Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Modified Pyrimidine Nucleobases and their Chemical Properties

-

A Theoretical and Experimental Study

Fabian Leonhard Zott

aus

Bobingen, Deutschland

2023

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Hendrik Zipse betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 03.02.2023 .....

Fabian Leonhard Zott

Dissertation eingereicht am 06.02.2023

1. Gutachter: Prof. Dr. Hendrik Zipse

2. Gutachter: Prof. Dr. Lena Daumann

Mündliche Prüfung am 14.03.2023

Acknowledgment

Mein Dank gilt **Prof. Dr. Hendrik Zipse**, der mir die Möglichkeit gegeben hat, unter seiner Ägide aber mit vielen Freiheiten ein interessantes, wissenschaftliches Feld zu beleuchten.

Außerdem gilt mein Dank **Prof. Dr. Lena Daumann** mit der ich nicht nur in den letzten vier Jahren erfolgreich kooperieren durfte, sondern sich auch dazu bereit erklärt hat als Zweitgutachterin meiner Dissertation aufzutreten.

Es ist unwürdig, die Zeit von hervorragenden Leuten
mit knechtischen Rechenarbeiten zu verschwenden,
weil bei Einsatz einer Maschine auch der Einfältigste
die Ergebnisse sicher hinschreiben kann.

- G.W. VON LEIBNIZ -

... FOR AFTERWARDS A MAN FINDS PLEASURE IN HIS PAINS,
WHEN HE HAS SUFFERED LONG AND WANDERED FAR.

- HOMER -

List of publications

Parts of this thesis have been published as follows:

Niko S. W. Jonasson, Rachel Janßen, Annika Menke, Fabian L. Zott, Hendrik Zipse, Lena J. Daumann; **"TET-Like Oxidation in 5-Methylcytosine and Derivatives: A Computational and Experimental Study"**; *ChemBioChem* **2021**, 22, 3333 – 334.

Fabian L. Zott, Vasily Korotenko, Hendrik Zipse, **"The pH-Dependence of the Hydration of 5-Formylcytosine: an Experimental and Theoretical Study"**; *ChemBioChem* **2022**, 23, e2021006.

Other publications:

Marta Marin-Luna, Benjamin Poelloth, Fabian Zott, H. Zipse; **"Size-Dependent Rate Acceleration in the Silylation of Secondary Alcohols: the Bigger the Faster"**; *Chem. Sci.* **2018**, 9, 6509 – 6515.

Tobias Blockhaus, Christian Klein-Heßling, Peter M. Zehetmaier, Fabian L. Zott, Harish Jangra, Konstantin Karaghiosoff, Karlheinz Sünkel; "**Ferrocenes with a Persulfurated Cyclopentadienyl Ring: Synthesis, Structural Studies, and Optoelectronic Properties"**, *Chem. Eur. J.* **2019**, 25, 12684 – 12688.

Tobias Blockhaus, Fabian L. Zott, Peter M. Zehetmaier, Christian Klein-Heßling, Karlheinz Sünkel; **"Tricyanated Ferrocenes; Isolation, structure, electrochemistry and DFT calculations"**; *Z. Anorg. Allg. Chem.* **2022**, e202200277.

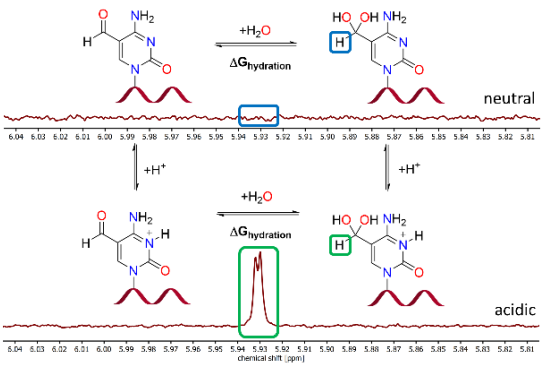
Tobias Blockhaus, Fabian Zott, Karlheinz Sünkel; **"1,1'-Dibromo-2,2',5,5'-tetrakis(methylthio)ferrocene"**; Molbank **2023**; M1542.

CONTENTS

1 SUMMARY

This thesis was performed in the DFG-funded collaborative research centre ("Sonderforschungsbereich: SFB 1309"). The "SFB 1309" focuses on chemical biology of epigenetic modifications, trying to understand the biochemical processes related to the field of epigenetics as the second layer of genetic information. In part this involves the development of molecules that can manipulate and influence these in order to establish new medical intervention possibilities.

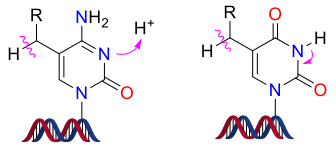
Chapter 3. The *pH*-Dependence of the Hydration of 5-Formylcytosine



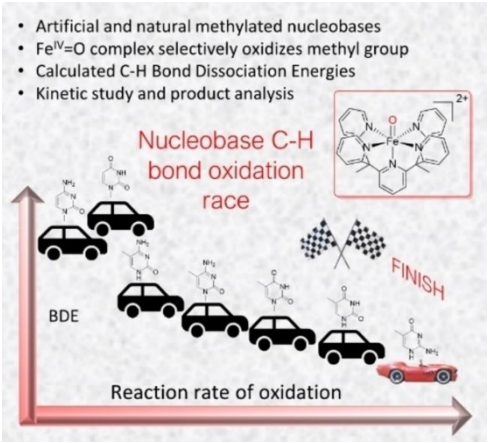
5-Formylcytosine is an important nucleobase in epigenetic regulation, whose hydrate form has been implicated in the formation of 5-carboxycytosine as well as oligonucleotide binding events. The hydrate content of 5-formylcytosine and its uracil derivative has now been quantified using a combination of 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.

Chapter 4. Acid-Base Properties of Cytosine and Uracil Derivatives

The acid and base properties of modified pyrimidine bases are of great importance to understand many enzymatic processes as well as genome stability influencing epigenetic regulation. Precise theoretical determination of pK_a values in combination with bond dissociation energy calculations (BDE) can shed light on the intrinsic chemical reactivity of 5-methylcytosine as well as its oxidated products. Special emphasis is given to the hydrate form of 5-formylcytosine, believed to play an important role as intermediate in the TET oxidation of 5-formylcytosine towards 5-caboxycytosine.



Chapter 5. TET-Like Oxidation in 5-Methylcytosine and Derivatives



The epigenetic marker 5-methylcytosine (5mC) is an important factor in DNA modification and epigenetics. It can be modified through a three-step oxidation performed by ten-eleven-translocation (TET) enzymes and it has been previously reported that the iron(IV)-oxo complex $[Fe(O)(Py_5Me_2H)]^{2+}$ **1** can oxidize 5mC. Here, we report the reactivity of this iron(IV)-oxo complex towards a wider scope of methylated cytosine and uracil derivatives relevant for synthetic DNA applications, such as 1-methylcytosine (**1mC**), 5-methyl-*iso*-cytosine (**5m*i*C**) and thymine (**T**/**5mU**). The observed kinetic parameters are

corroborated by calculation of the C−H bond energies at the reactive sites which was found to be an efficient tool for reaction rate prediction of **1** towards methylated DNA bases. We identified oxidation products of methylated cytosine derivatives using HPLC-MS and GC-MS. Thereby, we shed light on the impact of the methyl group position and resulting C−H bond dissociation energies in TET-like oxidation reactions.

2 INTRODUCTION

2.1 EPIGENETICS

While the term epigenetics was first introduced in 1942 by Conrad Waddington describing the causal interactions between genes and its products after phenotypic expression, this term has changed and somewhat been redefined over the years. [1] Today a consensus exits that "epigenetics is the collective heritable changes in phenotype due to processes that arise independent of primary DNA sequence".[2] This phenomenon is not only restricted to a single generation since it has been found that epigenetic processes can be translated over one generation.[3-4] Epigenetics have profound impact onto many aspects of life, like embryonic development, aging, cancer, to mention a few.[2] Besides histone and chromatin modification, one particularly interesting field is the DNA methylation, which plays a key role in differentiation as well as maintenance of cell type identity.[5-7] DNA methylation is a major epigenetic factor influencing gene activities.

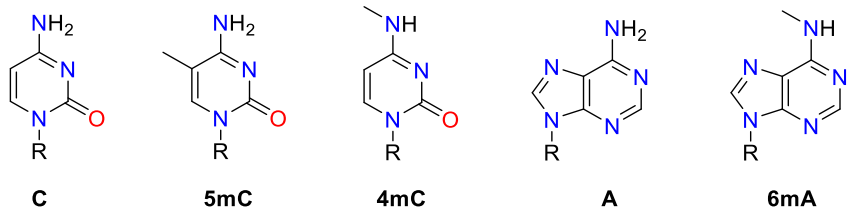


Figure 1: Representation of DNA bases **C** and **A** that get methylated by DNA methyltransferases (DNMT) to afford **5mC**, **4mC** and **6mA**.

The methylation of DNA in the genome by the DNA methyltransferase (DNMT) family of enzymes is one of the most important epigenetic modifications and represents a direct chemical modification.[8] Naturally, only cytosine (**C**) and adenine (**A**) nucleobases are methylated in DNA to afford 5-methylcytosine (**5mC**), N^4-methylcytosine (**4mC**) and N^6-methyladenine (**6mA**, see Figure 1). The brain, containing the highest level on DNA methylation of any tissue in the human body, it contains ~1% of **5mC** in human brain cell genome.[9] The majority of these methylated cytosines are also known to be located preceding a guanidine nucleoside, so called CpG sites.[10]

2.2 5-Methylcytosine and its Active and Passive Demethylation

There are three different classes of enzymes that establish, recognize and remove DNA bases: writers, erasers and readers. Writers can catalyse the methylation of cytosine residues. Erasers modify and remove the methyl groups. Readers recognize and bind to the methyl modifications to eventually influence gene expression, resulting in epigenetic regulation.[11]

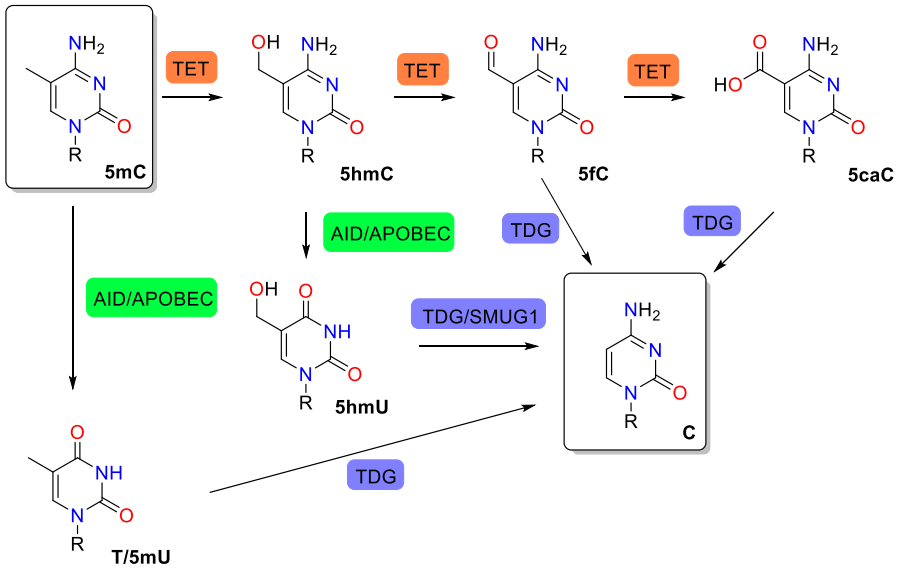


Figure 2: Possible pathways of active demethylation of **5mC** to **C** by modification of the methyl group at 5-position with the TET enzyme (TET), by active deamination of the 4-position with activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex (AID/APOBEC) and by base excision repair (BER) mechanism using the DNA glycosylase (TDG).

The demethylation of DNA can either be active or passive. Passive demethylation occurs as the Dnmt1 actively maintains DNA methylation during cell replication. Active DNA methylation however can occur in dividing as well as non-dividing cells. This process involves active chemical modifications by enzymatic reactions at the nucleoside level in order to remove the methyl group at the 5-position of **5mC.**[12-15] Since the direct cleavage of the strong and covalent C-C bond between the cytosine backbone and methyl group is thermodynamically very unfavourable, up to now no mechanism in mammalian cells is known that can achieve this. Due to this fact, other pathways have been explored by evolution (see Figure 2). Active demethylation can occur through a series of sequential oxidation steps of **5mC** to **5caC** via **5hmC** and **5fC**, catalysed by the ten-eleven translocation (TET) family of enzymes (see Figure 2), followed by the base excision repair (BER) mechanisms that recognizes **5caC** and **5fC** via the DNA repair protein thymidine DNA glycosylase (TDG).[16-18] Among these oxidation products of **5mC**, **5hmC** is considered a stable modification and accounts for around 1-10% of **5mC** in different cell types, while **5fC** and **5caC** are 100- and 1000-fold less abundant.[19]

Other mechanisms that involve a BER step at the end is the chemical modification of the amine group of **5mC** and **5hmC** by the activation-induced cytidine deaminase/apolipo-protein B mRNA-editing enzyme complex (AID/APOBEC) to form thymine (**T**/**5mU**) and 5-hydroxymethyluracil (**5hmU**).[20]

Another possible pathway for the active demethylation involves direct decarboxylation and deformylation of **5caC** and **5fC**, thus avoiding the possible damage of DNA strands due to genome instability by the BER mechanism. Evidence for the direct deformylation in mammalian cells has been reported recently.[18,21-24] Feng *et al.* claim to have uncovered the direct decarboxylation of **5caC** mammalian cells as a rapid process by treating **5caC**-containing genomic DNA with nuclear extracts as well as utilizing 2'-fluorinated **F-5caC** metabolic labelling together with mass spectroscopy analysis.[25] An enzyme facilitating this process has yet to be discovered.

