Proton-bound dimers of 1-methylcytosine and its derivatives: vibrational and NMR spectroscopy

Published in:
Physical Chemistry Chemical Physics

DOI:
10.1039/c3cp52260a

Citation for published version (APA):
Proton-bound dimers of 1-methylcytosine and its derivatives: vibrational and NMR spectroscopy†

Hou U. Ung,a Aaron R. Moehlig,a,b Ryan A. Kudla,a Leonard J. Mueller,a Jos Oomens,soft Giel Berden† and Thomas Hellman Morton*a

Vibrational spectroscopy and NMR demonstrate that the proton-bound dimer of 1-methylcytosine, 1, has an unsymmetrical structure at room temperature. In the gas phase, investigation of isolated homodimer 1 reveals five fundamental NH vibrations by IR Multiple Photon Dissociation (IRMPD) action spectroscopy. The NH–N stretching vibration between the two ring nitrogens exhibits a frequency of 1570 cm⁻¹, as confirmed by examination of the proton-bound homodimers of 5-fluoro-1-methylcytosine, 2, and of 1,5-dimethylcytosine, 3, which display absorptions in the same region that disappear upon deuterium substitution. 13C, and 15N NMR of the solid iodide salt of 1 confirm the nonequivalence of the two rings in the anhydrous proton-bound homodimer at room temperature. IRMPD spectra of the three possible heterodimers also show NH–N stretches in the same domain, and at least one of the heterodimers, the proton-bound dimer of 1,5-dimethylcytosine with 1-methylcytosine, exhibits two bands suggestive of the presence of two tautomers close in energy.

Introduction

Dimerization of nucleobases by H⁺ bridging represents a non-Watson–Crick form of association whose biological relevance has emerged in recent years. Many questions remain regarding the structure and dynamics of the proton bridge. This paper presents an examination of proton-bound dimers of 1-methylcytosine and its derivatives using several approaches – solid state NMR (ssNMR), vibrational spectroscopy, X-ray diffraction, and computation – in order to answer two questions. First, is the bridging proton shared equally by both bases in the equilibrium geometry? Second, does a normal modes treatment provide a suitable approximation to describe the vibrational structure? As presented below, the experimental data provide negative answers to both questions.

Background

Single strand self association within DNA has been recognized only comparatively recently as pertinent to biological processes.1 Patterns of self association between identical bases include the i-motif, which consists of proton-bound dimers of cytosine intercalated with one another,2,4 and G-quadruplexes, which bind the Watson–Crick face of one purine to the Hoogsteen face of another.1 Fig. 1A schematically depicts the intercalation of proton-bound deoxyC dimers (also called hemiprotonated C-dimers, represented by ellipses) within a single strand in the i-motif, while Fig. 1B represents the proton-bridged association of the two Watson–Crick faces of a pair of cytosine residues within each ellipse.

Schwalbe and coworkers recently published a detailed analysis of the i-motif structure of a 21-deoxynucleotide C-rich repeat unit from telomeric DNA.2 By means of NMR of labeled single strands, as well as circular dichroism (CD) and other techniques, they uncover a number of important structural features of the i-motif, such as the equilibrium and rates of interconversion between isomeric forms. Intermolecular association of oligo-C strands has also been demonstrated by CD studies, which give an idea of how many bases are needed to overcome the entropic barrier to H⁺-promoted pairing.3
Recent work raises the possibility that intrastrand association takes place not only at the ends of chromosomes (the telomeric region) but also within double helical domains deep inside duplex DNA. Such aggregation requires that C.G-rich regions of the Watson–Crick duplex separate from one another and form the i-motif and a G-quadruplex, respectively. One model proposes that promoters of genes such as c-Myc or VEGF undergo such a conformational change while underwound and that this influences transcription.

Many crystallographic structures of the proton-bound dimer of cytosine and its derivatives have been published, as well as of a number of aggregates of C-rich oligonucleotides that bind together via the i-motif. Association of N-substituted cytosines with their conjugate acids has recently been demonstrated in solution, as well. On the one hand, the majority of crystal structures show that the distances \( r_{NO} \) and \( r_{ON} \) shown in Fig. 1 differ by between 0.15 and 0.2 Å in the equilibrium geometry, consistent with the solid phase NMR spectra reported below. On the other hand, calculations suggest that proton transfer from one base to the other should occur easily, interchanging the two distances. The calculated barrier for the proton-bound dimer of cytosine does appear to have a value higher than that for the proton-bound dimer of 8-aminopurine, for which theory predicts a low-barrier ionic hydrogen bond.

Published data do not unambiguously resolve the question of the symmetry of hemiprotonated cytosines at room temperature. The barrier may depend on environment and the identity of the counterion. A careful X-ray investigation by Bošnjaković-Pavlović and Spasajević-de Biře reveals a temperature-dependent example. Above 200 K the decavanadate salt of the proton-bound dimer of cytosine possesses an inversion point of symmetry, meaning that \( r_{NO} \) has the same value as \( r_{ON} \), but at 100 K the symmetry vanishes. In another case, the two nitrogen–oxygen distances become nearly equal in the middle of a tetrameric stack of proton-bound oligodeoxynucleotides, but not at the ends.

