www.chembiochem.org

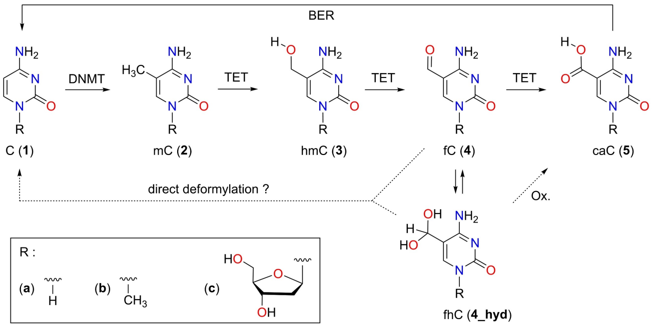
The pH-Dependence of the Hydration of 5-Formylcytosine: an Experimental and Theoretical Study

Fabian L. Zott+,[a] Vasily Korotenko+,[a] and Hendrik Zipse*[a]

5-Formylcytosine is an important nucleobase in epigenetic regulation, whose hydrate form has been implicated in the formation of 5-carboxycytosine as well as oligonucleotide binding events. The hydrate content of 5-formylcytosine and its uracil derivative has now been quantified using a combination of NMR and mass spectroscopic measurements as well as theoretical studies. Small amounts of hydrate can be identified for the protonated form of 5-formylcytosine and for neutral 5-formyluracil. For neutral 5-formylcytosine, however, direct detection of the hydrate was not possible due to its very low abundance. This is in full agreement with theoretical estimates.

Introduction

Epigenetic modifications add a further layer of information to the genetic code based on the linear sequencing of the four canonical DNA bases (A, C, G, T). This information may be encoded into the DNA by chemical modifications of the canonical nucleobases. For example, epigenetic modifications in the canonical nucleobase cytosine are known to control gene regulation in human cells and furthermore have implications in the development of cancer and other diseases.[1–4] 5-Methylcytosine (mC, **2**) as the most common modification is generated by an enzyme-catalyzed methylation at the C5 position of the cytosine base (**1**) and accounts for approximately 1% of all DNA bases in the human genome (Scheme 1).[5,6] The active removal of the methyl group from mC (demethylation) in the human genome is an active field of research. Since direct C–C bond cleavage in **2** is highly unfavorable from a thermochemical point of view, no known mammalian enzyme employs this pathway for the demethylation of mC.[7,8] Instead, the active DNA demethylation pathway known today employs a sequential oxidation of mC to caC (**5**) via hmC (**3**) and fC (**4**), catalyzed by the ten-eleven translocation (TET) family of enzymes (Scheme 1).[9–24] Many other pathways have been proposed, that lead to direct decarboxylation and deformylation of caC and fC, and evidence for the direct deformylation in mammalian cells has been reported recently.[25–28] This deformylation pathway would elegantly avoid the possible damage of DNA strands by the base excision repair (BER, Scheme 1) mechanisms for the active demethylation of caC and fC via the DNA repair protein thymidine DNA glycosylase (TDG).[4,26,29] Since these oxidized cytosine derivatives have also been described as stable epigenetic markers, their susceptibility towards (spontaneous) oxidation is an important field of interest.[30,31] Deaminated derivatives of fC such as 5-formyluracil (fU) can be formed by oxidative stress at thymine (T) sites via 5-hydroxymethyluracil (hmU), which is known to be toxic in mammalian cells.[32–34] In mouse embryotic stem cells, it was found that deamination does not substantially contribute to hmU levels, and that TET enzymes facilitate the oxidation of thymine to hmU.[35] During the discovery of fC in embryonic stem cell DNA, the authors reported evidence for the presence of its hydrated form fhC (**4_hyd**).[10] This *gem*-diol was detected in positive ion MS experiments and quantified at approximately 0.5% at the single nucleotide level. Whether the hydrated form of 5-formyl cytidine plays a role in structural or functional aspects of this base appears to depend on the specific system at hand.[27–29] Burrows et al. reported on two unique formation events for fC-containing DNA duplexes when studying the dynamics of DNA mismatch kinetics. These two events have a ratio of 5:1 and led to the proposal that fC exists in equilibrium with its hydrate fhC, each of which having different base-flipping kinetics.[36] Assuming the hydration reaction to be very slow, this implies



Scheme 1. Possible pathways for methylation and oxidative demethylation of cytosine (DNMT: DNA methyltransferases; TET: ten-eleven translocation enzymes; BER: base excision repair).

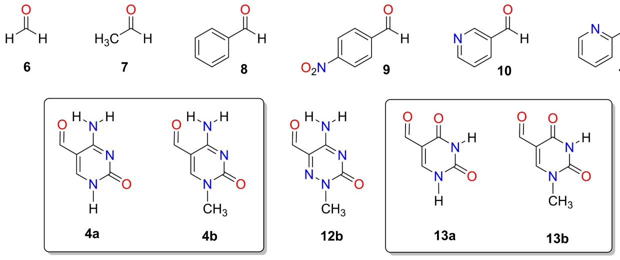
[a] F. L. Zott,+ V. Korotenko,+ H. Zipse
Department of Chemistry, LMU München
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: zipse@cup.uni-muenchen.de

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202100651 an equilibrium constant for the hydration (K_{hyd}) of 0.2. This value is similar to known hydration constants of aldehydes carrying electron-withdrawing substituents.[39,40] Experimental studies may, in some cases, also be impacted by the known conformational *syn/anti* dynamics of **4**.[41,42] In a recent NMR study on the melting kinetics of 5-formylcytosine in dsDNA no evidence for the respective geminal diol form was found in ^{1}H or ^{13}C NMR measurements.[37] This does, of course, not exclude transient hydrate formation in TET2-mediated fC oxidation reactions.[38] Direct measurements of the hydration equilibrium of fC have not yet been reported. Using a combined experimental/theoretical approach, we will show in the following that this may be difficult to achieve.

