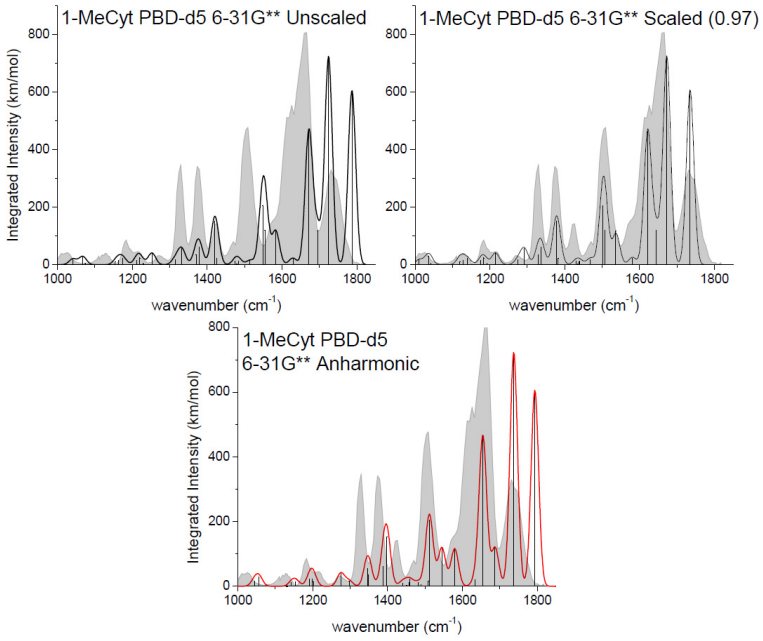


Figure 3.30. Experimental IRMPD spectrum of the d_5 analogue of the proton-bound dimer of 1-methylcytosine compared with normal modes of structure 2 (unscaled and scaled by 0.97) calculated at B3LYP/6-31G (20 cm^{-1} Gaussian broadening).**

1-Methylcytosine-d_5 (2)

1-Methylcytosine was dissolved in a mixture of D_2O and methanol-OD in order to replace all the exchangeable hydrogens with deuterium. The IRMPD of the deuterium-bound dimer between two molecules of 1-methylcytosine-d_2 (m/z 256, **2)** is shown in Figure 3.30 along with its unscaled and scaled harmonic and unscaled anharmonic DFT frequencies at the SBS**.** The four highest frequency unscaled predicted vibrations at 1787 cm^{-1}, 1725 cm^{-1}, 1672 cm^{-1}, and 1549 cm^{-1} all overestimate their respective experimental peaks by more than 40 cm^{-1}. Scaling the SBS harmonic predictions by 0.97 provides a good match overlap

with the IRMPD of **2**. Scaling the four peaks mentioned above (1733 cm^{-1}, 1673 cm^{-1}, 1622 cm^{-1}, 1502 cm^{-1}) results in bands that fit under their respective experimental peaks and which can be easily assigned (Table 3.24). If the peak predicted by the SBS anharmonic calculation at 1793 cm^{-1}, which corresponds to the N-D-N asymmetric stretch, is ignored the anharmonic predictions do an adequate job fitting the experimental bands. The four highest frequency peaks (1735 cm^{-1}, 1687 cm^{-1}, 1634 cm^{-1}, and 1509 cm^{-1}) are an improvement over the fit provided by the unscaled vibrations and comparable to the scaled harmonic predictions. The N-D-N asymmetric stretch band that appears in the anharmonic theoretical spectrum (that lacks a corresponding experimental band) is predicted by the SBS unscaled harmonic calculations to occur at 2007 cm^{-1}, scaling by 0.97 gives a band at 1947 cm^{-1}, both of which are out of the range of the experiment. The only difference between the overlap provided by the SBS scaled harmonic spectrum and the anharmonic spectrum with the experiment arise between 1300 and 1400 cm^{-1}. The experimental bands centered at 1330 cm^{-1} and 1374 cm^{-1} are better matched by the SBS scaled harmonic vibrations (1329/1337 cm^{-1} and 1382/1378 cm^{-1}) than the anharmonic vibrations (1346/1349 cm^{-1} and 1388/1398 cm^{-1}). Both basis sets assign every experimental band, but the scaled harmonic vibrations from the SBS provides the best overlap with the experimental spectrum.

The highest frequency experimental band of the h_5 analogue in Figure 3.27, 1761 cm^{-1}, combines the carbonyl stretch of the charged monomer with the

in plane NH bend of the bridging proton. This band shifts to 1732 cm^{-1} when all five exchangeable hydrogens have been replaced by deuterium, which agrees with the 30 cm^{-1} shift predicted by DFT computations. Deuteration also decouples the carbonyl stretch of the unprotonated monomer from the ND bends, thus causing the pair of bands at 1600 and 1665 cm^{-1} to shift to a pure carbonyl stretch at 1667 cm^{-1} and a much lower frequency ND bend at 1214 cm^{-1} for the deuterated dimer. The band at 1567 cm^{-1} in the h_5 experimental spectrum is absent in the d_5 spectrum, while the band near 1600 cm^{-1} merely shifts, as discussed above, supporting the assignment of the 1567 cm^{-1} band as the N-H-N asymmetric stretch.

In order to confirm that the N-H-N asymmetric stretch in the proton bound dimer between two molecules of 1-methylcytosine is responsible for the band at 1567 cm^{-1} other PBDs were synthesized and their IRMPD spectra compared to theory and isotopically substituted analogs.

Table 3.24. Comparison of experimental and theoretical band positions (cm^{-1}) for the 1-methylcytosine-d_5 proton-bound dimer.

Exptl band	SBS Harmonic	SBS Scaled Harmonic	SBS Anharmonic	Assignment
	2007	1947	1793	N-D-N asymmetric stretch
1732	1787	1733	1735	C=O stretch & in-plane ND bends
1667	1725	1673	1687	C-N, C=C, & C=O stretches
1626	1695	1644	1655	C=C stretch & H-C=C-H i.p. bends
1620	1672	1622	1634	C=O, C=C, N-D stretches, & H-C=C-H i.p. bends
1581	1630 1584	1581 1536	1580 1546	C-N, C=C stretches & H-C=C-H i.p. bends
1509	1554 1549	1508 1502	1512 1509	C-C-N stretches & CH3 bends
1471	1514 1514	1469 1468	1483 1459	CH_3 bends
1427	1479 1474	1435 1430	1451 1440	CH_3 umbrella motions
1374	1425 1420	1382 1378	1398 1388	CH_3 umbrella motions H-C=C-H asymmetric i.p. bends
1330	1378 1370	1337 1329	1349 1346	C-N-C bends, ND, H-C=C i.p. bend
1269	1314	1275	1275	N-C(O)-N stretches, ND_2 scissoring
1214	1252	1215	1224	ND_2 &ND i.p. bends
1183	1230 1220 1212	1193 1184 1176	1202 1199 1192	N-CH_3 stretch, H-C=C bends, ND_2 scissoring
1127	1175 1162	1140 1128	1154 1143	ND_2 scissoring, ND i.p. bend H-C=C-H i.p. bends & N-methyl rocking
1120	1154 1154 1152	1120 1119 1118	1132 1132 1129	N-CH_3 stretch, ND i.p. bend methyl torsions
1040	1074 1067	1042 1035	1057 1045	CH_3 rocking, H-C=C-H i.p. bends
1010	1042	1010	1024	N-C=O stretch, CH_3 rocking

5-Fluoro-1-methylcytosine (3)

5-Fluoro-1-methylcytosine was synthesized in the same manner as 1-methylcytosine, starting from 5-fluorocytosine. A solution of 5-fluoro-1-methylcytosine in a mixture of water and methanol was electrosprayed, its proton-bound dimer ion isolated in an ICR cell, and the mass-selected ions subjected to the free electron laser. The experimental IRMPD spectrum for the unlabeled (h_5) PBD between two molecules of 5-fluoro-1-methylcytosine (**3**) is plotted along with its scaled harmonic, unscaled harmonic, and unscaled anharmonic DFT predicted frequencies performed at the small and large basis sets (Figure 3.31). As with the 1-methylcytosine PBD, both the small and large basis set unscaled harmonic predictions do not overlap with the experimental spectrum well. Unscaled SBS (1815 cm^{-1}) and LBS (1791 cm^{-1}) predict the carbonyl stretch coupled with the N-H-N in plane bend at positions that are greater than the highest frequency band (1752 cm^{-1}) in the observed spectrum. The unscaled harmonic frequencies for both basis sets also predict a peak slightly above 1400 cm^{-1}, corresponding to the ring sp^3 nitrogen-sp^2 carbon bond coupled with the methyl umbrella motions which appears in a peak-free region of the experimental spectrum. Furthermore, near the region where the N-H-N asymmetric stretch was preliminarily assigned for the 1-methylcytosine PBD (1567 cm^{-1}) the SBS unscaled harmonic prediction fits a peak (1585 cm^{-1}) that disappears upon replacement of the five exchangeable hydrogen with deuterium, suggesting that this peak is the N-H-N asymmetric stretch (discussed below).

The unscaled harmonic calculations also predict the N-H-N asymmetric stretch to occur at 2620 cm^{-1}, which lies outside the domain of the free-electron laser. Although the peak predicted near 1585 cm^{-1} does involve exchangeable hydrogens (in the form of N-H-N and N-H in plane bends), it also involves carbon-carbon stretches, which would also shift when the five hydrogens are replaced by deuterium. The discrepancies between the observed bands and the unscaled harmonic frequencies from the large and small basis set in the other regions of the spectrum make them poor predictors of IR spectrum for these molecules.

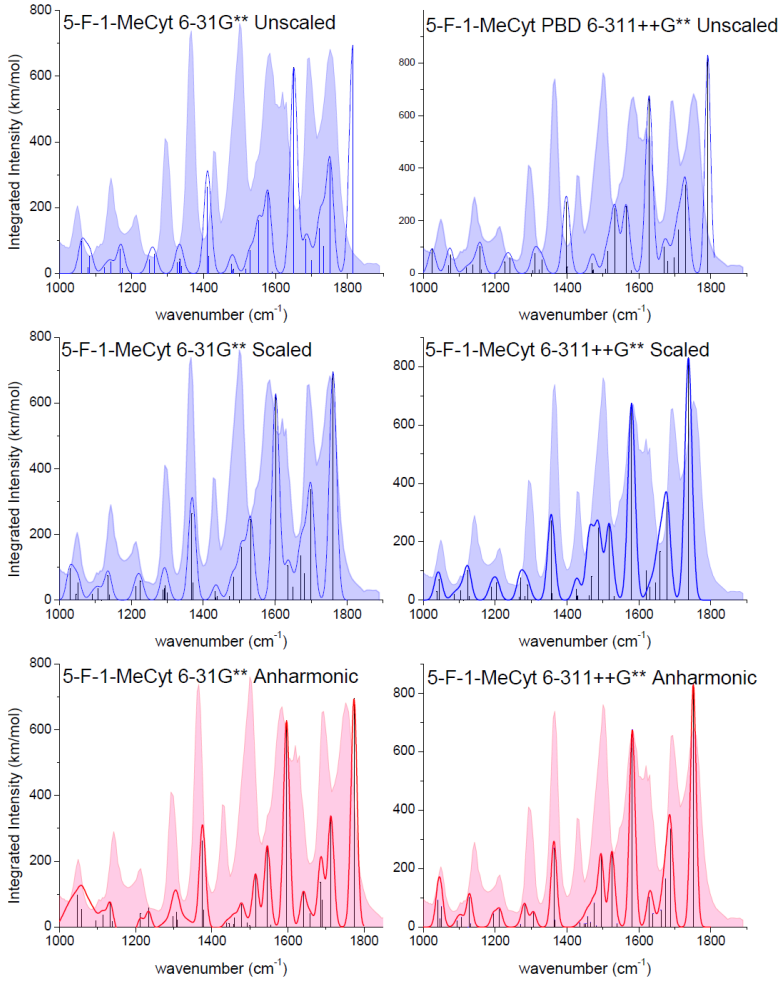


Figure 3.31. Experimental IRMPD spectrum of the proton-bound dimer of 5-fluoro-1-methylcytosine (in silhouette) compared with normal modes of structure 3 calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

Scaling the harmonic calculations by 0.97 gives a much better fit for the observed peaks at both levels of theory. The SBS scaled harmonic predictions do a better job fitting the two largest experimental bands at 1752 cm^{-1} and 1694 cm^{-1} by predicting vibrations at 1761 cm^{-1} (carbonyl stretch coupled with N-H in-plane bends) and 1699 cm^{-1} (carbon-amino nitrogen stretch coupled with N-H in-plane bends). The large basis red-shifts these to 1738 cm^{-1} and 1678 cm^{-1},

differing from the corresponding experimental bands by 14 cm^{-1} and 16 cm^{-1}, respectively. Both basis sets assign 15 of the 16 observed peaks (Table 3.25). The band unassigned by the SBS scaled harmonic calculation, at 1585 cm^{-1}, lies near the band for the N-H-N asymmetric stretch previously assigned for the proton-bound homodimer of h_5-1-methylcytosine. The LBS anharmonic prediction has a theoretical band at 1580 cm^{-1} that directly overlaps the experimental peak near 1585 cm^{-1} (this peak disappears in the d_5 spectrum). This predicted band corresponds to the carbonyl stretch of the unprotonated monomer coupled with the scissoring motions of the amino hydrogens and the in-plane bend of the proton shared between the two ring nitrogens. The smaller basis set predicts this vibration should occur at 1602 cm^{-1} and overlaps with the experimental band centered at 1615 cm^{-1}. Although the larger basis set seems to fit the experiment better than the smaller one above 1600 cm^{-1}, it does not do an adequate job in the region where N-H-N asymmetric stretches have been inferred for the previous PBDs. Moreover, the anharmonic frequencies from the LBS have almost no overlap with the observed band at 1428 cm^{-1}, reinforcing the notion that this level of theory is a poor choice for comparison with these IRMPD spectra.