The literature does not contain much detail regarding the vibrational structure of proton-bound dimers of cytosine. The only band characteristic of the proton-bound dimer has been reported to occur at 1890 cm\(^{-1}\), which lies outside the domain of the gas phase experiments described below. The recently developed ability to take IR spectra of organic ions in the gas phase has greatly increased knowledge of ionic hydrogen bonding. Our preliminary communication regarding the proton-bound dimer of 1-methylcytosine in the fingerprint region (300–1800 cm\(^{-1}\)) shows a feature at 1570 cm\(^{-1}\) that we have assigned to the motion of the bridging proton from one base to the other. The present work exhibits the same band in the substituted 1-methylcytosine homo- and heterodimers depicted in Chart 1, which

![Chart 1](image-url)
provides a confirmation of this assignment. In heterodimers, where the two bases have proton affinities close to one another, the band appears to be doubled, possibly indicating the presence of two tautomers.

**Experimental section**

1-Methylcytosine was prepared as described in the literature.\(^1\)\(^8\) 5-Fluoro-1-methyl-cytosine and 1,5-dimethylcytosine were prepared from commercially available 5-fluorocytosine and 5-methylcytosine, respectively, using the same procedure. Crystalline samples of the iodide salt of 1 were prepared by adding 0.5 equivalents of 47% aqueous hydrogen iodide to a saturated solution of 1-methylcytosine in absolute ethanol, removal of solvent by distillation under reduced pressure, and recrystallization of the resulting solid 7 times from absolute ethanol. After repeated recrystallizations the mixture of polymorphs resolved principally into crystal habit B, fine needles whose solid phase NMR spectra are reproduced in Fig. 2 and in the ESI.\(^1\) The isotopomer in which the five NHs have been replaced by deuterium were prepared by repeated recrystallization from ethanol-\(\text{O-d}_3\). 1-Methyl-\(\text{d}_3\) cytosine was prepared using a method previously described\(^6\) but with CD\(\text{D}_3\)I in place of CH\(\text{D}_3\)I.

Magic angle spinning (MAS) solid state NMR (ssNMR) experiments were performed at 14.1 T (\(\text{^1H}\) frequency 600.01 MHz) on a Bruker AVANCE III spectrometer equipped with a triple-resonance 1.3 mm Biosolids (\(\text{^1H/^{13C/^{15N}}\)) MAS probe, spinning at a MAS rate of 50 kHz. Solid state proton spectra were generated by Fourier tranformation of a time-domain FID in response to a 3 \(\mu\)s excitation pulse, with 8192 complex time-domain data points digitized with a dwell of 40 \(\mu\)s (spectral width 25 kHz, total acquisition time 328 ms). 16 transients were averaged with a recycle delay of 3 s. Water served as an external standard calibrated to the chemical shift of TMS. The peaks in the proton ssNMR all have approximately the same linewidth (568/174 Hz full width at half maximum).

\(^{13}\text{C}\) cross-polarization (CP) MAS ssNMR experiments were performed at 14.1 and 9.4 T (\(\text{^1H}\) frequency 600.01 and 400.37 MHz respectively) using a double-resonance 4 mm MAS probe, spinning at a MAS rate of 8 kHz. 9.4 T experiments were performed on the Bruker AVANCE III spectrometer equipped with a double-resonance 2.5 mm MAS probe, spinning at a MAS rate of 20 kHz. 83 kHz \(\text{^1H\pi/2}\) and decoupling pulses were used along with a 2 ms CP and high power (83 kHz) \(\text{^1H}\) decoupling during acquisition. During CP the \(^{13}\text{C}\) nutation rate was set to 41 kHz and the \(^{1H}\) nutation rate ramped from 58–77 kHz. For each spectrum, 2048 complex data points with a dwell of 20 \(\mu\)s (spectral width 50 kHz, total acquisition time 41 ms) were acquired with a recycle delay of 3 s (14.1 T) and 4 s (9.4 T). \(^{15}\text{N}\) CP-MAS experiments on the iodide salt of 1 were performed at 9.4 T as described above, with 1024 complex data points with a dwell of 30 \(\mu\)s (spectral width 33.3 kHz, total acquisition time 30 ms) acquired with a recycle delay of 4 s. During CP the \(^{15}\text{N}\) nutation rate was set to 50 kHz and the \(^{1H}\) nutation rate ramped from 58–77 kHz. \(^{15}\text{N}\) CP-MAS of the neutral and protonated monomers were performed at 14.1 T under conditions described above.
DFT calculations used the Gaussian09 program suite. Unless otherwise specified, geometry optimizations and normal modes computations were performed at the B3LYP/6-31G** level, with a scaling factor of 0.97 applied to normal mode frequencies above 800 cm\(^{-1}\). GIAO calculations of NMR chemical shifts were performed on the gaseous ions. Barrier top geometries were optimized by imposing \(C\equiv_{2h}\) symmetry on the proton-bound homodimers.

Infrared powder spectra were recorded on a Perkin-Elmer One FT-ATR, Raman (1064 nm exciting line) on a Thermo Scientific Nicolet 6700 FT-IR with an XNR FT-Raman module, and single crystal IR on a Bruker Equinox 55 equipped with a microscope.

