An IRMPD Spectroscopic and Computational Study of Protonated Guanine-Containing Mismatched Base Pairs in the Gas Phase

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Physical Chemistry Chemical Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CP-ART-11-2019-006393</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>26-Nov-2019</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Cheng, Ruodi; Memorial University of Newfoundland, Department of Chemistry Loire, Estelle; Laboratoire Chimie Physique -CLIO Martens, Jonathan; Radboud University Nijmegen Faculty of Science, Institute for Molecules and Materials, FELIX Laboratory Fridgen, Travis; Memorial University of Newfoundland, Department of Chemistry</td>
</tr>
</tbody>
</table>
An IRMPD Spectroscopic and Computational Study of Protonated Guanine-Containing Mismatched Base Pairs in the Gas Phase

Ruodi Cheng, a Estelle Loire, b Jonathan Martens, c and Travis D. Fridgen a, *

a: Department of Chemistry, Memorial University, St. John’s, NL, A1B 3X7, Canada
b: Laboratoire Chimie Physique – CLIO, Batiment 201, Porte 2, Campus Universite d’Orsay, 91405 France
c: Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7c, Nijmegen, 6525ED, The Netherlands

* to whom all correspondence should be addressed at tfridgen@mun.ca
Abstract

Infrared multiple photon dissociation (IRMPD) spectroscopy has been used to probe the structures of the three protonated base-pair mismatches containing 9-ethylguanine (9eG) in the gas phase. Computational chemistry has been used to determine the relative energies and compute the infrared spectra of these complexes. By comparing the IRMPD spectra with the computed spectra, in all cases, it was possible to deduce that what was observed experimentally were the lowest energy computed structures. The protonated complex between 9eG and 1-methyl thymine (1mT) is protonated at N7 of 9eG—the most basic site of all three bases in this study—and bound in a Hoogsteen type structure with two hydrogen bonds. The experimental IRMPD spectrum for the protonated complex between 9eG and 9-methyladenine (9mA) is described as arising from a combination of the two lowest energy structures, both which are protonated at N1 of adenine and each containing two hydrogen bonds with 9eG being the acceptor of both. The protonated dimer of 9eG is protonated at N7 with an N7-H⁺--N7 ionic hydrogen bond. It also contains an interaction between a C-H of protonated guanine and the O6 carbonyl of neutral guanine which is manifested in a slight red shift of that carbonyl stretch. The protonated 9eG/9mA structures have been previously identified by X-Ray crystallography in RNA and are contained within the protein database.
1. Introduction.

There are approximately 6 billion nucleobases composing the human genome, existing as base pairs and constituting the genes which replicate during the process of cell division. The nucleobases in DNA double-stranded helices and the folded single-stranded RNA typically form canonical base pairs by non-covalent interactions; guanine (G) matches with cytosine (C) and adenine (A) matches with thymine (T) in DNA and with uracil (U) in RNA (Scheme 1). Nearly 1 out of $10^7$ nucleobases is estimated to occur as natural mismatches, which can cause mutagenesis, carcinogenesis or cell death. While checking mechanisms exist for incorrect base insertion, spontaneous mutagenesis has been shown to be influenced by the mismatched base pairs (ie. GT or CA) adopting geometries other than Watson-Crick (WC) geometries, but which mimic the shape of the WC base-pair, and go unnoticed. Sometimes tautomerization of nucleobases can be the cause for these mismatches. Because the frequency of mutations in DNA and RNA molecules is susceptible to the stabilities and shapes of these nucleobases’ mismatches, there is extensive scientific literature on mismatched base pairs which show that they don’t follow the complementary principles discovered by Watson and Crick.

Among all five nucleobases, guanine has been shown to have a tendency to self-assemble and form G-quartets, G-quadruplexes, guanine-guanine (GG) mismatches, and some other guanine-containing adducts. This is possibly due to guanine involved base pair mismatches and self-assemblies being more thermodynamically stable than others which is supported by the higher melting temperature of high guanine containing double-stranded helices. DNA duplexes are reported to be efficient charge carriers due to π–π interactions in their close stacked base pairs. The efficiency of long-distance electron transfer in DNA was considered to be sensitive to guanine and G-containing mismatches because the electrons were found to be generated along with the guanine radical cation, and mismatches might disrupt the integrity of π–π stacking of regular base pairs leading to lower electron transfer rates. However, recently there have been claims that neither guanine nor G-containing mismatches affect the rate of electron migration over long distances in DNA.