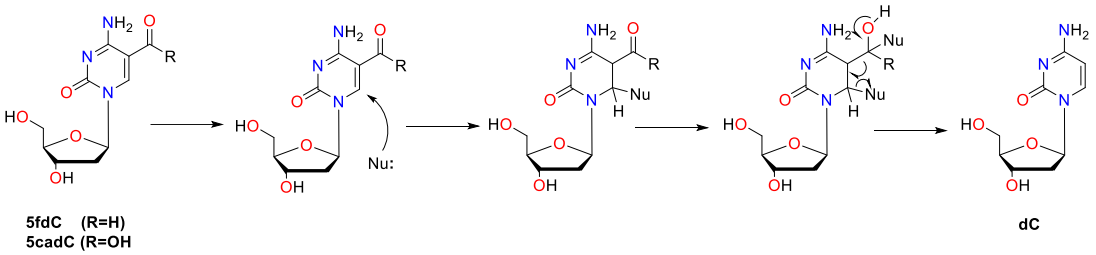


Figure 3: Proposed mechanism for the direct decarboxylation of deoxy-5-formylcytidine (**5fdC**) and deoxy-5-caboxycytidine (**5cadC**) to deoxycytidine (**C**) by nucleophilic attack at the 6-positon of the pyrimidine base.[21]

A proposed mechanism (Figure 3) for the direct deformylation as well as decarboxylation requires the activation of the nucleobase by a nucleophilic attack S_NAr at the C6 position of the pyrimidine base.[21] The C6 attack at pyrimidine bases with a helper nucleophile is well known as the central mechanistic process during the DNA methylation of **C** by DNA methyl transferases (DNMT).[26-27] The DNMT methylation is enabled by C6-attack of cysteine, protonation by a glutamate residue and consequent methylation by S-adenosyl-L-methionine (SAM, see Figure 4). In fact, when treating **5fdC** with bisulfite, being a cysteine (cys) attack model, rapid deformylation and deamination was observed.[21, 28] This deamination process is known for **5mC** and actively used in so called "bisulfite sequencing".[29]

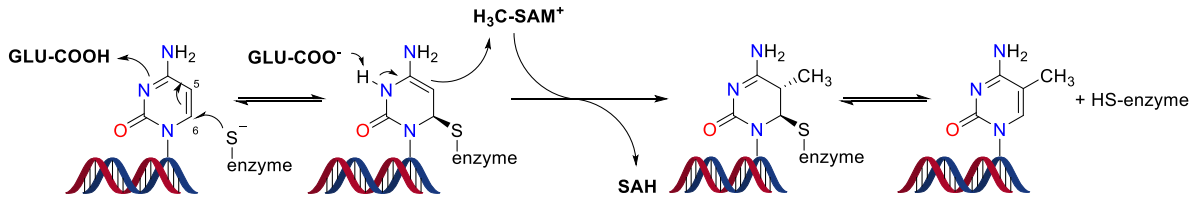


Figure 4: Mechanism of the DNMT enzyme: Attack at the C6 position by cysteine facilitates the methylation at the C5 position of **C** with S-adenosyl-L-methionine (SAM) to afford **5mC**. Glutamate (GLU-COOH) residue acts as acid/base catalyst.[30]

A simplified mechanism for DNMT enzymes is shown in Figure 4, emphasizing the importance of protonation steps in the catalytic cycle of biochemical processes. The pK_a values of nucleobases and their modifications at the N3-position is thus of high importance, enabling and accelerating enzymatic reactions.

2.3 The TET Enzymes and their Function

In 2003 the ten-eleven translocation (TET) methylcytosine dioxygenases were first described in a study on acute myeloid leukaemia.[31] Since then, they have been associated with a variety of solid cancers and other myeloid and lymphoid malignancies.[32] They also play a key role in other oncogenic transformations and are regulating the self-renewal capacities of a wide range of stem cell types.[33] The main catalytic function was uncovered in 2009 by Rao *et al.* describing the conversion of **5mC** in three consecutive oxidative steps towards **5caC** (Figure 2) and play a key role in the active demethylation of **5mC** already described above (paragraph 2.2).[34] It is known that TET enzymes can also oxidize unnatural cytosine derivatives, like 5-ethylcytosine (**5etC**), 5-vinylcytosine (**5vC**) and 5-ethinylcytosine (**5etyC**).[35-37]

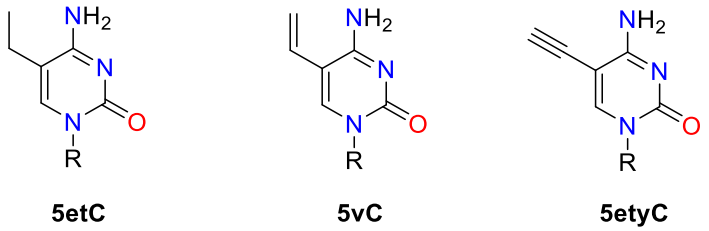


Figure 5: Structure of unnatural cytosine derivatives that can be oxidized by the TET enzyme family (R = ribose).

The structure of a TET enzyme was first revealed in 2013 and its mechanism is well described in the literature.[38] The active site involves a α-ketoglutarate dependent iron centre that coordinates two histidine and one asparagine residue (Figure 6).

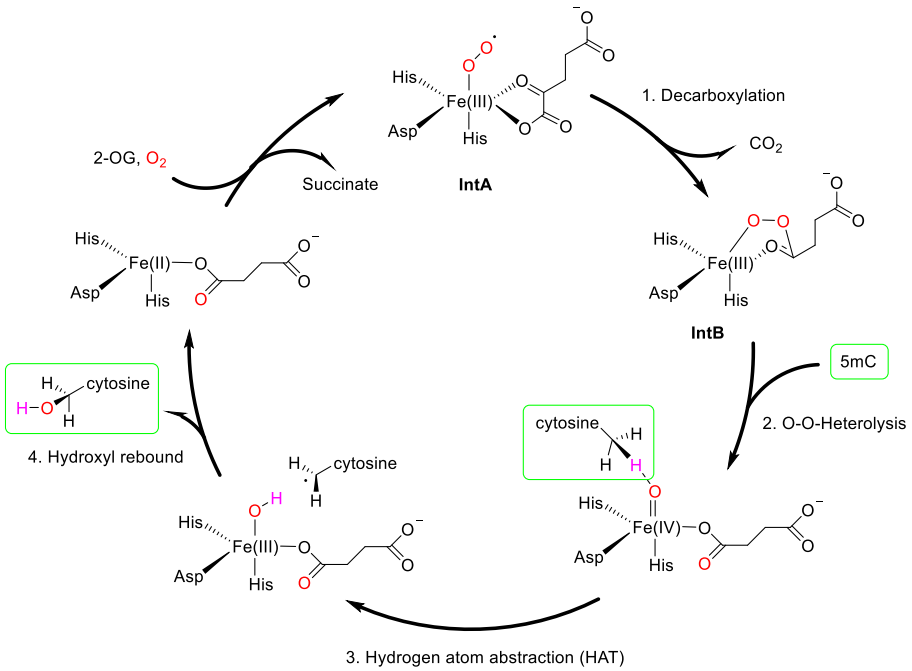


Figure 6: Model for the oxidative reactions catalysed by TET proteins (Asp = asparagine, His = histidine).[38,39]

In step one of the catalytic cycle of the TET-mediated oxidation of **5mC** the α-ketoglutarate and O_2 coordinating iron(II) intermediate **IntA** undergoes decarboxylation to form a peroxyacid coordinating succinate residue (Figure 6).[39] This intermediate **IntB** complex then undergoes a O-O-bond heterolysis to afford an reactive Fe(IV)-oxo complex, which then is followed by a rate determining hydrogen atom abstraction (HAT) step from the target substrate (**5mC**). The resulting Fe(III)-OH and substrate radical then undergo a final hydroxyl rebound step, forming the oxidised product. The resulting Fe(II) complex is subsequently loaded again by α-ketoglutarate and oxygen under release of succinate. The different substrates, **5mC**, **5fC** and **5caC**, are having no effect on the formation of the active Fe(IV)=O species, but rather the hydrogen abstraction step during product formation.[39] The intermediates **IntA** and **IntB** have never been observed directly but are proposed in the literature.[40-41] Experimental results of the steady-state kinetic analyses for the TET2-mediated oxidation of **5mC**/**5hmC**/**5fC**-DNA substrates reveals that **5mC**-DNA is more active than **5hmC**/**5fC**-DNA. This is a result of the distinct intrinsic properties of **5mC**/**5hmC**/**5fC** within the catalytic cavity of TET proteins during oxidation.[42] The experimental barriers $\Delta G^{\ddagger}$ that can be derived from the k_{cat} and K_m values by Michaelis-Menten analysis show only small differences for the three different substrates. The theoretically determined barriers for the rate determining HAT step, however, show that **5fC** is the most unfavourable substrate with a up to +36.0 kJ mol^{-1} higher barrier difference compared to **5hmC** and **5fC**.[39]

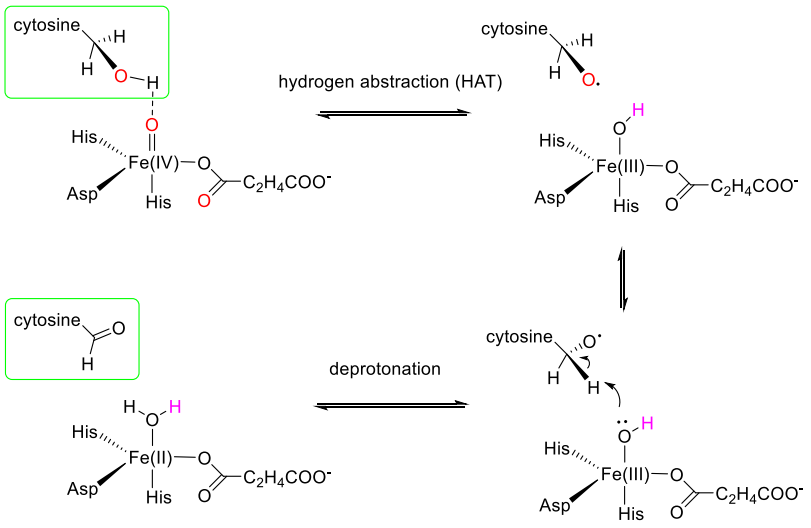


Figure 7: Proposed substrate specific catalytic pathway for the oxidation of **5hmC** with the wild type of TET enzyme. [43]

It should be mentioned that in a computational investigation into the TET oxidation of **5hmC** to **5fC** as alternative catalytic pathway with hydrogen atom abstraction at the hydroxy group of **5hmC** was proposed (Figure 7).[43] The highly reactive iron oxo(IV) species activates the O-H bond of the substrate, after which the substrate reorients in the active site. This reduces the distance between the second hydrogen atom and the basic Fe(III)-OH species, which induces a H-atom transfer from the methyl-oxo radical moiety in a subsequent step to form the product **5fC**.

2.4 Modified Cytosine Derivatives in Epigenetics

Figure 8 lists all possible oxidation products of 5-methylcytosine (**5mC**) as well as those of thymine (**T**, **5mU**). The modified uracil bases represent formal deamination products of their cytosine analogues.

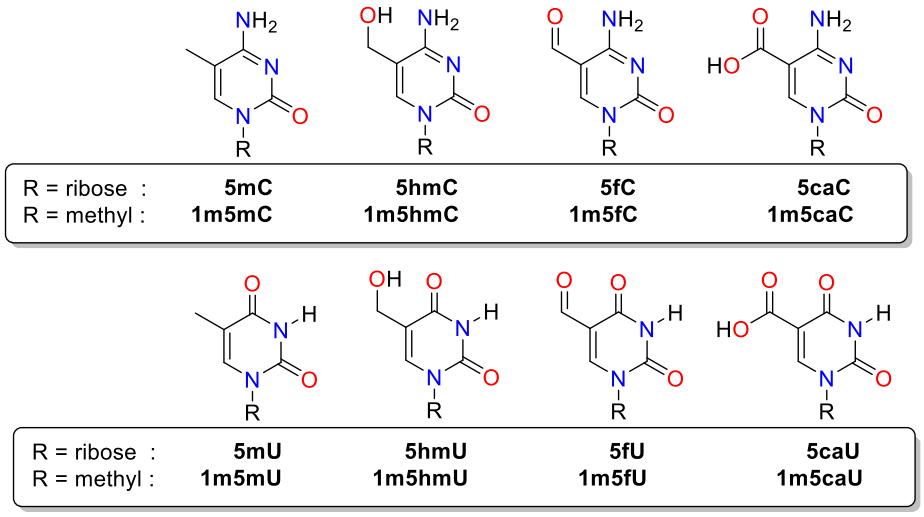


Figure 8: Epigenetic modifications of cytosine (top) and its deamination products (bottom). The nomenclature of N1-methylated pyrimidines will be used in later paragraphs as model systems in computational studies.

With **5mC** being the most prevalent modified nucleoside accounting between 1-10% in different human cell types, its epigenetic function is well described.[19] Sometimes described as the fifth DNA base, it is important in silencing the transcription in the genome. **5mC** in presence of promoters can recognize methyl CpG sites, leading to a stable and long-lived transcriptional silencing.[44] Its associated with gene expression, repression of nearby genes, mediating genomic imprinting as well as silencing viral DNA elements.[45-48] In certain cases **5mC** is linked with positive effects in transcription.[49] With embryonic stem cells having a very high level of DNA methylation, it has been shown that up to 25% of **5mC** is not linked to a CpG site, suggesting an even more important role in early life development that has yet to be explored.[49-50]

Less is known about its oxidation product 5-hydroxmethylcytosine (**5hmC**) being 100 times less abundant than **5mC**. Highest levels of **5hmC** are recorded in mouse brain cells as well as in mouse embryonic stem cells being ~7-10% as abundant as **5mC** for the latter.[51-52] Sometimes proclaimed to be the sixth DNA base, **5hmC** plays an important role as intermediate for the active demethylation pathway by TET oxidation and following BER mechanism described in paragraph 2.2. However, **5hmC** was found to be more sensitive to deamination by AID, opening up its own demethylation pathway followed by base excision (BER) of 5-hydroxymethyluracil (**5hmU**).[53] By directing the dynamic remodelling and organization of chromatin structures it plays a profound role in gene transcription.[54] In 2010 it was reported that specific methyl-CpG-binding domain proteins exist that do not bind to **5mC**, but **5hmC** *in vitro*.[55] All these findings indicate that **5hmC** is not merely an intermediate of DNA methylation, but a stable and independent epigenetic marker directly influencing DNA structure and function. With its regulatory functions, it is critical in neurodevelopment and diseases in mice as well

as human.[56] It should be mentioned that **5hmC** can also be formed by oxidative stress on **5mC** sites *in vitro* induced by hydroxyl radicals and others.[57]

With **5fC** and **5caC** being by another factor of 10 less prevalent in various cell types, not much is known if they have distinct functions or are rather results of iterative oxidations of **5hmC**.[58] There is initial evidence that **5fC** could be a stable or at least semipermanent epigenetic marker involved in gene regulation.[59-62] As mentioned before **5fC** as well as **5caC** can be formed via oxidative damage (Figure 9) or as intermediate of TET-mediated DNA demethylation.[63]

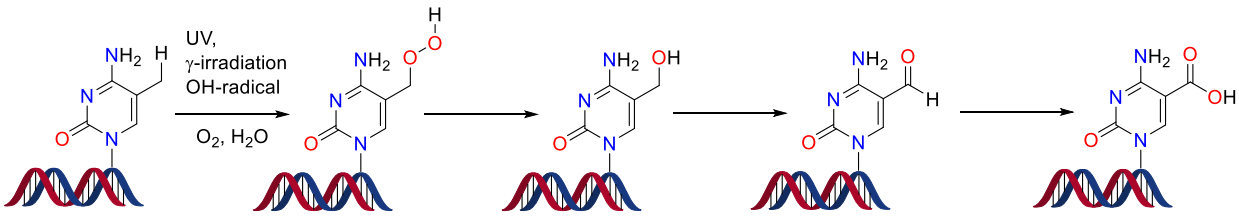


Figure 9: Transformation of **5mC** by radical-mediated autoxidation process leading to oxidative DNA damage. [63]