The SBS anharmonic calculations predict a spectrum similar to the LBS anharmonic frequencies, the notable exception being that the LBS anharmonic calculation better fits the observed peak at 1752 cm^{-1}. The LBS predicts this peak at 1751 cm^{-1}, while the SBS anharmonic prediction places it at 1774 cm^{-1}. Neither

the small or large basis sets provide overlap for the experimental peak at 1428 cm^{-1}. They both predict small amplitude bands near 1428 cm^{-1} (1439 cm^{-1} for the SBS and 1435 cm^{-1} for the LBS), while the small and large basis set scaled harmonic calculation predicts much larger amplitude vibrations at 1439 cm^{-1} (SBS) and 1429 cm^{-1} (LBS) that provide better overlap with the experimental band.

Akin to the comparison between theory and the 1-methylcytosine PBD experimental spectrum, the scaled harmonic SBS calculation provides better overlap with the observed peaks than the other predicted spectra, with the LBS anharmonic prediction being a close second. All of the other predictions do a poor job approximating the observed spectrum, especially in the region where the N-H-N stretch seems to appear. The IRMPD spectrum of the d_5 analogue of **3** was recorded in order to confirm the assignment of the band at 1585 cm^{-1} as the N-H-N asymmetric stretch.

Table 3.25. Comparison of experimental and theoretical band positions (cm^{-1}) for the 5-fluoro-1-methylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	LBS Anharmonic	Assignment
1752	1761	1774	1738	1751	C=O stretch & in-plane NH bend
1694	1699 1682 1671	1713 1689 1686	1678 1659 1647	1688 1674 1661	$C\text{-}NH_2$ stretch and HNH i.p. bends C=C stretch, NHN ip bend
1630	1649	1659	1631	1639	C=C stretch, NHN i.p. bend, HNH i.p. bend
1620	1634	1641	1620	1629	C=C stretch, NHN i.p. bend, HNH i.p. bend
1615	1602	1597	1580	1582	C=O stretch, HNH i.p. bends, NHN i.p. bend
1585					
1550	1544 1531	1555 1546	1531 1517	1539 1525	NHN ip bend, HNH i.p. bend, non C=C ring stretches
1500	1507 1484	1515 1478	1487 1468	1495 1476	HCH, HNH, NHN bends, N-C ring stretches
1428	1439 1439	1439 1446	1429 1429	1435 1447	HCH bends
1366	1372 1368	1379 1375	1359 1356	1366 1363	non C=C ring stretches, CH_3 umbrella
1293	1299 1294 1293 1288	1309 1308 1307 1298	1290 1283 1271 1266	1306 1307 1282 1271	C=CH in-plane bends HNH in-plane bend, N-CO-N ring stretches, C-F stretch
1213	1225 1212	1234 1213	1204 1190	1212 1195	N-CO-N ring stretch, HNH i.p. bend, C-F stretch, C=CH in-plane bends
1143	1139 1134	1139 1132	1127 1123	1131 1128	HNH, NHN, C=CH bends $N\text{-}CH_3$ stretch
1123	1118 1117	1132 1127	1113 1112	1125 1126	HCH torsions
1108	1107 1091	1114 1096	1104 1086	1106 1094	NHN, HNH,C=CH in-plane bends CH_3 rock
1049	1051 1047 1030	1058 1058 1047	1043 1038 994	1050 1046 1041	CH_3 rock, HNH i.p. bend, N-CH ring stretches NHN, HNH o.o.p. bends

5-Fluoro-1-methylcytosine-d_5 (4)

5-Fluoro-1-methylcytosine was dissolved in a mixture of D_2O and methanol-OD in order to replace all the exchangeable hydrogens with deuterium. The IRMPD of the deuterium-bound dimer between two molecules of 5-fluoro-1-methylcytosine-d_2 (m/z 292, **4)** is shown in Figure 3.32 with its unscaled and scaled harmonic DFT frequencies at the SBS**.** The unscaled harmonic vibrations appear to do a better job predicting the actual IR spectrum than in the previous cases. It appears to match many of the large peaks; however, four theoretical bands (1783 cm^{-1}, 1264 cm^{-1}, 1250 cm^{-1}, and 1085 cm^{-1}) paint a different picture. Each of those four peaks appear in a band-free region of the experimental spectrum, slightly red-shifted from the observed bands they purport to match. Scaling the harmonic vibrations for the d_5 predicted spectrum not only match those four peaks (1729 cm^{-1}, 1226 cm^{-1}, and 1212 cm^{-1} and 1053 cm^{-1}), but also shift many of the other unscaled bands to regions that better overlap with the experimental spectrum. All of the experimental bands in the scaled d_5 spectrum overlap well with (and correspond to) a predicted theoretical vibration (Table 3.26). The unassigned band for the h_5 spectrum (1585 cm^{-1}) by the two best-fit theoretical spectra in Figure 3.32 disappears upon the replacement of the five exchangeable hydrogens with deuterium. This disappearance of an unassigned experimental peak when the five exchangeable hydrogens have been replaced with deuterium is consistent with the assignment of that band (1585 cm^{-1}) as the N-H-N asymmetric stretch. This further suggests the region

from 1550-1650 cm^{-1} contains N-H-N asymmetric stretches of the bridging proton in cytosine-like PBDs.

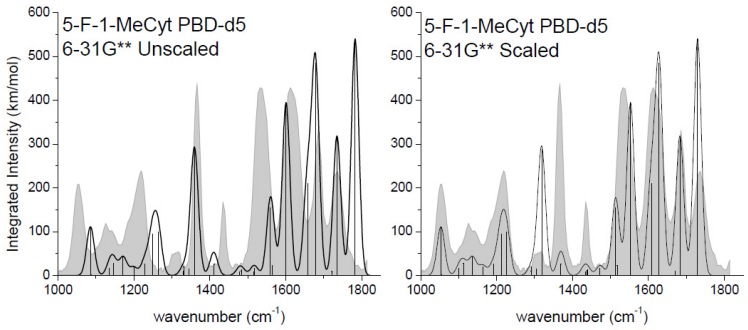


Figure 3.32. Experimental IRMPD spectrum of the d_5 analogue of the proton-bound dimer of 5-fluoro-1-methylcytosine compared with normal modes of structure 4 (unscaled and scaled by 0.97) calculated at B3LYP/6-31G (20 cm^{-1} Gaussian broadening).**

Table 3.26. Comparison of experimental and theoretical band positions (cm^{-1}) for the 5-fluoro-1-methylcytosine-d_5 proton-bound dimer.

Exptl band	SBS Harmonic	SBS Scaled Harmonic	Assignment
1737	1783	1729	C=O stretch, N-D stretches
1686	1736	1683	C-N, C=C, & C=O stretches
	1721	1670	C=C-H in-plane bends
1612	1678	1628	C-N, C=C, & C=O stretches, C=C-H i.p. bends
	1658	1608	& ND stretches
1533	1601	1553	C-ND_2 stretch
	1565	1518	N-C-CF stretches, CH_3 bends
	1560	1513	
1480	1517	1472	CH_3 bends
	1517	1471	
1435	1484	1439	CH_3 bends and umbrella motions
	1483	1439	
	1478	1433	
1367	1412	1369	C-N-C(O) stretches & CH_3 umbrella
	1410	1368	
1318	1360	1319	C-ND_2 stretch, C=C-H bends, ND_2 scissoring
	1346	1305	N-D-N in-plane bend
1219	1264	1226	C-F stretch, N-D-N i.p. bend
	1250	1212	N-C(O) stretch
	1229	1192	C=C-H in-plane bends
1126	1172	1136	ND_2 scissoring, NDN & C=C-H i.p. bends
	1152	1118	CH_3 rocking
	1151	1117	
	1147	1113	
1056	1087	1055	CH_3 rocking
	1085	1053	

1,5-Dimethylcytosine (5)

1,5-Dimethylcytosine was synthesized in the same manner as 1-methylcytosine, starting from 5-methylcytosine. A solution of 1-methylcytosine in a mixture of water and methanol was electrosprayed, the ion corresponding to its PBD (m/z 279, **5**) mass-selected in an ICR cell, and the isolated ions subjected to the free electron laser tuned to explore the IR fingerprint region. Figure 3.33 shows experimental IRMPD spectrum for the unlabeled (h_5) PBD between two molecules of 1,5-dimethylcytosine (**5**) along with its scaled harmonic, unscaled harmonic, and unscaled anharmonic DFT predicted frequencies done at the small and large basis sets. Both unscaled harmonic predictions do a poor job fitting the observed spectrum. Each predicts the carbonyl stretch coupled with the N-H in-plane bends in a region free of experimental bands (1815 cm^{-1} for the SBS and 1789 cm^{-1} for the LBS). The unscaled harmonic predictions also do a poor job overlapping with the three experimental bands between 1400 and 1500 cm^{-1}, or at 1329 cm^{-1}. The LBS unscaled harmonic vibrations do a slightly better job fitting the experimental spectrum than the SBS, but neither is adequate to assign the experimental bands.

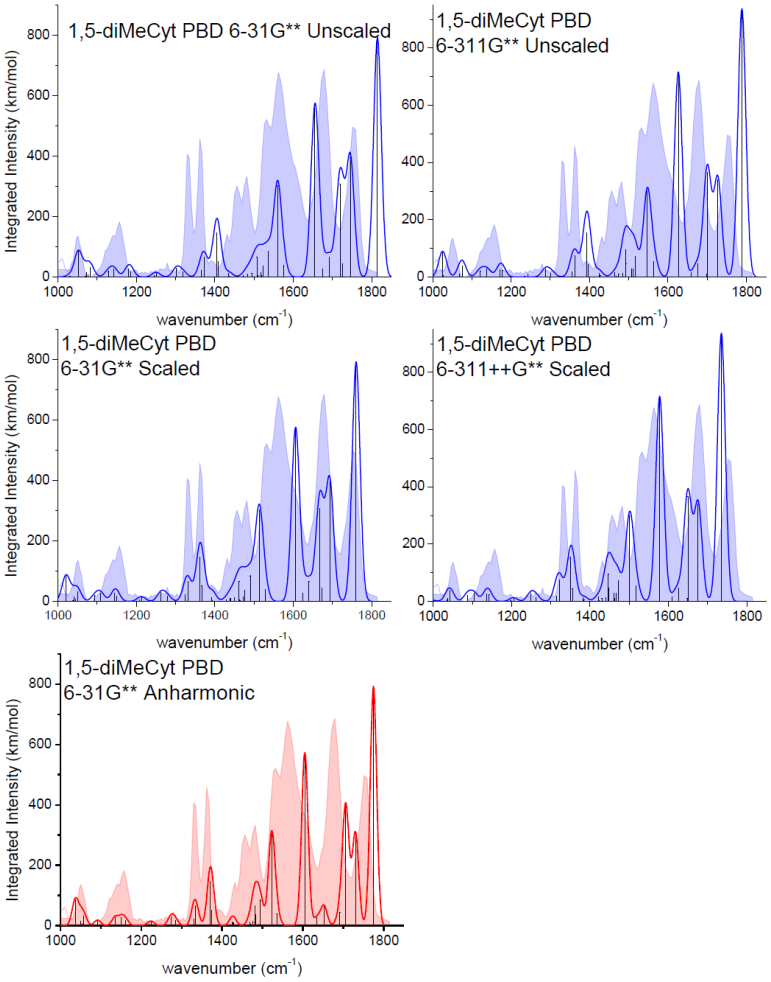


Figure 3.33. Experimental IRMPD spectrum of the proton-bound dimer of 1,5-dimethylcytosine (in silhouette) compared with normal modes of structure 5 calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

Scaling the harmonic frequencies by 0.97 provide a much better fit for the experimental spectrum. The SBS scaled harmonic spectrum almost perfectly shifts the three largest predicted frequencies. The predicted peak at 1760 cm^{-1}, corresponding to the carbonyl stretch coupled with the N-H in-plane bends, overlaps with the experimental band centered at 1752 cm^{-1}. The predicted bands

at 1693 cm^{-1} and 1674 cm^{-1}, corresponding to the protonated ring carbon-carbon and exocyclic carbon-nitrogen stretches coupled with the NH_2 scissoring, overlap with the experimental band centered at 1679 cm^{-1}. It appears that scaling the LBS harmonic vibrations by 0.97 is an overcorrection, as the three DFT vibrations predicted at 1735 cm^{-1}, 1675 cm^{-1}, and 1649 cm^{-1}, are slightly red-shifted from the experimental peaks centered at 1752 cm^{-1} and 1679 cm^{-1}. Other LBS scaled harmonic bands are red-shifted from their corresponding experimental peaks, including the 1501 cm^{-1} vibration, which sits between the experimental bands centered at 1533 cm^{-1} and 1482 cm^{-1}, making those bands difficult to assign. The SBS scaled harmonic vibrations at 1513 cm^{-1} (corresponding to the ring carbon-carbon and ring carbon-nitrogen stretches coupled with the rocking of the amino and methyl hydrogens), and 1491 cm^{-1} (corresponding to the ring carbon-amino nitrogen stretch coupled with the amino and methyl hydrogen bending motions) give better overlap with the experimental bands mentioned above. Both scaled harmonic spectra leave one experimental band unassigned in the 1550-1650 cm^{-1} region. The SBS fits an experimental shoulder peak at 1606 cm^{-1} (1606 cm^{-1}) and leaves the experimental peak at 1563 cm^{-1} unassigned. The LBS fits the experimental band at 1563 cm^{-1} (1578 cm^{-1}) but leaves the shoulder peak at 1606 cm^{-1} unfit. Both scaled harmonic calculations are improvements over the unscaled theoretical spectra, with the SBS scaled harmonic predictions giving the best fit to the experimental spectrum.

