PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

submitted to Phys.Chem.Chem.Phys.

Diverse Mixtures of 2,4-Dihydroxy Tautomers and O4 Protonated Conformers of Uridine and 2'-Deoxyuridine Coexist in the Gas Phase

R. R. Wu,[†] Bo Yang,[†] C. E. Frieler,[†] G. Berden,[‡] J. Oomens,[‡§] and M. T. Rodgers[†,*]

[†]*Department of Chemistry, Wayne State University, Detroit, Michigan 48202*
[‡]*Radboud University Nijmegen, Institute for Molecules and Materials, FELIX Facility, Toernooiveld 7, 6525 ED, Nijmegen, The Netherlands,*
[§]*van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands*

Abstract

The gas-phase conformations of protonated uridine, $[Urd+H]^+$, and its modified form, protonated 2'-deoxyuridine, $[dUrd+H]^+$, generated by electrospray ionization are investigated using infrared multiple photon dissociation (IRMPD) action spectroscopy techniques. IRMPD action spectra of $[Urd+H]^+$ and $[dUrd+H]^+$ are measured over the IR fingerprint and hydrogen-stretching regions. $[Urd+H]^+$ and $[dUrd+H]^+$ exhibit very similar IRMPD spectral profiles. However, the IRMPD yields of $[Urd+H]^+$ exceed those of $[dUrd+H]^+$ in both the IR fingerprint and hydrogen-stretching regions. The measured spectra are compared to the linear IR spectra predicted for the stable low-energy structures of these species computed at the B3LYP/6-311+G(d,p) level of theory to determine the tautomeric conformations populated by electrospray ionization. Both B3LYP and MP2 methods find O4 and O2 protonated canonical as well as 2,4-dihydroxy tautomers among the stable low-energy structures of $[Urd+H]^+$ and $[dUrd+H]^+$. Comparison between the measured IRMPD and calculated linear IR spectra suggests that these species exist in their ring-closed forms and that both 2,4-dihydroxy tautomers as well as O4 protonated canonical conformers coexist in the population generated by electrospray ionization for both $[Urd+H]^+$ and $[dUrd+H]^+$. The 2'-deoxy modification of $[dUrd+H]^+$ reduces the variety of 2,4-dihydroxy tautomers populated in the experiments vs those of $[Urd+H]^+$.

Keywords: 2'-deoxyuridine, electronic structure theory, IRMPD action spectroscopy, protonation, tautomers, uridine.

Physical Chemistry Chemical Physics Accepted Manuscript

Introduction

The conformations of DNA and RNA nucleosides as the constituent building blocks of nucleic acids are important in determining their overall structures. In particular, the nucleobase orientation relative to the glycosidic bond (anti and syn) and the sugar puckering (typically C2'-endo and C3'-endo) of the nucleosides are the primary factors that determine the overall geometry. The nucleobases of the most commonly occurring B-form of double-stranded DNA adopt the anti orientation and the sugars exhibit C2'-endo puckering. A change in the sugar puckering from C2'-endo to C3'-endo leads to the A-form of double-stranded DNA.[1] In addition to the nucleobase orientation and sugar configuration, hydrogen-bonding interactions are also significant determinants of the overall structures of nucleic acids. Since the discovery of Watson-Crick base pairs,[2] the important role that hydrogen-bonding interactions play in preserving genetic information and controlling biochemical function of nucleic acids has been well recognized.[3,4] The reason that the hydrogen-bonding network prevails in nucleic acids is the specific hydrogen-bond donors and acceptors that the nucleobases possess. Protonation has been found to greatly influence hydrogen-bonding interactions within nucleic acids, which in turn alters their structures and biochemical properties.[5–8] Therefore, protonated DNA and RNA structures have been studied extensively.[9–19] Although still rather limited, structural characterization of the protonated forms of the canonical as well as modified DNA and RNA nucleosides is growing.[20–28] Knowledge of the structural features of the protonated DNA and RNA nucleosides as the constituents of nucleic acids can be used to enhance our comprehension of larger and more complex DNA and RNA structures and how they are influenced by the local environment, in this case, the local pH.

Uridine (Urd) is one of the pyrimidine nucleosides of RNA. Uracil, the nucleobase of uridine, is of great interest because it possesses two keto moieties that may be susceptible to tautomerization depending upon the local environment. Tautomerization of uracil occurs readily upon protonation, resulting in the 2,4-dihydroxy tautomer.[29–33] Because of this unique character of protonated uracil, a wide variety of conformations should be accessible to protonated uridine,

Physical Chemistry Chemical Physics Accepted Manuscript

[Urd+H]$^+$. Therefore, it is of great interest to comprehensively examine the conformational space of [Urd+H]$^+$, and to determine the influence of the sugar moiety on the tautomeric conformation(s) of the protonated uracil moiety. 2'-Deoxyuridne (dUrd), a modified form of Urd, is the DNA analogue of Urd, but is not one of the canonical DNA nucleosides. dUrd is highly susceptible to modification at the 5-position such that minor forms of dUrd with halogen, ethyl, methylthio, hydroxymethyl and allyl substituents are known, and these dUrd variants are used as antiviral drugs.[34–36] Tabet and coworkers[37] measured the proton affinities (PAs) of Urd and dUrd using the kinetic method[38] and complemented these measurements with theoretical calculations. They found that the PA of dUrd is ~10 kJ/mol greater than the PA of Urd, and that the sugar configuration exerts only a minor effect on the PA (~2 kJ/mol). In a theoretical study, Hoffmann and coworkers[39] examined the orientation of uracil relative to the glycosidic bond in 5'-deoxy- and 5'-O-methyl-uridine-2',3'-cyclic monophosphate in both gas phase and aqueous environments. Their results suggest minor energy differences between the syn and anti conformers for both nucleotides in the gas phase. In contrast, they found that the syn conformers of both nucleotides are ~60 kJ/mol more stable in aqueous solution than their anti counterparts. Most recently, Alonso and coworkers characterized the structure of neutral Urd in the gas phase by Fourier transform microwave spectroscopy and theoretical calculations. They found that the nucleobase in an anti orientation, the sugar is C2′-endo puckered, and that O3′H···O2′H···O2 dual and O5′···HC2 and O2′H···O5′ noncanonical hydrogen-bonding interactions play an important role in stabilizing neutral Urd in the gas phase.[40]

In this work, infrared multiple photon dissociation (IRMPD) action spectroscopy techniques and synergistic electronic structure calculations are used to characterize the conformations and relative stabilities of [Urd+H]$^+$ and [dUrd+H]$^+$. The IRMPD spectra were measured using a Fourier transform ion cyclotron resonance mass spectrometer coupled to the FELIX free electron laser or an OPO laser, whereas linear IR spectra are computed using electronic structure methods. Comparison between the measured IRMPD and calculated IR spectra allows the resonant vibrational modes, preferred sites of protonation, and the low-energy

Physical Chemistry Chemical Physics Accepted Manuscript

tautomeric conformations that are populated by electrospray ionization (ESI) to be determined. The differences in the conformations and the measured IRMPD spectra for $[Urd+H]^+$ vs $[dUrd+H]^+$ reveal the effects of the 2'-deoxy modification. Comparison to IRMPD studies reported previously for protonated uracil[32] provides insight into the impact of the sugar moieties on the tautomeric conformations of protonated uracil and the effects of hydrogen-bonding interactions with the sugar moieties on the IR signatures of the nucleobase. Comparison to Fourier transform microwave spectroscopy studies reported previously for neutral uridine[40] enables the effects of protonation on the gas-phase conformation to be elucidated.