The importance of complexes containing guanine has raised our interests in revealing their intrinsic molecular structures and has also attracted a plethora of theoreticians to study the structures of metal cation-mediated complexes and the electrostatic interactions governing their structures. Research on Ruthenium(III) (Ru$^{3+}$) containing complexes with regular base pairs and transversion mismatches, GG, AA, CC, and TT, was explored using density functional theory (DFT). It was illustrated that the GG mismatches with two different types of Ru$^{3+}$ complexes has the highest interaction energies compared to any other base pairs. It was also revealed that G-containing mismatches and the regular GC pair are relatively more stable than all the other base pairs both in the
absence of and when complexed with Ru$^{3+}$. G-quartets and mixed AGAG-quartets have also been studied in the presence of proton and metal cations. The interbase hydrogen bonds were found to be stabilized by two protons, and the [AGAG+2H]$^{2+}$ dication is considered as two protonated GA mismatches. The two protons in [AGAG+2H]$^{2+}$ are bound to N1 or N7 of adenine and overcome the repulsive interactions between the lone pairs on N7 of guanine and N1 or N7 of adenine. In addition, mismatches were found to be stabilized in acidic environment. It is crucial, therefore, to understand the influence of protonation on the structure of G-containing mismatches based on experimental evidence.

Vibrational spectroscopy, specifically infrared multiple photon dissociation (IRMPD), has been used to measure the vibrational spectra of various nucleobase adducts with cations, including the G-tetrad, uracil complexes and self-assemblies, i-motif structures, GC base pairs and many other biomolecules in the gas-phase. A recent study on alkali metal cationized 1-methylcytosine (1-mCyt) dimers revealed an asymmetric structure due to an interbase hydrogen bonding interaction to be a major contributor to the gas phase vibrational spectra of K$^+$, Rb$^+$ and Cs$^+$ complexes and likely even contributing to the Li$^+$ and Na$^+$ complexes. Also, the gas phase IRMPD spectra of the protonated GC base pairs were found to have either the Hoogsteen structures or the less thermodynamically stable Watson-Crick structure when electrospayed from solutions of pH 3.2 or 5.8, respectively.

In this paper, vibrational spectroscopy was employed to uncover the structures of protonated G-containing mismatches in gas-phase. Specifically, (9eG:9mA)H$^+$, (9eG:1mT)H$^+$, and (9eG:9eG)H$^+$ (where 9eG = 9-ethylguanine, 9mA = 9-methyladenine, 1mT = 1-methylthymine) were studied by IRMPD spectroscopy in the fingerprint region and computational methods. Methyl- and ethyl- groups on the nucleobases block the site that would be attached to the ribose/deoxyribose in nucleic acids.


2.1 Experimental Methods. The IRMPD spectroscopy experiments in the fingerprint region were performed at two different free electron laser (FEL) facilities coupled to ion trapping mass spectrometers. The (9eG:9mA)H$^+$ and (9eG:9eG)H$^+$ spectra were collected at the free electron laser for infrared experiments (FELIX) in the Netherlands which is coupled to a modified 3D quadrupole ion trap mass spectrometer (Bruker AmaZon Speed ETD). These ions were electrospayed from the solutions described below, which were first diluted 100-fold with acetonitrile. The trapped and isolated ions were irradiated with FEL radiation scanned at 3 cm$^{-1}$ intervals with 2 pulses of tunable infrared radiation at FELIX in the 900 – 1900 cm$^{-1}$ region. At the centre laser infrarouge d’Orsay (CLIO) a Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker Apex-Qe 7T) coupled
to the FEL was used to collect the spectrum for \((9eG:1mT)H^+\). These ions were isolated in the FTICR and irradiated with FEL radiation by for 250 ms at 5 cm\(^{-1}\) intervals in the 900 – 1900 cm\(^{-1}\) region.