Results and Discussion

For selected aldehydes carrying aromatic substituents (Scheme 2) the relevant equilibrium data for the hydration reaction has been collected in Table 1. We also include formaldehyde (**6**) and acetaldehyde (**7**) here as two well studied small reference systems. Benzaldehyde (**8**) was employed as a prototype for aldehydes carrying aromatic substituents to verify the measurement strategy. 1-Methyl-5-formyluracil (**13 b**), 5-formyluracil (**13 a**), 5-formylcytosine (**4 a**) and 1-methyl-5-formylcytosine (**4 b**) have been synthesized and purified following modified procedures as described below. These model nucleobases retain the essential functionality of nucleotides while facilitating quantitative experimental and theoretical studies.[42]



Scheme 2. Structures of aldehydes studied in this work.

^{18}O Isotopic exchange experiment

In order validate that the reversible addition of water to the aldehyde carbonyl group in **13 b** leads to transient formation of the hydrate form **13 b_hyd**, an ^{18}O isotope exchange experiment was performed under neutral conditions (Figure 1). The results show that the formyl group reacts readily with $H_2{}^{18}O$ to yield the ^{18}O-labelled nucleobase **^{18}O-13 b**, most likely via the hydrated form **13 b_hyd**. This latter conclusion is supported by analysis of the fragmentation patterns for **13 b** and **^{18}O-13 b**, and oxygen exchange of the other carbonyl groups present in **13 b** can be ruled out (see Supporting Information Figures S3 and S4). The ^{18}O isotope experiment for **4 b** performed under the same conditions shows fast oxygen exchange, with the same results in fragmentation pattern analysis.

^{1}H NMR identification and quantification

The ^{1}H NMR spectrum of **13 b** in D_2O measured at a concentration of 3.1×10^{-3} mol L^{-1} is shown in Figure 2. All ^{1}H resonances of aldehyde **13 b** are accompanied by additional resonances for its hydrate **13 b_hyd** at much lower intensities, which were also matched by NMR shift calculations. When measuring the same

Table 1. Equilibrium constants for the hydration of selected aldehydes (p_ for the protonated form).

System	K_w	K	T [°C]	$\Delta G_{(exp)}$ [kJ/mol]	Ref.
6	2.29×10^{3}	41.2	25	-9.2	[42]
	1.8×10^{3}	32.43	20	-8.5	[43]
7	1.06	0.0191	25	$+9.8$	[44]
	1.08	0.0194	25	$+9.8$	[45]
	1.50	0.0270	25	$+8.9$	[46]
8	0.011	0.98×10^{-3}	25	$+21.1$	[47]
	9.67×10^{-3}	1.74×10^{-4}	22	$+21.3$	this study
9	0.25 ± 0.1	4.50×10^{-3}	25	$+13.4$	[48]
10	0.115	2.07×10^{-3}	20	$+15.1$	[49]
p_10	5.1	0.920	25	$+5.9$	[49]
11	0.66	0.012	25	$+11.0$	[46]
p_11	199	3.58	25	-3.2	[50]
4 a	2.25×10^{-3}	4.05×10^{-5}	22	$>+24.8$[c]	this study
	6.75×10^{-4}	1.22×10^{-5}	22	$<+27.8$[h]	this study
4 b[c,d]	$<4.50\times10^{-4}$	$<8.11\times10^{-6}$	22	$>+28.8$	this study
p_4 b[a,b]	0.005	9.72×10^{-5}	22	$+22.7$	this study
	0.005	9.10×10^{-5}	30	$+22.7$	[14]
4 c[c]	$<4.40\times10^{-4}$	$<7.93\times10^{-6}$	22	$>+28.8$	this study
p_4 c[e]	0.007	1.23×10^{-4}	22	$+22.1$	this study
12 b[f]	0.25	0.0045	22	$+13.3$	[51]
13 a[g]	0.016	2.94×10^{-4}	22	$+20.0$	this study
13 b[g]	0.013	2.42×10^{-4}	22	$+20.5$	this study

[a] Protonated **4 b**. [b] pH = 2. [c] Derived from limit of detection (LOD, see Supporting Information, section S.5). [d] pH = 7.7. [e] pH = 2.6. [f] Presumably under neutral pH conditions. [g] pH = 5.9. [h] Derived from limit of quantification (LOQ, see Supporting Information, section S.5).

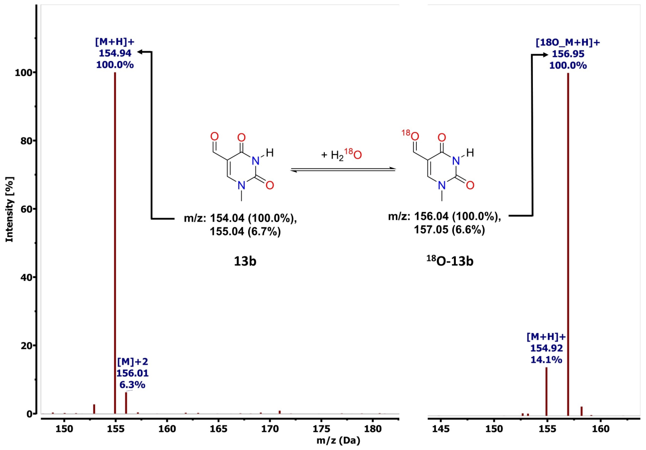


Figure 1. ^{18}O isotope exchange experiment with **13 b** (left side: **13 b** at natural abundance; right side: ^{18}O labelled **13 b** after equilibration with $H_2{}^{18}O$; only the $[M]^+$ peak region is shown).