The techniques for IR multiple photon dissociation (IRMPD) spectroscopy of gaseous ions have been described in detail elsewhere. Briefly, ions electrosprayed from a \(\approx 1\) mM solution in methanol–water containing a trace of acetic acid were injected into a home-built 4.7 T Fourier-transform ion-cyclotron resonance mass spectrometer via a quadrupole deflector and a 1 m long RF octopole ion guide. IRMPD spectra were monitored by expulsion of neutral fragments as well as by diminution of the intensity of the protonated parent ions or proton-bound dimer ions as a function of IR wavelength. Infrared radiation was produced either from a free-electron laser (in the domain \(\leq 1800\) cm\(^{-1}\)) or a LaserVision benchtop optical parametric oscillator (OPO) laser (in the domain 2450–3800 cm\(^{-1}\), whose maximum output is 20 mJ per pulse with a 10 Hz repetition rate). In cases where the extent of ion decomposition from irradiation in the 2450–3800 cm\(^{-1}\) domain was difficult to observe using the OPO by itself (Fig. 4 and the blue trace in Fig. 5), output from a 30 watt cw CO\(_2\) laser was used to irradiate the sample immediately after the pulse from the tunable laser (OPO).

**Results**

The data presented here embrace a variety of experimental and computational approaches. As reported below, X-ray analysis indicates two stable crystal habits (called A and B) of the iodide salt of 1, only one of which (form A) has been previously described. Ascertainning the symmetry of proton-bound dimer 1 at room temperature for form B represents one objective, which ssNMR resolves in favor of a structure where the bridging proton does not lie midway between the nitrogens. Similarities between vibrational spectra in the crystalline and the gas phases suggest that the dimer has the same molecular symmetry in both states of matter. While IR and Raman of crystals and IRMPD vibrational spectroscopy of gaseous 1–6 provide evidence for that interpretation, they reveal that DFT calculations of normal modes do not give an accurate picture of the motion of the bridging proton from one nitrogen to the other. A brief summary of computational results serves to introduce the experimental data.

**Computational results**

Previous investigators have calculated the barrier to proton transfer within the gaseous proton-bound dimer of cytosine. The potential surface for predicting vibrational frequencies makes use of the electronic energy barrier, while estimation of the enthalpy of activation includes a correction for the zero point energy difference between the equilibrium structure (\(C\equiv_{2h}\) symmetry with 3N-6 modes) and the barrier top (\(C\equiv_{2h}\) symmetry having 3N-7 modes with positive force constants). B3LYP/cc-pVTZ optimizations for 1 and the barrier top give an electronic energy barrier of 4.50 kcal mol\(^{-1}\), which diminishes to 1.87 kcal mol\(^{-1}\) with application of a 2.63 kcal mol\(^{-1}\) zero point energy correction, comparable to previously published results. An electronic energy difference of 5.27 kcal mol\(^{-1}\) results from single point calculations at CCSD/cc-pVTZ/B3LYP/cc-pVTZ. Experimental data from vibrational spectroscopy reported below suggest that normal modes calculations may seriously overestimate the zero point energy for the equilibrium geometry, owing to their inability to provide a reliable approximation for double-well potentials having low barriers. In any event, the calculated electronic energy barrier has a value higher than the harmonic zero point energy for the vibration having the highest degree of NH - N asymmetric stretch character (for which the normal modes calculation predicts an unscaled frequency of 2737 cm\(^{-1}\) at B3LYP/cc-pVTZ).

The calculated barrier for proton transit differs among the homodimers. The B3LYP/6-31G** electronic energy barriers for proton transfer between partners in proton-bound homodimers 1, 2, and 3 have values of 3.73, 3.25, and 3.90 kcal mol\(^{-1}\), respectively (as compared with the literature value for the proton-bound dimer of cytosine itself of 4.62 kcal mol\(^{-1}\) at that level\(^{11}\)). As noted in the previous paragraph, enlarging the basis set increases the calculated barrier.

Apart from the aforementioned motion in a double well potential, scaled normal mode frequencies give a good fit to observed IR and Raman bands, as will be outlined below. In other words, for each of the homodimers only one experimental band fails to match closely any of the calculated peaks, and this band vanishes in the gaseous ions with a D’ bridge. That outcome leads to assignment of the superimery band in the fingerprint region of the H’ bridged ions to the motion...
Experimental results

Repeated recrystallizations of salts of the proton-bound dimer of 1-methylcytosine from absolute ethanol yield two different crystalline compounds, depending on whether an excess of molecular iodine is present. On one hand, the triiodide salt of 1 (black needles) contains two molecules of water for every proton-bound dimer. Cursory X-ray analysis of a polycrystalline sample indicates that it crystallizes in the \( P2(1)/c \) space group. On the other hand, the mono-iodide salt of 1 contains no water but exists as a mixture of colorless polymorphs, one of which (crystal habit A) has the same unit cell dimensions as the previously reported structures (monoclinic \( P \); \( a = 7.1788 \) \( \text{Å} \), \( b = 8.6098 \) \( \text{Å} \), \( c = 11.4628 \) \( \text{Å} \), \( \beta \) angle = 97.241° at 100 K),\(^6\) while the other (crystal habit B) exhibits a monoclinic C-centered cell \( (a = 20.430 \) \( \text{Å} \), \( b = 6.473 \) \( \text{Å} \), \( c = 13.064 \) \( \text{Å} \), \( \beta \) angle = 125.395° at 100 K). The two crystal habits of the mono-iodide differ: A forms flakes, while B forms small needles.