Current studies towards **5fC** and **5caC** focus its effects on genome stability, with Matsuda *et al.* linking **5fC** levels with induced C-G to T-A or A-T mutations by shifting the tautomeric equilibrium of **5fC** from the amino to the imino form due to hydrogen bonding effects.[64-65] On the contrary, Willson showed that there is no effect on the efficiency and fidelity of DNA repair by human DNA polymerase β by **5fC** incorporation, suggesting that the amino/imino equilibrium does not shift and negating it to be mutagenic.[66] Local effects of **5fC** DNA modifications have been described by changing the DNA melting kinetics,[67] increasing DNA flexibility (together with **5caC**),[68] which may ultimately have fundamental effects on how **5fC** is recognized by TDG during the base flipping processes.[69-70] Due to its possibility to form imines with external amine nucleophiles, **5fC** can undergo DNA-protein crosslink (DPG) formation in cells.[71]

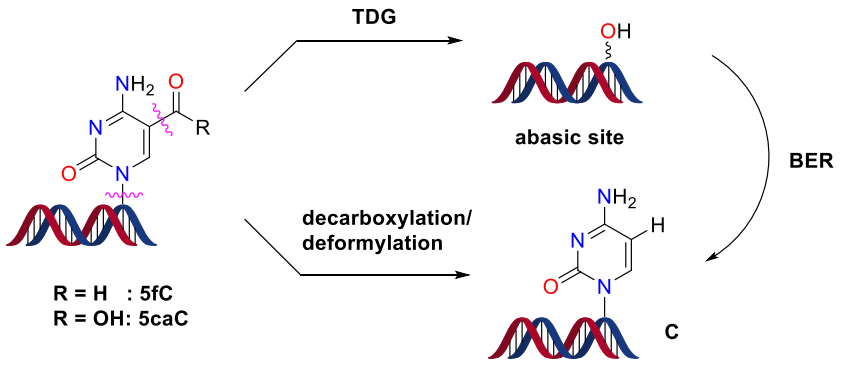


Figure 10: Two possibilities of active demethylation of **5fC** and **5caC** via deformylation and decarboxylation respectively (compare with Figure 2, where only the TDG pathway is shown).

The restoration of **C** via base excision repair (BER) pathway was already briefly addressed in paragraph 2.2. Since it was revealed the first time by Drohat *et al.* a relation between the stability of the N-

glycosidic bond and TDG activity was drawn, with **5fC** being excised at a greater rate than thymine (**T**/**5mU**).[72-73] This effect can be attributed to the electron-withdrawing effect of the carbonyl group, leading to a destabilization of the N-glycosidic bond and a tautomeric stabilization of the anionic **5fC** and **5caC** species generated by a nucleophilic attack at the glycosidic carbon atom (see Figure 11). The acidity at the N3 position of the modified nucleobases also play an important role in this process.

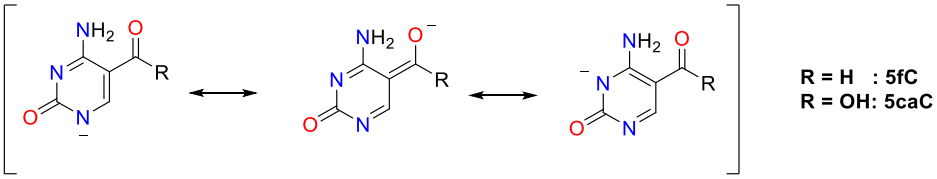


Figure 11: Tautomerism of the **5fC**- and **5caC**-anions formed during base excision repair.

Since **5caC** is also a substrate for **TDG**,[72] it should be noted that the oxidation by TET enzyme to **5fC,** followed by excision via BER are rather inefficient *in vitro*.[72,74] Since direct decarboxylation has already been mentioned before (paragraph 2.2), not much is known as of yet about the direct deformylation pathway (Figure 10, bottom). It has been suggested by Carell *et al.* showing that C-C cleavage occurs for **5fC** and **5caC** by treatment with high concentrations of thiols.[23] Due to these deformylation and decarboxylation reactions not being dependent on DNMT or TET, unknown pathways leading to C-C cleavage are suggested.[22]

With thymine being one of the four canonical nucleosides in DNA, less is known about its oxidative intermediate 5-hydroxymethyluracil (**5hmU**). Initially it was believed to be a product of endogenous oxidative DNA damage, although being less reactive towards hydroxyl radicals than **5mC**.[75] This DNA damage induced by reactive oxygen species (ROS) is associated with mutagenesis.[76-77] Newer findings show that TET enzymes can oxidize thymine to **5hmU** in mouse embryonic stem cells followed by BER, indicating an effect on transcription (see Figure 2).[78] It has also been suggested that **5hmU** might be a product of deamination leading to 5hmU:G mismatches in DNA.[79-82] However, due to the very low amount of **5hmC** in DNA as well as the low rate constant for the spontaneous deamination ($4.8 \times 10^{-10}\ s^{-1}$), it is unlikely a real source of 5hmU:G mismatches.[83-84] In fungi, like *Rhodotorula glutinis*, *Neurospora crassa* and *Aspergillus nidulans*, an active demethylation pathway exists, called the thymidine salvage pathway (*Figure 12*).[85]

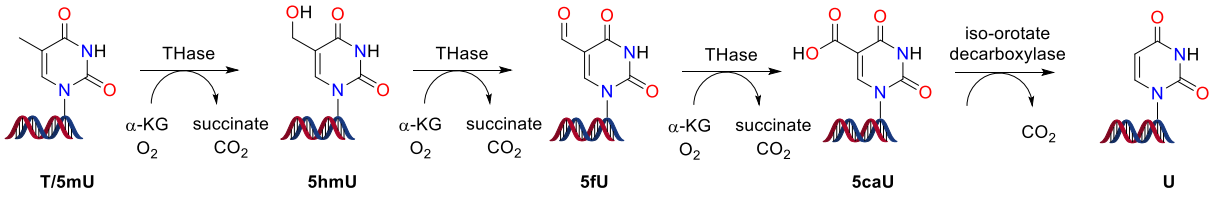


Figure 12: Thymidine salvage pathway oxidising thymine (**T**) to isoorotate (**5caU**) by thymine-7-hydroxylase (THase), followed by decarboxylation to uracil (**U**) by isoorotate decarboxylase. [86]

As seen above, thymidine-7-hydroxylase (THase) can oxidise thymidine similar to the TET enzyme, in three consecutive oxidation steps to form 5-caboxyuracil/isoorotate (**5caU**). In a last step **5caU** is decarboxylated, resulting in uracil to facilitate the active demethylation of thymidine.[86] As of yet, no

homologue of THase has been found in mammals. This is an example of how enzymes can catalyse the cleavage of an energetically difficult C-C bond, still to be uncovered in mammals for **5caC**.

2.5 Outline of this Thesis

This introductory chapter offers an overview of the active demethylation pathways of 5-methylcytosine (**5mC**) by the ten-eleven translocation (TET) enzyme, and the potential epigenetic significance of its oxidative derivatives. This study seeks to expand upon the existing knowledge of **5mC**, as well as 5-methyluracil (**5mU**) and their respective oxidative derivatives. Utilizing both theoretical and experimental approaches, we aim to characterize these known substrates to better understand their interactions with the TET enzyme. Key properties, such as bond dissociation energies (BDE), (intrinsic) dissociation constants, and hydration equilibria, will be examined and compared with experimental data to draw conclusions regarding substrate-specific interactions.

Chapter 3 introduces an intriguing hydrate form of 5-formylcytosine (**5fC**) and 5-formyluracil (**5fU**), denoted as 5-dihydroxymethylcytosine (**5dhmC**) and 5-dihydroxymethyluracil (**5dhmU**), respectively. The hydrate form of **5fC** may serve as potential substrate in the TET enzyme-mediated oxidation of **5fC** to 5-carboxylcytosine (**5caC**).

Chapter 4 provides further evidence supporting this hypothesis by determining crucial bond dissociation energies for this hydrate form (**5dhmC**) and comparing its dissociation constants with those of the non-hydrated form (**5fC**). Additionally, the controversial site-specific deprotonation of **5caC** is investigated.

Chapter 5 employs both theoretical and experimental techniques in a kinetic study to describe and predict the oxidative chemistry of a TET-like iron-oxo complex with benzylic and N-methylated cytosine and uracil bases. Furthermore, the stability of the oxidized N-methyl substituents is investigated.

2.6 REFERENCES

[1] C. H. Waddington, *Nature* **1942**, *150*, 563-565.
[2] T. O. Tollefsbol, in *Handbook of Epigenetics (Third Edition)* (Ed.: T. O. Tollefsbol), Academic Press, **2023**, pp. 3-8.
[3] L. Liu, Y. Li, T. O. Tollefsbol, *Curr. Issues Mol. Biol.* **2008**, *10*, 25-36.
[4] S. Chong, E. Whitelaw, *Curr. Opin. Genet. Dev.* **2004**, *14*, 692-696.
[5] K. Shiota, Y. Kogo, J. Ohgane, T. Imamura, A. Urano, K. Nishino, S. Tanaka, N. Hattori, *Genes Cells* **2002**, *7*, 961-969.
[6] G. Liang, M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird, P. A. Jones, *Mol. Cell. Biol.* **2002**, *22*, 480-491.
[7] M. Bibikova, E. Chudin, B. Wu, L. Zhou, E. W. Garcia, Y. Liu, S. Shin, T. W. Plaia, J. M. Auerbach, D. E. Arking, R. Gonzalez, J. Crook, B. Davidson, T. C. Schulz, A. Robins, A. Khanna, P. Sartipy, J. Hyllner, P. Vanguri, S. Savant-Bhonsale, A. K. Smith, A. Chakravarti, A. Maitra, M. Rao, D. L. Barker, J. F. Loring, J. B. Fan, *Genome. Res.* **2006**, *16*, 1075-1083.
[8] L. Shen, R. A. Waterland, *Curr. Opin. Clin. Nutr. Metab. Care.* **2007**, *10*.
[9] M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nuc. Ac. Res.* **1982**, *10*, 2709-2721.
[10] A. Bird, M. Taggart, M. Frommer, O. J. Miller, D. Macleod, *Cell* **1985**, *40*, 91-99.
[11] L. D. Moore, T. Le, G. Fan, *Neuropsychopharmacol.* **2013**, *38*, 23-38.
[12] L. Shen, Y. Kondo, Y. Guo, J. Zhang, L. Zhang, S. Ahmed, J. Shu, X. Chen, R. A. Waterland, J. P. Issa, *PLoS Genet.* **2007**, *3*, 2023-2036.
[13] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Curr. Biol.* **2000**, *10*, 475-478.
[14] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502.
[15] Z. Paroush, I. Keshet, J. Yisraeli, H. Cedar, *Cell* **1990**, *63*, 1229-1237.
[16] E.-A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, *Nat. Rev. Chem.* **2017**, *1*, 0069.
[17] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kuśnierczyk, C. B. Vågbø, D. Schuermann, A. Klungland, P. Schär, *Nat. Commun.* **2016**, *7*, 10806.
[18] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338.
[19] C. J. Lio, X. Yue, I. F. Lopez-Moyado, M. Tahiliani, L. Aravind, A. Rao, *J. Biosci.* **2020**, *45*.
[20] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *Cell* **2008**, *135*, 1201-1212.
[21] A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591-5594.
[22] 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.
[23] 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.
[24] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLOS ONE* **2010**, *5*, e15367.
[25] Y. Feng, J.-J. Chen, N.-B. Xie, J.-H. Ding, X.-J. You, W.-B. Tao, X. Zhang, C. Yi, X. Zhou, B.-F. Yuan, Y.-Q. Feng, *Chem. Sci.* **2021**, *12*, 11322-11329.
[26] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 2916-2921.
[27] A. Jeltsch, *Chembiochem* **2002**, *3*, 274-293.
[28] E. Kriukienė, Z. Liutkevičiūtė, S. Klimašauskas, *Chem. Soc. Rev.* **2012**, *41*, 6916-6930.
[29] J. C. Susan, J. Harrison, C. L. Paul, M. Frommer, *Nuc. Ac. Res.* **1994**, *22*, 2990-2997.
[30] Q. Du, Z. Wang, V. L. Schramm, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 2916-2921.
[31] R. B. Lorsbach, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, J. R. Downing, *Leukemia* **2003**, *17*, 637-641.
[32] K. D. Rasmussen, K. Helin, *Genes. Dev.* **2016**, *30*, 733-750.

[33] L. Cimmino, O. Abdel-Wahab, Ross L. Levine, I. Aifantis, *Cell Stem Cell* **2011**, *9*, 193-204.
[34] 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.
[35] S. O. Waheed, A. Varghese, S. S. Chaturvedi, T. G. Karabencheva-Christova, C. Z. Christov, *ACS Catal.* **2022**, *12*, 5327-5344.
[36] S. Kavoosi, B. Sudhamalla, D. Dey, K. Shriver, S. Arora, S. Sappa, K. Islam, *Chem. Sci.* **2019**, *10*, 10550-10555.
[37] U. Ghanty, J. E. DeNizio, M. Y. Liu, R. M. Kohli, *J. Am. Chem. Soc.* **2018**, *140*, 17329-17332.
[38] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545-1555.
[39] 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.
[40] C. W. John, R. P. Hausinger, D. A. Proshlyakov, *J. Am. Chem. Soc.* **2019**, *141*, 15318-15326.
[41] S. D. Wong, M. Srnec, M. L. Matthews, L. V. Liu, Y. Kwak, K. Park, C. B. Bell, 3rd, E. E. Alp, J. Zhao, Y. Yoda, S. Kitao, M. Seto, C. Krebs, J. M. Bollinger, Jr., E. I. Solomon, *Nature* **2013**, *499*, 320-323.
[42] 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.
[43] H. Torabifard, G. A. Cisneros, *Chem. Sci.* **2018**, *9*, 8433-8445.
[44] P. A. Defossez, I. Stancheva, *Prog. Mol. Biol. Transl. Sci.* **2011**, *101*, 377-398.
[45] A. M. Deaton, A. Bird, *Genes. Dev.* **2011**, *25*, 1010-1022.
[46] A. P. de Koning, W. Gu, T. A. Castoe, M. A. Batzer, D. D. Pollock, *PLoS Genet.* **2011**, *7*, e1002384.
[47] M. G. Butler, *J. Assist. Reprod. Genet.* **2009**, *26*, 477-486.
[48] M. Gardiner-Garden, M. Frommer, *J. Mol. Biol.* **1987**, *196*, 261-282.
[49] D. H. Yu, C. Ware, R. A. Waterland, J. Zhang, M. H. Chen, M. Gadkari, G. Kunde-Ramamoorthy, L. M. Nosavanh, L. Shen, *Mol. Cell. Biol.* **2013**, *33*, 1845-1858.
[50] 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.
[51] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930.
[52] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Cell Cycle* **2009**, *8*, 1698-1710.
[53] T. P. Gu, F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabó, G. P. Pfeifer, J. Li, G. L. Xu, *Nature* **2011**, *477*, 606-610.
[54] L. Tan, Y. G. Shi, *Development* **2012**, *139*, 1895-1902.
[55] S. G. Jin, S. Kadam, G. P. Pfeifer, *Nuc. Ac. Res.* **2010**, *38*, e125.
[56] K. E. Szulwach, X. Li, Y. Li, C. X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, P. Jin, *Nat. Neurosci.* **2011**, *14*, 1607-1616.
[57] J. R. Wagner, J. Cadet, *Acc. Chem. Res.* **2010**, *43*, 564-571.
[58] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472-479.
[59] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797-11800.
[60] 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.
[61] Cornelia G. Spruijt, F. Gnerlich, Arne H. Smits, T. Pfaffeneder, Pascal W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, Arie B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159.
[62] C.-X. Song, Keith E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678-691.
[63] Y. Zhang, C. Zhou, *DNA Repair (Amst.)* **2019**, *81*, 102649.