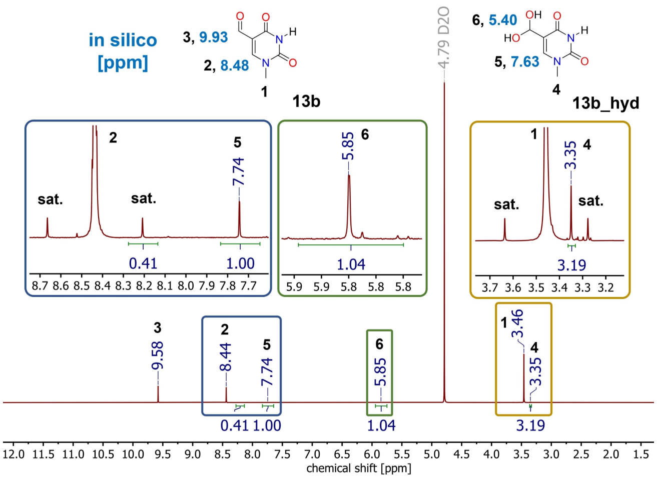


Figure 2. ^1H NMR spectra of **13 b** and its hydrated form **13 b_hyd** in D_2O at a concentration of 3.1×10^{-3} mol L^{-1} together with the relevant signal assignments and calculated chemical shifts (in light blue).

sample of **13 b** in anhydrous DMSO-d_6, no signals for the hydrates can be observed (see Supporting Information Figure S12). Quantification of the formyl- and the hydrated forms was achieved by using the $^{13}C(^1H)$ satellite signals of **13 b** as a reference. Assuming that a single $^{13}C(^1H)$ satellite signal corresponds to 0.535 % of the intensity of the parent ^1H signal, we find a 100:1.3 ratio for aldehyde **13 b** and its hydrate **13 b_hyd** by a standardized procedure (see Supporting Information for details).[52] At a reaction temperature of 23 °C this corresponds to a free energy difference between the two forms of $\Delta G_{exp} = +20.5$ kJ mol^{-1} (Table 1).

Studying 1-methyl-5-formylcytosine (**4 b**) under the same conditions, no ^1H NMR signals could be detected at the theoretically calculated shift region for hydrate **4 b_hyd** at neutral pH. Still, the ^{18}O isotope experiment showed fast oxygen exchange under the same conditions. When acidifying the NMR sample to pH = 2, distinct signals for the hydrate form (**p_4 b_hyd**) arise at shift regions predicted *in silico* with an abundance of 0.5 % (Figure 3). This process is reversible upon neutralization excluding a kinetically controlled equilibrium, while deamination can be ruled out due to differences in chemical shifts (see Supporting Information). These observations are consistent with previous work by Carell et al., where levels of 0.5 % **p_4 c_hyd** have been detected by LC–MS measurements with water/acetonitrile (2 mM NH_4HCOO) under acidic conditions.[14] Whether or not these conclusions are also valid at the full nucleoside level was subsequently studied for 5-formyl-2′-deoxycytidine (**4 c**) through ^1H NMR measurements in D_2O. Under unbuffered conditions (pH = 8.3) the hydrate signals proved too small for quantitative evaluation. Acidification to pH = 2.6 leads to hydrate signals closely similar to those observed before for **4 b_hyd**, and a free energy of hydration of $\Delta G_{exp} = +22.1$ kJ mol^{-1} was measured for protonated **4 c** (**p_4 c**). In contrast to **4 b**, however, slow hydrolysis of nucleoside **4 c** can be observed under acidic conditions, which also implies that the hydration energy for **4 b** may be somewhat more reliable (see Supporting Information Figure S21). In any case we can conclude that protonation has a significant influence on the hydration equilibrium of 5-formylcytosine derivatives. In a more general sense this may also imply that the aldehyde/hydrate equilibrium of 5 fC can be shifted through specific environmental effects.In Table 1 all experimentally determined free energies of hydration ΔG_{hyd} are listed along with important references for theoretical calculations. For systems where the ΔG_{hyd} value could not be determined experimentally, the limits of detection and quantification (LOD and LOQ, see Supporting Information section S.5) are stated.

Theoretical determination of ΔG_{hyd}

The hydration of aldehydes has been studied repeatedly using theoretical methods, but a reliable approach for the direct prediction of hydration energetics has not yet emerged.[53] The performance of various theoretical approaches can be demon-

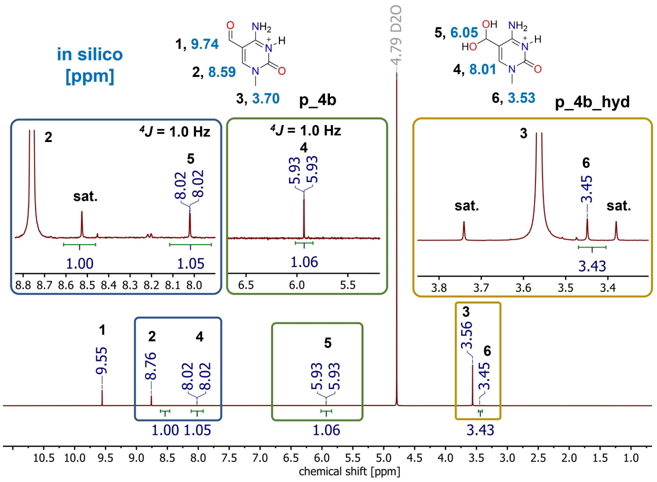


Figure 3. ^1H NMR spectra of the protonated form of **4 b** and its hydrated form **4 b_hyd** in D_2O at pH = 2 and a concentration of 5.8×10^{-3} mol L^{-1} together with the relevant signal assignments and calculated chemical shifts (in light blue).