Solid state NMR (ssNMR) spectra of a sample of the mono-iodide salt of 1 that contains predominantly the B form shows five proton resonances. The furthest downfield occurs at 16.1 ppm (see ESI†), which matches almost exactly the chemical shift predicted by GIAO calculations for the bridging proton in between the two ring nitrogens. The solid-phase probe does not have a wide range of temperature variability, but the proton ssNMR near 50 °C is superimposable upon the one at room temperature.

Were the bridging proton held midway between two ring nitrogens, with \( r_{\text{NO}} \) equal to \( r_{\text{ON}} \), the proton NMR should...
Table 1: CP-MAS $^{13}$C and $^{15}$N chemical shifts (ppm) of crystalline samples of the iodide salt of 1 and its monomeric constituents at 150.87 MHz (carbon), 40.5 MHz (nitrogen) for the iodide salt of 1, and 60.8 MHz (nitrogen) for neutral 1-methylcytosine and the iodide of its conjugate acid. External standard for carbon is adamantane referenced to TMS. External standard for nitrogen is ammonium chloride referenced to liquid ammonia.

<table>
<thead>
<tr>
<th></th>
<th>Neutral MeCyt</th>
<th>Conjugate acid iodide salt</th>
<th>Proton-bound dimer monoiodide salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl</td>
<td>40.0</td>
<td>42.8</td>
<td>39.5, 40.8</td>
</tr>
<tr>
<td>C5 vinyl</td>
<td>94.9</td>
<td>92.6</td>
<td>93.6, 98.4</td>
</tr>
<tr>
<td>C6 vinyl</td>
<td>147.5</td>
<td>148.3</td>
<td>148.7, 150.0</td>
</tr>
<tr>
<td>Carbonyl</td>
<td>157.5</td>
<td>152.8</td>
<td>155.0</td>
</tr>
<tr>
<td>C-NH$_2$</td>
<td>168.7</td>
<td>159.8</td>
<td>163.7, 164.2</td>
</tr>
<tr>
<td>Nitrogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-Me</td>
<td>139.7</td>
<td>144.0</td>
<td>135.2, 136.4</td>
</tr>
<tr>
<td>H$_2$N</td>
<td>93.6</td>
<td>101.2</td>
<td>102.6, 107.7</td>
</tr>
<tr>
<td>Ringer</td>
<td>207.8</td>
<td>147.3</td>
<td>169.7</td>
</tr>
</tbody>
</table>

Neutral MeCyt, Conjugate acid iodide salt, and Proton-bound dimer monoiodide salt exhibit only five resonances, as observed. However, heteronuclear NMR demonstrates that the iodide salt of 1 does not possess such a high degree of symmetry. Fig. 2 reproduces the $^{13}$C and $^{15}$N ssNMR spectra of the sample. The resonances predicted for the isolated dimer cation in the gas phase enable the assignments summarized in Fig. 2. Table 1 lists the $^{13}$C and $^{15}$N chemical shifts and assignments for the constituent monomers as well as for the proton-bound dimer.

IR of a single needle containing predominantly the B form displays the 600–2000 cm$^{-1}$ absorption profile reproduced at the top of Fig. 3, which exhibits slightly better resolution than the powder spectra of a mixture containing both polymorphs, of which both the IR and Raman are also reproduced. As noted above, a previous report cites one absorption as having special importance: the broad band with medium intensity at 1890 cm$^{-1}$, which appears in the IR of the iodide salt of 1 and also as a weak band in the Raman spectrum, shifted slightly to the red. This band persists with unchanged intensity in the IR of the salt in which all NHs have been exchanged for deuterium (see ESI†), which argues against its assignment as an NH stretch. An absorption near 1570 cm$^{-1}$, assigned in the gas phase spectra below to the motion of the proton that bridges from one ring nitrogen to the other, appears as a shoulder in the IR of the iodide salt of 1 but does not show up in the Raman.

Our previously published IRMPD spectrum of gaseous 1 exhibits this 1570 cm$^{-1}$ band, which disappears upon replacement of the five exchangeable hydrogens with deuterium. That result leads to the assignment of the band as having predominantly the character of the asymmetric stretching motion of the bridging proton between the two ring nitrogens (the red H in Fig. 1B). Because of the difficulty of accessing frequencies above 1850 cm$^{-1}$ at the free-electron laser beam setting used here, it has not yet proven possible to see if the gaseous ion also exhibits an 1890 cm$^{-1}$ band in the gas phase. Apart from the 1570 cm$^{-1}$ band, which is weak in the IR of the crystalline salt and absent in its Raman, the bands observed by IRMPD spectroscopy of the gaseous ion occur at nearly the same frequencies as the IR absorptions of the solid (see ESI†). The spectra reproduced in Fig. 5 below permit assignment of the other NH stretches in the gaseous ion 1.