Table 3.27. Comparison of experimental and theoretical band positions (cm^{-1}) for the 1,5di-methylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	Assignment
1752	1760	1775	1735	C=O stretch & in-plane NH bend
1679	1693	1706	1675	C-NH_2, C=O, C=C stretches
	1674	1692	1649	NH_2 scissoring & N-H
	1668	1730	1647	in-plane bends
1606	1606	1605		C=C stretch & NH scissoring
1563			1578	C=C stretch & NH scissoring
1533	1528	1535	1517	ring C-C & C-N stretch,
	1513	1523	1501	N-H and C-H rocking
1482	1491	1495	1472	C-NH_2, ring C-N stretches, &
	1477	1483	1466	NH_2 scissoring
	1473	1471	1462	HCH, HNH, NHN bends,
	1472	1476	1461	
1457	1464	1490	1448	C-NH_2 stretch
	1462	1482	1447	HCH, HNH, NHN bends,
	1450	1481	1440	methyl torsions
	1450	1468	1439	
1432	1440	1426	1430	HCH scissoring
	1439	1429	1430	
	1431	1447	1423	CH_3 umbrella motions
	1429	1449	1421	
1390	1394	1425	1385	CH_3 umbrella motions
	1392	1425	1383	
1362	1367	1374	1355	Ring C-N stretches &
	1362	1371	1351	CH_3 umbrella motions
1330	1332	1334	1322	C=C-H bends
	1325	1331	1314	C-N ring stretches
1279	1279	1284	1262	Ring N-CO & N-CN stretches
	1263	1273	1249	
1157	1150	1161	1143	C=C-H bends & N-CH_3 stretch
1142	1144	1151	1136	C=C-H bends & N-CH_3 stretch
1109	1119	1117	1113	N-CH_3 rocking
	1118	1116	1112	
	1109	1135	1104	H-N-H rocking
1051	1051	1057	1043	H-N-H rocking,
	1045	1053	1036	C-CH_3 bends
	1041	1049	1036	C-CH_3 rocking
	1038	1052	1034	
1020	1021	1037	997	C-CH_3 rocking & NH_2 stretches

Unfortunately, the LBS anharmonic calculations were not completed for this molecule. The anharmonic frequencies resulting from the LBS set calculation were often greater than 4000 cm^{-1} suggesting that the calculation was suffering from an error that could not be corrected. The SBS anharmonic spectrum shown in Figure 3.33 does provide a match to the experimental spectrum below 1400 cm^{-1}, but not for the region above 1400 cm^{-1}. The two predicted vibrations above 1700 cm^{-1} (1775 cm^{-1} and 1706 cm^{-1}) are blue shifted by about 20 cm^{-1} from the corresponding experimental bands, unlike the scaled harmonic predictions done at the SBS, which fit those peaks well. There are two experimental peaks between 1400 cm^{-1} and 1650 cm^{-1} with which the SBS anharmonic predictions do not overlap. One at 1457 cm^{-1}, for which the scaled harmonic vibrations do fit, and one at 1563 cm^{-1} that the LBS scaled harmonic prediction does fit, but the SBS does not.

1,5-dimethylcytosine-d_5 (6)

1,5-Dimethylcytosine was dissolved in a mixture of D_2O and methanol-OD in order to replace all the exchangeable hydrogens with deuterium. The IRMPD of the d_5 deuterium-bound dimer between two molecules of 1,5-dimethylcytosine (m/z 284, **6)** is shown in Figure 3.34 with its unscaled and scaled harmonic DFT frequencies at the SBS**.** The unscaled SBS spectrum does a poor job fitting the d_5 spectrum. The unscaled spectrum predicts three vibrations that are not observed in the experimental spectrum. Bands at 1784 cm^{-1}, corresponding to

carbonyl stretch coupled with the N-D-N in-plane bend, the 1551 cm^{-1}, corresponding to the sp^2 ring carbon-sp^3 ring carbon stretch coupled with the sp^3 ring carbon-ring nitrogen stretch, and 1401 cm^{-1}, which corresponds to the sp^3 ring nitrogen-sp^2 ring carbon stretch coupled with the methyl umbrella motions, show up in band-free regions of the experimental spectrum (Table 3.28). Scaling the harmonic predictions by 0.97 gives vibrations at 1730 cm^{-1}, 1504 cm^{-1}, and 1359 cm^{-1} that overlap with experimental bands centered at 1729 cm^{-1}, 1498 cm^{-1}, and 1360 cm^{-1}. The scaled harmonic predictions provide a much better fit than does the unscaled spectrum. Of the two peaks left unassigned in the h_5 spectrum (1562 cm^{-1} by the SBS scaled harmonic and both anharmonic predictions, and 1606 cm^{-1} by the LBS scaled harmonic predictions) only the peak at 1562 cm^{-1} is absent in the d_5 spectrum. This suggests that the peak at 1562 cm^{-1} is the N-H-N asymmetric stretch and that scaled harmonic vibrations from the SBS provide the best choice to assign the spectra.

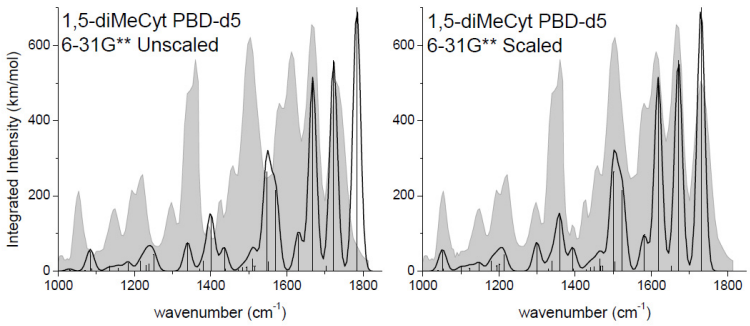


Figure 3.34. Experimental IRMPD spectrum of the d_5 analogue of the proton-bound dimer of 5-fluoro-1-methylcytosine compared with normal modes of structure 6 (unscaled and scaled by 0.97) calculated at B3LYP/6-31G (20 cm^{-1} Gaussian broadening).**

Table 3.28. Comparison of experimental and theoretical band positions (cm^{-1}) for the 1-methylcytosine-d_5 proton-bound dimer.

Exptl band	SBS Harmonic	SBS Scaled Harmonic	Assignment
1729	1784	1730	C=O stretch, N-D-N in-plane bend
1665	1722 1704	1671 1653	C=O, C=C, & C-ND_2 stretches C=C-H in-plane bends
1612	1668	1618	C=O, C=C, N-D stretches & CH bends
1581	1631	1582	C-ND_2 stretch & ND_2 scissoring
1504	1551 1547	1504 1501	N=C-C stretches & CH_3 bends
1459	1511 1510 1495 1494	1466 1464 1450 1449	CH_3 bends
1428	1484 1475 1473	1439 1431 1429	CH_3 bends and umbrella motions
1360	1401 1381	1359 1339	C=C-N stretches & CH_3 umbrella
1297	1340 1329	1299 1289	N-C(O) stretch, ND_2 scissoring, & N-D-N in-plane bend
1222	1250	1213	N-C(O) stretch, ND and CH i.p. bends
1193	1231 1216	1194 1180	C-CH_3 stretch, ND_2 scissoring, CH_3 rocking, & CH bends
1149	1183	1148	ND_2 scissoring, CH & NDN i.p. bend
1053	1087	1054	N-CH_3 rocking

HETERODIMERS

As outlined in the background section of this chapter, the PBDs that form between cytosine base pairs in double stranded regions of DNA do not have a uniform structure. The PBDs at the edge of the crystal structure of $d(ACCCT)_4$ are asymmetric, while the PBDs at the middle of the crystal are almost symmetric. This suggests that the medium plays a role in the structure of these PBDs. One advantage of recording gas phase IR spectra is that the medium can be excluded. By isolating only the ion of interest the effects of the medium or counterion can be avoided. The lack of a corresponding theoretical vibration in the 1550-1650 cm^{-1} region allows for identification of N-H-N asymmetric stretches for these homo PBDs. One would expect only one band for the N-H-N asymmetric stretch, regardless of whether the homo PBDs have an inversion point of symmetry, because both asymmetric tautomers of these homo PBDs (the proton on one side or the other) are identical. Gas phase vibrational spectra on proton-bound heterodimers can provide insight into whether or not the proton prefers to be shared equally between the two monomers. As DFT predictions do not provide an accurate prediction of these N-H-N stretches, it is impossible to compare the N-H-N stretches that are observed in the 1550-1650 cm^{-1} domain with DFT calculations of the symmetric (proton in the middle) and asymmetric (proton situated on one side) PBDs. In order to determine the symmetry of the PBDs in the gas phase the IRMPD of intrinsically asymmetric, hetero, PBDs **(7-9)** were recorded.

Different levels of theory (SBS versus LBS, scaled harmonic versus anharmonic) provide varying degrees of fit to the different proton-bound dimers observed experimentally. One constant feature is that fewer peaks are predicted than the number of bands observed by IRMPD. Along with the fact that none of the DFT spectra for the homodimers predict a band that ought to disappear upon deuteration, that trend implicates the supernumerary band as arising from the asymmetric N-H-N stretch. The following experiments on heterodimers confirm that interpretation.

1-Methylcytosine: 5-Fluoro-1-methylcytosine (7)

A solution of a 50:50 mixture of 1-methyl-5-fluorocytosine and 1-methylcytosine dissolved in a mixture of water and methanol was electrosprayed, the proton-bound hetero dimer ion (m/z 269, **7**) mass-selected in an ICR cell, and the isolated ions subjected to the free electron laser. The IRMPD spectra of the proton-bound dimer between these two molecules (**7**) along with its unscaled harmonic, scaled harmonic, and unscaled anharmonic DFT predicted frequencies done at the small and large basis sets are shown in Figure 3.35. The unscaled harmonic calculations do a poor job fitting the observed spectrum. In particular, they both overestimate the experimental peak at 1767 cm^{-1} by 50 cm^{-1} (SBS) and 25 cm^{-1} (LBS) that corresponds to the carbonyl stretch coupled with the N-H-N in-plane bend. The SBS unscaled harmonic calculation also predicts three peaks (1657 cm^{-1}, 1414 cm^{-1}, and 1262 cm^{-1}) in regions where no

experimental peaks are displayed. These peaks correspond to the NH_2 scissoring motions coupled with carbonyl stretch, the sp^3 ring nitrogen-sp^2 ring carbon stretch coupled with the methyl umbrella motions, and the sp^2 ring nitrogen-sp^2 ring carbon stretch coupled with the carbon fluorine stretch and N-H rocking.

When the SBS harmonic frequencies are scaled by 0.97 the four discrepancies listed above now overlap well with the experimental spectrum. The scaled values for those four peaks are 1763 cm^{-1}, 1607 cm^{-1}, 1371 cm^{-1}, and 1224 cm^{-1} which overlap with the experimental bands centered at 1759 cm^{-1}, 1605 cm^{-1}, 1368 cm^{-1}, and 1219 cm^{-1}, respectively. Along with the predicted peak at 1792 cm^{-1}, the LBS unscaled frequencies have two other peaks that do not overlap with the experimental spectrum: 1) the predicted peak at 1728 cm^{-1}, corresponding to the ring carbon-amino nitrogen stretch coupled with the NH_2 scissoring motions, and 2) the predicted peak at 1241 cm^{-1}, corresponding to the sp^2 ring nitrogen-sp^2 ring carbon stretch coupled with the carbon fluorine stretch and N-H rocking. After scaling these peaks by 0.97 (1676 cm^{-1} and 1203 cm^{-1}) they overlap better with the observed peaks at 1690 cm^{-1} and 1219 cm^{-1}, but neither overlap the observed spectrum as well as the SBS scaled harmonic calculations.

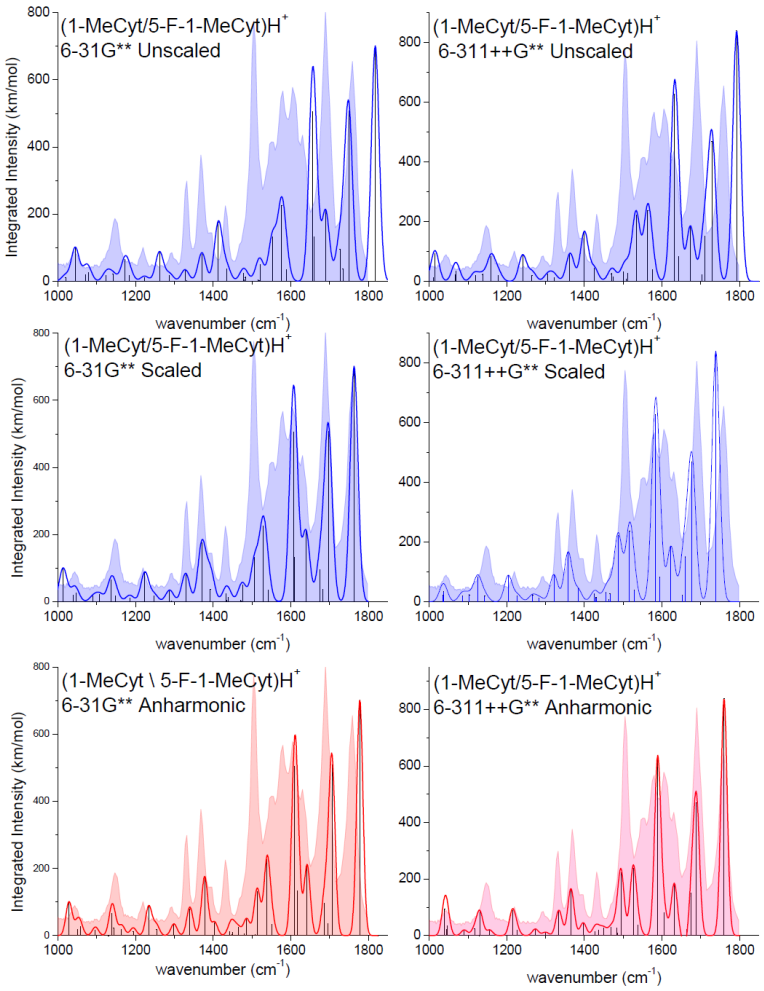


Figure 3.35. Experimental IRMPD spectrum of the proton-bound heterodimer of 1-methylcytosine and 5-fluoro-1-methylcytosine (in silhouette) compared with normal modes of the most stable tautomer of structure 7 calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

Both the large and small basis set anharmonic calculations do a better job than their unscaled harmonic counterparts, specifically with the higher frequency peaks. The LBS anharmonic predictions, at 1760 cm^{-1} and 1689 cm^{-1}, overlap particularly well with the two strong experimental bands at 1759 cm^{-1} and 1690 cm^{-1}.