Table 2. Hydration Gibbs free energies (ΔG_{298}, in kJ mol^{-1}) for acetaldehyde (**7**) and benzaldehyde (**8**) in the gas phase and in aqueous solution.

	Gas phase			Water (SMD model)			Exp.
	DFT[a]	CCSD(T)/CBS[b,c]	G3B3	DFT[a]	CCSD(T)/CBS[b,c,d]	G3B3[e]	
7	+17.5	+15.0	+15.8	+21.5	+16.2	+15.1	+9.8
7(w1)	+12.2	+12.6	+11.5	+20.9	+17.6	+13.5	+9.8
7(w2)	+0.5	+4.5	+1.5	+18.3	+14.7	+10.9	+9.8
8	+34.4	+27.3	+29.1	+44.1	+34.8	+32.4	+21.1
8(w1)	+29.8	+24.2	+25.2	+42.6	+35.9	+35.7	+21.1
8(w2)	+17.8	+13.9	+14.1	+37.8	+37.7	+32.5	+21.1

[a] B3LYP-D3/6-31+G(d,p). [b] Using gas phase B3LYP-D3/6-31+G(d,p) geometries. [c] Based on DLPNO-CCSD(T) single point calculations with the cc-pVTZ and cc-pVQZ basis sets. [d] SMD solvation energies calculated at SMD(H2O)/B3LYP-D3/6-31+G(d,p) level. [e] SMD solvation energies calculated at SMD(H2O)/B3LYP/6-31G(d) level.

strated for acetaldehyde (**7**) as a well-characterized small reference system (Table 2), the hydration reaction of this system being endergonic by +9.8 kJ mol^{-1} at 298.15 K.[42,44] In order to address the effects of aqueous solvation appropriately, we employ a combination of continuum solvation models (here SMD) with different numbers of explicit water molecules (Figure 4). Analysis of the hydration energies of **7** with theoretical methods known to work well for the prediction of thermochemical data such as G3B3 or DLPNO-CCSD(T)/CBS shows this to be an endergonic process of around $\Delta G_{298} = +15$ kJ mol^{-1} (see Supporting Information for additional validation studies). The B3LYP-D3/6-31+G(d,p) hybrid DFT method employed here for geometry optimizations gives, in this case, a closely similar value. The addition of explicit water molecules as in **7(1 w)** or **7(2 w)** makes the reaction systematically less endergonic, and leads to a basically thermoneutral process in the presence of two explicit water molecules. This finding indicates that the hydration equilibrium in non-aqueous (or non-homogeneous) environments may be altered by specific hydrogen bonding interactions and may also provide a rationalization for the comparatively high levels of 5-formylcytosine hydrate in DNA duplex systems reported by Burrows et al.[36] The effects of bulk aqueous solvation have then been added with aid of the SMD continuum solvation model. This decreases the overall hydration energy and approaches the experimental value for the combination of the G3B3 compound scheme and two explicit water molecules. Similar validation steps have also been performed for benzaldehyde (**8**) as an aldehyde carrying an aromatic substituent and having a significantly less favorable hydration energy of $\Delta G_{298} = +21.1$ kJ mol^{-1}. Again, the gas phase hydration energy becomes more favorable with each explicitly considered water molecule, while the additional consideration of bulk solvation with the SMD model leads to a notable increase. We note, however, that all theory combinations considered here predict the hydration reaction to be less favorable than observed experimentally by approximately 10 kJ mol^{-1}. Using the same theoretical methods and solvation strategies as before, hydration energies have been calculated for the aldehydes shown in Figure 2 (Table 3). In addition to neutral 1-methyl-5-formylcytosine (**4 b**), this also includes its protonated form (**p_4 b**) (Figure 5).

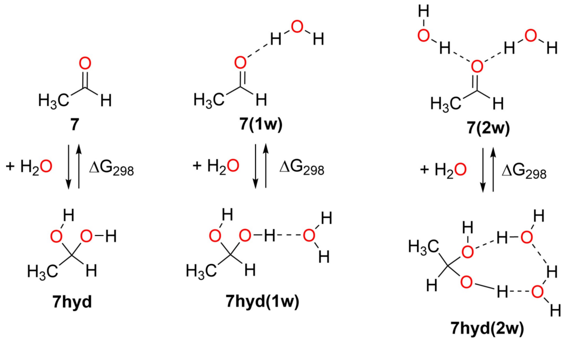


Figure 4. The hydration of acetaldehyde (**7**) in the absence and presence of solvating water molecules.