IRMPD of the gaseous proton-bound homodimer of 1-methylcytosine, 1, leads to expulsion of a neutral 1-methylcytosine molecule. The protonated monomer (m/z 126) that remains consists of a mixture of O-protonated and N-protonated monomers, as previously reported, despite the fact that the parent proton-bound dimer must be exclusively protonated on the ring nitrogen (as drawn for the m/z 251 ion in Fig. 5 as well as in Chart 1). The published IRMPD spectrum of the m/z 126 fragment from dissociation of the dimer exhibits a profile identical to that of the protonated monomer ion produced directly by electrospray from aqueous solution.

IRMPD spectroscopy of the protonated monomer in the 3150–3800 cm$^{-1}$ domain, reproduced in Fig. 4, shows a prominent band at 3555 cm$^{-1}$, confirming the presence of the O-protonated structure, as previously inferred from IRMPD spectra of the protonated monomer in the fingerprint region. The frequency is very close to the that of the OH stretch of the carboxyl group in protonated Ala-Ala as well as to the C≡O–H band predicted for O-protonated formamide.

In order to assign the NH vibrations of gaseous 1 it becomes necessary to look at a partially deuterated ion. The IRMPD spectrum of the undeuterated proton-bound homodimer 1 displays more bands in the 2500–3650 cm$^{-1}$ domain than one might expect (Fig. 5). The spectrum was recorded under two conditions, one using only the output from a tunable OPO laser to dissociate the ions (red trace in panel A) and the other with an additional pulse from a CO$_2$ laser to enhance dissociation (blue trace in panel A). The blue trace shows a pair of absorption bands at 2560 and 2600 cm$^{-1}$ that barely emerge from the noise level, which compares with a single absorption predicted at B3LYP/6-31G** of 2815 cm$^{-1}$ (harmonic frequency scaled by 0.97), corresponding to the stretch of the hydrogen-bonded proton of the NH$_2$-group of the positively charged partner. Neither scaled normal modes at B3LYP/6-311+G** nor unscaled anharmonic frequencies at B3LYP/6-31G** and at B3LYP/6-311+G** match band positions any better, so this paper compares experimental vibrational band positions with scaled B3LYP/6-31G** normal modes throughout.

To clarify the spectra, as well as to diminish harmonics and combination bands, the experiment was repeated by electrosparying the ions from D$_2$O-methanol-O-d and isolating the d$_4$ ions for IRMPD spectroscopy, as panel B shows. Such an experiment would present serious challenges in condensed phase spectroscopy, but in a gas phase study isolation of an ion possessing a designated m/z value becomes straightforward. While complete replacement of the five exchangeable hydrogens with deuterium produces a d$_5$ ion, the object of this experiment was to observe the spectra of a mixture of the four d$_4$ isomers drawn in panel B. Two bands that are not fundamentals (indicated by the open arrows in panel A) disappear in the d$_4$ spectra. Six prominent bands remain in the domain above 2500 cm$^{-1}$: the two bands assigned to the non-hydrogen bonded NH stretches near 3500 cm$^{-1}$ (not as well resolved in
the $d_4$ spectra in panel B as in panel A; predicted for the $d_0$ at 3552 and 3595 cm$^{-1}$ and for two of the $d_4$ isomers at 3550 and 3580 cm$^{-1}$, respectively, the NH stretch just below 3400 cm$^{-1}$ corresponding to the hydrogen bonded amino group belonging to the uncharged partner in the dimer (predicted at 3360 cm$^{-1}$ for $d_0$ and at 3385 cm$^{-1}$ for one of the $d_4$ isomers), the CH stretches of the methyl groups just below 3000 cm$^{-1}$, and a pair of bands near 2600 cm$^{-1}$ corresponding to the hydrogen bonded amino group of the charged partner in the dimer. The frequency of this pair of bands shifts slightly to the blue in panel B (2600 and 2640 cm$^{-1}$) relative to panel A, but the separation between them remains approximately 40 cm$^{-1}$. As noted above, DFT predicts a single band at 2815 cm$^{-1}$ for $d_0$, as compared with a single band predicted at 2830 cm$^{-1}$ for one of the $d_4$ isomers. The experimental result would appear to rule out tunneling splitting as a cause of the doubling of this absorption. One possibility, discussed below, invokes a combination band with a butterfly vibration of the two rings. The NH–N stretch of the bridging proton between the two ring nitrogens, which DFT calculations predict at much higher frequencies than observed, occurs at a much lower frequency, as noted above.

Examination of the proton-bound dimer of 1-methyl-$d_3$-cytosine, in which the 3000 cm$^{-1}$ band vanishes (ESI†), provides confirmation of the assignment of the methyl CH band. With the exception of the 1570 cm$^{-1}$ band, the observed band positions agree not only with predictions from scaled DFT normal modes calculations, but also with the positions of the absorptions of the solid salt in Fig. 3 (ESI†).

Looking at the IRMPD spectra of homodimers 2 and 3 reveals many of the same features seen in the published IRMPD spectrum of ion 1 in the fingerprint region.17 Rather than show the published spectra of 1 once again, this paper reproduces the IRMPD spectra of 2 and 3 and the $d_4$-analogues from replacement of all the exchangeable hydrogens with deuterium. Fig. 6–8 reproduce gas phase IRMPD spectra of the proton-bound homodimers of substituted analogues of 1-methylcytosine: 5-fluoro-1-methylcytosine (2) and 1,5-dimethylcytosine (3) and their deuterated analogues.