Table 3.29. Comparison of experimental and theoretical band positions (cm^{-1}) for the most stable tautomer of the 1-methylcytosine/5-fluoro-1-methylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	LBS Anharmonic	Assignment
1759	1763	1778	1739	1760	C=O stretch & in-plane NH bends
1690	1697	1707	1676	1689	C-NH_2 stretch & NH bends
1662	1682 1675	1695 1687	1660 1652	1675 1663	HNH in-plane bends, C=C & C=O stretches
1630	1639	1642	1623	1632	HNH & NH i.p. bends, CO & C=C stretches
1605	1610 1607	1618 1610	1593	1605	HNH & NH in plane bends, C=C, C=O, & C-NH_2 stretches
1580			1583	1589	HNH & NH in-plane bends, CO stretch
1551	1541	1552			NH i.p. bends, NC-CH & NH_2C-CF stretches
1523	1529	1514	1516	1526	NH i.p. bends, NC-CH & NH_2C-CF stretches
1505	1507	1508	1488	1494	NH_2C=N stretch & NH_2 scissoring
1479	1476 1473 1470	1486 1480 1464	1466 1463 1456	1483 1485 1469	NH & CH_3 in-plane bends, C-NH_2 stretches
1431	1439 1439 1435 1433	1450 1448 1443 1404	1430 1429 1426 1425	1427 1428 1437 1450	methyl CH_3 bends & umbrella motions
1392	1393	1378	1383	1397	HC=CH bends, CH_3 umbrella, CN-CH stretch
1368	1371	1340	1358	1365	FC=CH bends, CH_3 umbrella CN-CH stretch
1332	1330	1307	1321	1334	HC=CH bends, CH_3 rocking, CN-CH stretch
1294	1293	1299	1283	1300	FC=CH bends, CH_3 rocking CN-CO stretch
1277	1288	1254	1267	1273	FC=CH bends, NH_2C-N-CO stretches
1219	1248 1224	1234 1194	1227 1203	1226 1214	H_3CN-CO-N-CNH_2 stretches & NH_2 rocking N-CO-N & FC-CNH_2 stretches, NH_2 bends

Both anharmonic predictions provide very good overlap with the experimental spectrum at lower frequencies as well. Both anharmonic calculations leave one experimental band unassigned in the region where N-H-N stretches have previously been identified. Similar to the previously discussed dimers, the SBS calculation does a better job fitting this region, as the experimental bands at 1551 cm^{-1}, 1605 cm^{-1}, and 1630 cm^{-1} all overlap well with the predicted spectrum (1552 cm^{-1}, 1610 cm^{-1}, and 1618 cm^{-1}). The LBS anharmonic calculations are not as easily assigned to experimental peaks in this region. They do predict a vibration at 1632 cm^{-1} and 1605 cm^{-1} that match the observed bands at 1630 cm^{-1} and 1605 cm^{-1} better than does the SBS.

Each of the scaled harmonic and anharmonic predicted vibrations assign every experimental band except for one. The SBS scaled harmonic and anharmonic vibrations fit an experimental band at 1551 cm^{-1} (1541 cm^{-1}/ 1552 cm^{-1}) and leave the experimental peak at 1580 cm^{-1} unassigned. The LBS scaled harmonic and anharmonic predicted vibrations overlap with the experimental band at 1580 cm^{-1} (1583 cm^{-1}/ 1589 cm^{-1}) and leave the peak at 1551 cm^{-1} unassigned. All four of these theoretical spectra leave only a single experimental band unassigned. Although they do not agree on which peak is unassigned, the fact that each theoretical spectrum has a single unassigned peak suggests that only a single N-H-N asymmetric stretch is present.

Combining the spectrum from the SBS scaled harmonic calculations above 1650 cm^{-1} with the SBS anharmonic calculations below 1650 cm^{-1} gives the best fit for the experimental spectrum and leaves one band (1580 cm^{-1}) unfit, for which the N-H-N asymmetric stretch is assigned (Table 3.29). The appearance of only a single N-H-N asymmetric stretch is indicative of either only a single tautomer of this dimer being present under IRMPD conditions or that the proton has enough energy to overcome the energetic barrier to transit and is oscillating between the two tautomers. The two tautomers of the proton-bound heterodimer between 1-methylcytosine and 5-fluoro-1-methylcytosine, one where the proton is on the side of 5-fluoro-1-methylcytosine and the other where the proton is on the side of 1-methylcytosine, have a calculated equilibrium geometry electronic energy difference of 8.8 kJ/mol and a binding enthalpy difference of 18 kJ/mol (Figure 3.36) in favor of the tautomer with the proton on 1-methylcytosine (B3LYP/6-311++G**). Proton transit from the 1-methylcytosine side to the 5-fluoro-1-methylcytosine side is calculated, with zero-point energy corrections, to be over 11 kJ/mol (B3LYP/6-311++G**) endothermic. This suggests that the proton is not equally shared between the two monomers, as it should not have enough energy to overcome this thermodynamic barrier at room temperature. The magnitude of these energy differences suggest that only the more stable tautomer of this PBD is present under IRMPD conditions and, thus, only the N-H-N asymmetric stretch corresponding to the most stable tautomer is observed.

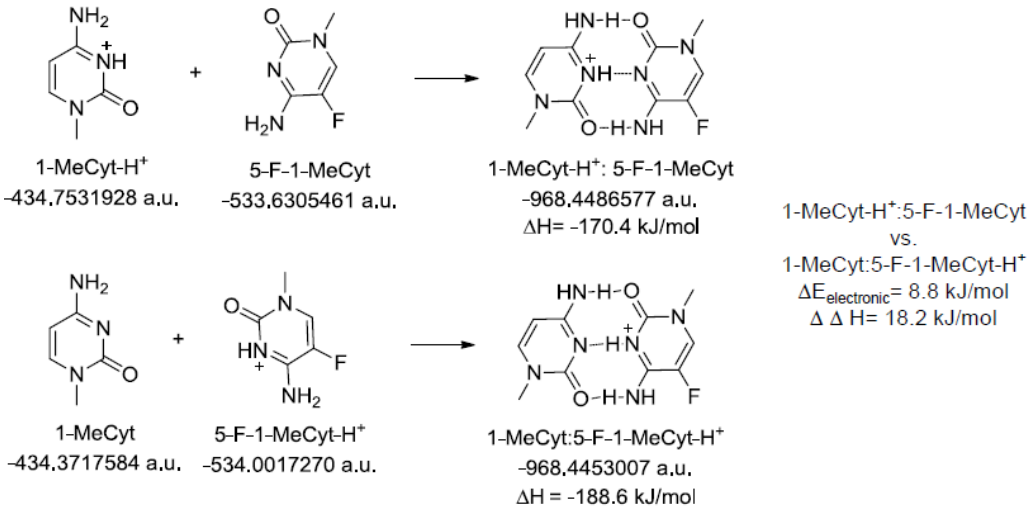


Figure 3.36. Calculated energies (B3LYP/6-311++G) of neutral and protonated monomers of the 1-methylcytosine:5-fluoro-1-methylcytosine proton-bound heterodimer. The binding enthalpies (BE) are the sum of the calculated energy of the monomers subtracted from the calculated energy of the proton-bound dimer.**

5-Fluoro-1-methylcytosine/1,5-Dimethylcytosine (8)

A solution of a 50:50 mixture of 5-fluoro-1-methylcytosine and 1,5-dimethylcytosine dissolved in a mixture of water and methanol was electrosprayed, the proton-bound hetero dimer (m/z 283, **8**) mass-selected in an ICR cell, and the isolated ions subjected to the free electron laser. The IRMPD spectra of the proton-bound dimer formed between these two molecules (**8**), along with its scaled harmonic, unscaled harmonic, and unscaled anharmonic DFT predicted frequencies done at the small and large basis sets, are shown in Figure 3.37. Both unscaled harmonic spectra blue-shift the carbonyl stretch coupled with the N-H in plane bends. The SBS predicts this vibration at 1815 cm^{-1} and the LBS predicts it to be at 1789 cm^{-1}, neither of which match the

experimental spectrum's highest frequency peak, observed at 1759 cm^{-1}. Although the SBS appears to fit the experimental band at 1759 cm^{-1} with a predicted peak at 1745 cm^{-1} (corresponding to the ring carbon-amino nitrogen stretch coupled with the NH_2 scissoring) it is more likely that this motion corresponds to the experimental peak centered at 1686 cm^{-1}, because the carbonyl stretch is always the highest frequency observed band. The SBS unscaled harmonic spectrum also has three vibrations that appear in peak-free regions of the experimental spectrum (1655 cm^{-1}, 1414 cm^{-1}, and 1404 cm^{-1}). The corresponding peaks in the unscaled harmonic LBS spectrum (1631 cm^{-1}, 1400 cm^{-1}, and 1392 cm^{-1}) also appear in peak-free regions of the experimental spectrum. These discrepancies between the predicted and experimental spectra make the unscaled harmonic predictions a poor choice to assign the observed bands.

Scaling the harmonic predictions by 0.97 shifts the four unmatched peaks in the SBS harmonic predicted spectrum to regions that overlap with experimental bands. The scaled harmonic peaks now at 1761 cm^{-1}, 1606 cm^{-1}, 1371 cm^{-1}, and 1362 cm^{-1} fit the experimental bands at 1759 cm^{-1}, 1619 cm^{-1}, and 1367 cm^{-1}, respectively. The scaled SBS harmonic spectrum predicts zero peaks to appear in band-free regions of the experimental spectrum and leaves only the experimental band at 1575 cm^{-1} unassigned (Table 3.30). The LBS scaled harmonic spectrum also fits the observed spectrum better than its unscaled spectrum, although the two high frequency vibrations at 1735 cm^{-1} and

1674 cm^{-1} are red-shifted 24 cm^{-1} and 12 cm^{-1}, respectively, from their corresponding experimental bands. Other peaks, including the ones predicted at 1124 cm^{-1}, 1205 cm^{-1}, 1351 cm^{-1}, and 1358 cm^{-1}, are slightly red-shifted from their corresponding experimental bands (Table 3.30). The LBS scaled harmonic spectrum fits the band at 1575 cm^{-1} that the SBS leaves unassigned, but does not fit the band observed at 1544 cm^{-1} that the SBS does assign.

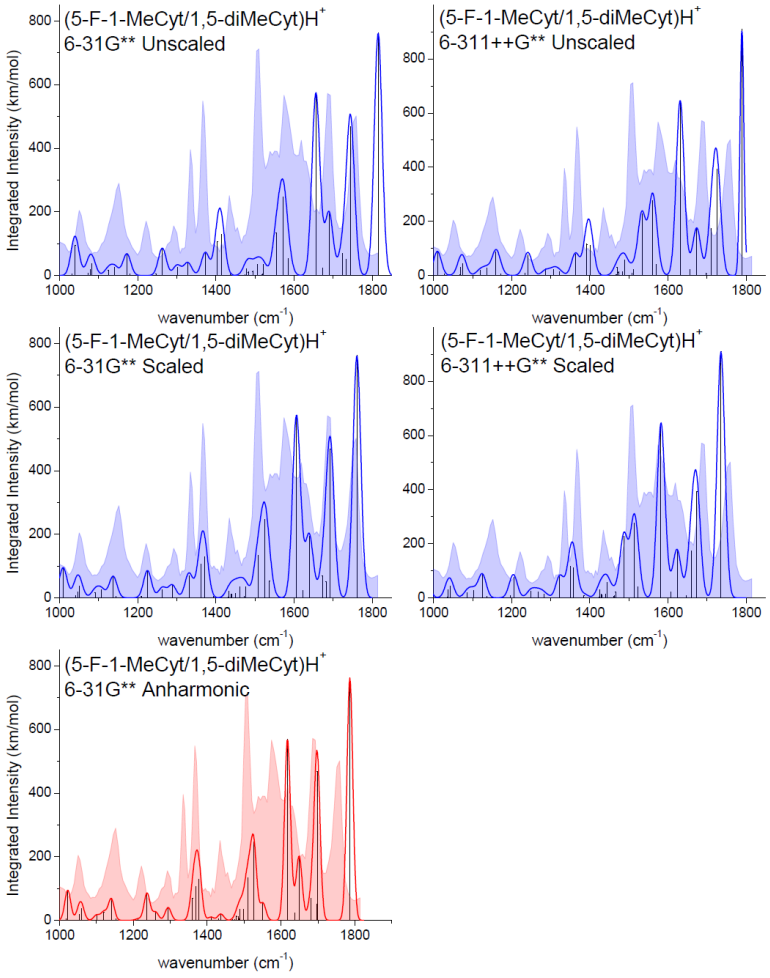


Figure 3.37. Experimental IRMPD spectrum of the proton-bound heterodimer of 5-fluoro-1-methylcytosine and 1,5-dimethylcytosine (in silhouette) compared with normal modes for the most stable tautomer of structure 8 calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

The LBS anharmonic spectrum is not shown due to many predicted bands appearing above 4000 cm^{-1}, suggesting an error Gaussian could not correct. The SBS anharmonic spectrum leaves three experimental bands uncovered at 1759 cm^{-1}, 1575 cm^{-1}, and 1435 cm^{-1}. The predicted peak at 1787 cm^{-1} most likely corresponds to the 1759 cm^{-1} band, and a weak predicted vibration at 1437 cm^{-1} could correspond to the 1435 cm^{-1}, but the observed band at 1575 cm^{-1} has no corresponding predicted vibration. Although the anharmonic spectrum provides an assignment for every experimental band, save for one, it does not overlap with the experimental spectrum as well as do the scaled harmonic predictions. The lack of assignment of the observed band at 1575 cm^{-1} by the SBS scaled harmonic predictions suggests that this is the N-H-N asymmetric stretch for this proton-bound hetero dimer.