[64] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782-13788.
[65] N. Karino, Y. Ueno, A. Matsuda, *Nuc. Ac. Res.* **2001**, *29*, 2456-2463.
[66] M. J. Howard, K. G. Foley, D. D. Shock, V. K. Batra, S. H. Wilson, *J. Biol. Chem.* **2019**, *294*, 7194-7201.
[67] Q. Dai, P. J. Sanstead, C. S. Peng, D. Han, C. He, A. Tokmakoff, *ACS Chem. Biol.* **2016**, *11*, 470-477.
[68] T. T. M. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, T. Ha, *Nat. Com.* **2016**, *7*, 10813.
[69] B. J. Dow, S. S. Malik, A. C. Drohat, *J. Am. Chem. Soc.* **2019**, *141*, 4952-4962.
[70] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *J. Am. Chem. Soc.* **2013**, *135*, 15813-15822.
[71] S. Ji, H. Shao, Q. Han, C. L. Seiler, N. Y. Tretyakova, *Angew. Chem. Int. Ed.* **2017**, *129*, 14318-14322.
[72] A. Maiti, A. C. Drohat, *J. Bio. Chem.* **2011**, *286*, 35334-35338.
[73] M. T. Bennett, M. Rodgers, A. S. Hebert, L. E. Ruslander, L. Eisele, A. C. Drohat, *J. Am. Chem. Soc.* **2006**, *128*, 12510-12519.
[74] L. Zhang, W. Chen, L. M. Iyer, J. Hu, G. Wang, Y. Fu, M. Yu, Q. Dai, L. Aravind, C. He, *J. Am. Chem. Soc.* **2014**, *136*, 4801-4804.
[75] S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nuc. Ac. Res.* **1997**, *25*, 553-558.
[76] A. Klungland, R. Paulsen, V. Rolseth, Y. Yamada, Y. Ueno, P. Wiik, A. Matsuda, E. Seeberg, S. Bjelland, *Toxicolo. Lett.* **2001**, *119*, 71-78.
[77] Z. Djuric, L. K. Heilbrun, S. Lababidi, E. Berzinkas, M. S. Simon, M. A. Kosir, *Cancer Epidemiol. Biomarkers Prev.* **2001**, *10*, 147-149.
[78] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, *Nat. Chem. Biol.* **2014**, *10*, 574-581.
[79] G. Rangam, K.-M. Schmitz, A. J. Cobb, S. K. Petersen-Mahrt, **2012**.
[80] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758.
[81] J. U. Guo, Y. Su, C. Zhong, G.-l. Ming, H. Song, *Cell* **2011**, *145*, 423-434.
[82] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, *Cell* **2011**, *146*, 67-79.
[83] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
[84] V. Rusmintratip, L. C. Sowers, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14183-14187.
[85] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes Sr, A. A. Axhemi, *Biochim Biophys Acta Gen Subj* **2005**, *1723*, 256-264.
[86] B. Warn-Cramer, L. Macrander, M. Abbott, *J. Biol. Chem.* **1983**, *258*, 10551-10557.

3 THE pH-DEPENDENCE OF THE HYDRATION OF 5-FORMYLCYTOSINE

Fabian L. Zott, Vasily Korotenko, and Hendrik Zipse

ChemBioChem, **2022**, 23, e202100651. - Published by Wiley-VCH GmbH.

DOI and link to article: https://doi.org/10.1002/cbic.202100651

Author contributions: F.Z and V.K. contributed equally to this work. The project was conceived by F.Z., V.K. and H.Z. The experimental study was performed by F.Z. The computational study was performed by V.K and assisted by F.Z. The dissociation constants were calculated by V.K. and F.Z. The manuscript was jointly written by F.Z., V.K. and H.Z. The experimental part of the SI was prepared by F.Z., the computational part was prepared by V.K.

Additional Information: The supporting information of this article is presented in an altered version, to fit into the structure of this thesis. The original file can be accessed under the flowing link: https://chemistry-europe.onlinelibrary.wiley.com/202100651sup.pdf

Note: This section contains individual numbering of compounds, figures, tables and references.

ChemBioChem

Research Article
doi.org/10.1002/cbic.202100651

Chemistry Europe
European Chemical Societies Publishing

www.chembiochem.org

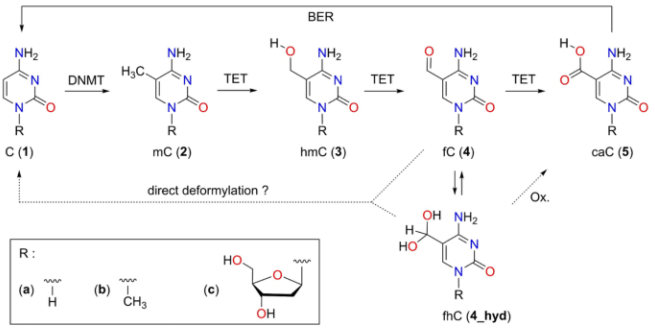
The pH-Dependence of the Hydration of 5-Formylcytosine: an Experimental and Theoretical Study

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



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

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

[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

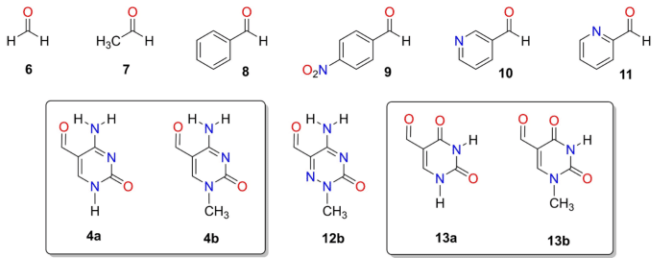
[+] *These authors contributed equally to this work.*

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 (**13b**), 5-formyluracil (**13a**), 5-formylcytosine (**4a**) and 1-methyl-5-formylcytosine (**4b**) 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 **13b** leads to transient formation of the hydrate form **13b_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-13b**, most likely via the hydrated form **13b_hyd**. This latter conclusion is supported by analysis of the fragmentation patterns for **13b** and **^{18}O-13b**, and oxygen exchange of the other carbonyl groups present in **13b** can be ruled out (see Supporting Information Figures S3 and S4). The ^{18}O isotope experiment for **4b** 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 **13b** 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 **13b** are accompanied by additional resonances for its hydrate **13b_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]
4a	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
4b[c,d]	$<4.50\times10^{-4}$	$<8.11\times10^{-6}$	22	> +28.8	this study
p_4b[a,b]	0.005	9.72×10^{-5}	22	+22.7	this study
	0.005	9.10×10^{-5}	30	+22.7	[14]
4c[c]	$<4.40\times10^{-4}$	$<7.93\times10^{-6}$	22	> +28.8	this study
p_4c[e]	0.007	1.23×10^{-4}	22	+22.1	this study
12b[f]	0.25	0.0045	22	+13.3	[51]
13a[g]	0.016	2.94×10^{-4}	22	+20.0	this study
13b[g]	0.013	2.42×10^{-4}	22	+20.5	this study

[a] Protonated **4b**. [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).

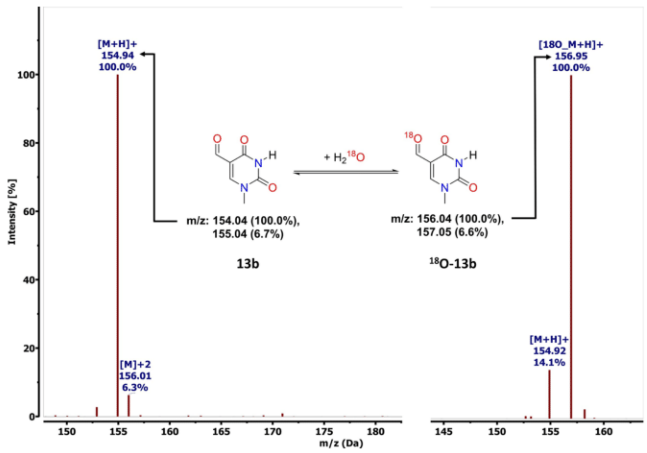


Figure 1. ^{18}O isotope exchange experiment with **13 b** (left side: **13 b** at natural abundance; right side: ^{18}O labelled **13 b** after equilibration with $H_2{}^{18}O$; only the $[M]^+$ peak region is shown).

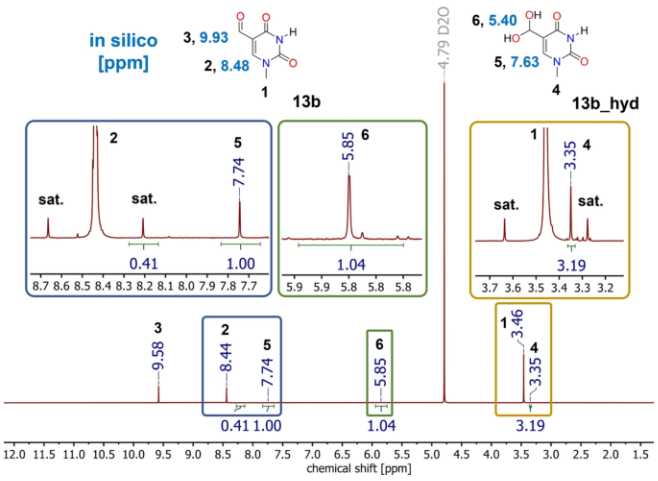


Figure 2. ^1H NMR spectra of **13 b** and its hydrated form **13 b_hyd** in D_2O at a concentration of 3.1×10^{-3} mol L^{-1} together with the relevant signal assignments and calculated chemical shifts (in light blue).

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-

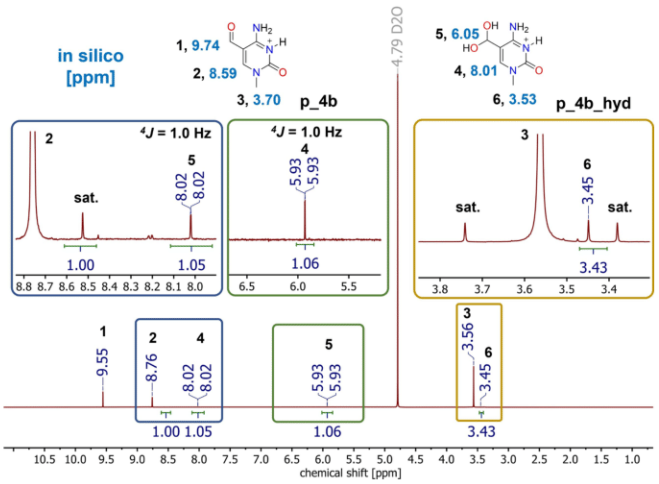


Figure 3. ^1H NMR spectra of the protonated form of **4 b** and its hydrated form **4 b_hyd** in D_2O at pH = 2 and a concentration of 5.8×10^{-3} mol L^{-1} together with the relevant signal assignments and calculated chemical shifts (in light blue).

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).

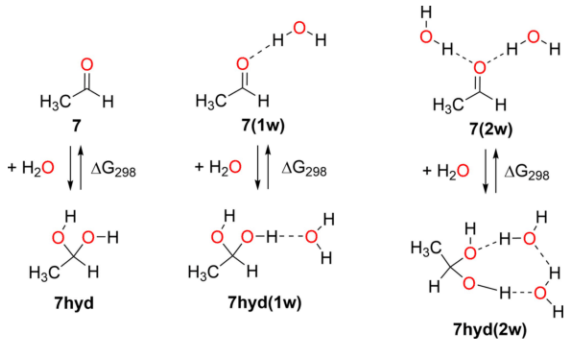


Figure 4. The hydration of acetaldehyde (**7**) in the absence and presence of solvating water molecules.

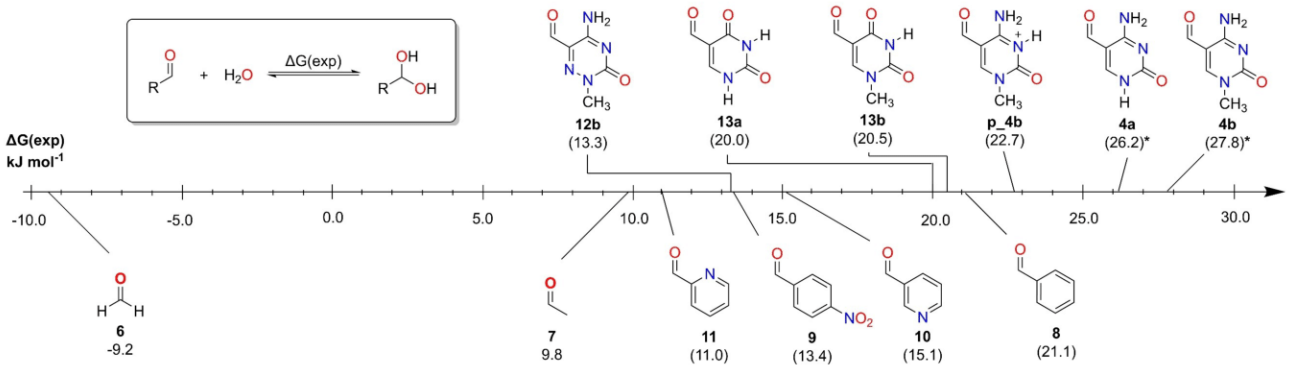


Figure 5. Experimental hydration free energies of selected aldehydes (* based on combination of the G3B3 $\Delta\Delta G_{298}$ values with the experimentally measured value for **p_4 b**) as the reference).

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.