Fig. 6 compares the spectra of undeuterated 2 with a scaled B3LYP/6-31G** normal modes calculation, while Fig. 7 provides a comparison of the $d_4$ dimer produced by electrospray from D$_2$O–methanol-O–d with the same level of DFT. Fig. 7 displays a good fit between experimental and calculated band positions in the 1000–1800 cm$^{-1}$ domain, but Fig. 6 exhibits strong bands that do not coincide with theory. As noted above, the spectra in the 1000–1800 cm$^{-1}$ domain closely resemble the previously published IRMPD spectra of 1 and its $d_4$-analogue in the same domain.17 The extra bands in the 3200–3300 cm$^{-1}$ domain of 2 are assigned to the same type of overtones as indicated by an open arrow in Fig. 5A. The extra band in the 1000–1800 cm$^{-1}$ domain is assigned to a fundamental at 1580 cm$^{-1}$ corresponding the NH–N stretch between the two ring nitrogens. For homodimer 3, comparing experiment with theory for the deuterated $d_4$-dimer in lower panel of Fig. 8 again shows an excellent fit between observed and predicted band positions, while the upper panel of Fig. 8 again exhibits a poorer fit between experiment and the DFT normal modes. In looking at Fig. 6–8, it becomes apparent that both homodimers 2 and 3 display one extra peak in the 1500–1600 cm$^{-1}$ domain beyond what theory predicts. As in the case of the previously published IRMPD spectra of 1, the extra peak is assigned to the stretching vibration of the bridging proton (the H shown in red in Fig. 1B).

Fig. 9–11 reproduce the IRMPD spectra of the three proton-bound heterodimers 4–6 that can be formed from 1-methylcytosine and its aforementioned analogues. Once again, comparing experimental spectra with DFT normal modes calculations show more experimental bands than DFT predicts. In Fig. 10 and 11, overtones occur in the 3200–3300 domain, just as observed in Fig. 5A and 6. And, once again, extra peaks occur in the

![Fig. 6 IRMPD spectra of ion 2 recorded using a free-electron laser (blue trace, left hand panel) and an OPO laser (blue trace, right hand panel) compared with normal mode frequencies calculated at B3LYP/6-31G** (red traces, scaled by 0.97; 30 cm$^{-1}$ Gaussian broadening). On the basis of Fig. 5, the band observed around 3230 cm$^{-1}$ and the shoulder at 3300 cm$^{-1}$ are assigned as overtones. Following Fig. 5, the band around 3350 cm$^{-1}$ is assigned to the stretch of the hydrogen-bonded NH from the primary amino group of the uncharged partner and the 3470–3520 cm$^{-1}$ bands to the non-hydrogen-bonded NHs of the primary amino groups. The arrow indicates the band in the fingerprint region assigned to motion of the bridging H$^+$.](image-url)
1500–1600 cm$^{-1}$ domain in all of the heterodimers. The existence of two tautomers of the heterodimers provides a possible explanation, which the Discussion section below will briefly address.

**Discussion**

The symmetry of the proton-bound dimer of cytosine and its derivatives has occasioned much discussion.\(^6\),\(^7\) The \(^{13}\)C and \(^{15}\)N solid state NMR spectra of solid samples of the iodide salt of the proton-bound dimer of 1-methylcytosine, 1, (predominantly the B crystal habit) confirm the asymmetry of the dimer at room temperature. The bridging proton preferentially associates with one or the other of the two ring nitrogens, as Fig. 1B illustrates, permitting differentiation between the charged and the uncharged partners in dimer ion 1. Vibrational spectroscopy of gaseous ion 1 at ambient temperature and the IR of its crystalline iodide salt show such similarities as to suggest that 1 has the same structure in both phases.

The widths of solid phase NMR absorptions (>100 Hz, even with magic angle spinning) preclude measurement of scalar spin–spin coupling constants and also make it difficult to separate closely spaced resonances. On one hand, the proton NMR of the crystalline iodide salt of 1 shows only five peaks, possibly consistent with a symmetric structure. On the other hand, heteronuclear solid phase NMR spectra (\(^{15}\)N and \(^{13}\)C) display five and eight peaks, respectively, which require the dimer to have an asymmetric structure with \(r_{\text{NO}} \neq r_{\text{ON}}\) at room temperature. If the dimer did have \(C_2\) symmetry, only three \(^{15}\)N and five \(^{13}\)C peaks should have been seen.

The \(^{15}\)N natural abundance NMR exhibits smaller differences than predicted by DFT for the gaseous cation. The two primary amino groups exhibit a difference of 5.5 ppm (calculated 14 ppm; see ESI\(^\dagger\)) and are both downfield of the experimental values for the neutral and protonated monomers. A 6 ppm difference separates the resonances of the methylated ring nitrogens (calculated 9 ppm), which are both upfield from the experimental values for the neutral and protonated monomers. The resonances of proton-bridged ring nitrogens overlap one another and have a value nearly equal to the algebraic mean of the experimental values for neutral and monoprotonated monomers. Although the literature reports that, in the hydrated iodide salt of 1, \(^{14}\)N nuclear quadrupole resonance distinguishes between the proton-bridged ring nitrogens,\(^2\),\(^2\) this proves not to be the case in the \(^{15}\)N NMR experiments reported here. Although the charged and uncharged partners remain distinct in the anhydrous iodide salt of 1, NMR resolves neither the two proton-bridged ring nitrogens nor the two carbonyl carbons.