Much like the proton-bound hetero dimer between 1-methylcytosine and 5-fluoro-methylcytosine, the 5-fluoro-1-methylcytosine/1,5-dimethylcytosine PBD only has one unassigned band in the experimental spectrum, suggesting that only the most stable tautomer is present under IRMPD conditions. The two tautomers have a calculated (B3LYP/6-311++G**) equilibrium geometry electronic energy difference of about 11.3 kJ/mol and a binding enthalpy difference of about 25.1 kJ/mol (Figure 3.38) in favor of the tautomer with the proton on the side of 1,5-dimethylcytosine. The magnitude of these energy differences further suggest that only one tautomer is present under IRMPD

conditions and, therefore, only one N-H-N asymmetric stretch is observed in the experimental spectrum.

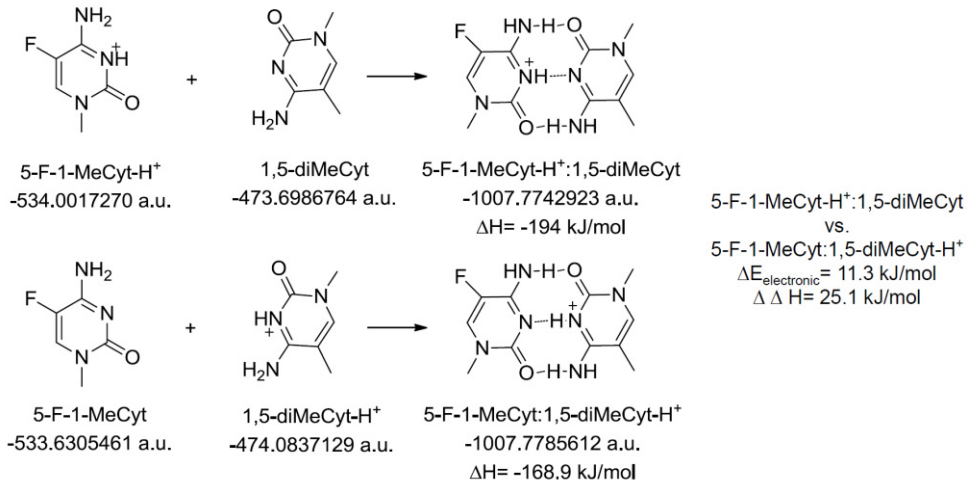


Figure 3.38. Calculated energies (B3LYP/6-311++G) of neutral and protonated monomers of the 5-fluoro-1-methylcytosine:1,5-dimethylcytosine proton-bound heterodimer. The binding enthalpies (ΔH) are the sum of the calculated energy of the monomers subtracted from the calculated energy of the proton-bound dimer.**

Table 3.30. Comparison of experimental and theoretical band positions (cm^{-1}) for the most stable tautomer of the 5-fluoro-1-methylcytosine/1,5-dimethylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	Assignment
1759	1761	1787	1735	C=O stretch & in-plane NH bend
1686	1693 1682 1672	1698 1697 1681	1674 1660	C-NH_2 Stretch and HNH i.p. bends C=C stretch, NHN ip bend
1659	1639	1650	1647	C=C stretch, NHN i.p. bend, HNH i.p. bend
1619	1623 1606	1638 1618	1624 1607	C=C stretch, NHN i.p. bend, HNH i.p. bend C-NH_3 stretch
1575			1582	
1544	1537	1551		NH i.p. bends, non C=C ring stretches
1510	1525 1508	1526 1510	1523 1513	C-C-N ring stretches NH in-plane bends
1456	1461 1450	1474 1469	1466 1463	C-NH_2 stretch, NH in-plane bends & CH_3 bends
1435	1440 1439 1433 1431	1437 1430 1410	1440 1430 1429 1425	N-CH_3 bending & umbrella motions
1367	1371 1362	1378 1360	1358 1351	N-CH stretch & CH_3 umbrella CH in-plane bend
1335	1332		1323	C-H in-plane bend, C-N-C stretches
1292	1293 1288	1294 1290	1283 1267	C-H in-plane bend, C-N-C stretches
1222	1225 1209	1237 1213	1248 1205	N-C=O stretch & NH_2 rocking C-CH_3 stretch
1152	1145 1137	1153 1141	1136 1124	N-CH_3 stretch, C-H bends & NH_2 rocking
1050	1050 1046	1059 1054	1043 1037	NH_2 and CH_3 rocking

1-Methylcytosine/1,5-Dimethylcytosine (9)

A solution of a 50/50 mixture of 1-methylcytosine and 1,5-dimethylcytosine dissolved in a mixture of water and methanol was electrosprayed, the proton-bound hetero dimer (m/z 265, **9**) mass-selected in an ICR cell, and the isolated ions subjected to the free electron laser. The IRMPD spectra of the proton-bound heterodimer, **9**, along with its scaled harmonic, unscaled harmonic, and unscaled anharmonic DFT predicted frequencies calculated at the small and large basis sets are shown in Figure 3.39. Unscaled harmonic frequencies from both basis sets produce theoretical vibrations in peak-free regions of experimental spectrum. They both overestimate the highest frequency band observed at 1752 cm^{-1} by predicting the carbonyl stretch coupled with the N-H in-plane bends at 1816 cm^{-1}(SBS) and 1790 cm^{-1} (LBS). The SBS predicts a peak at 1713 cm^{-1} that corresponds to an empty region of the experimental spectrum, and the LBS does the same at 1633 cm^{-1}. These unscaled DFT harmonic calculations do a better job overlapping the experimental spectrum than does the unscaled theoretical spectra for the previously discussed PBDs. Scaling the harmonic frequencies by 0.97 results in better overlap with the experimental spectrum in both the large and small basis set cases.

The scaled harmonic SBS prediction provides excellent overlap with the largest frequency, predicting that the carbonyl stretch coupled with the N-H in plane bends at 1761 cm^{-1} match the experimental peak at 1752 cm^{-1}. Both the

1693 cm^{-1} DFT scaled harmonic frequency, corresponding to the ring carbon-amino nitrogen stretch coupled with the NH_2 scissoring motions, and the 1662 cm^{-1} predicted frequency, which corresponds to ring carbon-carbon double bond stretch coupled with the NH_2 scissoring motions, fit underneath the observed band centered at 1679 cm^{-1}. Perhaps unsurprisingly at this point, the scaled harmonic frequencies from the LBS do not overlap well with the experimental spectrum as do those from the SBS. The theoretical peaks at 1736 cm^{-1}, 1643 cm^{-1}, and 1512 cm^{-1} are all slightly red shifted to lower frequencies than the corresponding SBS calculations and do not overlap with the experimental spectrum. The corresponding SBS scaled harmonic predictions at 1761 cm^{-1}, 1662 cm^{-1}, and 1525 cm^{-1} are all slightly blue shifted from their LBS equivalents and overlap well with the observed spectrum.

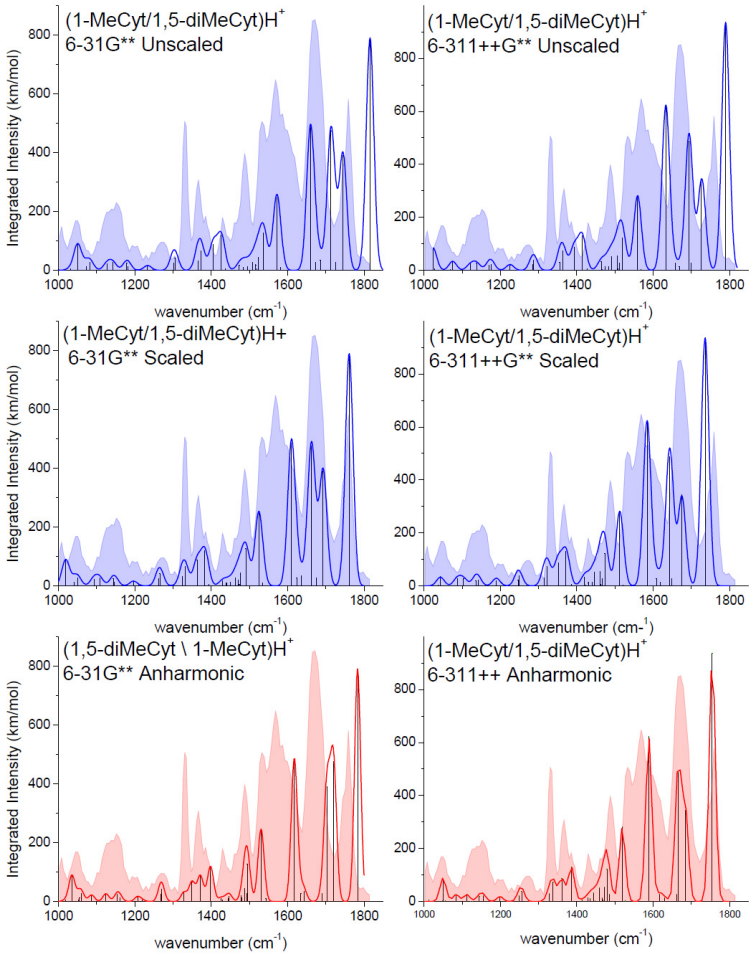


Figure 3.39. Experimental IRMPD spectrum of the proton-bound heterodimer of 1-methylcytosine and 1,5-dimethylcytosine (in silhouette) compared with normal modes for the most stable tautomer of structure 9 calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

The SBS anharmonic spectrum does a poor job fitting the experimental peaks above 1650 cm^{-1}. The three predicted large amplitude motions (1783 cm^{-1}, 1720 cm^{-1}, and 1703 cm^{-1}) are all blue shifted by least 20 cm^{-1} from the experimental bands. The LBS anharmonic predictions (1755 cm^{-1}, 1684 cm^{-1},

and 1666 cm^{-1}) do a much better job fitting the corresponding observed bands, differing by about 5 cm^{-1} from the experimental peak centers. Both anharmonic calculations do an adequate job fitting all the peaks below 1550 cm^{-1}, but differ in their assignment of the peaks between 1550 cm^{-1} and 1650 cm^{-1}. The LBS overlaps with two of the four bands in this region at 1587 cm^{-1} and 1625 cm^{-1}, while leaving the observed bands at 1569 cm^{-1} and 1600 cm^{-1} unassigned (Table 3.31). The SBS anharmonic vibrations overlap with only one of the experimental bands between 1550-1600 cm^{-1}, leaving three observed peaks unfit. Shifting the SBS anharmonic peaks slightly in order to assign them to experimental peaks allows for all but two of the experimental peaks to be assigned (Table 3.31). The SBS scaled harmonic vibrations leave the experimental bands centered at 1569 cm^{-1} and 1587 cm^{-1} unassigned, while the LBS leaves the peaks at 1569 cm^{-1} and 1625 cm^{-1} unassigned. For this PBD (**9**) the SBS scaled harmonic vibrations does the best job of fitting the experimental spectrum. Each of the theoretical spectra leave two unassigned peaks in the 1550-1650 cm^{-1} region where previous N-H-N asymmetric stretches have been identified. This suggests both tautomers of this PBD are present under IRMPD conditions.

Table 3.31. Comparison of experimental and theoretical band positions (cm^{-1}) for the most stable 1,5-dimethylcytosine/1-methylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	LBS Anharmonic	Assignment
1752	1761	1783	1736	1755	C=O stretch coupled with N-H in plane bends
1679	1693 1675 1662	1703 1691 1720	1674 1649 1643	1684 1661 1666	NH_2 in plane bend with C-NH_2 , C=C, & C=O stretches
1638	1635	1643	1618	1629	C=C stretch with NHN, NH_2,
1625	1624	1635		1616	HC=CH, & HC-CCH_3 in plane bends
1600	1610	1617	1608		C=O & C=C stretches, N-H in plane bends
1587			1584	1589	C=O & C=C stretches, N-H in plane bends
1569					
1539	1534 1525	1543 1530	1520 1512	1526 1519	C-C stretches & NH_2 rocking NHN in-plane bend
1488	1491 1477	1494 1487	1473 1466	1479 1485	HCH bends, i.p. NH bends, N=C-NH_2 stretches
1471	1473 1471	1487 1479	1462 1461	1469 1472	Methyl CH_3 bends, N-H ip bends, HC=CHbends
1430	1440 1439	1444 1426	1430 1430	1429 1434	N-Methyl H-C-H bends
1386	1383	1398	1371	1386	N-methyl CH_3 umbrella HC=CH bends
1367	1363	1371	1352	1361	CH_3 umbrella, HC=CH bends, OCN-CH stretch
1335	1332 1324	1350 1329	1322 1314	1338 1327	H_3CC=CH & HC=CH bends (O)C-N-C ring stretches
1252	1266 1263	1270 1268	1248 1246	1256 1250	ring N-CO stretches, methyl rocking ring CH bends,
1200	1210	1219	1198	1202	ring C-H i.p. bends, N-C(O) & C-CH_3 stretches
1159	1196 1147	1207 1161	1189 1141	1199 1155	HC=CH in-plane bends, N-CH3 stretch C-CH_3 stretch, H_3CC=CH bends
1145	1144	1154	1136	1144	HC=CH in-plane bends & methyl rocking
1133	1118 1118	1134 1118	1113 1113	1127 1122	CH_3 methyl torisons
1108	1108	1124	1103	1112	CH_3 and NH2 i.p. rock
1090	1094	1087	1088	1084	CH_3 and NH2 i.p. rock
1056	1050 1041	1059 1053	1043 1037	1050 1049	N-CH_3 i.p. bends CH ring bends, & NH_2 rocking

Figure 3.40 shows the IRMPD spectrum of structure **9** (m/z 265) compared with the scaled harmonic, unscaled harmonic and unscaled anharmonic DFT predicted frequencies calculated at the small and large basis sets for the less stable tautomer of **9**, where the proton resides on the 1-methylcytosine side of the dimer. Both unscaled spectra have peaks clearly blue-shifted from the corresponding experimental bands, just as does the more stable tautomer of **9** discussed above. Both cases show peaks in empty regions of the experimental spectrum, particularly at frequencies above 1600 cm^{-1}.