Methods

IRMPD Spectroscopy. IRMPD action spectra of $[Urd+H]^+$ and $[dUrd+H]^+$ were measured using a 4.7 T Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS)[41–43] coupled to the FELIX free electron laser[44] or an OPO/OPA laser system. 1 mM of Urd or dUrd, purchased from Sigma-Aldrich, and 10 mM hydrochloric acid were dissolved in 50%:50% methanol-water solutions. The nucleobase solution was delivered to a Micromass "Z-spray" ESI source at a flow rate in the range of 3.0–8.8 μL/min. $[Urd+H]^+$ or $[dUrd+H]^+$ were accumulated in an rf hexapole ion trap for several seconds to bring about thermalization of the trapped ion population before being pulsed-extracted into a quadrupole deflector, and injected into the FT-ICR MS via a 1-m long rf octopole ion guide. Electrostatic switching of the dc bias of the octopole (in the absence of a pulse of buffer gas) was employed to avoid collisional heating of the ions and to ensure transfer of the approximately room temperature distribution of ions generated by the ESI source to the ICR cell.[42] The $[Urd+H]^+$ or $[dUrd+H]^+$ ions were isolated using stored waveform inverse Fourier transform (SWIFT) techniques. The ions were irradiated for 2.5−3 s by the FEL over the IR fingerprint region between ~550 and ~1920 cm^{-1} or for 5−10 s by the OPO/OPA laser system over the range between ~2800 and ~3800 cm^{-1} to characterize the photodissociation behavior of the protonated nucleosides and to generate the IRMPD action spectra. However, the very low intensities and extreme broadening of the features

Physical Chemistry Chemical Physics Accepted Manuscript

observed in the range between ~2800 and ~3300 cm^{-1} makes this region of poor diagnostic quality and thus is not shown in the figure or discussed further.

Electronic Structure Calculations. Figure 1 shows the chemical structures of neutral Urd and dUrd. In both structures, the uracil moiety is depicted in the anti orientation relative to the glycosidic bond. The most favorable protonation sites for Urd and dUrd were examined including O2 and O4 as well as the 2,4-dihydroxy protonated tautomers. In addition, the sugar moieties of these species were examined in both the ring-closed as well as ring-opened forms. Candidate structures for each protonated form of Urd and dUrd were generated by simulated annealing using HyperChem software[45] with the Amber 2 force field and the same protocol employed in our previous IRMPD studies of the protonated forms of the guanine, adenine and cytosine DNA and RNA nucleosides.[22–24] Structures chosen for higher level optimization were based primarily on the relative stabilities of the candidate structures found via simulated annealing, where the 20-30 most stable structures found for each protonated form were subject to higher level optimization. However, to ensure that conformational space was comprehensively examined, additional structures that include all possible combinations of protonation site (O2, O4 and the 2,4-dihydroxy tautomers), nucleobase orientation (anti and syn), and sugar puckering (C2′-endo, C3′-endo, C2′-exo and C3′-exo) as well as ring-opened forms were also subjected to higher level optimization. Geometry optimizations, frequency analyses, and single point energy calculations of all candidate structures were performed using the Gaussian 09 suite of programs.[46] To facilitate convergence of the geometry optimization, all candidate structures were first optimized at the B3LYP/6-31G(d) level of theory. The stable structures thus determined were then re-optimized using the 6-311+G(d,p) basis set to improve the description of the intramolecular hydrogen-bonding interactions that stabilize these systems. The 6-311+G(d,p) basis set was also used to perform frequency analyses of the optimized structures. This basis set has been found to be reliable for geometry optimization as well as describing the frequencies of the protonated adenine, guanine and cytosine DNA and RNA nucleosides.[22–24] Single point energies were calculated at the B3LYP and MP2 levels of theory using the 6-311+G(2d,2p)

Physical Chemistry Chemical Physics Accepted Manuscript

extended basis set to more accurately predict the relative stabilities of the low-energy conformers. Zero-point energy (ZPE) and thermal corrections to 298 K based on the vibrational frequencies calculated at the B3LYP/6-311+G(d,p) level of theory were included in the relative stabilities determined.

Linear IR spectra based on these vibrational frequencies, scaled by a factor of 0.98 and broadened using a 20 cm^{-1} fwhm Gaussian line shape over the IR fingerprint region, and scaled by a factor of 0.954 with 15 cm^{-1} broadening for the hydrogen-stretching region, along with the computed IR intensities were generated for all stable conformations computed. When interpreting the IRMPD spectra by comparing with the calculated IR spectra, one should keep in mind that the measured nonlinear IRMPD and calculated linear IR spectra are not identical because multiple photons are involved in the IRMPD processes, leading to varying degrees of anharmonicity from various vibrational modes. Therefore, there are inevitable discrepancies between the measured IRMPD and calculated IR spectra, such as shifts in the band positions, broadening, and changes in the relative intensities of IR bands due to variable anharmonicities and efficiency of intramolecular vibrational redistribution (IVR) for the various modes of the system . However, these discrepancies are generally sufficiently limited that the IRMPD spectrum still provides a good reflection of the linear IR absorption spectrum.[47]

Results

IRMPD Action Spectroscopy. When photodissociation is induced by the FEL, the primary dissociation pathway for $[Urd+H]^+$ and $[dUrd+H]^+$ involves N-glycosidic bond cleavage, producing protonated uracil as the ionic product detected. A very minor product at m/z = 193, resulting from loss of two H_2O molecules from $[dUrd+H]^+$ is also observed. When the OPO laser is used, protonated uracil is the only ionic product observed for these protonated nucleosides. The observation of the minor dual water loss pathway with the FEL is the result of its higher photon flux, which leads to greater excitation before dissociation occurs. The IRMPD yield was determined for each protonated nucleoside, $[Nuo+H]^+$ = $[Urd+H]^+$ or $[dUrd+H]^+$, from its

Physical Chemistry Chemical Physics Accepted Manuscript

intensity, and the sum of the intensities of the product ions, after laser irradiation at each vibrational frequency as shown in eq (1):

$$\text{IRMPD Yield} = \sum_i I_{\text{product}_i} \Big/ \left(\sum_i I_{\text{product}_i} + I_{[\text{Nuo+H}]^+} \right) \quad (1)$$

The IRMPD yield was normalized linearly to correct for changes in the output of the laser as a function of the frequency of the FEL or OPO lasers. IRMPD action spectra were obtained for $[\text{Urd+H}]^+$ and $[\text{dUrd+H}]^+$ over the range of vibrational frequencies extending from ~550 to 1920 cm^{-1} and ~2800 to 3800 cm^{-1} and are compared in Figure 2. The IRMPD spectra profiles of $[\text{Urd+H}]^+$ and $[\text{dUrd+H}]^+$ are very parallel in both regions. Small shifts in the band positions are observed, however, the influence of the 2'-deoxy modification on the spectral features is primarily seen in the relative intensities of the measured bands. The IRMPD yield of $[\text{Urd+H}]^+$ exceeds that of $[\text{dUrd+H}]^+$ in both spectral regions. Differences in the IRMPD yields of the protonated forms of the adenine, cytosine and guanine DNA vs. RNA nucleosides were also observed.[22–24] However, the IRMPD yields of these protonated RNA nucleosides exceed those of their DNA analogues in the IR fingerprint region, whereas the yields of the DNA nucleosides exceed their RNA analogues in the hydrogen-stretching region.

Theoretical Results. Table 1 lists the relative enthalpies and Gibbs free energies of the stable ring-closed O2, O4 and 2,4-dihydroxy protonated low-energy conformers of $[\text{Urd+H}]^+$ and $[\text{dUrd+H}]^+$ found after B3LYP/6-31+G(d,p) optimization. Values for the ring-open forms are listed in Table S1 of the Electronic Supplementary Information. The nomenclature used to describe the stable ring-closed low-energy conformers is based on the protonation site/tautomeric conformation (**T** for the 2,4-dihydroxy tautomers, **O4** or **O2)** and is followed by a capital letter, lowercase letter, or lowercase Roman numeral for each conformer. Capital letters are used for conformers that are found for both $[\text{Urd+H}]^+$ and $[\text{dUrd+H}]^+$, lowercase Roman numerals are used for conformers that are only found for $[\text{Urd+H}]^+$, i.e., **Ti**, **O4i**, **Tii**, etc., whereas lowercase letters are used for conformers that are only found for $[\text{dUrd+H}]^+$, i.e., **O2a** and **O2b**. Conformations analogous to the **O2a** and **O2b** conformers are not found for $[\text{Urd+H}]^+$ because

Physical Chemistry Chemical Physics Accepted Manuscript

the 2'-hydroxyl substituent interacts with the protonated O2 atom of the nucleobase, and leads to the **O2i** conformer instead. The ordering of the low-energy conformers is based on the B3LYP 298 K relative Gibbs free energies of $[Urd+H]^+$; parallel conformations for $[dUrd+H]^+$ are named in the same fashion even when the stability order differs. Because the 2'-deoxy modification of dUrd decreases the opportunities for hydrogen-bonding interactions with the nucleobase, and multiple favorable rotational orientations of the 2'- and 3'-hydroxyl substituents of Urd exist, fewer low-energy conformers are found for $[dUrd+H]^+$ than $[Urd+H]^+$. The optimized structures of all of the stable ring-closed low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ computed along with their relative free energies at 298 K are shown in Figures S1 and S2 of the Supporting Information. The same nomenclature is also used to describe the stable ring-open low-energy conformers except that their names are appended by an underscore RO4′ or RO1′ to indicate that upon ring opening, the ring oxygen is bound to the 4′ or 1′ carbon, respectively. Optimized structures and relative Gibbs free energies at 298 K for the stable ring-open low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ are shown in Figure S3.