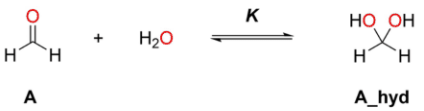
For most aldehydes considered here, the calculated hydration energies are overestimated to a similar extent as already observed for benzaldehyde (**8**). Due to the systematic nature of this phenomenon, good linear correlations can be observed between experimentally measured and theoretically calculated 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)

$$K = \frac{[A_{hyd}]}{[A] \times [H_2O]} \quad (2)$$

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

$$K = \frac{K_w}{[H_2O]} \quad (4)$$

$$\Delta G = -RT \ln(K) \quad (5)$$

^{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 10000 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 **13b** and **4b** 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 $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91\ kJ\,mol^{-1}$. 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 4b: 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 13b: 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 (**4c**). 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

3.1 General Information and Techniques

General Methods: All reactions sensitive to air and moisture were conducted under a nitrogen atmosphere and the glassware as well as magnetic stir bars were dried overnight in a dry oven at 110°C.

Solvents: If not further specified, solvents were obtained from the companies Acros Organics, Sigma Aldrich, Fluka or Merck and purified with a rotary evaporator. NMR used for the NMR experiments were additionally dried with calcium hydride and stored under nitrogen atmosphere. Anhydrous DMSO-D6 is purchased in glass ampullas. Molar sieves were prepared by washing in acetone, then storing them for several days in the dry oven at 110°C and before use, heating them for 20 min with a heat gun up to 500°C under vacuo.

Reagents and Catalysts: Commercially available reagents were obtained from the companies TCI, Sigma Aldrich or ABCR and used without further purification.

Chromatograpy: Silica gel for column chromatography was purchased from Acros Organics (mesh 35-70). Thin-layer chromatograpy was done by using TLC plates purchased by Merck (silica gel 60 F254, thickness 0.2 mm). Preparative layer chromatography (PLC) was carried out by using Merck TLC glass plates (silica gel 60 F254, thickness 2 mm).

Analytical Methods

NMR spectroscopy: All ^{1}H-NMR spectra were recorded on Mercury 200, Varian INOVA 400 and 600 machines in $CDCl_3$ at 400 MHz or 600MHz at room temperature. All ^{13}C-NMR spectra were recorded respectively at 101 MHz and 151 MHz. The chemical shifts are reported in ppm (δ), relative to the resonance of $CDCl_3$ (at δ = 7.27 ppm for ^{1}H and δ = 77.16 ppm for ^{13}C), DMSO-D6 (at δ = 2.50 ppm for ^{1}H and δ = 39.52 ppm for ^{13}C) and D_2O (at δ = 4.79 ppm for ^{1}H, the ^{13}C signals are automatically referenced to the deuterium lock of the solvent signal). Spectra were imported and processed in the MestreNova 11.0.4 program. Coupling constants are reported in Hertz [Hz] and for NMR splitting patterns the following abbreviations are used: singlet (s), doublet (d), triplet (t), quartet (q), multiplet (m), broad singlet (br-s).

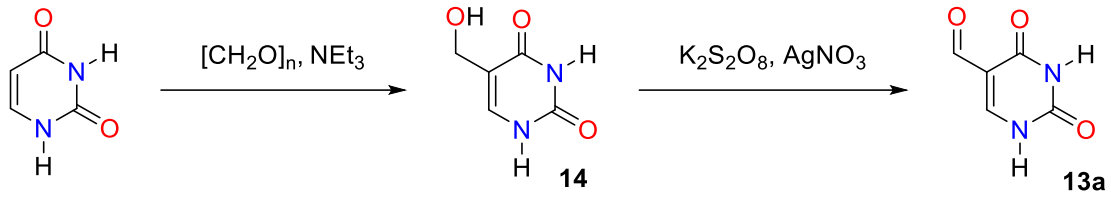
GC-FID measurements: GC measurements were obtained at a Shimadzu GC-2010 Plus Gas Chromatograph with AOC-20i autosampler (with temperature-controlled sample holder) and an Optima 1701-0.25 µm (25 m * 0.25 mm) column.

Mass spectroscopy: HR-MS spectra were obtained by using a Thermo Finnigan LTQ FT machine of the MAT 95 type with a direct exposure probe (DEP) and electron impact ionization (EI, 70 eV).

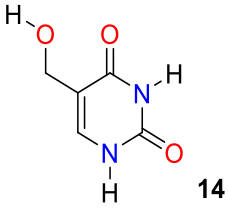
Isotope exchange experiments were performed with an Advion expressionL compact mass spectrometer (CMS) by using the atmospheric solid analysis probe (ASAP) technique. 350 m/z, and acquisition speed 10 000 m/z units per second. The ion source settings “typical” were used, corresponding to: capillary temperature 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 **13b** and **4b** 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.

3.2 SYNTHESIS

Synthesis of 5-Formyluracile (13a)



Synthesis of 5-Hydroxymethyluracil (14)



Following the procedure by Hudson et al.[1], uracil (89.2 mmol, 10.0 g, 1.0 eq.) and paraformaldehyde (187.4 mmol, 5.36 g, 3.0 eq.) were suspended in water (300 mL). Triethylamine (133.8 mmol, 13.54 g, 18.64 mL, 1.5 eq.) was added and the mixture was heated to 65°C, at which point it turned clear. It was stirred at this temperature for 12 hours, then the solvent was removed, and the crude product was re-solved in water (50 mL). Ethanol (50 mL) was added, and after cooling to 4 °C the product was collected by filtration, washing with cold water. The filtrate was evaporated, the residue re-solved and the product precipitated in the same manner two more times to yield 5-hydroxymethyluracil (10.2 g, 71.6 mmol, 81%) as a white solid.

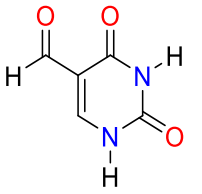
^{1}H-NMR (400 MHz, DMSO-d6, ppm): δ = 11.05 (br-s, 1H), 10.72 (br-s, 1H), 7.24 (s, 1H), 4.85 (t, J = 5.6 Hz, 1H), 4.10 (d, J = 3.6 Hz, 2H).

^{13}C-NMR (100 MHz, DMSO-d6, ppm): δ = 163.85, 151.39, 138.24, 112.76, 55.84.

EA: Calculated [%]: C: 42.26 N: 19.71 H: 4.26; Found [%]: C: 40.87 N: 19.08 H: 4.38.

HR-MS (EI, 70 eV, M+): [C5H6N2O3+], calculated: 142.0373, found: 142.0376.

Synthesis of 5-Formyluracil (13a)



13a

Following the procedure by Hudson et. al [1], 5-hydroxymethyluracil (24.6 mmol, 3.5 g, 1.0 eq.) was dissolved in water (130 mL) by heating to 90°C. The solution was cooled to 45◦C and potassium persulfate (45.1 mmol, 12.2 g, 1.8 eq.) and silver nitrate (0.74 mmol, 3 mol%, 12.53 mmol) were added. The mixture was stirred at 45◦C for 20 minutes, after cooling to 0 °C the precipitated product was collected by filtration, washing with cold water. The filtrate was concentrated, and the precipitate was collected in this manner two more times to yield 5-formyluracil (17.5 mmol, 2.45 g, 72 %) as a white solid.

^{1}H-NMR (400 MHz, DMSO-d6, ppm): δ = 11.91 (s, 1H), 11.51 (s, 1H), 9.73 (s, 1H), 8.14 (s, 1H).

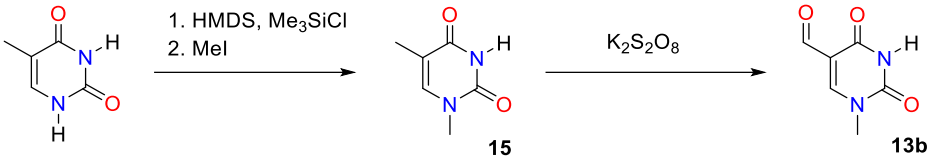
^{13}C-NMR (100 MHz, DMSO-d6, ppm): δ = 186.49, 162.53, 150.53, 149.41, 110.13.

EA: Calculated [%]: C: 42.87 N: 20.00 H: 2.88; Found [%]: C: 40.32 N: 19,87 H: 2.84.

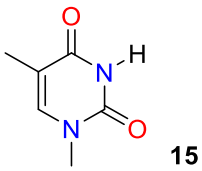
HR-MS: (EI, 70 eV, M+): [C5H4N2O3+], calculated: 140.0217, found: 142.0210.

Synthesis of 1-Methyl-5-Formyluracil (13b)

Independent Route 1:



Synthesis of 1-N-Methylthymine (1,5-dimethyluracil, 15)

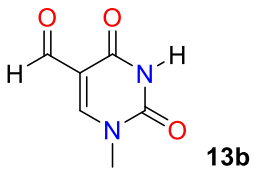


According to a modified literature procedure [2]. In an oven dried Schlenk flask under nitrogen, a suspension of thymine (1.90 g, 15.0 mmol), HMDS (35 mL) and Me_3SiCl (0.75 mL) was refluxed for 5h. After cooling down to 40 °C iodomethane (15 mL) was added to the reaction mixture and the suspension was refluxed overnight. All volatile compounds were removed under high vacuum with the addition of a cooling trap, to yield a brown solid. Now 35 mL of 6N acetic acid were added, the reaction mixture was stirred for 20 min at room temperature and the solvent was removed under reduced pressure. The crude product was recrystallized in water to yield 1,5-dimethyluracil (1.47 g, 10.5 mmol, 70%) as a slightly brown solid.

^{1}H-NMR (400 MHz, DMSO-*d6*, ppm): δ = 11.22 (br-s, 1H), 7.50 (q, *J* = 1.2 Hz, 1H), 3.19 (s, 3H), 1.73 (d, *J* = 1.2 Hz, 3H).

^{13}C-NMR (100 MHz, DMSO-*d6*, ppm): δ = 164.49, 151.24, 142.37, 108.10, 34.96, 11.92.

Synthesis of 1-Methyl-5-Formyluracil (13b)



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 1h at 85°C and the reaction mixture was stirred for 16h.

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 = 11.73 (br-s, 1H, N-*H*), 9.76 (s, 1H, formyl-*H*), 8.48 (s, 1H, C_6-*H*), 3.36 (s, 3H, -C*H*$_3$).

^{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, -C*H*$_3$).

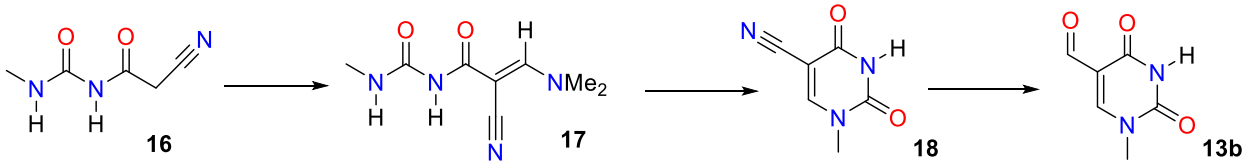
^{13}C-NMR (100 MHz, DMSO-D6, ppm): δ = 186.24, 162.50, 152.66, 150.53, 109.82, 36.29.

^{13}C-NMR (100 MHz, D_2O, ppm): δ = 186.70, 162.95, 153.14, 150.98, 110.26, 36.75.

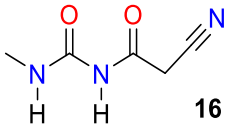
EA: Calculated [%]: C: 46.76 N: 18.18 H: 3.92; Found [%]: C:48.49 N: 18.20 H: 3.98.

HR-MS (EI, 70 eV, M+): [C6H6N2O3+], calculated: 154.0373, found: 154.0372.

Independent Route 2:



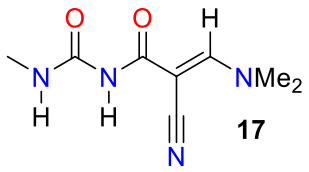
Synthesis of 1N-Methyl-5-cyanouracil (16)



Following a procedure by Notani et al.[3], N-methylurea (14.81 g, 0.2 mmol, 1 eq) and 2-cyanoacetic acid (17.01 g, 0.2 mmol, 1 eq) were dissolved in 100 mL of acetic anhydride and heated to 80°C for 45 mins. After cooling down to room temperature the precipitate was filtered off and recrystallized in EtOH to afford 12.25 g (0.09 mmol, 43 %) of white crystals.

^{1}H-NMR (400 MHz, DMSO-d6, ppm): 10.52 (br-s, 1H, N-*H*), 7.88 (br-s, 1H, N-*H*), 3.87 (s, 2H, CH_2-CN), 2.67 (d, 3H, *J* = 4.6 Hz, N-CH_3).

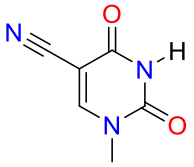
^{13}C-NMR (100 MHz, DMSO-d6, ppm): 164.75, 152.95, 115.21, 26.72, 26.18.



Following a procedure by Notani et al.[3], to DMF (5,54 mL, 5.26 g, 7.2 mmol, 1.2 eq) in a Schlenk flask with reflux condenser and drop funnel under nitrogen, dimethyl sulfate (6.813 mL, 9.08 g, 7.2 mmol, 1.2 eq) was added dropwise at 40 °C and the reaction mixture was stirred for another 2h at 80 °C, which resulted in DMS-DMF adduct in a slightly yellow oil. Now to 1-methyl-3-cyanoacetyl urea (8.47 g, 6.0 mmol, 1eq) in 40 mL of dry DMF under nitrogen at 0 °C the DMF-DMS adduct was added quickly, after which dry triethylamine (9.98 mL, 7.29 g, 7.2 mmol, 1.2 eq) slowly over 20 min to form a white precipitate after complete solvation of the reactants. The reaction mixture was stirred for 2h at room temperature, then the product filtered, washed with Et_2O and dried over vacuum to afford 5.788 g (2.95 mmol, 49 %) of pure product.

^{1}H-NMR (400 MHz, DMSO-d6, ppm): δ = 9.09 (s, 1H), 8.30 (d, *J* = 4.9 Hz, 1H), 7.89 (s, 1H), 3.35 (s, 2H), 3.26 (s, 3H), 3.20 (s, 3H), 2.70 (d, *J* = 4.7 Hz, 3H).

^{13}C-NMR (100 MHz, DMSO-d6, ppm): δ = 166.35, 156.40, 154.04, 118.47, 70.77, 47.17, 38.24, 26.07.



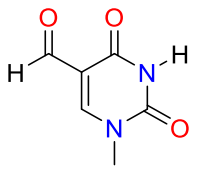
18

Following a procedure by Notani et al.[3], 1-methyl-3-cyano-(dimethylaminomethylene)acetyl urea (5.79 g, 29.5 mmol) was added slowly to 33 ml of 10 % aqueous NaOH until complete dissolution. Now a concentrated HCl solution was added slowly at 0 °C until pH = 3 was reached to form a precipitate and stirred for further 60 mins. The product was filtered off, washed with H_2O, and dried by coevaporation with toluene under vacuum to afford 4.429 g (29.3 mmol, 99%) of white product.