Scaling the harmonic frequencies by 0.97 sufficiently red-shifts many of the theoretical vibrations so they overlap with the experimental spectrum. The LBS scaled spectrum appears to be over red-shifted, as the predicted vibrations at 1738 cm^{-1} and 1525 cm^{-1} are both 14 cm^{-1} lower than their corresponding experimental peaks. The SBS scaled harmonic vibrations at 1763 cm^{-1} and 1537 cm^{-1} differ by less than 11 cm^{-1} from their corresponding experimental bands. Unlike the theoretical spectrum for the more stable tautomer, the LBS scaled harmonic vibrations leaves only one experimental peak, centered at 1569 cm^{-1}, unassigned. The SBS scaled harmonic peaks, like the more stable tautomer's theoretical spectrum, leaves the experimental bands at 1569 cm^{-1} and 1587 cm^{-1} unassigned.

The SBS anharmonic predicted spectrum suffers from the same deficiencies as does the more stable tautomer predictions. Specifically, it blue

shifts the predicted peaks from its corresponding experimental bands above 1650 cm^{-1}. The LBS anharmonic prediction does a very good job overlapping with the experimental spectrum, just as it did for the more stable tautomer. The SBS scaled harmonic vibrations leave the experimental bands centered at 1569 cm^{-1} and 1587 cm^{-1} unassigned, while the LBS anharmonic spectrum leaves the peaks at 1569 cm^{-1} and 1625 cm^{-1} unassigned (Table 3.32).

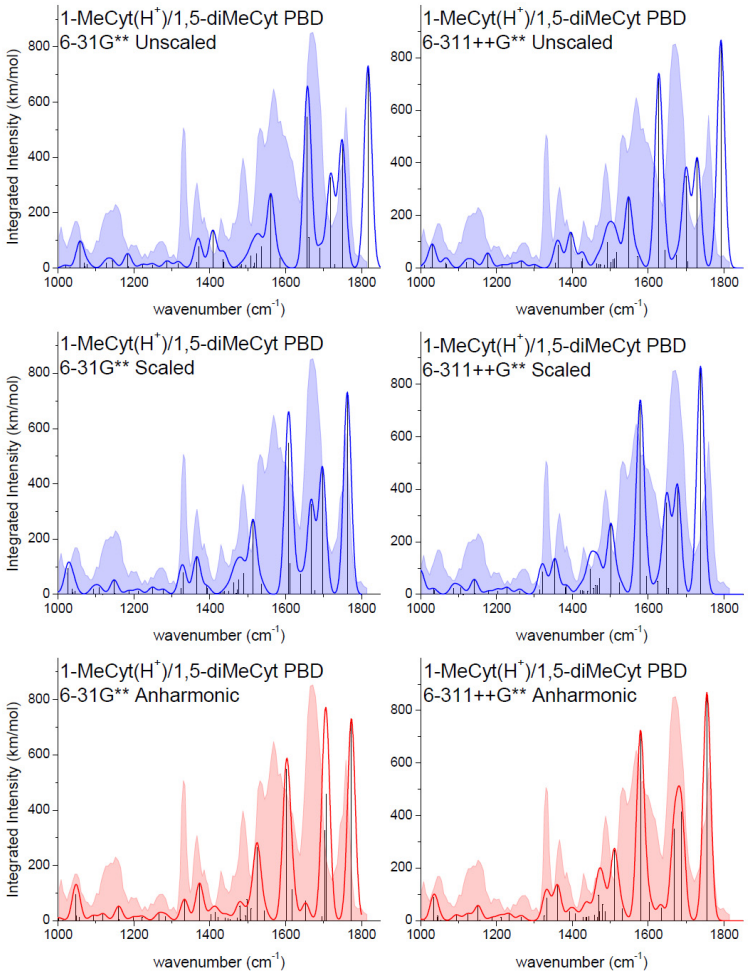


Figure 3.40. Experimental IRMPD spectrum of the proton-bound heterodimer of 5-fluoro-1-methylcytosine and 1,5-dimethylcytosine (in silhouette) compared with normal modes for less stable tautomer of structure 9 calculated at B3LYP/6-31G and 6-311++G** both unscaled and scaled by 0.97 (in blue) and anharmonic modes (in red) with 20 cm^{-1} Gaussian broadening.**

Table 3.32. Comparison of experimental and theoretical band positions (cm^{-1}) for the less stable 1,5-dimethylcytosine/1-methylcytosine proton-bound dimer.

Exptl band	SBS Scaled Harmonic	SBS Anharmonic	LBS Scaled Harmonic	LBS Anharmonic	Assignment
1752	1763	1773	1738	1755	C=O stretch coupled with N-H i.p. bends
1679	1697 1677 1667	1708 1704 1695	1677 1653	1688 1669 1668	NH_2 in plane bend with C-NH_2 , C=C, & C=O stretches
1638	1640	1653	1649	1636	C=C stretch with NHN, NH_2,
1625	1611	1618	1625		HC=CH, & HC-CCH_3 in plane bends
1600	1607	1602	1595	1603	C=O & C=C stretches, N-H in plane bends
1587			1579	1580	C=O & C=C stretches, N-H in plane bends
1569					
1539	1537 1514	1544 1525	1525	1532 1511	C-C stretches & NH_2 rocking NHN in-plane bend
1488	1490 1477	1509 1498	1502	1487 1480	HCH bends, i.p. NH bends, N=C-NH_2 stretches
1471	1472 1470	1479 1472	1471 1466	1472 1469	methyl CH_3 bends, N-H ip bends, HC=CH bends
1430	1435 1429	1448 1441	1430 1430	1435 1429	N-methyl H-C-H bends
1386	1394 1392	1415 1402	1384 1383	1393	N-methyl CH_3 umbrella HC=CH bends, HC=CH bends, OCN-CH stretch
1367	1366	1373	1354	1360	H_3CC=CH & HC=CH bends
1335	1330 1325	1334 1333	1321 1314	1333 1326	(O)C-N-C ring stretches ring N-CO stretches, methyl rocking
1252	1278 1249	1266 1222	1261 1228	1266 1233	ring CH bends, C-H i.p. bends, N-C(O) & C-CH_3 stretches
1200	1212 1187	1199 1161	1203 1179	1212 1189	HC=CH in-plane bends, N-CH_3 stretch C-CH_3 stretch, H_3CC=CH bends
1159	1150	1160		1157	HC=CH in-plane bends & methyl rocking
1145	1148		1144	1152	CH_3 methyl torsions
1133	1119 1118	1124 1119	1142 1114	1130 1124	CH_3 and NH_2 i.p. rock
1108	1109	1113	1112	1120	CH_3 and NH_2 i.p. rock
1090	1094	1093	1087	1094	N-CH_3 i.p. bends
1056	1045 1038	1056 1048	1036 1034	1053 1045	CH ring bends, & NH_2 rocking

The two tautomers produce very similar theoretical spectra, with only twelve frequencies differing by more than 10 cm^{-1} among the four harmonic spectra. While the anharmonic spectra from each tautomer do differ more drastically from one another, their assignments are very similar (Tables 3.31 and 3.32). Both tautomers leave two unassigned experimental bands in the putative N-H-N asymmetric stretch region, suggesting that neither prediction fits the N-H-N asymmetric stretch from these PBDs and that both tautomers are present under IRMPD conditions.

This notion is further supported by DFT large basis set electronic energy calculations. The electronic energy for the tautomer in which the bridging proton resides on 1-methylcytosine is calculated to be 2.4 kJ/mol less favorable than the other tautomer. The 1-methylcytosinium tautomer has a calculated heat of formation at zero Kelvin of 1.24 kJ/mol higher than that of the other tautomer and a binding enthalpy difference of 7 kJ/mol in favor of the 1,5-dimethylcytosinium tautomer (Figure 3.41). This corresponds to a ratio of 5:3 of the two tautomers present at room temperature. This ratio implies that both tautomers should be present under the IRMPD conditions. The calculated barrier to proton transit, including zero-point corrections, is 8.5 kJ/mol, suggesting that both tautomers are present as the PBD is unable to overcome this barrier on the IR timescale. Moreover, the energy and binding enthalpy difference between the two tautomers are lower than the barrier to transit. The energy differences between the two

tautomers of **9** are much lower than the differences between the tautomers of **7** and **8**, but the barrier to proton transit is comparable.

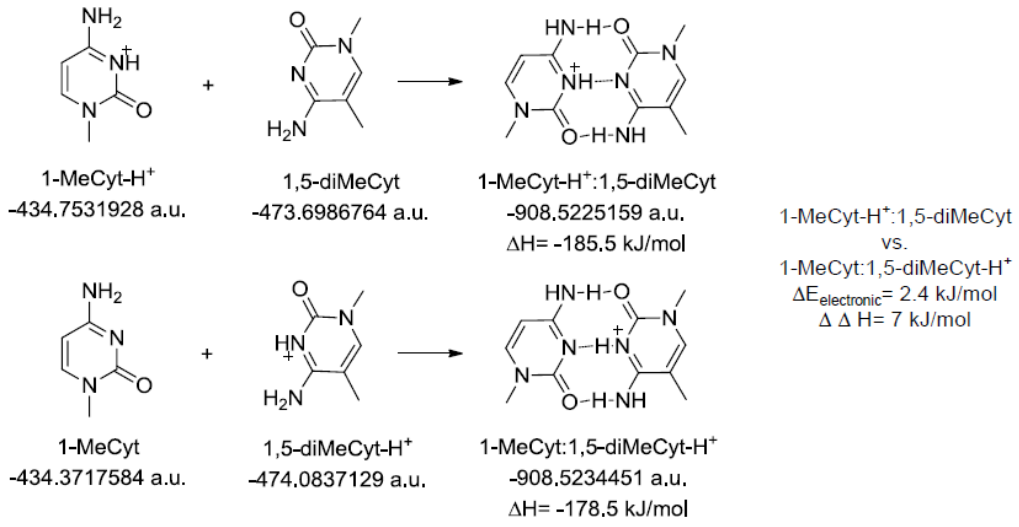


Figure 3.41. Calculated energies (B3LYP/6-311++G) of neutral and protonated monomers of the 1-methylcytosine:1,5-dimethylcytosine proton-bound heterodimer. The binding enthalpies (ΔH) are the sum of the calculated energy of the monomers subtracted from the calculated energy of the proton-bound dimer.**

An exact assignment of all the theoretical peaks is not necessary to conclude that two N-H-N asymmetric stretches are present under IRMPD conditions. Although each basis set does not agree on which band in the 1550-1650 cm^{-1} region corresponds to an N-H-N asymmetric stretch, they all leave two bands unassigned. The small(er) difference in energy between the two tautomers compared to **7** and **8** makes it more likely that both tautomers are present in **9.** The two unassigned bands from each comparison with theory suggest that this is the case.

Equilibrium Constants Determination by Mass Spectrometry

Heterodimers **7**-**9** were formed in the gas phase by mixing solutions containing the individual 5-substituted-1-methylcytosine monomers and injecting the combined solution into an electrospray ionization source connected to a mass spectrometer. Five ions are observed in the mass spectrum for each mixture of neutral cytosine analogues. Two correspond to the protonated monomers, two correspond to the proton-bound homodimers, and one to the proton-bound heterodimer. If there were no preference in the formation of the dimers the ratio between the abundances seen in the MS would give a statistical distribution. That is, (% of monomer)^2 for each homodimer to 2x(% of monomer 1 x % of monomer 2) for the heterodimer. For example, if there was no preference in dimer formation for the 70:30 mixture of 1-methylcytosine and 1,5-dimethylcytosine one would expect to see $(0.7)^2 = 0.49$ for the 1-methylcytosine homodimer to $(0.3)^2 = 0.9$ for the 1,5-dimethylcytosine homodimer to 2x(0.7*0.3) = 0.42 for the heterodimer. Dividing the expected values by a common denominator results in a reduced ratio of 5.4 : 4.7 : 1 of 1-methylcytosine homodimer to 1-methylcytosine/1,5-dimethylcytosine heterodimer to 1,5-dimethylcytosine homodimer. Computations show that the binding enthalpies for each PBD are not equal (Figures 3.36, 3.38, and 3.41) and therefore, the abundance ratios should not follow a statistical distribution.

The peak intensity ratios of the PBDs can be used to determine the equilibrium constant (K_{eq}) for the conversion of one dimer to another. Using the

generic scheme below (Figure 3.42) with **A** denoting the neutral cytosine analogue with the smaller m/z and **B** denoting the neutral cytosine analogue with the larger m/z, an equation for K_{eq} can be obtained.

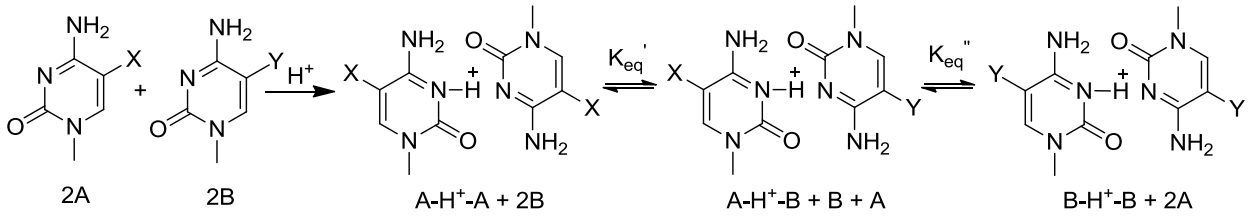


Figure 3.42. Generic scheme for the ionization of two equivalents of different 5-substituted-1-methylcytosines to form proton-bound homodimers (A-H-A and B-H-B) and heterodimer (A-H-B).