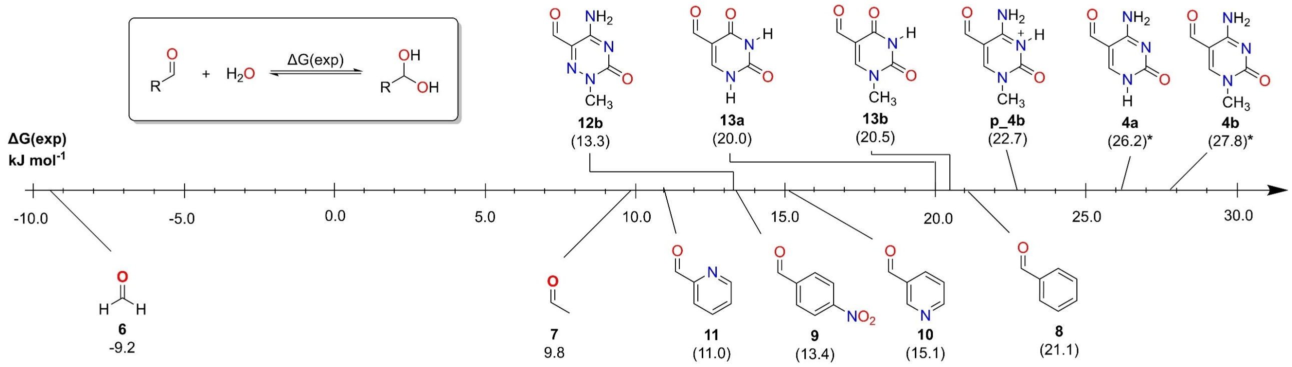


Figure 5. Experimental hydration free energies of selected aldehydes (* based on combination of the G3B3 $\Delta\Delta G_{298}$ values with the experimentally measured value for **p_4 b**) as the reference).

Table 3. Hydration Gibbs free energies (ΔG_{298}, in kJ mol^{-1}) for the aldehydes shown in Figure 2.

	Water (SMD model)						Exp.
	No explicit water molecules			One explicit water molecule			
	DFT[a,d]	CCSD(T)/CBS[b,c,d]	G3B3[e]	DFT[a,d]	CCSD(T)/CBS[b,c,d]	G3B3[e]	
6	−3.7	−4.3	−2.4	−13.6	−12.0	−5.1	−9.2
7	21.5	16.2	15.1	20.9	17.6	13.5	9.8
8	44.1	34.8	32.4	42.6	35.9	35.7	21.1
9	32.9	26.2	23.6	27.2	23.0	23.8	13.4
10	38.9	29.9	27.7	37.4	31.4	26.5	15.1
11	28.0	22.3	24.8	31.1	27.1	20.1	11.0
4a	52.0	40.0	45.7	53.6	44.6	39.3	>24.8/<27.8[f]
4b	54.4	42.3	46.6	55.3	46.1	42.7	>28.8
p_4b	42.3	36.3	36.3	42.5	35.8	36.1	22.7
12b	41.9	31.0	31.4	36.3	27.8	19.6	13.3
13a	38.5	34.3	29.0	39.1	35.6	31.3	20.0
13b	39.3	34.8	29.1	42.7	38.3	35.3	20.5

[a] B3LYP-D3/6-31+G(d,p). [b] Using gas phase B3LYP-D3/6-31+G(d,p) geometries. [c] Based on DLPNO-CCSD(T) single point calculations. [d] SMD solvation energies calculated at SMD(H2O)/B3LYP-D3/6-31+G(d,p) level. [e] SMD solvation energies calculated at SMD(H2O)/B3LYP/6-31G(d) level. [f] Calculated from LOQ and LOD.

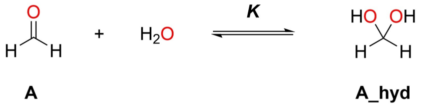
For most aldehydes considered here, the calculated hydration energies are overestimated to a similar extent as already observed for benzaldehyde (**8**). Due to the systematic nature of this phenomenon, good linear correlations can be observed between experimentally measured and theoretically calculated hydration energies in aqueous solution at SMD(H2O)/CCSD(T)/CBS or SMD(H2O)/G3B3 level with $R^2 = 0.95$ - 0.97. These correlations can be employed for an accurate estimate of the hydration energy difference between **4b** and its protonated form **p_4b**. Based on the values reported in Table 3 at the CBS or G3B3 level, this difference falls into the range of $\Delta\Delta G_{298} = 6.0$–10.3 kJ/mol. However, as already noted above, these values are generally somewhat too large and the correlations can be employed to scale these down to a more realistic value of $\Delta\Delta G_{298} = +5.1$ kJ mol^{-1}. Combination with the experimentally measured value of ΔG_{298}(**p_4b**) = +22.7 kJ mol^{-1}, this then yields ΔG_{298}(**4b**) = +27.8 kJ mol^{-1}, which is closely similar to the limiting value of ΔG_{298}(**4b**) = >28.8 kJ mol^{-1} derived from the ^{1}H NMR measurements. The same approach yields a theoretically predicted value for the free base of ΔG_{298}(**4a**) = +26.2 kJ mol^{-1}, being close to the experimental approximation derived from LOQ and LOD between +24.8–27.8 kJ mol^{-1}.