^{1}H-NMR (400 MHz, DMSO-d6, ppm): δ = 11.97 (s, 1H), 8.66 (s, 1H), 3.27 (s, 3H).

^{13}C-NMR (100 MHz, DMSO-d6, ppm): δ = 160.91, 155.63, 150.07, 114.62, 86.72, 36.48.

HR-MS (EI, 70 eV, M+): [C6H5N3O2+], calculated: 151.0377, found: 151.0375.

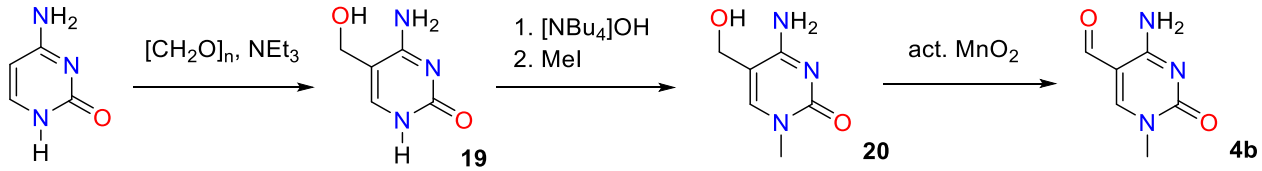


13b

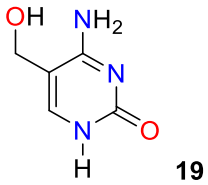
Damp Raney-Ni was placed in a Schlenk tube under nitrogen, after which 1-N-methyl-5-cyanouracil (0.50 g, 3.31 mmol) dissolved in 17 mL of formic acid (98%) was added quickly. The reaction mixture was stirred for 1h at room temperature and then refluxed for 1.5 hours. After cooling, the reaction mixture was filtered hot over celite and evaporated under reduced pressure. Now the crude mixture was dispersed with acetic acid (2 mL) in 40 mL of MeOH and refluxed for 1h. The solvent was removed under reduced pressure and the crude dimethoxy acetal was purified by flash column chromatography over silica gel (2%MeOH/2%AcOH/DCM to 3%MeOH/2%AcOH/DCM). The pure methoxy acetal was dissolved in 10 mL of 5% AcOH in water and stirred overnight at room temperature. The solvent was removed under reduced pressure to afford the clean compound as a white solid (76 mg, 0.5 mmol, 15%).

Analytical data: see Independent Route 1

Synthesis of 1-Methyl-5-Formylcytosine (4b)



Synthesis of 5-Hydroxymethylcytosine (19)



Following a procedure by Wojciechowski et al.[1], to a solution of cytosine (5.27 g, 47.4 mmol) in water (140 mL) at 92 °C, paraformaldehyde (3.28 g, 109 mmol) and triethylamine (16,5 mL, 118 mmol) were added, and the reaction was refluxed for 72h. After 24h and 32h, a first (1.14 g, 38.0 mmol) and a second batch (0.57 g, 19.0 mmol) of paraformaldehyde were added. After completion of the reaction, all volatile compounds were removed to give a dark-coloured residue. The crude product was recrystallized in a solution of 95% ethanol in water to receive 3.48 g (24.6 mmol, 52%) of a slightly brown solid.

^{1}H-NMR (400 MHz, D_2O, ppm): δ = 7.52 (s, 1H, C_6-*H)*, 4.40 (s, 2H, CH_2-OH).

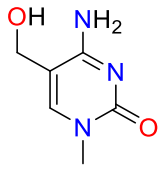
^{1}H-NMR (400 MHz, DMSO-D6, ppm): δ = 10.40 (br-s, 1H, N-*H*), 7.26 (s, 1H, C_6-*H)*, 7.18 (br-s, 1H, -NH_2), 6.47 (br-s, 1H, -NH_2), 4.94 (t, *J* = 5.4 Hz, 1H, -O*H*), 4.15 (d, *J* = 4.5 Hz, 2H, CH_2-OH).

^{13}C-NMR (100 MHz, D_2O, ppm): δ = 165.95, 158.87, 142.19, 105.46, 57.43.

^{13}C-NMR (101 MHz, DMSO-D6, ppm): δ = 165.47, 156.72, 140.47, 104.61, 57.17.

HR-MS (EI, 70 eV, [M+]): [C6H7N3O2+], calculated: 141.0533, found: 141.0534.

Synthesis of 1-N-Methyl-5-hydroxymethylcytosine (20)



20

At room temperature and under nitrogen, to a dispersion of 0.5 g of 5-hydroxymethylcytosine (3.54 mmol) in 33 mL of anhydrous DMF, 34.5 mL of a 0.1 M solution of NH_4OH in MeOH/iPrOH (3.54 mmol, 1 eq) was added slowly. Now methyl iodide (603 mg, 4.25 mmol, 1.2 eq) in 5 mL of anhydrous DMF were added slowly over 15 min and the resulting reaction mixture was stirred for 3h. After the reaction was completed, all volatiles were removed under reduced pressure to afford a slightly yellow crude product, which was purified by flash column chromatography over silica gel (20% -> 25% MeOH/5% NH_4OH/DCM) to afford a clean white powder (250 mg, 1.61 mmol, 46%).

R_f (25% MeOH/5% NH_4OH/DCM) = 0.39

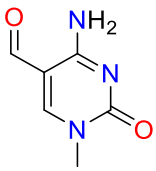
^{1}H-NMR (400 MHz, DMSO-D6, ppm): δ = 7.54 (s, 1H, C_6-*H*), 7.11 (br-s, 1H, N*H*), 6.44 (br-s, 1H, N*H*), 5.00 (s, 1H, -O*H*), 4.15 (s, 2H, -C*H*$_2$-OH), 3.21 (s, 3H, -C*H*$_3$).

^{13}C-NMR (100 MHz, DMSO-D6, ppm): δ = 164.74, 156.25, 144.67, 105.07, 57.14, 36.50.

EA: Calculated [%]: C: 46.45 N: 27.08 H: 5.85; Found [%]: C: 46.00 N: 26.00 H: 5.82.

HR-MS (EI, 70 eV, M+): [C6H9N3O2+], calculated: 155.0690, found: 155.0694.

Synthesis of 1-N-methyl-5-formylcytosine (4b)



4b

A solution of 1-methyl-5-hydroxymethylcytosine (119 mg, 0.77 mmol) and activated MnO_2 (333 mg, 3.84 mmol, 5 eq) in 12 ml of anhydrous acetonitrile was stirred at room temperature for 20 h. Now the reaction mixture was diluted with methanol (10 mL) and filtered (washed with methanol). The crude product was purified by flash column chromatography over silica gel (12% -> 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, D_2O, ppm): δ = 9.45 (s, 1H, formyl-*H*), 8.46 (s, 1H, C_6-*H*), 3.49 (s, 1H, -C*H*$_3$).

^{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).

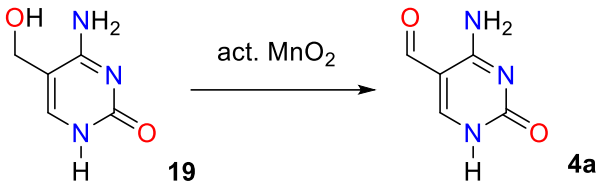
^{13}C-NMR (100 MHz, D_2O, ppm): δ = 189.93, 163.04, 160.37, 156.95, 105.36, 38.32.

^{13}C-NMR (100 MHz, DMSO-D6, ppm): δ = 188.08, 162.68, 160.18, 153.90, 104.27, 37.73.

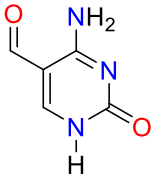
EA: Calculated [%]: C: 47.06 N: 27.44 H: 4.61; Found [%]: C: 47.90 N: 26.09 H: 4.36.

HR-MS (EI, 70 eV, M+): [C6H7N3O2+], calculated: 153.0533, found: 153.0533.

Synthesis of 5-Formylcytosine (4a)



Synthesis of 5-Formylcytosine (4a)



A solution of 5-hydroxymethylcytosine (268 mg, 1.9 mmol) and activated MnO_2 (8.37 mg, 39.63 mmol, 5 eq) in 25 ml of anhydrous acetonitrile was stirred at room temperature for 48 h. Now the reaction mixture was diluted with methanol (20 mL) and filtered (washed with methanol and water). The crude product was purified by flash column chromatography over silica gel (20% -> 25% MeOH/5% NH_4OH/DCM) to afford a clean slightly yellow powder (52 mg, 0.37 mmol, 21%).

R_f (25% MeOH/5% NH_4OH/DCM) = 0.33

^{1}H-NMR (400 MHz, D_2O, ppm): δ = 9.48 (s, 1H, formyl-*H*), 8.37 (s, 1H, C_6-*H*).

^{1}H-NMR (400 MHz, DMSO-D6, ppm): δ = 11.61 (br-s, 1H; N-*H*) 9.44 (s, 1H, formyl-*H*), 8.41 (s, 1H, C_6-*H*), 7.98 (br-s, 1H, -NH_2), 7.84 (br-s, 1H, -NH_2).

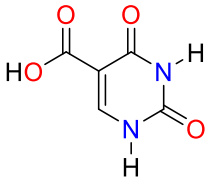
^{13}C-NMR (100 MHz, DMSO-D6, ppm): δ = 188.81, 162.95, 157.63, 154.34, 104.01.

EA: Calculated [%]: C: 43.17 N: 30.21 H: 3.62; Found [%]: C: 41.69 N: 28.54 H: 3.47.

HR-MS (EI, 70 eV, M+): [C5H5N3O2+], calculated: 139.0377, found: 139.0364.

Other Synthesis

Synthesis of 5-Carboxyuracil (21)



21

Following the procedure by Hudson et al.[1], 5-formyluracil (8.5 mmol, 1.2 g, 1.0 eq.) was dissolved in water (100 mL) by heating to reflux. Potassium persulfate (45.1 mmol, 12.2 g, 1.8 eq.) and silver nitrate (0.74 mmol, 3 mol%, 12.53 mmol) were added and the solution was stirred for 30 minutes while cooling to 65 °C. The mixture was cooled to 4 °C, and the precipitated crude product was collected by filration, setting aside the filtrate and washing the crude product with cold water. To purify, it was then suspended in water (30 mL) and 2 M sodium hydroxide solution was added until the suspension turned clear. The product was then re-precipitated by adding concentrated hydrochloric acid until p*H* 2 was reached and cooling to 4 °C. This procedure was repeated after removing the solvent from the first filtration to yield 5-carboxyluracil (3.33 mmol, 0.52 g, 39 %) as greyish solid.

^{1}H-NMR (400 MHz, DMSO-d6, ppm): δ = 12.74 (s, 1H), 12.15 – 11.85 (m, 2H), 8.26 (s, 1H).

^{13}C-NMR (100 MHz, DMSO-d6, ppm): δ = 165.22, 163.55, 150.44, 150.23, 101.27.

EA: Calculated [%]: C: 38.47 N: 17.95 H: 2.58; Found [%]: C: 35.53 N: 17.03 H: 2.80.

HR-MS: (EI, 70 eV, M+): [C5H4N2O3], calculated: 156.0171, found: 156.0174.

3.3 ^{18}O Isotopic Exchange Experiment

3.3.1 1-N-Methyl-5-formyluracil (13b)

In the following section the mass spectra for the ^{18}O-isotopic exchange experiments of the N-methylated nucleobases are shown and analyzed. All measurements are performed with a Advion expression CMS as stated under the section **3.1**.

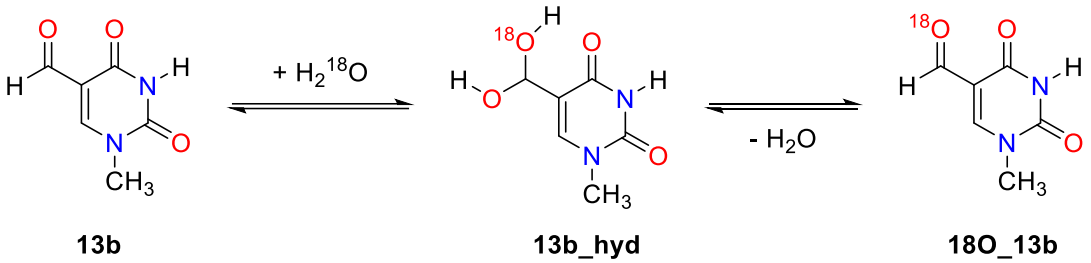


Figure S1: Reaction scheme for the ^{18}O-isotopic exchange reaction of **13b** via the hydrate form **13b_hyd**.

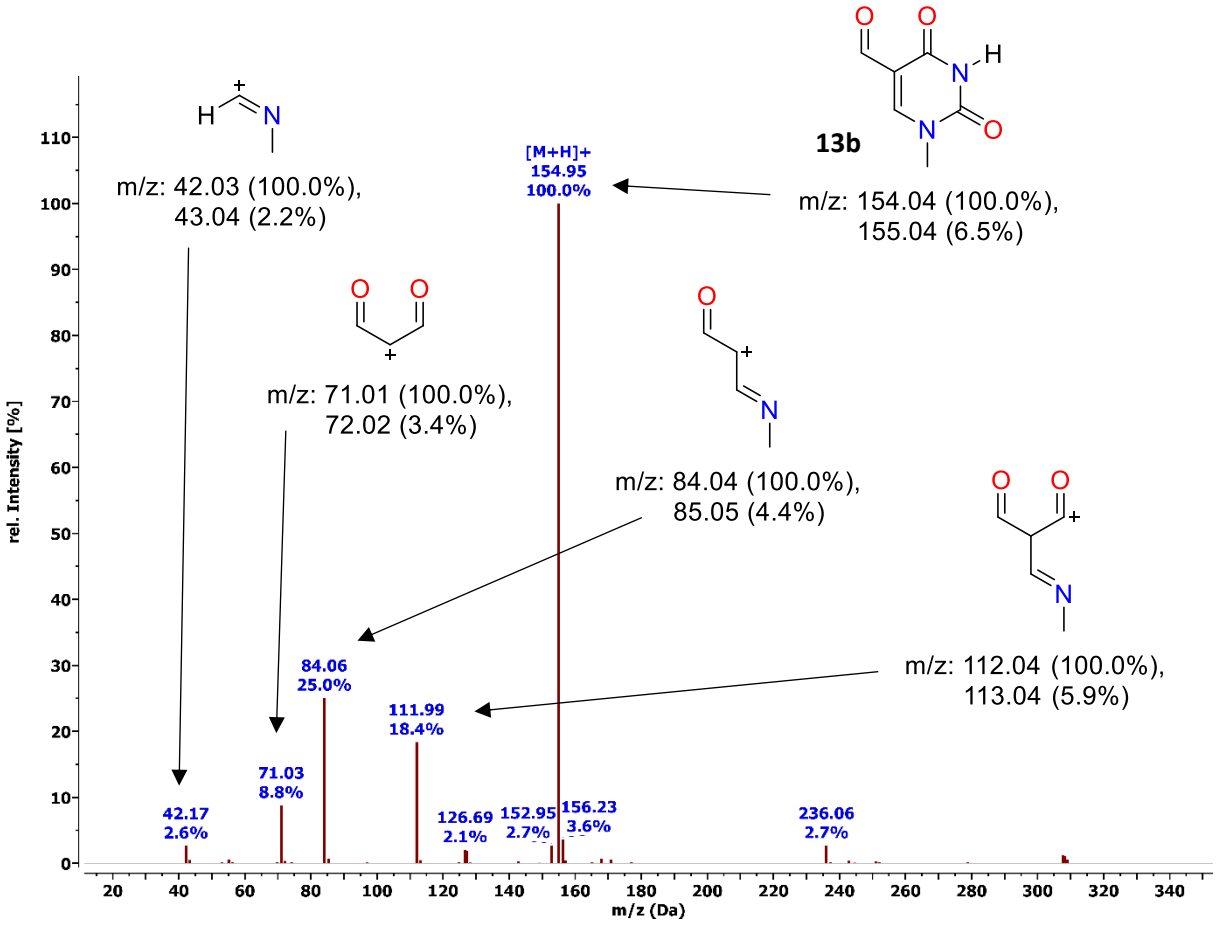


Figure S2: Mass spectra of 1-N-methyl-5-formyluracil (**13b**) measured with Advion expression CMs with atmospheric solid analysis probe (ASAP). All relevant fragments are assigned.