All chemicals were purchased from Sigma-Aldrich and used without further purification. All solvents described below were a 50/50 mix of 18 M\(\Omega\)•cm (Millipore) water and methanol. Solutions for the \((9eG)\_2H^+\) were prepared by adding two drops of 1\% formic acid into 0.75 mmol \(L^{-1}\) 9-ethylguanine solution. The \((9eG:1mT)H^+\) complex was prepared by adding 3-4 drops 1\% fresh formic acid to a solution containing 0.25 mmol \(L^{-1}\) 9eG and 1 mmol \(L^{-1}\) 1mT. The \((9eG:1mA)H^+\) complex was prepared by adding 3-4 drops of 1\% fresh formic acid into a solution containing 0.25 mmol \(L^{-1}\) 9eG and 1 mmol \(L^{-1}\) 1mA. Several attempts to prepare \((9eG:1mU)H^+\) with different concentrations of 9eG, 1mU, and 1% formic but were unsuccessful.

The IRMPD efficiencies (intensities) are calculated as the negative of the logarithm of the product ion intensities divided by the sum of the total ion intensities.

### 2.2 Computational Methods

B3LYP density functional theory and 6-31+G(d,p) split-valence basis set has been used to reliably model and compare the thermochemistries of isomeric bioorganic and bioinorganic systems with hydrogen bonds successfully.\(^{42,46,57-60}\) Geometry optimizations and frequency calculations were performed using the Gaussian 09/16 suite of programs.\(^{61,62}\) All B3LYP calculations an empirical dispersion correction was included using Grimme’s D3 version with the original D3 damping function, B3LYPD3.\(^{63}\) The computed vibrational frequencies were corrected by a factor of 0.975 in the fingerprint region (900-1900 cm\(^{-1}\)),\(^{46,48,64}\) and convoluted with a Lorentzian profile with a full width at half max of 15 cm\(^{-1}\) to compare with the experimental IRMPD spectra. For comparison, single point energy calculations were performed using 6-311+G(3df, 3pd) on all the optimized structures of B3LYP/6-31+G(d,p) and are reported as B3LYPD3/6-311+G(3df,3pd)//6-31+G(d,p). M06-2X\(^{65,66}\) is also considered to perform well for the thermodynamic calculations of complexes containing non-covalent interactions. For comparison, M06-2X/6-31+G(d,p) was used to optimize the lowest energy structures and calculate the thermochemical parameters. The thermochemistries reported are all 298 K values and in kJ mol\(^{-1}\). Finally, CBS-QB3 was used to compute proton affinities. No attempt was made to correct for basis set superposition error in these calculations.

### 3. Results and Discussion

All the complexes are named giving the site of protonation first then the hydrogen bond interactions with the donor first and the acceptor second. For example, GN7\_GN7-TO4\_TN3-GO6 shows that the 9eG is protonated at N7 and there is the hydrogen bond from N7 of 9eG to O4 of 1mT and from N3 of 1mT to O6 of 9eG.
In all figures showing computed and experimental vibrational spectra, the bolded grey traces are experimental IRMPD spectra while the black traces are computed IR spectra for the indicated structure.

### 3.1 (9eG:1mT)H⁺

The experimental IRMPD spectrum of (9eG:1mT)H⁺ in the fingerprint region collected at CLIO is illustrated in Figure 1 and is compared with computed IR spectra of the five lowest energy isomers. The computed IR spectrum of the lowest energy structure, GN7_GN7-TO4_TN3-GO6, a Hoogsteen type structure, is clearly consistent with the experimental spectrum. The observed absorption band at 1772 cm⁻¹ is reproduced by the predicted free C=O₂ stretch of 1mT at 1757 cm⁻¹ and the weakly hydrogen bonded C=O₆ stretch of 9eG at 1740 cm⁻¹. The experimental absorption centred at 1641 cm⁻¹ can be attributed to three predicted vibrations at 1644, 1639, and 1666 cm⁻¹ which are dominated by NH₂ bending of 9eG, C₅=C₆ stretching of 1mT, as well as hydrogen bonded C=O₄ stretching, and HNC bending of 1mT. The computed TN3-GO6 and GN7-TO4 hydrogen bonds are predicted to be 1.90 and 1.48 Å, respectively. The 1593 cm⁻¹ shoulder is well reproduced by the calculated 1586 cm⁻¹ band which is composed of predominantly NH₂, HN1C2, and HN7C₈ bending in 9eG. The weaker absorption at 1328 cm⁻¹ is also very nicely reproduced by the calculations and consists of numerous complicated stretching and bending motions involving ring atoms of both bases.