Conclusion

All experimental and theoretical studies presented here indicate that the hydrates of 5-formylcytosine (**4a**) and its N1-methylated derivative **4b** are just beyond the limit of what can be quantified reliably by ^{1}H NMR spectroscopy. The abundances of these species are expected to amount to less than 0.05 % under unbuffered standard conditions in water at ambient temperature. Protonation at the N3 position under acidic conditions will increase hydrate formation such that its direct detection through ^{1}H NMR spectroscopy becomes feasible at an abundance of 0.53 %. The hydrate form of 5-formyluridine as the formal deamination product of 5-formylcytosine is more abundant at 1.3 % under neutral aqueous conditions. Despite these seemingly low values, the aldehyde hydrate forms may nevertheless play an essential role in oxidation reactions to the respective 5-carboxy derivatives in a way well established for aldehyde oxidations mediated by chemical oxidants or dehydrogenase enzymes.[38,54–56] In both areas evidence for the stabilization of hydrate intermediates through directed hydrogen bonding interactions has been found, which is in full support of the gas phase calculations with explicit water molecules in the current study. This may also provide a rational basis for the proposed high abundance of 5-formylcytosine hydrates reported by Burrows et al. in base-flipping kinetics studies.[36] A potential TET-mediated oxidation of fC through the respective hydrate[38] moves this process mechanistically closer to that of hmC, where recent theoretical studies have established similar reaction barriers for initial O–H vs. C–H hydrogen abstraction steps.[19,57]

Experimental Section

Energy of hydration ΔG_{hyd}: The reaction of aldehydes (**A**) with water in aqueous solution yields the respective hydrate **A_hyd** according to Eq. (1). The position of this equilibrium is given through equilibrium constant *K*, which is defined through the equilibrium concentrations of reactants and products according to Eq. (2). In dilute solutions it is practical to consider the concentration of water as a constant with $[H_2O] = 55.5$ mol l^{-1} and combine this value with *K* into a new equilibrium constant K_w (sometimes also called K_{hyd}) according to Eq. (3). The true equilibrium constant *K* can then be obtained from experimentally measured K_w values according to Eq. (4). According to the law of mass action, the equilibrium constant *K* relates to the free energy of the reaction shown in Eq. (1) as defined in Eq. (5).



(1)

$$K = \frac{[A_{hyd}]}{[A] \times [H_2O]} \tag{2}$$

$$K \times [H_2O] = K_w = \frac{[A_{hyd}]}{[A]} \tag{3}$$

$$K = \frac{K_w}{[H_2O]} \tag{4}$$

$$\Delta G = -RT \ln(K) \tag{5}$$

^{18}O Isotopic exchange experiments: Isotope exchange experiments were performed with an Advion ExpressionL compact mass spectrometer (CMS) by using the atmospheric solid analysis probe (ASAP) technique; 350 m/z, and acquisition speed 10 000 m/z units per second. The ion source settings correspond to a capillary temperature of 250 °C, capillary voltage 110 V, source offset voltage 16 V, APCI source gas temperature 350 °C, and corona discharge voltage 4 μA. The obtained spectra were analyzed by using Advion CheMS Express software version 5.1.0.2. A trace amount of **13 b** and **4 b** was dissolved in 50 μL of ^{18}O isotopically labeled water (97 atom %) under nitrogen atmosphere in an oven-dried GC vial and shaken for 1 h. The glass capillary of the ASAP probe was used quickly under hot conditions to exclude ambient moisture contamination. After background subtraction, the corresponding MS spectrum was obtained. For details see Supporting Information (S.3).

Quantum chemical calculations: Geometry optimization was performed at the B3LYP-D3/6-31+G(d,p) level of theory in gas phase.[58–63] The solution state was modelled both through addition of explicit water molecules and through the implicit continuum solvation model (SMD).[60] Free energies in solution are referenced to a standard state of 1 M through consideration of a standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91\ \mathrm{kJ\,mol^{-1}}$. Single point energies were calculated for the optimized geometries using the DLPNO-CCSD(T) method.[64–66] Two-point (cc-pVTZ and cc-pVQZ) extrapolation was employed at the DLPNO-CCSD(T) level of theory to estimate a result obtained using a complete (infinitely large) basis set.[66] The isotropic chemical shielding values were calculated at the SMD(H_2O)/B3LYP/pcS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.[59] The ^{1}H chemical shifts were referenced relative to chemically and structurally similar molecules (see Supporting Information). All calculations were performed using Gaussian 09, Revision D.01.[68] To identify the conformations of diol molecules, a relaxed potential energy surface scan on two dihedral angles H–O–C–C (two hydroxyl groups) was performed at the SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level. The conformations with the lowest energies on the potential energy surface were then fully optimized at the SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level. The optimized water-complexed geometries have been located by a stochastic search procedure. This procedure generates an ensemble of initial random arrangements of water molecules around the respective structure, whose optimization at B3LYP-D3/6-31+G(d,p) level then generates the minima used for all quantitative work.[69,70]

Synthesis of 4 b: A solution of 1-methyl-5-hydroxymethylcytosine (119 mg, 0.77 mmol) and activated MnO_2 (333 mg, 3.84 mmol, 5eq) in 12 ml of anhydrous acetonitrile was stirred at room temperature for 20 h. The reaction mixture was then diluted with methanol (10 mL) and filtered (washed with methanol). The crude product was purified by flash column chromatography over silica gel (12 % to 15 % MeOH/5 % NH_4OH/DCM) to afford a clean white powder (35 mg, 0.23 mmol, 30 %). R_f (15 % MeOH/5 % NH_4OH/DCM) = 0.54. ^{1}H-NMR (400 MHz, DMSO–D6, ppm): δ = 9.41 (s, 1H, formyl-*H*), 8.64 (s, 1H, C_6-*H*), 7.98 (br-s, 1H, N-H_2), 7.79 (br-s, 1H, N-H_2), 3.37 (s, 1H, –CH_3). ^{13}C-NMR (100 MHz, DMSO–D6, ppm): δ = 188.1, 162.7, 160.2, 153.9, 104.3, 37.7. ^{1}H-NMR (400 MHz, D_2O, ppm): δ = 9.45 (s, 1H, formyl-*H*), 8.46 (s, 1H, C_6-*H*), 3.49 (s, 1H, –CH_3). ^{13}C-NMR (100 MHz, D_2O, ppm): δ = 189.9, 163.0, 160.4, 157.0, 105.4, 38.3. EA: Calculated [%]: C: 47.06 N: 27.44 H: 4.61; Found [%]: C: 47.90 N: 26.09 H: 4.36. HR-MS (EI, 70 eV, M+): [C6H7N3O2+], calculated: 153.0533, found: 153.0533.