K_{eq} can be expressed as a ratio of the concentration of the products of a reaction divided by the concentration of its reactants. The equilibrium constant for the transition from a A-H^+-A homodimer to a A-H^+-B heterodimer is shown in Equation 2 below. Equation 2 can be re-arranged to give equation 3, which is in the form of a linear equation (y= mx + b) where K_{eq}' is the slope of the line (m).

$$K_{eq}{}' = \frac{[A{-}H^+{-}A][A][B]}{[A{-}H^+{-}A][B]^2} \qquad \textbf{(2)}$$

$$\frac{[A{-}H^+{-}B]}{[A{-}H^+{-}B]} = K_{eq}{}' \frac{[B]}{[A]} \qquad \textbf{(3)}$$

The mild nature of electrospray ionization suggests that the concentration of **A** and **B** should not be drastically affected by the addition of H^+ to the system

in the course of the ionization process. That is, the ionization of the mixture of the two cytosine analogues will ionize a relatively small percentage of the molecules, leaving the concentration of **A** and **B** basically unchanged from their initial ratios, which can be controlled by preparing samples with specific ratios of each neutral monomer. The slope of the trendline (as determined by Microsoft Excel) corresponds to the equilibrium constant for the conversion from one dimer to another when the ratio of the proton-bound dimer peak intensities observed in the mass spectrum is plotted against the original concentration ratio of **A** and **B**.

0.05 M stock solutions of three different cytosine analogues (1-methylcytosine, 1,5-dimethylcytosine, and 5-fluoro-1-methylcytosine) were prepared in 50:50 methanol:water. From these stock solutions, samples were prepared for mass spectrometry experiments to examine the ratio at which proton-bound heterodimers are present under electrospray conditions. Five mixtures of varying percentage of monomers (70/30, 60/40, 50/50, 40/60, and 30/70) for each possible combination of cytosine analogs were prepared. Each sample has an overall concentration of 5 mM of combined cytosine analogues. Four scans were recorded for each of the five mixtures for each of the combinations of neutral monomers. Five ions are observed in each mass spectrum corresponding to the two protonated monomers, the two homodimers, and the heterodimer.

1-Methylcytosine:5-Fluoro-1-methylcytosine

Table 3.33 tabulates the average ratio of the proton-bound dimer intensities observed in the mass spectrum relative to the 5-fluoro-1-methylcytosine homodimer for each of the five mixtures of monomers. As expected, as the concentration of one neutral monomer is increased the relative intensity of the proton-bound dimer peak containing that monomer also increases. In the case of the 50/50 mixture of 1-methylcytosine/ 5-fluoro-1-methylcytosine, the 1-methylcytosine proton-bound homodimer is over 6 times as intense as the 5-fluoro-1-methylcytosine proton-bound homodimer. At equal concentrations of neutral monomer, the heterodimer peak is slightly more intense than the 1-methylcytosine homodimer peak and 7 times as intense as the 5-fluoro-1-methylcytosine homodimer peak.

Table 3.33. Average peak intensity ratios of the proton-bound dimers formed between mixtures of 1-methylcyotinse and 5-fluoro-1-methylcytosine relative to the 5-fluoro-1-methylcytosine homodimer.

Mixture (A/B)	1-MeCyt PBD (A-H^+-A) Ratio	1-MeCyt:5-F-1-MeCyt PBD (A-H^+-B) Ratio	5-F-1-MeCyt PBD (B-H^+-B) Ratio
70/30	24.4	15.6	1.0
60/40	12.0	9.8	1.0
50/50	6.1	7.0	1.0
40/60	2.6	4.3	1.0
30/70	1.0	2.5	1.0

As discussed in the beginning of this chapter, the electron withdrawing nature of the fluorine at the 5-position of cytosine make the imino nitrogen of 5-fluoro-1-methylcytosine less basic than the other neutral monomers. The proton-

bound dimers containing 1-methylcytosine should be more stable, as the basicity of the imino nitrogen makes them better equipped to handle the proton. Figure 3.43 shows the plot of proton-bound dimer peak intensity ratios versus the concentration ratio of the neutral monomers injected into the mass spectrometer. The equilibrium constant between the 1-methylcytosine proton-bound homodimer and the 1-methylcytosine:5-fluoro-1-methylcytosine heterodimer (K_{eq}') is the slope of the blue line and is calculated to be 0.94. As this value is less than 1, the 1-methylcytosine homodimer is slightly preferred over the heterodimer under electrospray conditions. The equilibrium constant between the heterodimer and and the 5-fluoro-1-methylcytosine homodimer (K_{eq}'') is the slope of the red line and is calculated to be much less than 1 (0.18). This means that the heterodimer is vastly preferred over the 5-fluoro-1-methylcytosime homodimer. This supports the notion that the basicity of the cytosine ring nitrogen plays a major role in the stability of the PBDs. That is, as calculated by DFT and confirmed by AIM analysis, the proton is more stable on 1-methylcytosine than on 5-fluoro-1-methylcytosine and, therefore, the PBD between two 1-meythylcytosine molecules is preferred over the proton-bound heterodimer with one 1-methylcytosine, which is preferred over the PBD with zero 1-methylcytosine molecules.

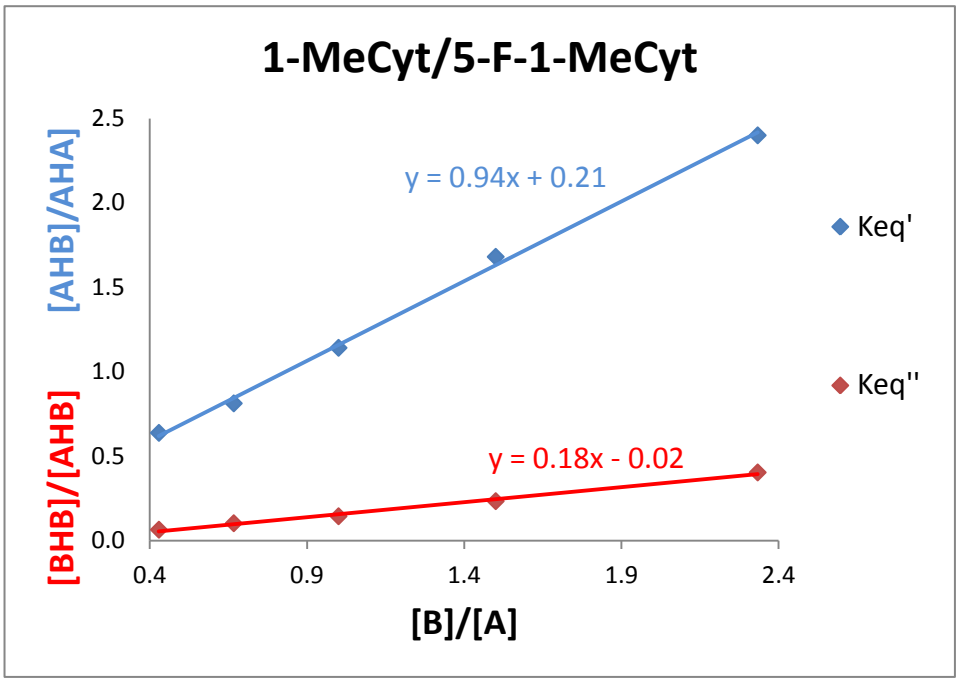


Figure 3.43. Plot of proton-bound dimer average peak intensity ratios versus the concentration ratio of monomers A (1-methylcytosine) and B (5-fluoro-1-methylcytosine) in the solution injected into the mass spectrometer.

1,5-Dimethylcytosine:5-Fluoro-1-methylcytosine

Table 3.34 tabulates the average ratio of the proton-bound dimer intensities observed in the mass spectrum relative to the 5-fluoro-1-methylcytosine homodimer for each of the five mixtures of monomers. In the case of the 50/50 mixture of 1,5-dimethylcytosine/ 5-fluoro-1-methylcytosine, the 1,5-dimethylcytosine proton-bound homodimer is over 10 times as intense as the 5-fluoro-1-methylcytosine proton-bound homodimer. At equal concentrations of neutral monomer the 1,5-dimethylcytosine homodimer peak is slightly more intense than the heterodimer peak, which is 9 times as intense as the 5-fluoro-1-methylcytosine homodimer peak. Even at the highest concentration of 5-fluoro-1-

methylcytosine the 1,5-dimethylcytosine homodimer peak is greater than that of the 5-fluoro-1-methylcytosine homodimer peak.

Table 3.34. Average peak intensity ratios of the proton-bound dimers formed between mixtures of 1,5-dimethylcyotinse and 5-fluoro-1-methylcytosine relative to the 5-fluoro-1-methylcytosine homodimer.

Mixture (A/B)	1,5-diMeCyt PBD (A-H^+-A) Ratio	1,5-diMeCyt:5-F-1-MeCyt PBD (A-H^+-B) Ratio	5-F-1-MeCyt PBD (B-H^+-B) Ratio
70/30	24.4	13.8	1.0
60/40	13.0	10.0	1.0
50/50	10.2	9.0	1.0
40/60	5.6	6.3	1.0
30/70	1.6	3.0	1.0

Figure 3.44 shows the plot of proton-bound dimer peak intensity ratios versus the concentration ratio of the neutral monomers injected into the mass spectrometer. The equilibrium constant between the 1,5-dimethylcytosine proton-bound homodimer and the 1,5-dimethylcytosine:5-fluoro-1-methylcytosine heterodimer (K_{eq}') is the slope of the blue line and is calculated to be 0.66. As this value is less than 1, the 1,5-dimethylcytosine homodimer is preferred over the heterodimer under electrospray conditions. The equilibrium constant between the heterodimer and the 5-fluoro-1-methylcytosine homodimer (K_{eq}'') is the slope of the red line and is calculated to be much less than 1 (0.13). This means that the heterodimer is vastly preferred over the 5-fluoro-1-methylcytosine homodimer. This corroborates the notion that the basicity of the cytosine ring nitrogen plays a major role in the stability of the PBDs. That is, as calculated by

DFT and confirmed by AIM analysis, the proton is more stable on 1,5-dimethylcytosine containing dimers than on 1-methylcytosine-containg dimers and much more stable than when its situated on 5-fluoro-1-methylcytosine containing dimers.

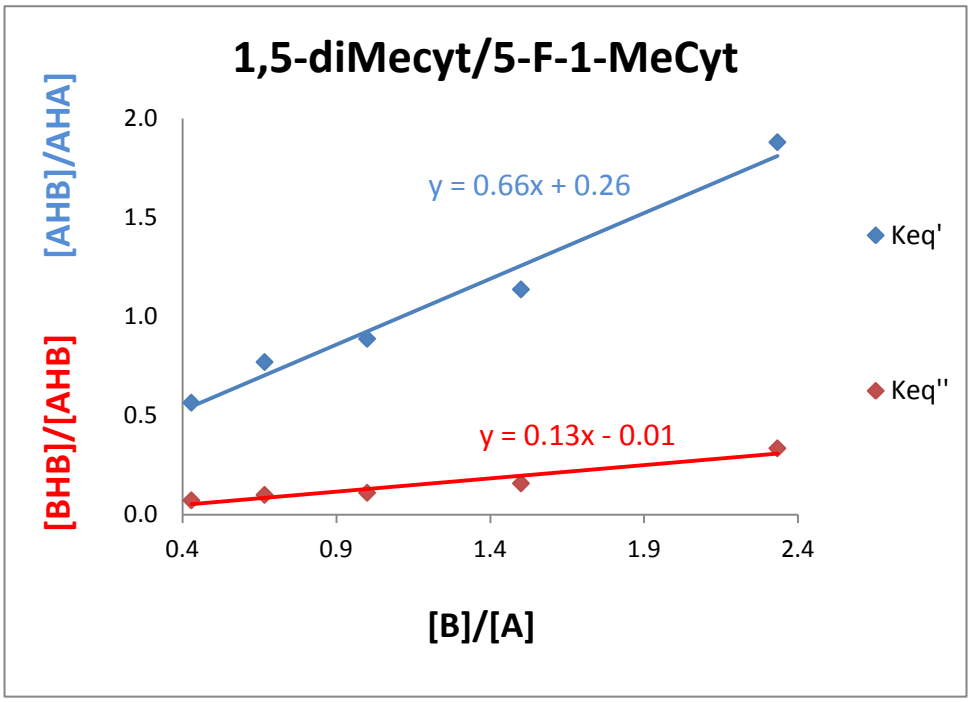


Figure 3.44. Plot of proton-bound dimer average peak intensity ratios versus the concentration ratio of monomers A (1,5-dimethylcytosine) and B (5-fluoro-1-methylcytosine) in the solution injected into the mass spectrometer.

1-Methylcytosine:1,5-Dimethylcytosine

Table 3.35 tabulates the average ratio of the proton-bound dimer intensities observed in the mass spectrum relative to the1,5-dimethylcytosine homodimer for each of the five mixtures of monomers. In the case of the 50/50 mixture of 1-methylcytosine/1,5-dimethylcytosine, the 1,5-dimethylcytosine

proton-bound homodimer peak is 30% more intense than the 1-methylcytosine proton-bound homodimer peak. At equal concentrations of neutral monomer, the heterodimer peak is slightly more intense than either homodimer peak.

Table 3.35. Average peak intensity ratios of the proton-bound dimers formed between mixtures of 1-methylcyotinse 1,5-dimethylcytosine relative to the 1,5-dimethylcytosine homodimer.

Mixture (A/B)	1-MeCyt PBD (A-H^+-A) Ratio	1-MeCyt:1,5-diMeCyt PBD (A-H^+-B) Ratio	1,5-diMeCyt PBD (B-H^+-B) Ratio
70/30	3.4	3.5	1.0
60/40	1.7	2.3	1.0
50/50	0.7	1.5	1.0
40/60	0.3	1.1	1.0
30/70	0.2	0.8	1.0

Figure 3.45 shows the plot of proton-bound dimer peak intensity ratios versus the concentration ratio of the neutral monomers injected into the mass spectrometer. The equilibrium constant between the 1-methylcytosine proton-bound homodimer and the 1-methylcytosine:1,5-dimethylcytosine heterodimer (K_{eq}') is the slope of the blue line and is calculated to be 2.24. As this value is greater than 1, the heterodimer is greatly favored over the 1-methylcytosine homodimer under electrospray conditions. This follows the same trend as the previously discussed equilibrium constants, namely, heterodimers are more stable than the homodimers of the less basic monomer. The equilibrium constant between the heterodimer and the1,5-dimethylcytosine homodimer (K_{eq}'') is the slope of the red line and is calculated to be 0.48.This indicates that the

heterodimer is favored over the 1,5-dimethylcytosine homodimer even though it has more basic monomers than the heterodimer. This dispels the notion that the equilibrium constant between heterodimers and homodimers is governed by the basicity of the monomers, and, therefore, the K_{eq} observed is not purely a result of the gas-phase equilibrium between the dimers. The reported K_{eq} values reflect an equilibrium constant that is a hybrid between the gas-phase and the solution-phase equilibriums.