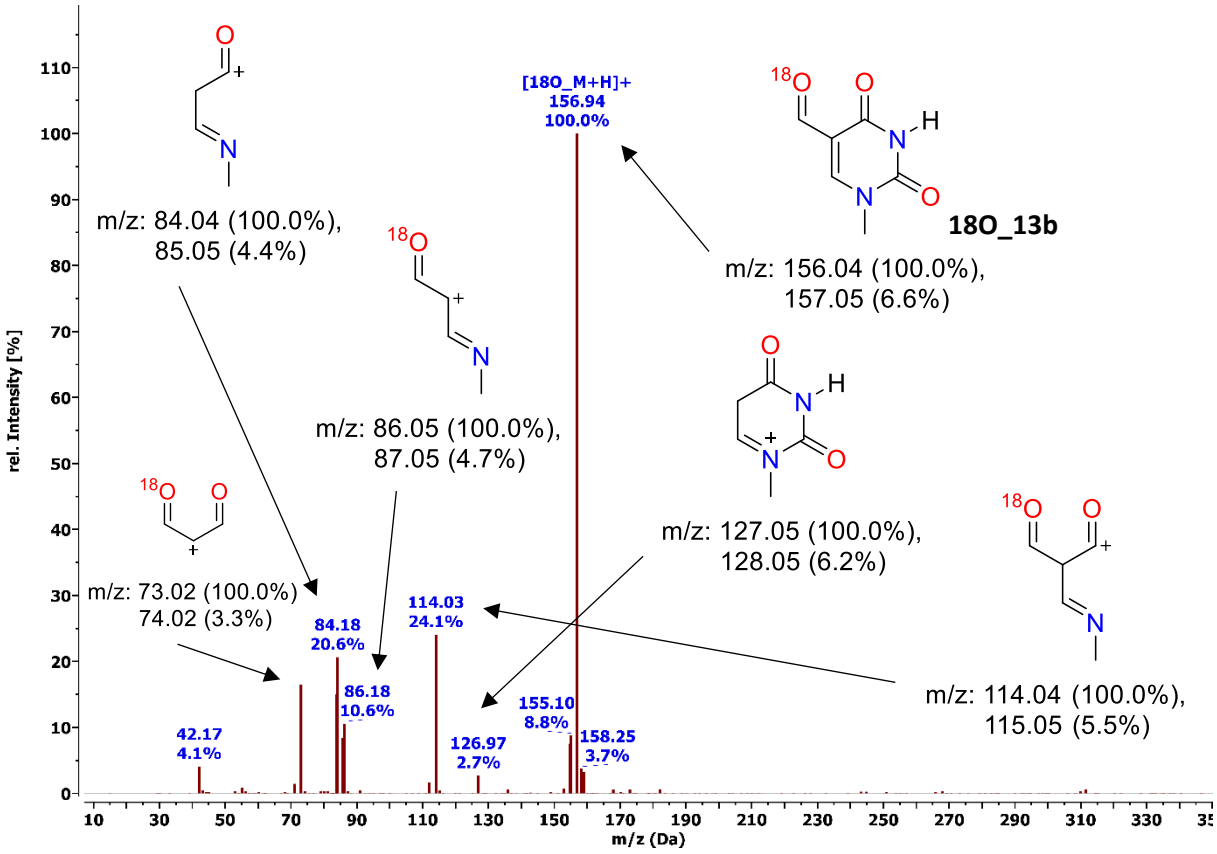


Figure S3: Mass spectra of ^{18}O-labelled 1-N-methyl-5-formyluracil (**13b**) measured with Advion expression CMs with atmospheric solid analysis probe (ASAP). All relevant fragments are assigned.

The fragmentation pattern of the **13b** (Figure S2) and ^{18}O-labeled **18O_13b** (Figure S3) show, that the oxygen exchange can be pinpointed to the formyl oxygen through a hydration mechanism via **13b_hyd** depicted in Figure S1. Please note that the accuracy of the Advion expression CMS is ± 0.1 m/z (HR-MS identification of **13b** see in Synthesis section 3.2).

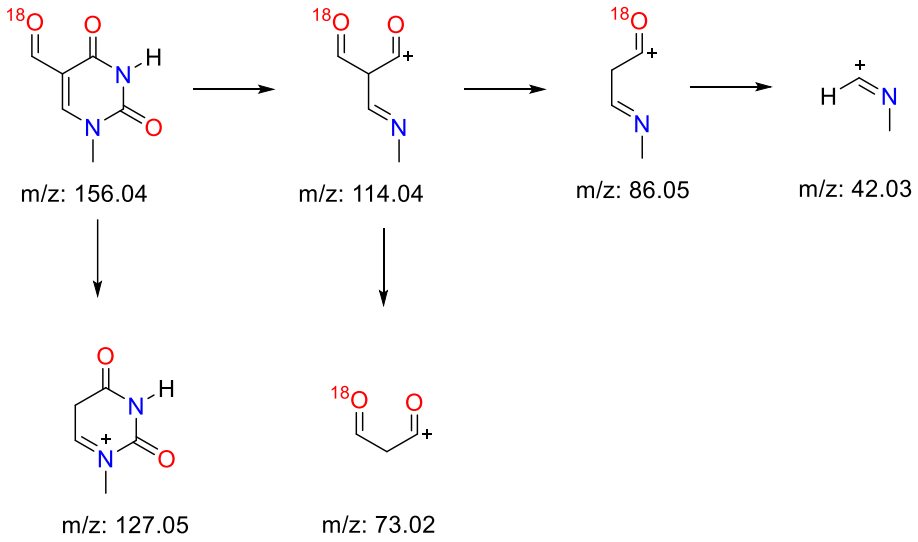


Figure S4: Main fragmentation pattern and its m/z values seen in the MS spectrum of **13b** (Figure S3).

3.3.2 1-N-Methyl-5-formylcytosine (4b)

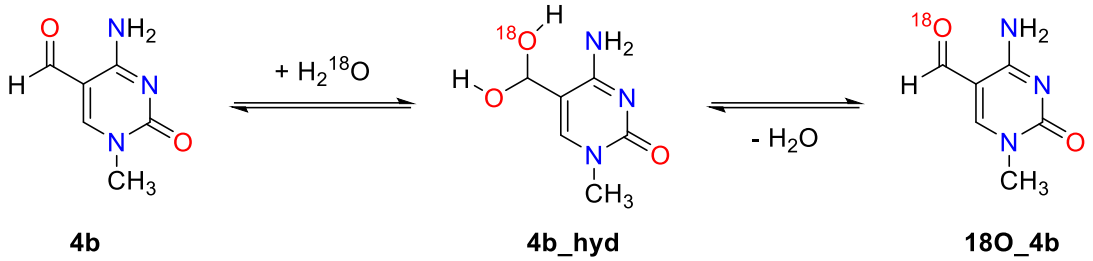


Figure S5: Reaction scheme for the ^{18}O-isotopic exchange reaction of **4b** via the hydrate form **14b_hyd**.

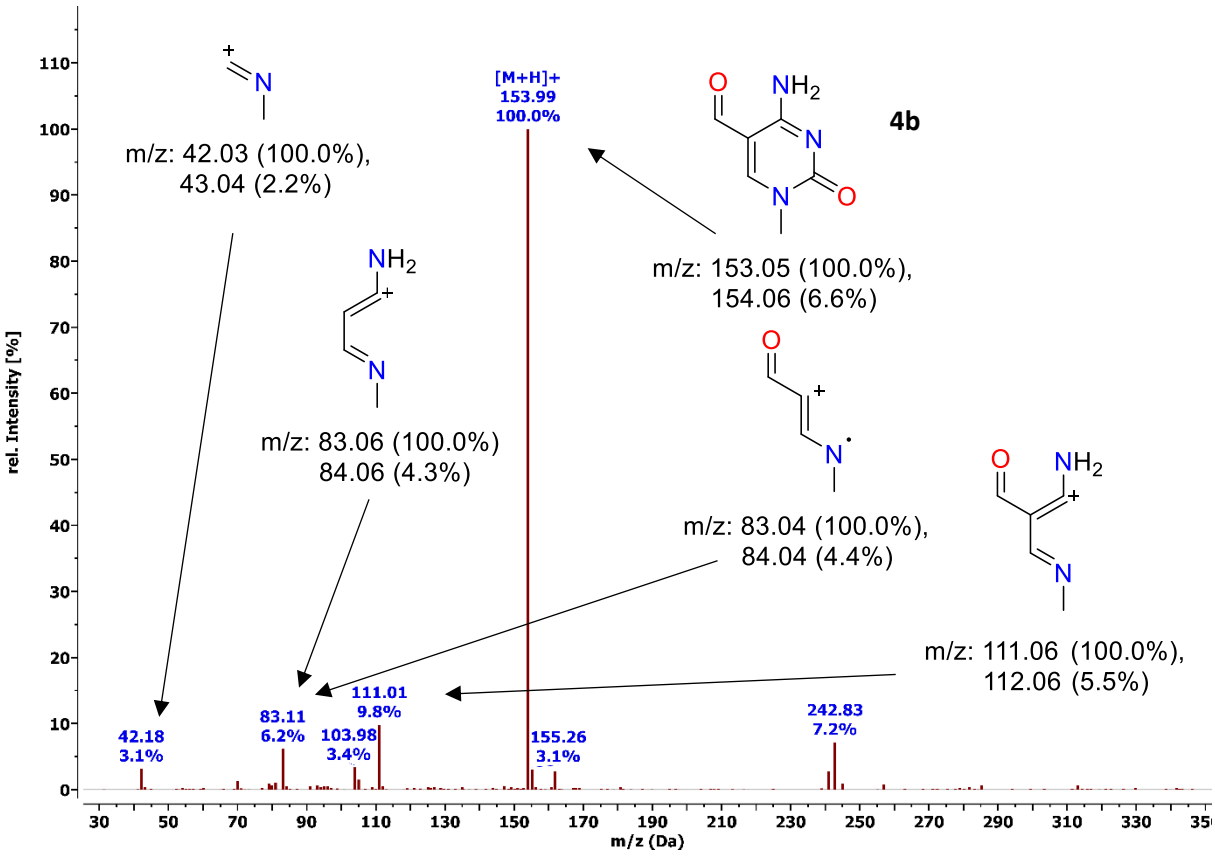


Figure S6: Mass spectra of 1-N-methyl-5-formylcytosine (**4b**) measured with Advion expression CMs with atmospheric solid analysis probe (ASAP). All relevant fragments are assigned.

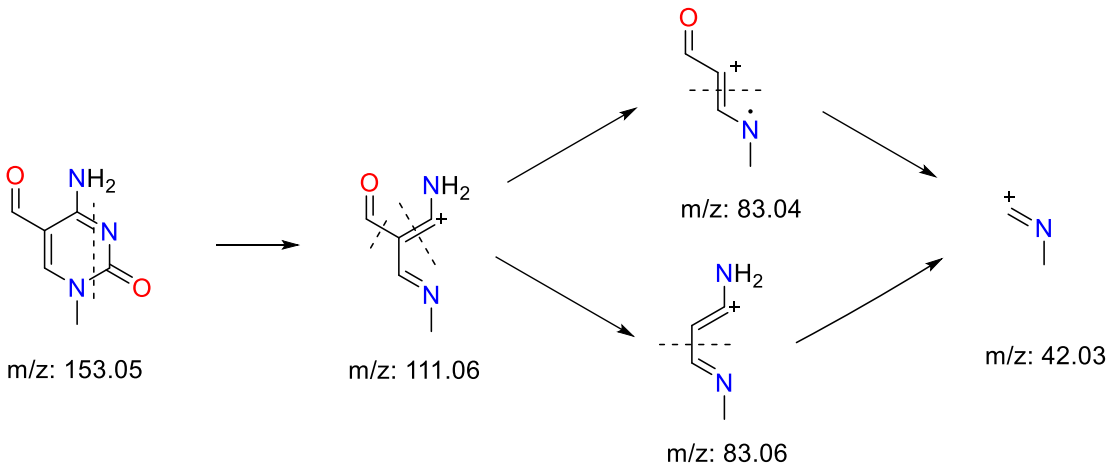


Figure S7: Main fragmentation pattern and its m/z values seen in the MS spectrum of **4b** (Figure S6).