In spite of the diversity of candidate structures subjected to higher level optimization, only a limited number of stable structures were found as many of the initial structures converged to the same optimized structures and virtually all exhibit C2′-endo or C3′-endo puckering. C2′-exo and C3′-exo conformations found via simulated annealing are generally found to lie much higher in energy (> 100 kJ/mol than the most stable C2′-endo or C3′-endo conformers) and generally convert to their more stable C2′-endo or C3′-endo counterparts upon higher level geometry optimization. However, in the case of $[Urd+H]^+$ two stable low-energy C3′-exo conformations are found, **O4ii** and **O4v**, that are stabilized by O3′H···O2′H···O2 dual hydrogen-bonding interactions. A detailed description of the low-energy tautomeric conformers shown in Figures S1 - S3 is given in the Electronic Supplementary Information. The calculated Cartesian coordinates of the low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ shown in Figures S1-S3 are provided in Tables S2 – S5, whereas key geometrical parameters of these structures are provided in Tables S6 - S7 of the Electronic Supplementary Information.

Physical Chemistry Chemical Physics Accepted Manuscript

The most stable ring-closed conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ calculated using the B3LYP/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) and MP2(full)/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) computational protocols are displayed in Figure 1. B3LYP and MP2 predict the same ground-state structure for the ring-closed form of $[Urd+H]^+$, where protonation results in stabilization of the 2,4-dihydroxy tautomer of the uracil moiety, which takes on an approximately anti orientation with the 4-hydroxyl hydrogen atom directed toward and the 2-hydroxyl hydrogen atom oriented away from the adjacent N3 atom to form coupled O2H···O2'H···O3' hydrogen-bonding interactions that induce C2'-endo puckering of the ribose moiety. The presence of the 2'-hydroxyl substituent is clearly necessary for the coupled hydrogen-bonding interactions to occur for $[Urd+H]^+$, such that the analogous conformer does not exist for $[dUrd+H]^+$. The 2,4-dihydroxy tautomer of uracil, also in an anti orientation and with C2'-endo sugar puckering is predicted to be the ground-state conformer for $[dUrd+H]^+$ at the MP2 level of theory. The absence of the coupled hydrogen-bonding interactions makes it more favorable for the uracil moiety to rotate to a full anti orientation and form a weak noncanonical C6H···O5' hydrogen-bonding interaction between the uracil and 2-deoxyribose moieties. B3LYP suggests that this conformer lies 0.7 kJ/mol higher in free energy than the O4 protonated conformer where the nucleobase is in the anti orientation and the sugar exhibits C3'-endo puckering. In contrast, MP2 suggests that the B3LYP O4 protonated ground-state conformer is 8.6 kJ/mol less stable than the most stable 2,4-dihydroxy tautomeric conformation.

2,4-Dihydroxy Tautomers. The most stable 2,4-dihydroxy tautomer of $[Urd+H]^+$, **Ti**, has the nucleobase in an approximately anti orientation, and is preferentially stabilized by flipping the orientation of the 2-hydroxyl hydrogen atom toward the sugar to enable coupled O2H···O2'H···O3' hydrogen-bonding interactions and C2'-endo sugar puckering. In contrast, the most stable 2,4-dihydroxy tautomer of $[dUrd+H]^+$, **TA**, takes on a fully anti orientation of the uracil residue, and C2'-endo puckering of the sugar moiety, but instead has both the 2- and 4-hydroxyl hydrogen atoms pointed toward N3, similar to that found in the ground-state conformation of isolated protonated uracil.[32] The **TA** tautomer of $[Urd+H]^+$ is found to be 0.2

Physical Chemistry Chemical Physics Accepted Manuscript

kJ/mol (B3LYP) and 0.3 kJ/mol (MP2) higher in free energy than **Ti**, the most stable 2,4-dihydroxy tautomer. For both $[Urd+H]^+$ and $[dUrd+H]^+$, the 2,4-dihydroxy tautomers that adopt syn conformations of the uracil moiety and are stabilized by an O2H···O5' hydrogen-bonding interaction, lie 2–4 kJ/mol higher in free energy, whereas those stabilized by an O2···HO5' hydrogen bonding interaction lie 22–37 kJ/mol higher in free energy than the most stable anti-oriented 2,4-dihydroxy tautomers. Thus, the O2···HO5' hydrogen-bonding interaction is > 20 kJ/mol less favorable than the O2H···O5' hydrogen-bonding interaction, which is not unexpected as hydrogen-bonding interactions that involve the excess proton gain additional stabilization from ion-dipole and ion-induced dipole interactions.

O4 Protonation. Both B3LYP and MP2 predict the most stable O4 protonated conformers of $[Urd+H]^+$, **O4A**, **O4i**, and **O4B** to have the nucleobase in an anti orientation with the 4-hydroxyl hydrogen atom pointed away from N3. In these conformers, C2'-endo and C3'-endo puckering is essentially isoenergetic, but lead to different rotations of the 2'- and 3'-hydroxyl substituents. The most stable B3LYP O4 protonated conformers of $[dUrd+H]^+$, **O4A** and **O4B**, also have the nucleobase in the anti orientation with the 4-hydroxyl hydrogen atom pointed away from N3. However, B3LYP slightly favors C3'-endo sugar puckering, whereas MP2 slightly prefers C2'-endo puckering. For both $[Urd+H]^+$ and $[dUrd+H]^+$, rotation of the 4-hydroxyl hydrogen atom of the anti conformers toward N3 results in an increase of > 10 kJ/mol in free energy. The O4 protonated syn conformers stabilized by an O2H···O5' hydrogen-bonding interaction lie > 20 kJ/mol higher in free energy than the most stable O4 protonated conformer.

O2 Protonation. The most stable O2 protonated conformer of $[Urd+H]^+$, **O2i**, has the nucleobase in an approximately anti orientation, the 2-hydroxyl hydrogen atom rotated away from N3 and hydrogen bonded to the 2'-hydroxyl oxygen atom forming coupled O2H···O2'H···O3' hydrogen-bonding interactions, and C2'-endo puckering of the sugar. In contrast, the most stable O2 protonated conformer of $[dUrd+H]^+$, **O2A**, prefers a syn orientation of the uracil moiety with the 2-hydroxyl hydrogen atom hydrogen bonded to O5' and C2'-endo sugar puckering. The **O2A** conformer of $[Urd+H]^+$ is found to be only 1.0 kJ/mol (B3LYP) and

Physical Chemistry Chemical Physics Accepted Manuscript

0.2 kJ/mol (MP2) less stable than the **O2i** conformer. For both $[Urd+H]^+$ and $[dUrd+H]^+$, conformers with an O2 protonated nucleobase in the anti orientation that do not involve the O2 proton in a hydrogen-bonding interaction lie > 35 kJ/mol higher in free energy than the anti or syn oriented O2 protonated conformers in which the O2 proton is stabilized by a hydrogen-bonding interaction with the 2'- or 5'-hydroxyl oxygen atoms. This is again a consequence of the enhanced stability provided by ion-dipole and ion-induced dipole interactions to the system.