The second-lowest energy isomer, TO4_TO4-GO6_TN3-GN7, is protonated at O₄ of 1mT despite the significantly higher proton affinity of 9eG than that of O₄ of 1mT, by some 103.9 kJ mol⁻¹. This isomer is only 6.8 kJ mol⁻¹ higher in Gibbs energy than the lowest energy isomer. The lower than expected relative energy of this complex, due to being 1mT protonated, is most likely due to the strong ion-dipole and ion-induced dipole interactions formed between 1mTH⁺ and 9eG. The main disagreement between the experimental IRMPD spectrum and the computed spectrum for this isomer is the hydrogen bonded C=O₆ stretch of 9eG predicted at 1687 cm⁻¹, but which is not observed experimentally. The hydrogen bond to this carbonyl oxygen is computed to be 1.49 Å.

The GN7_GN7-TO2_TN3-GO6 isomer differs from the lowest energy isomer by a flip of 1mT, and is computed to be 9.8 kJ mol⁻¹ higher in Gibbs energy. Spectroscopically, it cannot be ruled out due to a very similar predicted IRMPD spectrum to the lowest energy isomer. However, that it is higher in energy than the second lowest energy isomer which is not contributing to the experimental spectrum, it is unlikely that this one is contributing either. Two other higher energy structure are also shown in Figure 1 are Watson-Crick type structures and are not consistent with the experimental spectrum. Other higher energy structures with their computed spectra are presented in Figure S1 and compared with the (9eG:1mT)H⁺ experimental IRMPD spectrum.
3.2 (9eG:9mA)H⁺. In Figure 2, the IRMPD spectrum of (9eG:9mA) is compared to the four lowest energy structures. Interestingly, the four lowest energy structures are all 9mA protonated, despite the significantly larger proton affinity of 9eG (Table 1). The two lowest energy structures, AN1_AN1-GO6_AN6-GN7 and AN1_AN1-GN7_AN6-GO6 are complexes where the Hoogsteen face of 9eG is interacting with the WC face of 9mA; both are protonated at N1 of 9mA which is predicted to have a proton affinity almost 25 kJ mol⁻¹ lower than N7 of guanine. This type of structural oddity has been discussed before with respect to protonated complexes containing one high dipole moment monomer⁶⁷,⁶⁸ and the protonated 1-methylcytosine/9-ethylguanine complex, (9eG:1mC)H⁺.⁵⁵ In these species the proton is covalently bound to the lower proton affinity monomer and then energy deficit is made up by forming a strong ion-dipole and/or ion-induced dipole interaction. In the present case, 9eG has a dipole moment of 7.36 D, almost three times that of 9mA (Table 1). The 25 kJ mol⁻¹ deficit by protonating 9mA over 9eG is more than made up by the very strong ion-dipole interaction between 9mAH⁺ and 9eG. The larger polarizability of 9eG than 9mA (18.4 Å³ vs 15.7 Å³) means that a stronger ion-induced dipole interaction would favour the 9mAH⁺ structure over the 9eG protonated one. Unlike (9eG:1mC)H⁺ where the 9eG protonated complex was an optimized local minimum, in the present example with (9eG:9mA)H⁺ it was not possible to locate a local minimum similar to AN1_AN1-GN7_AN6-GO6 where 9eG was protonated at N7 instead of 9mA being protonated at N1. Optimization calculations on the GN7 protonated complex were repeated using B3LYP/6-311+G(3df,3pd) and M06-2X/6-31+G(d,p), but were unsuccessful, the proton shifted resulting in the AN1_AN1-GN7_AN6-GO6 structure.

The AN1_AN1-GO6_AN6-GN7 and AN1_AN1-GN7_AN6-GO6 structures differ by only 2.1 kJ mol⁻¹ in Gibbs energy and neither can be ruled out spectroscopically, both are consistent with the experimental spectrum. At the bottom of Figure 2, the weighted average based on the difference in Gibbs energy of these two lowest energy structures is compared with the experimental IRMPD spectrum for (9eG:9mA)H⁺ and better reproduces the spectrum. The most pronounced experimental features in the IRMPD spectrum are in the 1500-1800 cm⁻¹ region. The band at 1705 cm⁻¹ is well reproduced by 9eG C=O stretch combined with 9mA NH₂ bending. The band at 1630 cm⁻¹ is the NH₂ bend combined with C2-N2 stretching, both of 9eG. The 1580 cm⁻¹ feature belongs predominantly to 9eG C2-N3/N3-C4 stretching. This latter band is better resolved experimentally than is predicted by the calculations.