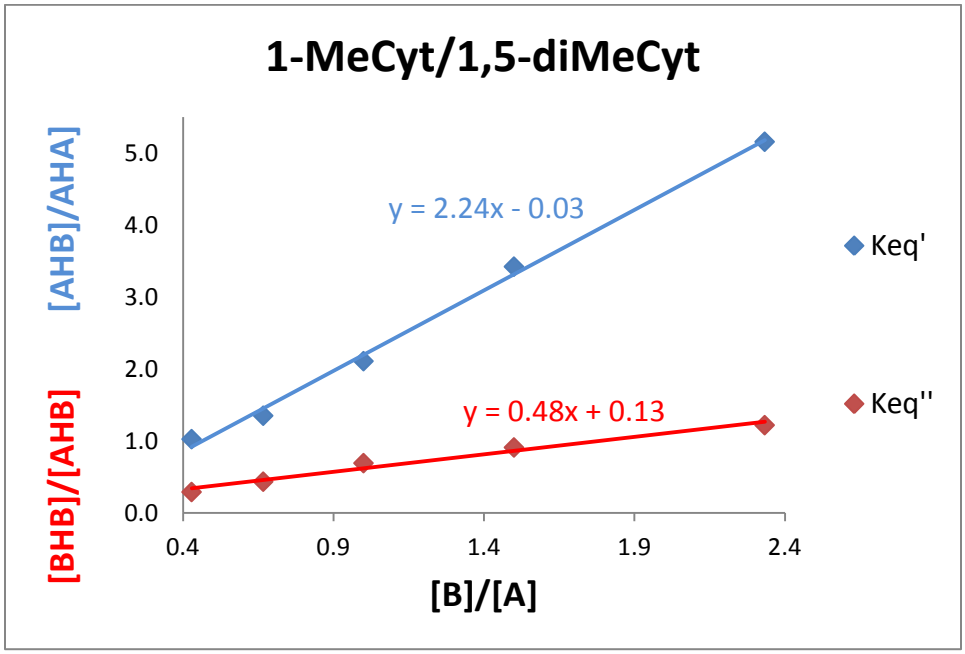


Figure 3.45. Plot of proton-bound dimer average peak intensity ratios versus the concentration ratio of monomers A (1-methylcytosine) and B (1,5-dimethylcytosine) in the solution injected into the mass spectrometer.

Conclusion

DFT-predicted IR spectra can be a powerful tool in describing and assigning experimental IR spectra in the gas phase, as there are no interfering media or counterions in the isolated sample. One must be careful which level of theory is used to assign the experimental bands, however. Unscaled harmonic predictions done at both the B3LYP/6-31G** and the B3LYP/6-311++G** level of theory consistently blue-shift the frequencies of their corresponding experimental peaks. Scaling the harmonic frequencies by a factor of 0.97 gives the predicted spectrum much better agreement with the observed spectra. The scaling factor often over red-shifts many of the predicted bands in the LBS harmonic spectra. This is especially true in the region above 1650 cm^{-1}, where bands corresponding to carbonyl and carbon-nitrogen stretches occur. The scaled SBS harmonic predictions do an excellent job fitting the experimental spectrum, specifically in the region above 1650 cm^{-1}. Anharmonic calculations done at both levels of theory give improvements over the unscaled harmonic spectra, but generally do not provide as much overlap with the experimental spectra as do their corresponding scaled harmonic spectra. The anharmonic predictions also do a poor job fitting the region above 1650 cm^{-1}, often blue-shifting the carbonyl and carbon-nitrogen stretches by 5-25 cm^{-1}. None of the DFT-calculated spectra assigns every peak in the experimental spectrum, each leaving one or two peaks unassigned. These unassigned peaks are assigned to the N-H-N asymmetric stretches, due to their disappearance upon deuteration. Although the scaled

harmonic and anharmonic spectra can assign every peak other than the N-H-N asymmetric peak(s), the SBS scaled harmonic vibrations provide the best fit for each spectra, therefore, its assignments are used in cases where theoretical predictions disagree. Anharmonic frequencies were calculated for the 1-methylcytosine-d_5 deuterium-bound dimer, but lacked the overlap that the SBS scaled harmonic predictions provided. No LBS or anharmonic frequencies were calculated for the two remaining d_5 deuterium-bound homodimers, because the SBS scaled frequencies provided suitable overlap and assignment of the experimental bands.

Comparisons with scaled harmonic and unscaled anharmonic DFT vibrations leave one peak unassigned in each spectrum of the h_5 homo PBDs (**1, 3, and 5**) near 1570 cm^{-1}. This unassigned peak disappears upon replacement of the exchangeable hydrogens with deuterium, implying that this region displays N-H-N asymmetric stretches.

The predicted vibrations from each level of theory place the N-H-N stretch between 2200 and 2800 cm^{-1}. While this is much lower than a normal N-H stretch, usually observed around 3400 cm^{-1}, it is far away from where N-H-N stretches are assigned in these PBDs. One experimental band is unassigned by theory in the N-H-N asymmetric stretch region for the proton-bound heterodimers between 1,5-dimethylcytosine and 5-fluoro-1-methylcytosine and 1-methylcytosine and 5-fluoro-1-methylcytosine, as well. The calculated electronic

energy differences between the two tautomers, along with the large binding enthalpy differences and high barrier to proton transit, suggest that only the more stable tautomer of these PBDs is present under IRMPD conditions. Only one N-H-N asymmetric stretch is observed in those spectra. The two tautomers of the proton-bound heterodimer between 1,5-dimethylcytosine and 1-methylcytosine have a much lower calculated electronic energy and binding enthalpy difference making it more likely that both tautomers are present. The two unassigned bands in the region around 1570 cm^{-1} appear to confirm this notion, as both N-H-N tautomers are observed in the IRMPD spectrum of this PBD. This suggests that, although only one N-H-N asymmetric stretch is observed in the h_5 proton-bound homodimer spectra, the bridging proton between PBDs prefers to be situated on one side and does not give an inversion point of symmetry.

Table 3.36 shows the predicted and identified N-H-N asymmetric stretches for each of the h_5 proton-bound dimers. The SBS scaled harmonic theoretical spectrum is the only DFT-calculated N-H-N stretch shown as it provided the most comprehensive overlap with the experimental spectrum. The vibrational transition from the ground state (E_0) to the first excited state above the barrier to proton transit (E_2) is closer to the assigned frequency for the observed N-H-N asymmetric stretches.

Table 3.36 Predicted and identified N-H-N asymmetric stretches for the h_5 proton-dound dimers 1-9.

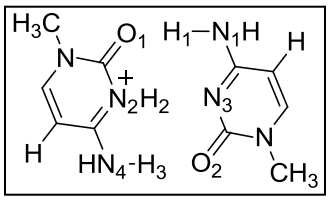
Dimer	SBS Scaled Harmonic N-H-N stretch	1-Dimensional PES E_0 to E_2 transition	Exptl. N-H-N stretch
1	2593	1159	1567
3	2541	1163	1585
5	2612	N/A	1563
7	2645	1394	1580
8	2666	N/A	1575
9	2612	1183	1569/1587

As discussed in the background section, the hydrogen-bonded neutral cytosine-cytosine dimer structure was discovered by comparing DFT-calculated vibrations for the lowest energy dimers with the IR spectrum using double resonance laser spectroscopy. Nir and colleagues made their assignment even though a extraneous peak was observed in the IR spectrum. A band at 3525 cm^{-1} was observed in the IR spectrum, but was not predicted for any of structures for which DFT-calculations were done. The peak was assigned as the anti-symmetric NH_2 vibration involved in a hydrogen bond with a carbonyl. DFT-calculated vibrations do an equally poor job predicting the N-H-N asymmetric stretch in the cytosine-like PBDs. In fact, DFT does a poor job predicting all N-H bonds involved in hydrogen bonds, particularly when those bonds involve a proton. The double-well potential that results from a neutral hydrogen or proton bound between two atoms is quite unlike a single-well harmonic oscillator, which harmonic DFT-calculations approximate, or a Morse potential, which anharmonic

DFT-calculations approximate. DFT cannot accurately estimate the N-H-N asymmetric stretch, but potential energy surface calculations, which are combinations of harmonic oscillators, should be able to.

As previously discussed, two-dimensional potential energy surfaces have been used to accurately predict the N-H-N asymmetric stretch for the proton between the two tertiary amines in *N,N,N',N'-tetramethylputrescine.* In that example only the proton and nitrogen positions were dramatically altered in going from the equilibrium geometry to the transition state and the two-dimensional potential energy surface, as a function of proton and nitrogen coordinates, was adequate to predict the N-H-N vibrational transition. In the case of the proton-bound dimers between substituted cytosines the motion of the proton from equilibrium to transition state is coupled with the movement of many other atoms. The distances between the hydrogen-bonded atoms of 1-methylcytosine in the equilibrium and transition state geometries are shown in Table 3.37.

Table 3.37. DFT-calculated (B3LYP/6-311++G) distances between hydrogen-bonded atom pairs for the 1-methylcytosine PBD in the equilibrium and transition state geometries.**



Atoms	Equilibrium Geometry Distance(Å)	Transition State Geometry Distance (Å)
O_1-H_1	1.95287	1.70082
N_1-H_1	1.01725	1.03018
O_1-N_1	2.96273	2.7246
N_2-H_2	1.05958	1.32153
N_3-H_2	1.78879	1.32153
N_2-N_3	2.84831	2.64307
N_4-H_3	1.04689	1.03018
O_2-H_3	1.65846	1.70082
N_2-O_2	2.70426	2.7246

Unlike the motion of the proton between the two tertiary nitrogens in *N,N,N',N'-tetramethylputrescine,* which involves only the motion of three atoms, the motion of the proton in the 1-methylcytosine PBD is coupled to the motion of nine atoms and six different coordinates. To perform an adequate vibrational transition calculation you would need a six-dimensional potential energy surface. Needless to say, that would take an enormous amount of computing power and time. The reason the one-dimensional potential energy surface could not accurately predict the N-H-N asymmetric stretch is that it did not take into account all of the atom movements coupled to the transfer of the proton from one side of the dimer to the other.

REFERENCES

1. Rajabi, K.; Theel, K.; Gillis, E. A.; Beran, G.; Fridgen, T. D. The structure of the protonated adenine dimer by infrared multiphoton dissociation spectroscopy and electronic structure calculations. *J. Phys. Chem. A* **2009,** *113,* 8099-8107.

2. Yaghmaei, S. *In Search of a Low Barrier Hydrogen Bond in Proton Bridged Diamines;* University of California, Riverside, 2008.

3. Pine, A. S.; Laffery, W. J.; Howard, B. J. Vibrational predissociation, tunneling, and rotational saturation in the HF and DF dimers. *J. Chem. Phys.* **1984,** *81*, 2939-2950.

4. Salpin, J.-Y.; Guilaumont, S.; Tortajada, J.; MacAleese, L.; Lemaire, J.; Maitre, P. Infrared spectra of protonated uracil, thymine and cytosine. *Chem Phys. Chem.* **2007,** *8,* 2235-2244.

5. Liu, M.; Li, T.; Cardoso, D. S.; Fu, Y.; Lee, J. K. Gas phase thermochemical properties of pyrimidine nucelobases. *J. Org. Chem.* **2008,** *73,* 9283-9291.

6. Purrello, R.; Molina, M.; Wang, Y.; Smulevich, G.; Fresco, J. R.; Spiro, T. G.; Fossela, J. Keto-Iminol Tautomerism of Protonated Cytidine Monophosphate Characterized by Ultraviolet Resonance Raman Spectroscopy: Implications of C+ Iminol Tautomer. *J. Am. Chem. Soc.* **1993,** *115,* 760-767.

7. Szczesniak, M.; Leszczynski, J.; Person, W. B. Identification of the Imino-Oxo Form of 1-Methylcytosine. *J. Am. Chem. Soc.* **1992,** *114,* 2731-2733.

8. Bowen, R. D. Ion-Neutral Complexes. *Acc. Chem. Res.* **1991,** *24,* 364-371.

9. Sivaramakrishnan, R.; Michael, J. V.; Klippenstein, S. J. Direct Observation of Roaming Radicals in the Thermal Decomposition of Acetaldehyde. *J. Phys. Chem. A* **2010,** *114,* 755-765.

10. Harding, L. B.; Georgievskii, Y.; Klippenstein, S. J. Roaming Radical Kinetics in the Decomposition of Acetaldehyde. *J. Phys. Chem. A* **2010,** *114,* 765-777.

11. Morton, T. H. The Reorientation Criterion and Positive Ion-Neutral Complexes. *Org. Mass Spectrom.* **1992,** *27,* 353-368.

12. Longevialle, P. Ion-Neutral Complexes in the Unimolecular Reactivity of Organic Cations in the Gas Phase. *Mass Spectrom. Rev.* **1992,** *11,* 157-192.

13. McAdoo, D. J.; Morton, T. H. Gas Phase Analogues of Cage Effects. *Acc. Chem. Res.* **1993,** *26,* 295-302.

14. Morton, T. H. Collisional Activation and Dissociation: Decomposition via Ion-Neutral Complexes. In *The Encyclopedia of Mass Spectrometry;* Armentrout, P. B., Ed.; Elservier: London, 2003; Vol. 1, pp 467-479.

15. Morton, T. H. Theoretical Models for Ion-Neutral Complexes in Unimolecular Ion Decompositions. In *The Encyclopedia of Mass Spectrometry;* Nibbering, N. M. M., Ed.; Elservier: London, 2004; Vol. 4, pp 165-173.

16. Julian, R. R.; Ly, T.; Finaldi, A. M.; Morton, T. H. Dissociation of a Protonated Secondary Amine in the Gas Phase via an Ion-Neutral Complex. *Int. J. Mass Spectrom.* **2007,** *265,* 302-307.