The Effect of the 2'-Deoxy Modification on Conformation. Previous comparisons of the gas-phase structures of the protonated DNA vs RNA nucleosides[22–24] provide insights into the effect of the 2'-hydroxyl substituent on gas-phase conformation. In analogy, comparisons of the conformations of $[Urd+H]^+$ vs $[dUrd+H]^+$ enable the effect of the 2'-deoxy modification to be elucidated. The 2'-deoxy modification reduces the number of favorable orientations of the 3'-hydroxyl substituent due to the lack of potential hydrogen-bonding interactions. The 2'-deoxy modification also eliminates the opportunity for formation of coupled O2H···O2'H···O3' and O3'H···O2'H···O2 hydrogen-bonding interactions, which stabilize the anti orientation of the nucleobase and C2'-endo and C3'-exo puckering of the sugar, respectively. Therefore, the 2'-deoxy modification provides additional freedom to the protonated uracil and sugar moieties such that when the 2-hydroxyl hydrogen atom is pointed away from the N3 atom, the nucleobase orientation and sugar configuration are not confined by hydrogen-bonding interactions.

Ring-Open Conformations. As described in the Electronic Structure Calculations subsection of the Methods section, we have also examined the ring-open forms of the ribose and 2′-deoxyribose sugars of $[Urd+H]^+$ and $[dUrd+H]^+$, respectively. In doing so, we considered the possibility that ring opening may occur at either the C1′-O4′ and C4′-O4′ bonds leading to the carbonyl moiety at the C4′ and C1′ atoms, respectively. For $[Urd+H]^+$, B3LYP suggests that the most stable ring-open conformers have the carbonyl moiety at the C4′ position and are found to be more stable than the most stable ring-closed conformers. In contrast, MP2 suggests that the most stable ring-open forms have the carbonyl at the C1′ position, but that the ring-closed forms are preferred. For $[dUrd+H]^+$, B3LYP suggest that the most stable ring-open conformers have

Physical Chemistry Chemical Physics Accepted Manuscript

the carbonyl at O1′ and again finds that the most stable ring-open conformers are more stable than the most stable ring-closed conformers. MP2 results for $[dUrd+H]^+$ are relatively parallel to those of B3LYP except that the preference for the ring-open forms is weaker. All stable ring-open forms computed (see Figure S3) are stabilized by one or two intramolecular hydrogen bonds between the various carbonyl and hydroxyl substituents of the nucleobase and sugar moieties.

Discussion

Comparison of Experimental IRMPD and Theoretical IR Spectra of $[Urd+H]^+$. The experimental IRMPD and theoretical IR spectra of the low-energy ring-closed conformers, **Ti**, **TA**, **TC**, **O4A**, **O4B**, **O4ii**, and **TD** of $[Urd+H]^+$ in the IR fingerprint and hydrogen-stretching regions are compared in Figure 3. As can be seen in Figure 3, the calculated IR spectra of the **T** and **O4** conformers are complementary to each other and exhibit very good agreement with the measured IRMPD spectrum. The measured band at ~1800 cm^{-1}, which is not predicted for the 2,4-dihydroxy tautomers, arises from the O4 protonated conformers. The measured bands at ~3615 and ~3395 cm^{-1} are also contributed by the O4 protonated conformers. The calculated spectra of the ground-state **Ti** conformer, which possesses coupled O2H···O2'H···O3 hydrogen-bonding interactions, exhibits reasonable agreement with the measured spectrum in the IR fingerprint region except that the calculated band at ~1480 cm^{-1} is shifted to a lower frequency than the measured band at ~1595 cm^{-1}. However, the calculated spectral features between ~1000 to ~1300 cm^{-1} of **Ti** better reproduce the measured spectrum than those of the other 2,4-dihydroxy tautomers. In addition, the small band measured at ~3510 cm^{-1}, which is not predicted for the other conformers, is contributed by **Ti**, thus establishing its presence in the experimental population. Therefore, these seven conformers may be populated in the experiments. However, the computed relative stability of **TD** suggests that its population should be immeasurably small.

The experimental IRMPD and theoretical IR spectra of the low-energy ring-closed conformers, **TB**, **O4i**, **Tii**, and **O4ii** $[Urd+H]^+$ in the IR fingerprint and hydrogen-stretching

Physical Chemistry Chemical Physics Accepted Manuscript

regions are compared in Figure S4. These conformers only differ in the orientations of the 2'- and 3'-hydroxyl substituents or sugar puckering from those shown in Figure 3. The calculated IR spectra of these four conformers also exhibit good agreement with the measured spectrum, indicating that the changes in the orientations of the 2'- and 3'-hydroxyl substituents and sugar puckering do not exert a significant impact on the spectral features. Therefore, these four conformers may also be populated in the experiments. Comparisons between the measured IRMPD and IR spectra predicted for the other low-energy conformers computed are shown in Figures S5−S6 and suggest that they are not populated by ESI. Spectral differences that enable these conformers to be eliminated are discussed in the Electronic Supplementary Information.

The experimental IRMPD and theoretical IR spectra of the stable low-energy ring-open conformers, **O4A_RO4′**, **TA_RO4′**, **O4A_RO1′**, **TA_RO1′**, and **Ti_RO1′** of $[Urd+H]^+$ in the IR fingerprint and hydrogen-stretching regions are compared in Figure S7. In all cases, the IR spectra of the ring-open conformers exhibit obvious differences from the measured IRMPD spectrum that clearly eliminate their presence in the experimental population. Predicted bands that exhibit obvious disagreement with the measured spectrum are highlighted. In particular, the carbonyl stretches associated with the ring-open sugar moieties are significantly shifted to either lower or higher frequencies. In contrast, the C2=O carbonyl stretches are well matched to the measured band, indicating that the band observed at ~1800 cm^{-1} is indeed contributed by the O4 protonated canonical conformers. Therefore, the ring-open conformations of $[Urd+H]^+$ are not populated by ESI.

$[Urd+H]^+$ Conformers Populated by ESI. In summary, a diverse mixture of low-energy conformers including 2,4-dihydroxy tautomers as well as O4 protonated conformers that exist exclusively in their ring-closed forms contribute to the experimentally accessed population. For the 2,4-dihydroxy tautomers, **Ti**, **TA**, **TB**, **TC**, **Tii**, and **TD** may be populated by ESI with **TD** expected to be in very low abundance. Among these 2,4-dihydroxy tautomeric conformers that may be populated, three are C2'-endo anti, one is C3'-endo anti, and two are C2'-endo syn conformers, indicating that these tautomers have significant conformation flexibility. The five

Physical Chemistry Chemical Physics Accepted Manuscript

O4 protonated conformers that may be populated, **O4A** and **O4iii** exhibit C2'-endo anti conformations, whereas **O4i** and **O4B**, both exhibit C3'-endo anti conformations, and **O4ii** exhibits a C3'-exo anti conformation, i.e., no O4 protonated syn conformers contribute to the measured IRMPD spectrum.

Resonant Vibrational Modes of [Urd+H]$^+$. Vibrational assignments of the measured IRMPD spectral features are based on the predicted IR spectra of the conformations populated. In the IR fingerprint region, the measured band at ~1800 cm^{-1} is contributed by the O4 protonated conformers and represents free O2 carbonyl stretches. The band observed at ~1650 cm^{-1} represents coupled C2–N3 and C5=C6 stretches, whereas the band at ~1595 cm^{-1} arises from C4–C5 stretching coupled with O4–H in-plane bending. The measured band at ~1505 cm^{-1} arises from stretching of the nucleobase ring coupled with C5–H and C6–H in-plane bending. The band at ~1210 cm^{-1} arises from O2–H and O4–H in-plane bending. The broad IR band at ~1115 cm^{-1} arises from stretching of the sugar ring. In the hydrogen-stretching region, the band observed at ~3665 cm^{-1} arises from coupled O3'–H and O5'–H stretches. The band at ~3615 cm^{-1}, primarily contributed by O4 protonated conformers, arises from O4–H and O2'–H stretches. The band at ~3565 cm^{-1} reflects O2–H, O4–H, and O2'–H stretches of the 2,4-dihydroxy tautomers. The small feature observed at ~3510 cm^{-1} is the result of O2'–H stretching of the **Ti** conformer, in which the 2'-hydroxyl is involved in coupled hydrogen-bonding interactions. The small IR band at ~3395 cm^{-1} is contributed by O4 conformers, and arises from N3–H stretching.