The next two structures shown in Figure 2 are 28.3 and 36.7 kJ mol⁻¹ higher in energy, and therefore are unlikely to be present in great abundance. Certainly, they cannot be ruled out as contributing to the experimental spectrum based on a comparison of their computed spectra. The lowest energy structure composed of canonical bases (ie. not tautomeric) with an N7 protonated 9eG,
GN7_GN1-AN1_GO6-AN6, is 54 kJ mol\(^{-1}\) higher in Gibbs energy than the lowest energy structure (Figure S2) and its computed IR spectrum is not consistent with the experimental spectrum. Other higher energy structures are compared to the experimental IRMPD spectrum in Figures S2 and S3; none of their computed IR spectra reproduce the observed spectrum.

The IRMPD dissociation mass spectra for (9eG:1mT)H\(^+\), (9eG:9mA)H\(^+\), and (9eG:9eG)H\(^+\) are presented in Figure 3. In all cases, the main dissociation product is protonated 9eG at m/z 180 which is expected based on the proton affinities and gas phase basicities of the bases (see computed values in Table 1), with N7 of 9eG having the highest proton affinity. Interestingly, however, for (9eG:9mA)H\(^+\) there is a small, but not insignificant amount of protonated 9mA. Based on the difference in gas basicities, for the two bases, the intensity of protonated 9mA should be no more than \(2 \times 10^{-5}\) compared to that of protonated 9eG. Experimentally, the ratio is 0.02:1. This is similar to the anomaly reported in previous work for the dissociation of protonated guanine-cytosine base pairs where protonated cytosine was observed in a significantly higher abundance than expected.\(^{54,55,69-71}\) For (9eG:1mC)H\(^+\), the observed 1mCH\(^+\):9eGH\(^+\) ratio was 0.3:1 despite an expected ratio of only 0.006:1 depending on the computed gas basicities. This was explained by the dynamics of dissociation of the energized (9eG:1mC)H\(^+\) system. (9eG:1mC)H\(^+\), is protonated at N3 of 1mC with a low-lying energy barrier for proton transfer to 9eG. The surface is very shallow around the minimum energy structure in both the dissociation and proton transfer degrees of freedom. The proton transfer energy barrier, however, grows significantly as distance between the two bases increase during dissociation. In order to transfer the proton from 1mC to 9eG, the energized complex must adopt a configuration where the distance between the two bases is similar to that of the minimum energy structure to make proton transfer energetically feasible. It is expected that a similar dynamics problem is at play in the present system such that much more 9mAH\(^+\) is observed in the dissociation of (9eG:1mA)H\(^+\) than expected based on gas basicities.

3.3 (9eG:9eG)H\(^+\). The IRMPD spectra for the (9eG:9eG)H\(^+\) complex and the computed IR spectra for the 5 lowest energy structures are compared in Figure 4. The lowest energy structure for (9eG:9eG)H\(^+\) (GN7_GN7-GN7_GC8-GO6) is one where the Hoogsteen faces of both 9eG are interacting but with only one classical hydrogen bond. It is protonated at N7 of one 9eG with a hydrogen bond to the N7 of the other 9eG and a hydrogen bonding-type interaction between C8 of the protonated 9eG and the O6 of the other. The computed spectrum for this structure is in excellent agreement with the experimental IRMPD spectrum. The predicted free C=O6 stretch of the protonated 9eG at 1751 cm\(^{-1}\) and the red-shifted shoulder predicted at 1718 cm\(^{-1}\) due to the C=O6 stretch of the neutral 9eG interacting with the C8-H is in excellent agreement with the experimental band at 1753 and the unresolved shoulder to the red of that band. The weak interaction, with a C-H—O=C bond distance of
2.25 is enough to slightly shift the C=O stretch to slightly lower energy. The strong absorptions at 1628 and 1579 cm\(^{-1}\) agree well with the NH\(_2\) bending absorptions predicted to occur at 1635, 1590, and 1584 cm\(^{-1}\). The observed bands at lower energy also agree well with the positions of predicted bands for modes involving ring stretches. This is the first reported dimeric base pair containing only a single classical hydrogen bond but not follow the regular Watson-Crick, Hoogsteen, or Wobble base pair principle, with the confident conclusion supported by both thermodynamic and spectroscopic results.