In the \(^{13}\)C NMR a difference of 1.3 ppm separates the two methyl resonances of the iodide salt of 1, substantially greater than theory predicts for the gaseous ion (a difference of only 0.2 ppm; see ESI\(^\dagger\)) but less than the difference between the
neutral precursor and the iodide salt of its conjugate acid (2.6 ppm). The upfield pair of vinyl resonances displays an opposite trend: a difference of 4.8 ppm separates the upfield pair of the proton-bound homodimer, as compared with a calculated difference 2.0 ppm and an experimental difference between neutral and conjugate acid of 2.4 ppm (with the neutral resonance more downfield). The downfield pair of vinylic carbons exhibits much better agreement with theory: an experimental difference of 1.3 ppm separates the pair from the homodimer, as compared with a calculated difference of 1.2 ppm and an experimental difference between neutral and conjugate acid of 0.8 ppm. The aminated carbons provide the furthest downfield absorptions, separated by only 0.4 ppm, \textit{versus} a predicted difference of 5.2 ppm and an experimental separation between neutral and conjugate acid of 8.9 ppm. The separation of the two carbonyl resonances from the dimer is less than the resolution of the instrument, while the predicted difference has a value of 8.8 ppm and the experimental separation between neutral and conjugate acid has a value of 4.7 ppm (with the neutral farther downfield). The disagreements between computation and theory suggest that, in the solid salts, interactions between the cations and the iodide counter-ion affect shielding.

Previous work has highlighted discrepancies between the normal modes approximation and experimental results for double well potentials.\textsuperscript{16} Double wells for proton-bridged conjugate acid ions with much simpler structure have been solved in reduced dimensionality. When the electronic energy barrier has a value comparable to the zero point energy for intramolecular proton transfer, observed $v = 0 \rightarrow v = 1$ transitions for N–H$\cdot\cdot\cdot$N transit have much lower values (550–600 cm$^{-1}$) than normal mode calculations predict (even with anharmonic corrections). In the present case, the electronic energy for the double-well potential of 1 has a value higher than the anticipated zero point energy, yet the observed vibrational transitions for 1–6 (around 1570 cm$^{-1}$) still exhibit frequencies lower than expected on the basis of DFT normal modes.

The IR absorption spectrum of the crystalline iodide salt of 1 shows a band at the same frequency as previously reported at 1890 cm$^{-1}$ (ref. 13) as well as a shoulder near 1570 cm$^{-1}$, the band assigned to the NH$\cdot\cdot\cdot$N stretch in the gaseous ions. Deuterium substitution of a crystalline sample argues against assignment of the 1890 cm$^{-1}$ band as an NH stretch, but the range of the free-electron laser used for the gas phase IRMPD spectra does not extend far enough to see if the 1890 cm$^{-1}$ band also occurs in the gaseous ion. The IRMPD spectrum of 1 in the
The IRMPD spectrum of gaseous 1 in the NH stretching region (Fig. 5) exhibits a band from the NH$_2$ hydrogen of the uncharged partner that enjoys hydrogen bonding with the carbonyl oxygen of the charged partner (3380 cm$^{-1}$) and a separate band from the NH$_2$ hydrogen of the charged partner that enjoys hydrogen bonding with the carbonyl oxygen of the uncharged partner (2560 cm$^{-1}$ with a band of nearly equal intensity 40 cm$^{-1}$ to the blue). Table 2 summarizes assignments for the NH stretches of proton-bound homodimers in the gas phase. The stretch of the hydrogen bonded NH$_2$ hydrogen of the neutral partner in ion 2 occurs at a slightly lower frequency (3345 cm$^{-1}$), as predicted by DFT (a 7 cm$^{-1}$ calculated red shift at B3LYP/6-31G**), corresponding to an NH distance 0.004 Å longer than calculated for 1. At the same time the NH···N stretch between the two ring nitrogens in 2 occurs at a slightly higher frequency (1580 cm$^{-1}$), corresponding to a distance between ring nitrogens 0.026 Å shorter than calculated for 1. Despite a calculated difference between $r_{\text{NO}}$ and $r_{\text{ON}}$ of 0.26 Å calculated for each of the homodimers (virtually the same as computed for 1 using the cc-pVTZ basis), DFT geometry optimization gives a N-H···N bond angle $\geq$ 179.5$^\circ$ for the ionic hydrogen bond between the ring nitrogens.

With regard to IR spectroscopy, the bands observed in the powder spectrum agree well with the absorption spectrum of a single needle of the B form. With the exception of the strong 1570 cm$^{-1}$ band in the previously published IRMPD spectrum of isolated 1, all of the bands in the crystalline sample in the 500–1700 cm$^{-1}$ domain correlate with those in the gas phase IRMPD spectrum (see ESI†). Both the IRMPD spectrum of gaseous 1 and the IR absorption spectrum of the solid iodide salt display bands that must be assigned to harmonics or combination bands. The method of partial deuteration (Fig. 5) reduces the number of bands in the gas phase spectrum sufficiently that assignments can be made in the 2500–3500 cm$^{-1}$ domain, as the first column in Table 2 summarizes.