Comparison of Experimental IRMPD and Theoretical IR Spectra of [dUrd+H]$^+$. The experimental IRMPD and theoretical IR spectra of the low-energy ring-closed conformers, **O4A**, **O4B**, **TA**, **TB**, **TC**, and **TD** of [dUrd+H]$^+$ in the IR fingerprint and hydrogen-stretching regions are compared in Figure 4. The calculated IR spectra of the 2,4-dihydroxy and O4 protonated conformers are complementary to each other and exhibit very good agreement with the measured spectrum. The bands observed at ~1800 and ~3610 cm^{-1}, which are not predicted for the 2,4-dihydroxy tautomers, are contributed by the **O4A** and **O4B** conformers. The weak IR feature observed at ~3395 cm^{-1} is difficult to identify due to the low S/N in that region. However,

Physical Chemistry Chemical Physics Accepted Manuscript

because the presence of **O4A** and **O4B** are indicated, and a moderate IR feature is predicted for both conformers at ~3395 cm^{-1}, it can be assumed that there is indeed an IR feature at this frequency. Comparisons between the measured IRMPD and IR spectra predicted for the other low-energy conformers calculated are shown in Figure S8 and suggest that they are not populated by ESI. Spectral differences that enable these conformers to be eliminated are discussed in the Electronic Supplementary Information.

The experimental IRMPD and theoretical IR spectra of the stable low-energy ring-open conformers, **O4A_RO1′**, **TA_RO1′**, **Ta_RO1′**, **O4A_RO4′**, and **TA_RO4′** of $[dUrd+H]^+$ in the IR fingerprint and hydrogen-stretching regions are compared in Figure S9. As found for $[Urd+H]^+$, the IR spectra of the ring-open conformers exhibit obvious differences from the measured IRMPD spectrum that clearly eliminate their presence in the experimental population. The carbonyl stretches associated with the ring-open sugar moieties are again significantly shifted relative to the measured spectrum. In contrast, the band positions of the C2=O carbonyl stretches of the O4 protonated species are again well matched with the measured band. Therefore as found for $[Urd+H]^+$, the ring-open conformations of $[dUrd+H]^+$ are not also populated by ESI.

$[Urd+H]^+$ Conformers Populated by ESI. In summary, a mixture of low-energy conformers including four 2,4-dihydroxy tautomers, **TA**, **TB**, **TC** and **TD**, may be present in the experiments with **TD** apparently in very low abundance. Among the 2,4-dihydroxy tautomeric conformations present, **TA** and **TB** are anti with C2'-endo and C3'-endo puckering respectively, and exhibit the same rotational orientations of the 2- and 4-hydroxyl hydrogen atoms as found for isolated protonated uracil, whereas both syn conformers, **TC** and **TD**, exhibit C2'-endo puckering. The most stable syn conformer **TC** has the 2-hydroxyl hydrogen flipped to enable an O2H···O5' hydrogen-bonding interaction to occur, whereas the much less stable syn conformer **TD** exhibits the same rotational orientations of the 2- and 4-hydroxyl hydrogen atoms as isolated protonated uracil, which leads to a much weaker O5'H···O2 hydrogen-bonding interaction and thus a significant loss in stability. The O4 protonated conformers populated, **O4A** and **O4B**

Physical Chemistry Chemical Physics Accepted Manuscript

strongly favor the anti conformation such that no O4 protonated syn conformers contribute to the measured spectrum.

Resonant Vibrational Modes of [dUrd+H]$^+$. Vibrational assignments of the measured IRMPD spectral features are again based on the conformations populated. In the IR fingerprint region, the band at ~1800 cm^{-1} is contributed by **O4A** and **O4B**, and represents C2=O carbonyl stretching. The band at ~1640 cm^{-1} arises from coupled C2–N3 and C5=C6 stretches. The strong band observed at ~1590 cm^{-1} represents C4–C5 stretching coupled with O4–H in-plane bending. The band at ~1510 cm^{-1} represents stretching of the nucleobase ring coupled with C5–H and C6–H in-plane bending. The broad band observed at ~1205 cm^{-1} reflects O2–H and O4–H in-plane bending. The strong band measured at ~1095 cm^{-1} represents stretches of the sugar ring. In the hydrogen-stretching region, the sharp absorption measured at ~3665 cm^{-1} represents coupled O3'–H and O5'–H stretching. The small band observed at ~3610 cm^{-1}, contributed by **O4A**, arises from O4–H stretching. The band observed at ~3575 cm^{-1} contributed by **TA**, **TB** and **TC** arises from O2–H and O4–H stretching. The feature observed at ~3395 cm^{-1} is contributed by **O4A** and **O4B**, and represents N3–H stretching.

Comparison of Experimental IRMPD, Maxwell-Boltzmann Weighted and Least Squares Fitted Theoretical IR Spectra. Comparisons between the measured IRMPD and calculated IR spectra for the low-energy conformers of [Urd+H]$^+$ and [dUrd+H]$^+$ suggest that a mixture of multiple low-energy conformers are populated in the experiments. In particular, the **Ti**, **TA**, **TB**, **TC**, **O4A**, **O4i**, **O4B**, **Tii**, **O4ii**, **O4iii** and **TD** conformers of [Urd+H]$^+$, and the **O4A**, **O4B**, **TA**, **TB**, **TC**, and **TD** conformers of [dUrd+H]$^+$ may contribute to the experimental population generated by ESI. To enhance the interpretation, a least squares fitting (LSF) of the measured spectra based on the calculated IR spectra of all low-energy conformers within 10 kJ/mol (B3LYP) as well as the **TD** conformers of both [Urd+H]$^+$ and [dUrd+H]$^+$ was performed. The best least squares fit found for [Urd+H]$^+$ suggest that the **Ti**, **T(O2/O4)A**, **T(O2/O4)C**, **T(O2/O4)D**, **O4A**, **O4B**, and **O4ii** are present and represent 0.4%, 18.7%, 18.6%, 36.2%, 5.1%, 6.4%, and 14.6% of the ESI population. Similarly, for [dUrd+H]$^+$, **O4A**, **O4B**, **TA**, **TB**, and **TC**

Physical Chemistry Chemical Physics Accepted Manuscript

are populated and represent 23.6%, 3.4%, 6.0%, 16.4%, and 50.6% of the ESI population. For $[Urd+H]^+$, the calculated IR spectra of **TA**, **TB** and **Tii** as well as **O4A** and **O4i**, and **O4B** and **O4iii**, respectively, are highly parallel, but the LSF results suggest that **TA**, **O4A**, and **O4B** may slightly better reproduce the measured spectrum. The LSF results also confirm that O2 protonated conformers are not populated in the experiments. The various fits that give rise to small residuals suggest that the ratios of the 2,4-dihydroxy tautomers vs. the O4 protonated conformers are 75±2 and 25±2 for $[Urd+H]^+$, and 68±7 and 32±7 for $[dUrd+H]^+$, respectively. This suggests that the 2,4-dihydroxy tautomers are somewhat more important for $[Urd+H]^+$. The measured IRMPD spectra of $[Urd+H]^+$ and $[dUrd+H]^+$ are compared to the linear IR spectra predicted for Maxwell-Boltzmann weighted (MBW) distributions at room temperature based on the B3LYP and MP2 relative stabilities as well as the best LSF results. These comparisons are shown in Figures S10 and S11. Comparisons in Figure S10 suggest that the intensities of the bands at ~1800, ~3390 and ~3625 cm^{-1} that arise from O4 protonated conformers are underestimated by the LSF, B3LYP and even more so by MP2, and suggests that both levels of theory underestimate the relative stabilities of the O4 protonated conformers for $[Urd+H]^+$. The O4 protonated conformers represent ~25% of the population according to B3LYP, which agrees with the ratio predicted from the LSF fits. Moreover, the small band measured at ~3510 cm^{-1} is well reproduced by the B3LYP and MP2 MBW spectra but not the LSF fits, suggesting that **Ti** is clearly present in the experiments. In contrast, the intensities of the three bands predicted at ~1800, ~3390 and ~3625 cm^{-1} are overestimated by B3LYP, but still underestimated by MP2 for $[dUrd+H]^+$, see Figure S11. This comparison suggests that B3LYP overestimates the relative stabilities of the O4 protonated conformers of $[dUrd+H]^+$, whereas MP2 provides underestimates. For $[Urd+H]^+$, B3LYP predicts the O4 protonated conformers to be ~3.0 kJ/mol less favorable than the most stable 2,4-dihydroxy tautomer, whereas for $[dUrd+H]^+$, B3LYP predicts an O4 protonated conformer as the ground conformation. Comparisons in Figures S10 and S11 suggest that the relative stabilities of the O4 protonated species should fall in the range between 0.0–3.0 kJ/mol in order to better reproduce the measured spectra. On the other hand, MP2 predicts

Physical Chemistry Chemical Physics Accepted Manuscript

relative stabilities that are similar to B3LYP for the 2,4-dihydroxy or O2 protonated conformers, but a much higher energy for the O4 protonated species vs B3LYP. These comparisons imply that MP2 systematically predicts the O4 protonated species to be less stable than suggested by the relative band intensities of the experimental spectrum.