The second lowest energy structure, GO6_GO6-GN7_GN1/GN2-GO6, has O6 protonated 9eG interacting via its WC face with the Hoogsteen face of neutral 9eG and is computed to be higher in Gibbs energy by 7.3 kJ mol\(^{-1}\), and clearly its calculated IR spectrum is not in good agreement with the experimental spectrum. Foremost, there is the absence of a free C=O in the computed structure, instead having C=O stretches computed to be at 1709 and 1687 cm\(^{-1}\) due to hydrogen bonding and protonation. The third lowest energy structure, GN7_GN1-GN7_GN2-GO6, is also one where N7 protonated 9eG interacts via its WC face to the Hoogsteen face of neutral 9eG and cannot be completely ruled out by spectroscopic means, but is 12.2 kJ mol\(^{-1}\) higher in Gibbs energy than the lowest energy structure. Its computed infrared spectrum is nearly as good a match to the experimental spectrum. The next two higher energy structures, the highest containing an N7-protonated N1 to O6 tautomer, are clearly not a good match to the experimental spectrum. Other structures are shown in Figure S4-S5. None of these higher energy conformers can reproduce the experimental IRMPD spectrum.

### 3.4 Computed Energies Comparison

Relative thermochemistries computed using B3LYPD3/6-31+G(d,p), M06-2X/6-31+G(d,p), and B3LYPD3/6-311+G(3df,3pd)//6-31+G(d,p) are compared in Tables S1, S2 and S3 for the lowest energy structures of (9eG:1mT)H\(^+\), (9eG:9mA)H\(^+\), and (9eG:9eG)H\(^+\), respectively. All computed energies provide a similar picture as those using B3LYP/6-31+G(d,p). One notable exception is in the TO4_TO6-GO4_TN3-GN7 structure for the (9eG:1mT)H\(^+\) complex which is nearly isoenergetic with the lowest energy GN7_GN7-TO4_TN3-GO6 structure using M06-2X/6-31+G(d,p). However, TO4_TO6-GO4_TN3-GN7 is ruled out spectroscopically as seen in Figure 1.

### 4. Conclusions

An IRMPD spectroscopic and computational study was undertaken to probe the structures of protonated guanine-containing mismatch complexes, specifically of (9eG:1mT)H\(^+\), (9eG:9mA)H\(^+\) and (9eG:9eG)H\(^+\). For (9eG:1mT)H\(^+\) and (9eG:9eG)H\(^+\), the lowest energy structures were sufficient to explain the IRMPD spectrum, while for (9eG:9mA)H\(^+\) the two lowest energy structures were only 2.1 kJ mol\(^{-1}\) apart in Gibbs energy and their weighted average is consistent with the IRMPD spectrum.
While the DNA bases have pK_a values which indicate that they are unlikely to be protonated, at physiological pH, it has also been determined that their chemical environments in nucleic acids—and potentially through interactions with metal cations—shift their pK_a’s to the physiological regime\textsuperscript{72,73} raising the interest toward understanding protonation of the nucleobases in nucleic acids.\textsuperscript{74–79} It would be interesting to compare our model protonated mismatch complexes of guanine to those known to occur in nucleic acids. Atomic coordinates for proteins and RNA contained within the protein data base, PDB, are from X-Ray crystallography measurements which do not all contain positions of hydrogens or protons.\textsuperscript{75} However, using computational chemistry and comparing with searches from the PDB potential structures for protonated base pairs have been identified in RNA.\textsuperscript{75} In a set of 19 base pairs where protonation is thought to occur,\textsuperscript{75} two of them are N1 protonated adenine bound to guanine, identical to the two structures found to be the lowest-energy structures in this study and to be consistent with the experimental vibrational spectrum. This is the first infrared spectroscopic evidence for the occurrence of a self-assembled protonated base pair, observed in the gas phase which has also been proposed to exist in cells.

While two protonated guanine dimers were also identified in the PDB, they were bound via the Hoogsteen face of neutral guanine to the sugar face of N3 protonated guanine, unlike that found in the present study. However, it is worth mentioning that the crystal structure of 9-ethylguanine hemihydrochloric acid\textsuperscript{80,81} was found to exist as protonated dimers such as that found here in the gas phase, albeit without the C-H---O=C interaction that the present spectroscopic evidence shows exists at least in the gas phase.