One combination band that partial deuteration of gaseous ion 1 appears not to remove is the higher frequency absorption band of 40 cm$^{-1}$ splitting does not change. A provisional assignment of the higher frequency band to a combination with the calculated 40 cm$^{-1}$ flapping motion of the two rings (also called a butterfly motion$^{23}$) seems probable. While IRMPD spectroscopy has not yet observed the 40 cm$^{-1}$ fundamental directly, vibronic spectroscopy of neutral molecules demonstrates the accuracy of unscaled normal modes for low frequency modes of ground electronic states when compared with the experimental values observed for electronically excited states.$^{24}$ For example, the lowest energy IR-active fundamentals of the neutral dimer of 1-methylcytosine with the enol form of cytosine have predicted harmonic frequencies of 30 and 48 cm$^{-1}$ (B3LYP/6-31G**), no

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Bands observed for NH stretches for gaseous homodimers 1–3 by IRMPD (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Non H-bonded hydrogens of –NH$_2$ groups</td>
<td>3500, 3520</td>
</tr>
<tr>
<td>–NH$_3$ H-bonded hydrogen of uncharged ring</td>
<td>3380</td>
</tr>
<tr>
<td>–NH$_3$ H-bonded hydrogen of charged ring</td>
<td>2560</td>
</tr>
<tr>
<td>H-bonded hydrogen between ring nitrogens</td>
<td>1570</td>
</tr>
<tr>
<td>Notes:</td>
<td>$^a$ Not recorded. $^b$ Too weak to observe.</td>
</tr>
</tbody>
</table>
more than 1–2 cm\(^{-1}\) away from the values from the reported vibronic spectrum of the neutral dimer in a hole-burning experiment.\(^{23}\)

The IRMPD spectra of the 5-fluoro and 5-methyl substituted homodimer cations 2 and 3 share nearly all of the features of the previously published spectrum of 1, including a band near 1570 cm\(^{-1}\) that is not predicted by theory, which vanishes upon replacement of all five exchangeable protons with deuterium. One would expect such a band to shift upon isotopic substitution, but in no case is it possible definitively to assign a band in the spectra of the deuterated ions that corresponds to a shifted band. Such behavior has been noted previously in proton-bridged diamines,\(^{16}\) and it may be that the intensity of the bands corresponding to shifted absorptions of the \(d_2\)-ions diminishes as a consequence of deuteration. In any event, just as for the ring NH proton, but exhibit the same absorptions of the tautomers of a heterodimer absorb at different frequencies.

The IRMPD spectra of all three proton-bound heterodimers (4, 5, and 6) in the 1000–1800 cm\(^{-1}\) domain appear more congested between 1500 and 1600 cm\(^{-1}\) than do the homodimer spectra. Occurrence of combination bands with very low frequency vibrations provides one possible explanation, analogous to the pair of bands near 2600 cm\(^{-1}\) in the IRMPD spectrum of 1. An alternative interpretation holds that two tautomers of a heterodimer absorb at different frequencies for the ring NH proton, but exhibit the same absorptions everywhere else. Whereas transit of the bridging \(H^+\) in a homodimer interchanges \(r_{NO}\) and \(r_{ON}\), leading to an ion having the same structure, proton shift within a heterodimer produces a tautomer ion with a different structure. Proton exchange between the charged and uncharged partners, particularly within heterodimer 5, could lead to appreciable concentrations of two tautomers for a heterodimeric ion at ambient temperature, accounting for an increase in the observed number of bands. DFT calculations suggest that only 0.5 kcal mol\(^{-1}\) separates the two tautomers of 5 (versus 2.4 kcal mol\(^{-1}\) for 4 and 1.9 kcal mol\(^{-1}\) for 6), both of which correspond to local minima.

Ring methylation affects binding. In biological systems, methylation at the 5-position of cytosine residues serves to repress gene transcription,\(^{26}\) and it appears likely that cytosine methylation occurs in promoter regions as well as in genes. DFT predicts that pairing within the i-motif becomes weakened by methylation of one of the cytosines. Although computational methods consistently overestimate the binding of cytosine to its protonated parent\(^{11}\) relative to the experimental value in the literature,\(^{27}\) DFT predicts that the favored tautomer of heterodimer 5 has a binding enthalpy 0.6 kcal mol\(^{-1}\) less than calculated for homodimer 1. Further investigation of proton-bound cytosine dimers can address substantiative effects on binding and tautomeric equilibria in greater detail.

Acknowledgements

The authors are grateful to Dr Fook Tham for X-ray crystallographic analysis, Bryan Tamashiro and Prof. G. J. O. Beran for performing anharmonic frequency calculations, and to the FELIX staff for their skillful assistance in helping perform IRMPD experiments, particularly Dr A. F. G. van der Meer and B. Redlich. This work was supported by the NSF (CHE-0848607 to L.J.M. and CHE-0848517 to T.H.M.), by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), and by the CNRS. H.U.U. and R.A.K. gratefully acknowledge support through US Department of Education GAANN Award #P200A120170.

References


Paper