Comparison to IRMPD and Theoretical Studies of Protonated Uracils. Previous IRMPD and theoretical studies[29,31–33] of protonated uracil and 5- and 6-halouracils suggest that protonation preferentially stabilizes the 2,4-dihydroxy tautomer in which both hydroxyl hydrogen atoms point toward the adjacent N3 atom. This type of tautomer is populated in dominant abundance along with a small population of the O4 protonated conformer, where the O4–H proton is rotated away from the adjacent N3 atom. For $[Urd+H]^+$ and $[dUrd+H]^+$, besides the 2,4-dihydroxy tautomers analogous to those populated for the free nucleobase and with uracil in the anti orientation, **Ti** and **TC** of $[Urd+H]^+$ and **TC** of $[dUrd+H]^+$, are also populated in measurable abundance, suggesting that the presence of the sugar moieties allows the 2-hydroxyl hydrogen to flip away from the N3 atom and hydrogen bond to the 2'-hydroxyl moiety, as in **Ti** of $[Urd+H]^+$, or hydrogen bond to the 5'-hydroxyl moiety, as in **TC** of $[Urd+H]^+$ and $[dUrd+H]^+$, thus leading to a syn orientation of the 2,4-dihyroxy uracil moiety. Therefore, the presence of the sugar moieties increases the variety of the conformers that may be accessed, and in particular allows a mixture of 2,4-dihydroxy tautomers of $[Urd+H]^+$ and $[dUrd+H]^+$ to be populated.

Comparison to the Gas-Phase Conformation of Neutral Uridine. The previous Fourier transform microwave spectroscopy and theoretical study of neutral Urd in the gas phase found that the nucleobase exist as the canonical diketo tautomer and takes on an anti orientation, the sugar is C2′-endo puckered, and both the 2′ and 3′-hydroxyls point up and toward the neutral uracil moiety and engage in dual O3′H···O2′H···O2 hydrogen-bonding interactions. Additional stabilization of Urd is also provided by O5′···HC2 and O2′H···O5′ noncanonical hydrogen-bonding interactions. As compared to our results for $[Urd+H]^+$, which suggest that 75±2% of the population is contributed by 2,4-dihydroxy tautomers, it is clear that similar to neutral uracil vs. protonated uracil,[29–32] protonation facilitates tautomerization of Urd and preferentially stabilizes

Physical Chemistry Chemical Physics Accepted Manuscript

the 2,4-dihydroxy tautomers. Similar to neutral Urd, the most stable 2,4-dihydroxy tautomers and O4 protonated conformers of $[Urd+H]^+$ populated in the experiments, **Ti**, **TA** and **O4A** all adopt anti-oriented nucleobase and C2'-endo sugar puckering. However, in all cases both the 2'- and 3'-hydroxyl substituents point down and away from the protonated uracil moiety indicating that protonation also alters the preferred orientations of the sugar hydroxyls.

IRMPD yields of $[Urd+H]^+$ vs $[dUrd+H]^+$. In our previous IRMPD studies of the protonated guanine, adenine, and cytosine DNA vs RNA nucleosides,[22–24] the IRMPD yields of the protonated DNA nucleosides were observed to exceed that of their RNA analogues in the hydrogen-stretching region. In contrast, the IRMPD yield of $[Urd+H]^+$ is greater than that of $[dUrd+H]^+$ in this region. The greater IRMPD yield observed previously for the protonated DNA nucleosides in the hydrogen-stretching region was interpreted as being due to the freedom of the 3'-hydroxyl substituent of the protonated DNA nucleosides.[22–24] In contrast, the greater IRMPD yield observed in this region for $[Urd+H]^+$ may be the result of free hydrogen stretches arising from the protonated uracil moiety rather than the free O3'–H stretch. This interpretation is consistent with the larger variety of 2,4-dihydroxy tautomers for $[Urd+H]^+$ being populated in the experiments than found for $[dUrd+H]^+$.

Implications of Diverse Mixtures of Protonated Uridine and 2'-Deoxyuridine Conformations. Despite the fact that uracil possesses two enolizable keto groups, the most stable conformation of neutral uracil in both the gas and solid phases is the canonical 2,4-diketo tautomer, whereas upon protonation, the 2,4-dihydroxy tautomer becomes the most stable conformation with the O4 protonated conformer in low abundance.[29–32] The presence of the ribose or 2'-deoxyribose moieties results in greater diversity of the protonated tautomeric conformations, including a larger variety of 2,4-dihydroxy tautomers and the coexisting O4 protonated conformers make up a larger portion of the population than for isolated protonated uracil. This suggests that the sugar moieties increase the conformational flexibilities of the protonated nucleoside analogues and may further enhance the tautomeric behavior of the uracil moiety. Because of the probable relationship between the tautomeric shifts and point mutation

Physical Chemistry Chemical Physics Accepted Manuscript

during nucleic-acid replication,[48] the thorough conformational investigation presented here for the protonated uracil nucleosides may provide better insights into the mechanisms of point mutation when the sugar moieties are considered. As compared to the previous IRMPD studies of the protonated guanine, adenine and cytosine nucleosides,[22–24] a much more diverse mixture of conformers of the protonated uracil nucleosides are populated in the experiments. The similar proton affinities (PAs) of the uracil and sugar moieties vs the larger differences in PAs between the other nucleobases and sugar moieties, and the similar PAs of the O2 and O4 sites of uracil,[29] allow multiple locations for the excess proton to reside, making the uracil nucleosides more likely to be involved in mutations and may explain why a much richer variety of modified uridines have been found in nature. Moreover, specific chemical reagents, such as acrylonitrile (AN), can be used as probes to examine the surface topography of tRNA molecules, and under the same reaction conditions only uridine reacts with AN. No reaction with AN was observed for adenosine, guanosine or cytidine.[49] These results suggest that uracil nucleosides are readily chemically distinguishable from the other nucleosides.

Conclusions

Comparisons between the measured IRMPD and the B3LYP/6-311+G(d,p) calculated spectra of $[Urd+H]^+$ and $[dUrd+H]^+$ in both IR fingerprint and hydrogen-stretching regions determine the favorable state(s) of protonation/tautomerization and the low-energy conformers populated by ESI. Both B3LYP and MP2 predict the same structure as the ground ring-closed conformer of $[Urd+H]^+$, where the additional 2'-hydroxyl substituent enables the formation of coupled O2H···O2'H···O3' hydrogen bonds that stabilize C2'-endo puckered anti conformers. In contrast, B3LYP predicts O4 as the most favorable protonation site for $[dUrd+H]^+$, whereas MP2 also predicts the 2,4-dihdroxy tautomer C2'-endo anti conformer as the ground species. The comparison between experiment and theory suggests that for both $[Urd+H]^+$ and $[dUrd+H]^+$, a diverse mixture of 2,4-dihydroxy tautomers and O4 protonated conformers are populated in the experiments. The presence of the sugar moieties leads to a larger variety of 2,4-dihydroxy

Physical Chemistry Chemical Physics Accepted Manuscript

tautomers of $[Urd+H]^+$ and $[dUrd+H]^+$ being populated in measurable abundance than found for isolated protonated uracil,[31–33] where only a single 2,4-dihydroxy tautomer is populated in significant abundance. The larger variety of $[Urd+H]^+$ conformers populated that possess hydroxyl stretches than found for $[dUrd+H]^+$ likely explains the higher IRMPD yield observed for $[Urd+H]^+$ in the hydrogen-stretching region, explaining this contrasting behavior vs that observed for other protonated RNA nucleosides investigated to date.[22–24] B3LYP and MP2 predict approximately equal stabilities for the 2,4-dihydroxy and O2 protonated conformers, whereas MP2 predicts O4 protonated species to be much less stable than found for B3LYP. Comparison between the experimental, Maxwell-Boltzmann weighted, and least squares fitted IR spectra of the low-energy conformers suggest that both levels of theory underestimate the stabilities of the O4 protonated conformers for $[Urd+H]^+$, whereas B3LYP overestimates, and MP2 underestimates the stabilities of the O4 protonated conformers for $[dUrd+H]^+$.