Only one protonated guanine (N7) complex with uracil has been identified\textsuperscript{74} but the hydrogen bond is between the N2 of guanine and O4 or uracil, very different than the structure concluded to exist for \((9eG:1mT)H^+\). Given that the \((9eG:9mA)H^+\) identified in this study and the \((9eG:1mC)H^+\) identified in a recent publication from the protein data bank, it would be interesting to see if \((9eG:1mT)H^+\) and \((9eG:9eG)H^+\) complexes like the ones found here will turn up in RNA.

In a previous study spectroscopic study from this group which included the gaseous protonated complex of 9-ethylguanine and 1-methylcytosine, \((9eG:1mC)H^+\), the lowest energy structure which was also consistent with the experimental vibrational spectrum was the Hoogsteen complex with the cytosine protonated at N3.\textsuperscript{55} A protonated complex with the same structure, between guanine and cytosine, has also been proposed to occur in RNA in the protein data base.\textsuperscript{75}
Acknowledgements

We are grateful for the excellent support from the FELIX and CLIO technical staff. This research was enabled in part by support provided by ACENET, Compute Ontario, Westgrid, and Compute Canada. Financial support from NSERC, Bruker, CFI, and Memorial University is greatly appreciated.
Figure Captions

**Figure 1.** Experimental IRMPD spectra (grey traces) of the protonated 9-ethylguanine/1-methylthymine complex compared to the B3LYPD3/6-31+G(d,p) computed IR spectra (black traces) of the lowest energy complexes. Relative enthalpies and Gibbs energies are 298 K values in kJ mol\(^{-1}\).

**Figure 2.** Experimental IRMPD spectra (grey traces) of the protonated 9-ethylguanine/9-methyladenine complex compared to the B3LYPD3/6-31+G(d,p) computed IR spectra (black traces) of the lowest energy complexes. Relative enthalpies and Gibbs energies are 298 K values in kJ mol\(^{-1}\).

**Figure 3.** IRMPD mass spectra showing fragmentation products for each of the three protonated 9-ethylguanine containing complexes.

**Figure 4.** Experimental IRMPD spectra (grey traces) of the protonated dimer complex of 9-ethylguanine compared to the B3LYPD3/6-31+G(d,p) computed IR spectra (black traces) of the lowest energy complexes. Relative enthalpies and Gibbs energies are 298 K values in kJ mol\(^{-1}\).
References

Table 1. 298 K proton affinities (PA) and gas basicities (GB) for 9-ethylguanine, 9-methyladenine, and 1-methylthymine. Protonation site is indicated in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>PA / kJ mol⁻¹</th>
<th>GB / kJ mol⁻¹</th>
<th>Dipole moment / D</th>
<th>Polarizability / Å³</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-ethylguanine (N7)</td>
<td>974.3</td>
<td>943.3</td>
<td>7.36</td>
<td>18.4</td>
</tr>
<tr>
<td>9-methyladenine (N1)</td>
<td>949.4</td>
<td>916.1</td>
<td>2.67</td>
<td>15.7</td>
</tr>
<tr>
<td>9-methyladenine (N7)</td>
<td>924.1</td>
<td>893.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-methylthymine (O4)</td>
<td>870.4</td>
<td>847.3</td>
<td>4.97</td>
<td>13.7</td>
</tr>
<tr>
<td>1-methylthymine (O2)</td>
<td>852.2</td>
<td>832.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a: CBS-QB3

b: B3LYP/6-31+G(d,p)
G:C

A:U(T)

254x190mm (96 x 96 DPI)
Some other higher energy structures of \((9eG:1mT)H^+\) along with their enthalpies and 298 K Gibbs energies relative to \(GN7\_GN7\_TO4\_TN3\_GO6\) and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in kJ mol\(^{-1}\).
Some other higher energy structures of (9eG:9mA)H⁺ along with their enthalpies and 298 K Gibbs energies relative to \textbf{AN1_AN1-G06_AN6-GN7} and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in kJ mol⁻¹.
Some other higher energy structures of \((9eG:9mA)H^+\) along with their enthalpies and 298 K Gibbs energies relative to \(AN1_AN1–GO6_AN6–GN7\) and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in kJ mol\(^{-1}\).
Some other higher energy structures of \((9\text{eG}:9\text{eG})H^+\) along with their enthalpies and 298 K Gibbs energies relative to \(\text{GN7}_\text{GN7-3GO6-9GO6}\) and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in kJ mol\(^{-1}\).
Some other higher energy structures of (9eG:9eG)H⁺ along with their enthalpies and 298 K Gibbs energies relative to GN7_GN7-GN7_GC8-GO6 and computed IR spectra (black trace) compared to the experimental IRMPD spectrum (grey trace). All energies are presented in kJ mol⁻¹.
Table S1 298K relative enthalpies and Gibbs energies, in kJ mol⁻¹, of five lowest energy isomers of (9eG:1mT)H⁺ by three different computational methods. Relative Gibbs energies are indicated in parentheses.