Synthesis of 13 b: 1,5-Dimethyluracil (0.50 g, 3.57 mmol) was dissolved in 170 mL distilled water, then $K_2S_2O_8$ (1.93 g, 7.14 mmol, 2 eq) was added portion-wise over 1 h at 85 °C and the reaction mixture was stirred for 16 h. After the TLC showed complete conversion of the reactant, the reaction mixture was cooled down to room temperature and the solvent was removed under high vacuum. The crude product was purified by flash column chromatography over silica gel (3 %MeOH/2 %AcOH/DCM to 10 % MeOH/2 %AcOH/DCM) to afford a clean white powder (302.63 mg, 1.96 mmol, 55 %). ^{1}H-NMR (400 MHz, DMSO–D6, ppm): δ = 9.76 (s, 1H, formyl-*H*), 8.48 (s, 1H, C_5-*H*), 3.65 (br-s, 1H, N-*H*), 3.36 (s, 3H, –CH_3). ^{13}C-NMR (100 MHz, DMSO–D6, ppm): δ = 186.7, 163.0, 153.1, 151.0, 110.3, 36.8. ^{1}H-NMR (400 MHz, D_2O, ppm): δ = 9.58 (s, 1H, formyl-*H*), 8.44 (s, 1H, C_6-*H*), 3.46 (s, 3H, –CH_3). ^{13}C-NMR (100 MHz, D_2O, ppm): δ = 186.7, 163.0, 153.1, 151.0, 110.3, 36.8. EA: Calculated [%]: C: 46.76 N: 18.18 H: 3.92; Found [%]: C:48.49 N: 18.20 H: 3.98. HR-MS (EI, 70 eV, [M+]): [C6H6N2O3+], calculated: 154.0373, found: 154.0372.

Acknowledgements

*We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for financial support via SFB1309 (PID 325871075). We also thank Felix Müller and Sophie Gutenthaler for helping with the purification of compounds and Stella Marie Bauer for the assistance in compound synthesis, as well as the Daumann group for a sample of 5fdC (**4 c**). Open Access funding enabled and organized by Projekt DEAL.*

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Keywords: aldehyde hydrates · computational chemistry · DNA methylation · epigenetics · modified nucleic acids · TET enzymes

[1] K. D. Robertson, *Nat. Rev. Genet.* **2005**, *6*, 597–610.

[2] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315–322.

[3] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010–1022.

[4] E.-A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, *Nat. Chem. Rev.* **2017**, *1*, 0069.