Electronic Supplementary Information

Tables summarizing the relative Gibbs free energies at 298 K of the stable low-energy ring-open conformers of $[Urd+H]^+$ and $[dUrd+H]^+$. Tables of the Cartesian coordinates and important geometrical parameters of all stable low-energy conformers computed for $[Urd+H]^+$ and $[dUrd+H]^+$. A detailed description of and figures showing the B3LYP/6-311+G(d,p) low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ and their relative Gibbs free energies at 298 K calculated at B3LYP/6-311+G(2d,2p) and MP2(full)/6-311+G(2d,2p) levels of theory. Comparisons of the measured IRMPD spectra and calculated linear IR spectra in the IR fingerprint and hydrogen-stretching regions of select low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$. Comparisons of the measured IRMPD, Maxwell-Boltzmann weighted, and least squares fitted IR spectra calculated for the low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ populated in the experiments and based on energetics calculated at the B3LYP and MP2 levels of theory. This information is available free of charge via the Internet at. . .

Physical Chemistry Chemical Physics Accepted Manuscript

Acknowledgements

Financial support of this work was provided by the National Science Foundation, Grants PIRE-0730072 and CHE-1409420. R. R. W. also received support from Wayne State University a Thomas C. Rumble Fellowship. The authors are indebted to WSU C&IT for the excellent computational resources and support provided. This work is part of the research program of FOM, which is financially supported by the Nederlandse Organisatie voor Wetenschappeklijk Onderzoek (NWO). The FELIX staff is gratefully acknowledged for their excellent maintenance and operation of the free electron and OPO/OPA lasers that enabled this work to be performed.

References

1 R. R. Sinden, DNA Structure and Function. Academic Press. 1994.
2 J. D. Watson and F. H. C. Crick, *Nature,* 1953, **171**, 737.
3 W. Saenger, Principles of Nucleic Acid Structure, Springer-Verlag, New York, 1984.
4 G. A. Jeffrey and W. Saenger, Hydrogen Bonding in Biological Structure, Springer-Verlag, New York, 1991.
5 R. Taylor and O. Kennard, *J. Mol. Struct.,* 1982, **78**, 1.
6 J. Šponer, J. Leszczyński, V. Vetterl and P. Hobza, *J. Biomol. Struct. Dyn.,* 1996, **13**, 695.
7 C. Colominas, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.,* 1996, **118**, 6811.
8 M. Chawla, P. Sharma, S. Hader, D. Bhattacharyya and A. Mitra *J. Phys. Chem. B,* 2011, **115**, 1469.
9 C. C. Hardin, M. Corregan, B. A. Brown and L. N. Frederick, *Biochemistry,* 1993, **32**, 5870.
10 G. J. Puppels, C. Otto, J. Greve, M. Robert-Nicoud, D. J. Arndt-Jovin and T. M. Jovin, *Biochemistry,* 1994, **33**, 3386.
11 V. A. Malkov, O. N. Voloshin, A. G. Veselkov, V. M. Rostapshov, I. Jansen, V. N. Soyfer and M. D. Frank-Kamenetskii, *Nucleic Acids Res.,* 1993, **21**, 105.
12 K. Gehring, J. L. Leroy and M. Gueron, *Nature,* 1993, **363**, 561.
13 J. Volker and H. H. Klump, *Biochemistry,* 1994, **33**, 13502.
14 S. E. Morse and D. E. Draper, *Nucleic Acids Res.,* 1995, **23**, 302.
15 S. B. Jang, L. W. Hung, Y. I. Chi, E. L. Holbrook, R. J. Carter and S. R. Holbrook, *Biochemistry,* 1998, **37**, 11726.
16 D. Collin and K. Gehring, *J. Am. Chem. Soc.,* 1998, **120**, 4069.
17 C. L. Tang, E. Alexov, A. M. Pyle and B. Honig, *J. Mol. Biol.,* 2007, **366**, 1475.
18 G. B. Goh, J. L. Knight and C. L. Brooks, *J. Chem. Theory Comput.,* 2012, **8**, 36.
19 G. B. Goh, J. L. Knight and C. L. Brook, *J. Chem. Theory Comput.,* 2013, **9**, 935.
20 J. Hauser, R. Keese, H, *Helv. Chim. Acta.* 2002, **85**, 2481.
21 A. Filippi, C. Fraschetti, F. Rondino, S. Piccirillo, V. Steinmetz, L. Guidoni and M. Speranza, *Int. J. Mass Spectrom.,* 2013, **354**, 54.

Physical Chemistry Chemical Physics Accepted Manuscript

22 R. R. Wu, B. Yang, G. Berden, J. Oomens and M. T. Rodgers, *J. Phys. Chem. B,* 2014, **118**, 14774.
23 R. R. Wu, B. Yang, G. Berden, J. Oomens and M. T. Rodgers, *J. Phys. Chem. B,* 2015, **119**, 2795.
24 R. R. Wu, B. Yang, C. E. Frieler, G. Berden, J. Oomens and M. T. Rodgers, *J. Phys. Chem. B,* 2014, **119**, 5773.
25 H. U. Ung, K. T. Huynh, J. C. Poutsma, J. Oomens, G. Berden and T. H. Morton, *Int. J. Mass Spectrom.*, 2015, **378**, 294–302.
26 E. Nir, P. Imhof, K. Kleinermanns, M. S. de Vries, *J. Am. Chem. Soc.*, 2000, **122**, 8091.
27 J. A. McCloskey, *Acc. Chem. Res.*, 1991, **24**, 81.
28 A. Filippi, C. Fraschetti, F. Rondino, S. Piccirillo, V. Steinmetz, L. Guidoni, M. Speranza, *Int. J. Mass Spectrom.,* 2013, **354**, 54.
29 J. K. Wolken and F. Turecek, *J. Am. Soc. Mass Spectrom.,* 2000, **11**, 1065.
30 E. S. Kryachko, M. T. Nguyen and T. Zeegers-Huyskens, *J. Phys. Chem. A,* 2001, **105**, 1288.
31 J-Y. Salpin, S. Guillaumont, J. Tortajada, L. MacAleese, J. Lemaire and P. Maitre, *ChemPhysChem,* 2007, **8**, 2235.
32 Y.-w. Nei, T. E. Akinyemi, J. D. Steill, J. Oomens and M. T. Rodgers, *Int. J. Mass Spectrom.,* 2010, **297**, 139.
33 K. T. Crampton, A. I. Rather, Y.-w. Nei, G. Berden, J. Oomens and M. T. Rodgers, *J. Am. Soc. Mass Spectrom.,* 2012, **23**, 1469.
34 P. J. Barr, T. A. Hamor and R. T. Walker, *Acta Crystallogr. Sect. B-Struct. Sci.,* 1978, **34**, 2799.
35 G. I. Birnbaum, R. Deslauriers, T. S. Lin, G. T. Shiau and W. H. Prusoff, *J. Am. Chem. Soc.,* 1980, **102**, 4236.
36 P. Herdewijn, *Antivir. Chem. Chemother.,* 1994, **5**, 131.
37 S. Mezzache, S. Alves, C. Pepe, M. Quelquejeu, F. Fournier, J.-M. Valery and J.-C. Tabet, *J. Mass Spectrom.,* 2005, **40**, 722.
38 P. B. Armentrout, *J. Am. Soc. Mass Spectrom.,* 2000, **11**, 371.
39 T. Grabarkiewicz and M. Hoffmann, *J. Mol. Model.,* 2006, **12**, 205.
40 I. Peña, C. Cabezas and J. L. Alonso, *Angew. Chem. Int. Ed.*, 2015, **54**, 2991.
41 J. J. Valle, J. R. Eyler, J. Oomens, D. T. Moore, A. F. G. van der Meer, G. von Heldon, G. Meijer, C. L. Hendrickson, A. G. Marshall and G. T. Blakney, *Rev. Sci. Instrum.,* 2005, **76**, 023103–1.
42 N. C. Polfer, J. Oomens, D. T. Moore, G. von Helden, G. Meijer and R. C. Dunbar, *J. Am. Chem. Soc.,* 2006, **128**, 517.
43 N. C. Polfer and J. Oomens, *Phys. Chem. Chem. Phys.,* 2007, **9**, 3804.
44 D. Oepts, A. F. G. van der Meer and P. W. van Amersfoort, *Infrared Phys. Techn.,* 1995, **36**, 297.
45 HyperChem Computational Chemistry Software Package, Version 5.0; Hypercube, Inc.: Gainsville, FL, 1997.
46 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A.