<table>
<thead>
<tr>
<th>Structures of (9eG:1mT)H⁺</th>
<th>B3LYP/6-31+G(d,p)</th>
<th>B3LYP/6-31+G(3df,3pd) //B3LYP/6-31+G(d,p)</th>
<th>M06-2X/6-31+G(d,p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN7_GN7-G1-GO6-TO4_TN3-GO6</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>TO4_TO4-GO6_TN3-GN7</td>
<td>2.6 (6.8)</td>
<td>3.4 (7.6)</td>
<td>-3.7 (0.2)</td>
</tr>
<tr>
<td>GN7_GN7-G2_TN3-GO6</td>
<td>10.7 (9.8)</td>
<td>10.2 (9.3)</td>
<td>9.9 (9.2)</td>
</tr>
<tr>
<td>GN7_GN1-GO6_TN3-GO6</td>
<td>11.6 (13.0)</td>
<td>11.5 (12.9)</td>
<td>13.6 (14.7)</td>
</tr>
<tr>
<td>GN7_GN2-G1_GN1-GO6</td>
<td>7.0 (16.5)</td>
<td>7.9 (17.4)</td>
<td>4.8 (13.2)</td>
</tr>
</tbody>
</table>

Table S2 298K relative enthalpies and Gibbs energies, in kJ mol⁻¹, of four lowest energy isomers of (9eG:1mA)H⁺ by three different computational methods. Relative Gibbs energies are indicated in parentheses.

<table>
<thead>
<tr>
<th>Structures of (9eG:1mA)H⁺</th>
<th>B3LYP/6-31+G(d,p)</th>
<th>B3LYP/6-31+G(3df,3pd) //B3LYP/6-31+G(d,p)</th>
<th>M06-2X/6-31+G(d,p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN1_AN1-GO6_AN6-GN7</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>AN1_AN1-GN7_AN6-GO6</td>
<td>2.1 (4.1)</td>
<td>3.8 (1.8)</td>
<td>4.5 (2.6)</td>
</tr>
<tr>
<td>AN7_AN7-GO6_AN6-GN7</td>
<td>28.3 (29.3)</td>
<td>28.8 (27.8)</td>
<td>30.7 (31.7)</td>
</tr>
<tr>
<td>AN7_AN7-GN7_AN6-GO6</td>
<td>35.6 (36.7)</td>
<td>34.9 (36.0)</td>
<td>33.2 (35.5)</td>
</tr>
</tbody>
</table>

Table S3 298K relative enthalpies and Gibbs energies, in kJ mol⁻¹, of five lowest energy isomers of (9eG:9eG)H⁺ by three different computational methods. Relative Gibbs energies are indicated in parentheses.

<table>
<thead>
<tr>
<th>Structures of (9eG:9eG)H⁺</th>
<th>B3LYP/6-31+G(d,p)</th>
<th>B3LYP/6-31+G(3df,3pd) //B3LYP/6-31+G(d,p)</th>
<th>M06-2X/6-31+G(d,p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN7_GN7-GN7_GC8-GO6</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>GO6_GO6-</td>
<td>4.1 (7.3)</td>
<td>3.2 (6.4)</td>
<td>-0.2 (3.9)</td>
</tr>
<tr>
<td>GN7_GN1-GN7_GN2-GO6</td>
<td>7.1 (12.2)</td>
<td>6.3 (11.5)</td>
<td>10.7 (15.7)</td>
</tr>
<tr>
<td>GN7_GN7-GO6_GC8-GN7</td>
<td>15.9 (14.4)</td>
<td>14.4 (12.8)</td>
<td>15.2 (9.8)</td>
</tr>
<tr>
<td>GN7_GN7-GO6_GO6-GN7</td>
<td>19.5 (21.6)</td>
<td>17.0 (19.0)</td>
<td>16.7 (15.0)</td>
</tr>
</tbody>
</table>