[5] A. Bird, *Genes Dev.* **2002**, *16*, 6–21.
[6] M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **2005**, *74*, 481–514.
[7] N. Bhutani, D. M. Burns, H. M. Blau, *Cell* **2011**, *146*, 866–872.
[8] E. Kriukienė, V. Labrie, T. Khare, G. Urbanavičiūtė, A. Lapinaitė, K. Koncevičius, D. Li, T. Wang, S. Pai, C. Ptak, J. Gordevičius, S.-C. Wang, A. Petronis, S. Klimašauskas, *Nat. Commun.* **2013**, *4*, 2190.
[9] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[10] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
[11] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1133.
[12] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
[13] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[14] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150.
[15] N. S. W. Jonasson, R. Janßen, A. Menke, F. L. Zott, H. Zipse, L. J. Daumann, *ChemBioChem* **2021**, *22*, 3333–3340.
[16] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472–479.
[17] L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo, Y. Xu, *Nature* **2015**, *527*, 118–122.
[18] D. J. Crawford, M. Y. Liu, C. S. Nabel, X.-J. Cao, B. A. Garcia, R. M. Kohli, *J. Am. Chem. Soc.* **2016**, 138, 730–733.
[19] J. Lu, L. Hu, J. Cheng, D. Fang, C. Wang, K. Yu, H. Jiang, Q. Cui, Y. Xu, C. Luo, *Phys. Chem. Chem. Phys.* **2016**, 18, 4728–4738.
[20] M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X.-J. Cao, B. A. Garcia, G. A. Cisneros, R. M. Kohli, *Nat. Chem. Biol.* **2017**, 13, 181–187.
[21] J. E. DeNizio, M. Y. Liu, E. M. Leddin, G. A. Cisneros, R. M. Kohli, *Biochemistry* **2019**, 58, 411–421.
[22] S. O. Waheed, S. S. Chaturvedi, T. G. Karabencheva-Christova, C. Z. Christov, *ACS Catal.* **2021**, 11, 3877–3890.
[23] M. B. Berger, A. R. Walker, E. A. Vazquez-Montelongo, G. A. Cisneros, *Phys. Chem. Chem. Phys.* **2021**, 23, 22227–22240.
[24] B. A. Caldwell, M. Y. Liu, R. D. Prasasya, T. Wang, J. E. DeNizio, N. A. Leu, N. Y. A. Amoh, C. Krapp, Y. Lan, E. J. Shields, R. Bonasio, C. J. Lengner, R. M. Kohli, M. S. Bartolomei, *Mol. Cell* **2021**, *81*, 859–869.
[25] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.
[26] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.
[27] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
[28] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72–78.
[29] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kuśnierczyk, C. B. Vågbø, D. Schuermann, A. Klungland, P. Schär, *Nat. Commun.* **2016**, *7*, 10806.
[30] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049–1055.
[31] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555–557.
[32] S. Waschke, J. Reefschläger, D. Bärwolff, P. Langen, *Nature* **1975**, *255*, 629–630.
[33] S. Bjelland, L. Eide, R. W. Time, R. Stote, I. Eftedal, G. Volden, E. Seeberg, *Biochemistry* **1995**, *34*, 14758–14764.
[34] K. Kemmerich, F. A. Dingler, C. Rada, M. S. Neuberger, *Nucleic Acids Res.* **2012**, *40*, 6016–6025.
[35] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574–581.
[36] R. P. Johnson, A. M. Fleming, R. T. Perera, C. J. Burrows, H. S. White, *J. Am. Chem. Soc.* **2017**, *139*, 2750–2756.
[37] R. C. A. Dubini, A. Schön, M. Müller, T. Carell, P. Rovó, *Nucleic Acids Res.* **2020**, *48*, 8796–8807.
[38] S. Sappa, D. Dey, B. Sudhamalla, K. Islam, *J. Am. Chem. Soc.* **2021**, *143*, 11891–11896.
[39] S. H. Hilal, L. L. Bornander, L. A. Carreira, *QSAR Comb. Sci.* **2005**, *24*, 631–638.
[40] S. Huang, A. K. Miller, W. Wu, *Tetrahedron Lett.* **2009**, *50*, 6584–6585.
[41] D. K. Rogstad, J. Heo, N. Vaidehi, W. A. Goddard, A. Burdzy, L. C. Sowers, *Biochemistry* **2004**, *43*, 5688–5697.
[42] R. E. Hardisty, F. Kawasaki, A. B. Sahakyan, S. Balasubramanian, *J. Am. Chem. Soc.* **2015**, *137*, 9270–9272.
[43] J. P. Guthrie, *Can. J. Chem.* **1978**, *56*, 962–973.
[44] R. P. Bell, in *Advances in Physical Organic Chemistry, Vol. 4* (Ed.: V. Gold), Academic Press, **1966**, pp. 1–29.
[45] J. L. Kurz, *J. Am. Chem. Soc.* **1967**, *89*, 3524–3528.
[46] L. C. Gruen, P. T. McTigue, *J. Chem. Soc.* **1963**, 5217–5223.
[47] E. Lombardi, P. B. Sogo, *J. Chem. Phys.* **1960**, *32*, 635–636.
[48] P. Greenzaid, *J. Org. Chem.* **1973**, *38*, 3164–3167.
[49] J. Sayer, *J. Org. Chem.* **1975**, *40*, 2545–2547.
[50] S. Cabani, P. Gianni, E. Matteoli, *J. Phys. Chem.* **1972**, *76*, 2959–2966.
[51] S. Barman, K. L. Diehl, E. V. Anslyn, *RSC Adv.* **2014**, *4*, 28893–28900.
[52] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591–5594; *Angew. Chem.* **2020**, *132*, 5639–5643.
[53] K. J. R. Rosman, P. D. P. Taylor, *Pure Appl. Chem.* **1998**, *70*, 217–235.
[54] R. Gómez-Bombarelli, M. González-Pérez, M. T. Pérez-Prior, E. Calle, J. Casado, *J. Phys. Chem. A* **2009**, *113*, 11423–11428.
[55] L. P. Olson, J. Luo, Ö. Almarsson, T. C. Bruice, *Biochemistry* **1996**, *35*, 9782–9791.
[56] A.-K. C. Schmidt, C. B. W. Stark, *Org. Lett.* **2011**, *13*, 4164–4167.
[57] A. J. K. Roth, M. Tretbar, C. B. W. Stark, *Chem. Commun.* **2015**, *51*, 14175–14178.
[58] H. Torabifard, G. A. Cisneros, *Chem. Sci.* **2018**, *9*, 8433–8445.
[59] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[60] F. Jensen, *J. Chem. Theory Comput.* **2008**, *4*, 719–727.
[61] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
[62] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[63] T. J. Zuehlsdorff, C. M. Isborn, *J. Chem. Phys.* **2018**, *148*, 024110.
[64] A. Nicolaides, A. Rauk, M. N. Glukhovtsev, L. Radom, *J. Phys. Chem.* **1996**, *100*, 17460–17464.
[65] F. Neese, E. F. Valeev, *J. Chem. Theory Comput.* **2011**, *7*, 33–43.
[66] M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2017**, *146*, 164105.
[67] A. Altun, F. Neese, G. Bistoni, *J. Chem. Theory Comput.* **2019**, *15*, 215–228.
[68] A. Altun, F. Neese, G. Bistoni, *Beilstein J. Org. Chem.* **2018**, *14*, 919–929.
[69] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 09, Rev. D.01, Wallingford, CT, **2009**.
[70] M. Saunders, *J. Comb. Chem.* **2004**, *25*, 621–626.
[71] D. Šakić, V. Vrček, *J. Phys. Chem. A* **2012**, *116*, 1298–1306.

Manuscript received: November 25, 2021
Revised manuscript received: January 24, 2022
Accepted manuscript online: January 27, 2022
Version of record online: February 10, 2022