Physical Chemistry Chemical Physics Accepted Manuscript

Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09, Revision C.01; Gaussian, Inc.: Wallingford, CT, 2009.
47 J. Oomens, A. G. G. M. Tielens, B. Sartakov, G. von Helden and G. Meijer, *Astrophys. J.*, 2003, **591**, 968.
48 I. Dabkowsha, M. Gutowski and J. Rak, *J. Am. Chem. Soc.*, 2005, **127**, 2238.
49 M. Yoshida, T. Ukita, *J. Biochem.*, 1965, **57**, 818.

Physical Chemistry Chemical Physics Accepted Manuscript

Table 1. Relative Enthalpies and Free Energies of Stable Low-Energy Conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ at 0 and 298 K in kJ/mol[a]

Species	Conformer	B3LYP			MP2(full)		
		ΔH_0	ΔH_{298}	ΔG_{298}	ΔH_0	ΔH_{298}	ΔG_{298}
$[Urd+H]^+$	**Ti**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	TA	**1.0**	**2.2**	**0.2**	**1.2**	**2.4**	**0.3**
	TB	**3.4**	**4.6**	**2.5**	**4.1**	**5.4**	**3.2**
	TC	**2.1**	**2.3**	**2.9**	**1.4**	**1.6**	**2.2**
	O4A	**4.6**	**6.3**	**2.9**	**11.7**	**13.4**	**10.1**
	O4i	**3.5**	**5.0**	**3.0**	**10.7**	**12.1**	**10.1**
	O4B	**5.6**	**7.2**	**4.4**	**12.5**	**14.1**	**11.4**
	Tii	**6.0**	**7.2**	**5.5**	**4.7**	**5.9**	**4.2**
	O4ii	**7.0**	**8.1**	**6.4**	**11.7**	**12.7**	**11.0**
	O2i	8.6	8.9	8.0	9.2	9.5	8.6
	Tiii	8.6	10.0	8.1	8.6	10.0	8.1
	O2A	8.5	8.8	9.0	8.3	8.7	8.8
	O4iii	**10.7**	**12.3**	**9.6**	**16.6**	**18.2**	**15.5**
	O4C	15.6	17.5	13.8	22.9	24.9	21.1
	O4iv	16.0	17.8	15.1	23.5	25.4	22.6
	O4D	17.3	19.2	15.8	24.7	26.6	23.3
	O4v	19.4	20.9	18.5	24.3	25.8	23.4
	O4E	23.8	25.6	20.3	31.7	33.6	28.3
	O4vi	21.7	23.5	20.4	27.7	29.5	26.3
	O4F	23.2	24.5	21.9	31.5	32.7	30.2
	TD	**29.9**	**31.6**	**26.8**	**32.7**	**34.4**	**29.5**
	O2B	44.2	46.5	42.0	43.8	46.0	41.6
	O2ii	49.0	51.2	47.2	46.8	49.1	45.1
	O2C	50.4	52.8	47.2	52.5	55.0	49.3
	O2iii	59.1	61.8	56.6	60.1	62.8	57.6
$[dUrd+H]^+$	**O4B**	**0.0**	**0.3**	**0.0**	**9.3**	**11.1**	**8.6**
	O4A	**1.9**	**2.1**	**0.4**	**9.7**	**11.5**	**7.4**
	TA	**0.9**	**0.7**	**0.7**	**0.9**	**2.3**	**0.0**
	TB	**1.2**	**1.1**	**1.7**	**3.5**	**5.0**	**3.3**
	TC	**1.5**	**0.0**	**4.3**	**0.0**	**0.0**	**2.0**
	O2A	6.3	4.9	8.8	5.5	5.7	7.2
	O4C	12.2	12.7	11.9	21.9	24.1	21.0

Physical Chemistry Chemical Physics Accepted Manuscript

Table 1. (cont'd) Relative Enthalpies and Free Energies in kJ/mol of Stable Low-Energy Conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ at 0 and 298 K in kJ/mol[a]

Species	Conformer	B3LYP			MP2(full)		
		ΔH_0	ΔH_{298}	ΔG_{298}	ΔH_0	ΔH_{298}	ΔG_{298}
$[dUrd+H]^+$	**O4D**	13.7	14.2	12.3	21.6	23.7	19.5
	O4E	20.1	20.7	17.3	30.9	33.0	27.3
	O4F	33.9	33.4	32.5	45.2	46.2	43.1
	TD	**34.8**	**34.9**	**33.7**	**36.5**	**38.1**	**34.6**
	O2B	46.2	47.0	45.4	45.7	48.0	44.1
	O2a	47.9	48.6	46.4	47.1	49.3	44.8
	O2C	49.8	51.0	48.2	53.4	56.1	51.0
	O2b	55.6	56.8	53.6	59.2	62.0	56.5

[a]Single point energies of the B3LYP/6-311+G(d,p) optimized structures determined at the B3LYP/6-311+G(2d,2p) and MP2(full)/6-311+G(2d,2p) levels of theory including ZPE and thermal corrections. [b]The stabilities of all conformers that may be populated in the experiments are indicated in boldface.

Figure Captions

Figure 1. Chemical structures of uridine (Urd) and 2'-deoxyuridine (dUrd). Most stable low-energy conformers of $[Urd+H]^+$ and $[dUrd+H]^+$ and relative free energies at 298 K predicted at the B3LYP/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) (shown in black) and MP2(full)/6-311+G(2d,2p)//B3LYP/6-311+G(d,p) (shown in red) levels of theory.

Figure 2. Infrared multiple photon dissociation (IRMPD) action spectra of $[Urd+H]^+$ and $[dUrd+H]^+$ in the IR fingerprint and hydrogen-stretching regions.

Physical Chemistry Chemical Physics Accepted Manuscript

Figure 3. Comparison of the measured IRMPD action spectrum of $[Urd+H]^+$ with the theoretical linear IR spectra for the low-energy conformers of $[Urd+H]^+$ that may be populated in the experiments and the corresponding optimized structures calculated at the B3LYP/6-311+G(d,p) level of theory. Also shown in black are the B3LYP/6-311+G(2d,2p) and in red the MP2(full)/6-311+G(2d,2p) relative Gibbs free energies at 298 K. The site of protonation/tautomeric conformation, nucleobase orientation, and sugar puckering are also indicated for each conformer. The measured IRMPD spectrum is overlaid with each theoretical spectrum and scaled to match the intensity of the most intense computed band in both IR fingerprint and hydrogen-stretching regions to facilitate comparisons.

Figure 4. Comparison of the measured IRMPD action spectrum of $[dUrd+H]^+$ with the theoretical linear IR spectra for the select low-energy conformers of $[dUrd+H]^+$ that may be populated in the experiments and the corresponding optimized structures calculated at the B3LYP/6-311+G(d,p) level of theory. Also shown in black are the B3LYP/6-311+G(2d,2p) and in red the MP2(full)/6-311+G(2d,2p) relative Gibbs free energies at 298 K. The site of protonation/tautomeric conformation, nucleobase orientation, and sugar puckering are also indicated for each conformer. The measured IRMPD spectrum is overlaid with each theoretical spectrum and scaled to match the intensity of the most intense computed band in both IR fingerprint and hydrogen-stretching regions to facilitate comparisons.

Physical Chemistry Chemical Physics Accepted Manuscript