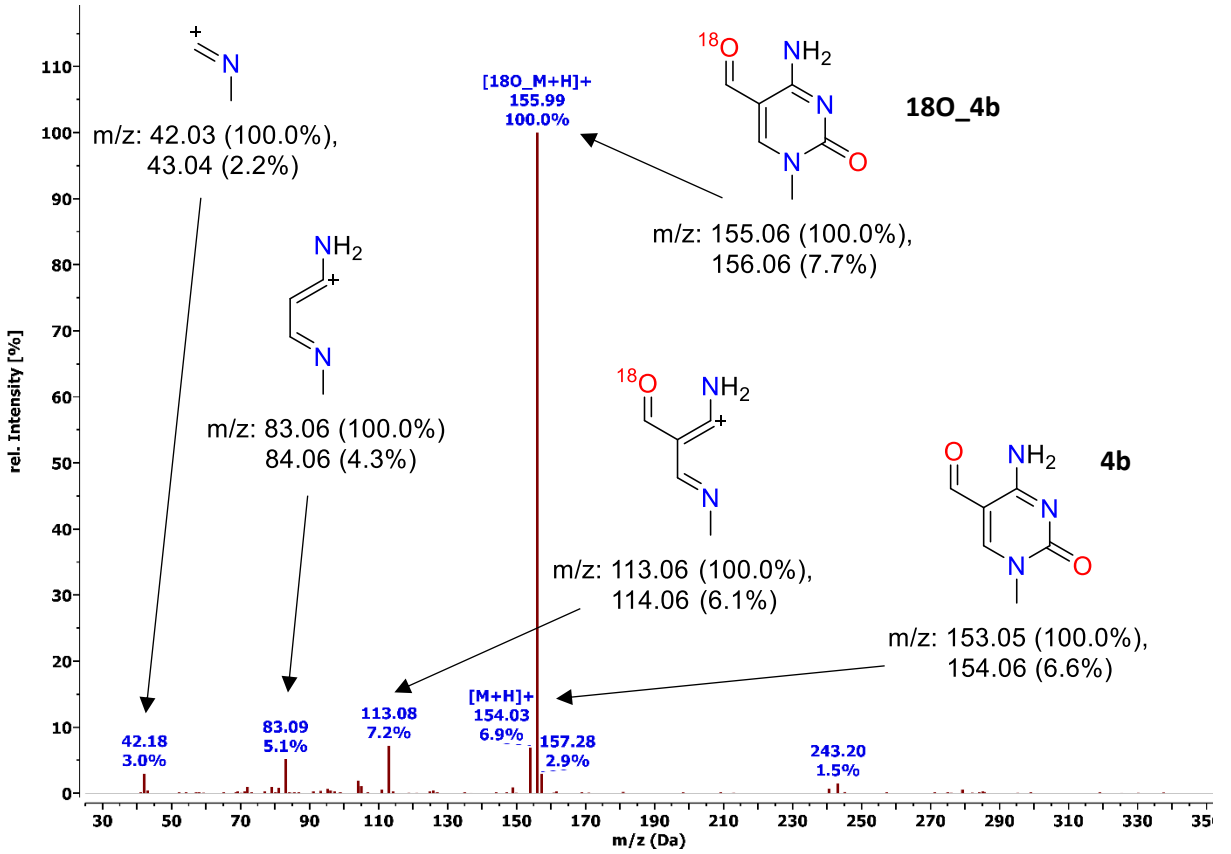


Figure S8: Mass spectra of ^{18}O-labelled 1-N-methyl-5-formylcytosine (4b) measured with Advion expression CMs with atmospheric solid analysis probe (ASAP). All relevant fragments are assigned.

The fragmentation patterns of the **4b** (Figure S6) and ^{18}O-labeled **18O_4b** (Figure S8) show, that the oxygen exchange can be pinpointed to the formyl oxygen through a hydration mechanism via **4b_hyd** depicted in Figure S5. Please note that the accuracy of the Advion expression CMS is ± 0.1 m/z (HR-MS identification of **4b** see in Synthesis section). In Figure S8 a residual signal of the unlabeled nucleobase can be seen.

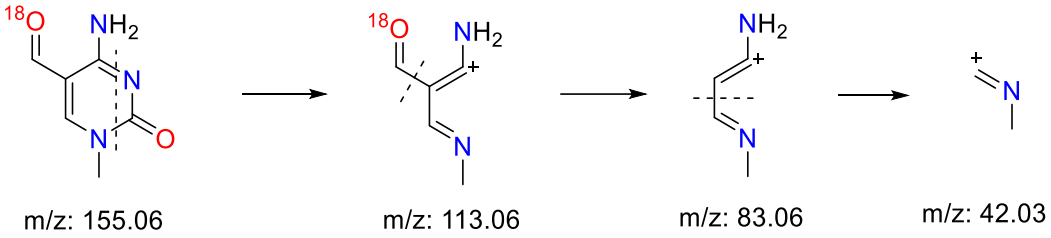


Figure S9: Main fragmentation pattern and its m/z values seen in the MS spectrum of **18O-4b** (Figure S8).

3.4 NMR Identification of Hydrated 5-Formyl Nucleobases

3.4.1 Purification of 5-Formyl Nucleobases

To obtain a sufficiently clean compound for further analysis an additional preparative HPLC purification step was performed.

For uracil compounds preparative RP-HPLC was performed on a HPLC Waters Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump) equipped with the column VP 250/32 C18 from Macherey Nagel using a flow of 5 ml/min, a gradient of 0-3% of buffer B in 60 min was applied for the purification. Nucleobases were purified using the following buffer system: buffer A: 2 mM $NH_4^+HCOO^-$ (p*H* 4.0) in H_2O and buffer B: 2 mM $NH_4^+HCOO^-$ (p*H* 4.0) in 80% (v/v) acetonitrile. The p*H* values of buffers were adjusted using a MP 220 p*H*-meter (Metter Toledo). Nucleobases were detected at a wavelength of 260 nm.

For cytosine compounds preparative RP-HPLC was performed on an Agilent 1260 II system (G1364E 1260 FCPS, G7165A 1260 MWD, G7161A Prep Bin Pump) equipped with the column Macherey-Nagel Nucleosil C-18 column (7 µm, 250mm×10 mm; flow rate: 5 mL/min). As eluents mixtures of water and acetonitrile supplemented with the ion-pair reagent TFA (0.01%–0.1%) were used. The detection wavelength was 254 nm.

3.4.2 5-Formyluracil Hydrate (13a_hyd)

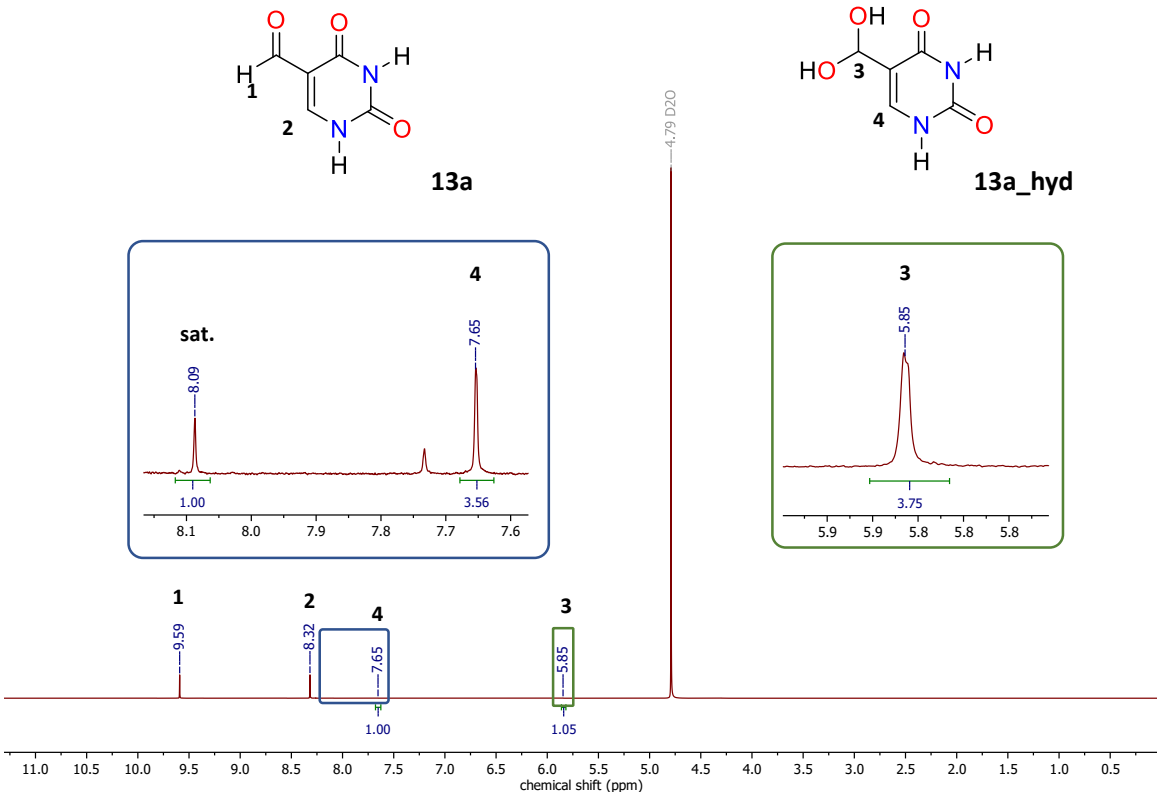


Figure S10: ^{1}H NMR spectra of 5-formyluracil (**13a**) in D_2O with magnification of the NMR signals of the geminal diol form (T = 295K, c = 4.6x10^{-3} mol l^{-3}).

In comparison, the literature ^{1}H NMR chemical shifts for 5-formyl-2'-deoxyuridine in D_2O are 9.67 and 8.80 ppm for the formyl and the H-6 proton.[4]

3.4.3 1-N-Methyl-5-formyluracil hydrate (13b_hyd)

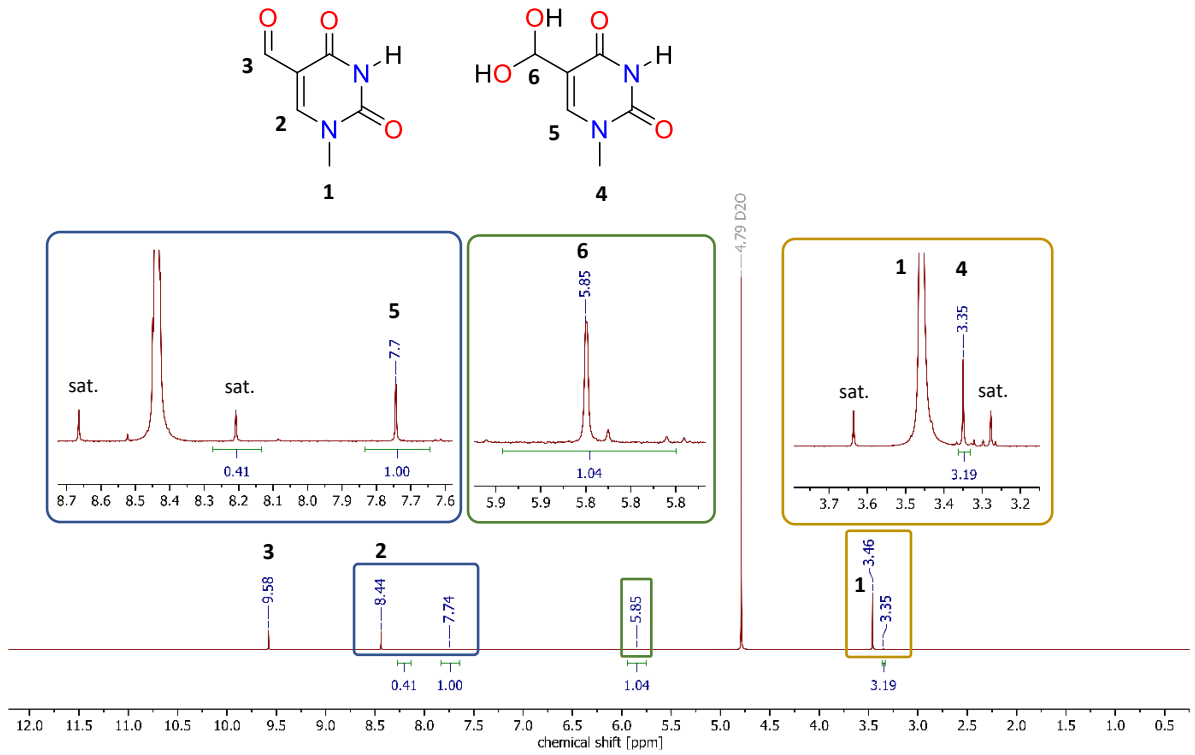


Figure S11: ^{1}H NMR spectra of **13b** in D_2O at a concentration of 3.1x10^{-3} mol l^{-1} together with the relevant signal assignments (T = 295K).

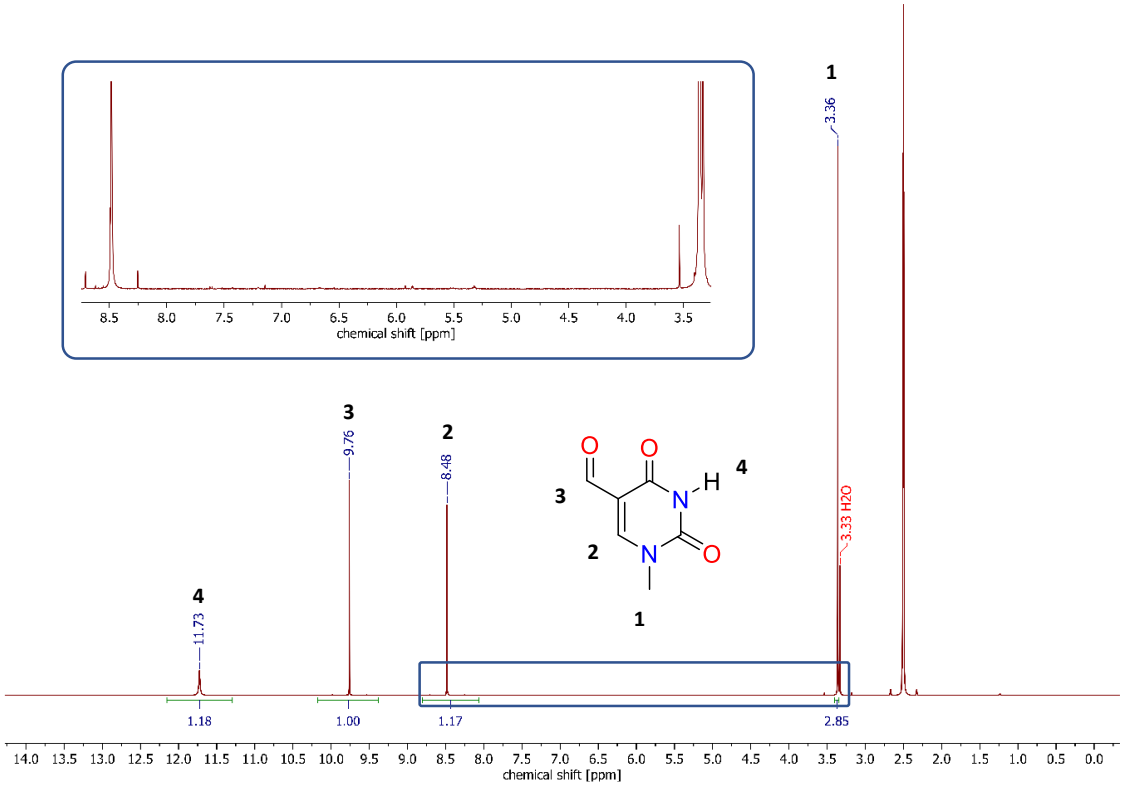


Figure S12: ^{1}H NMR spectra of **13b** in DMSO-D6.

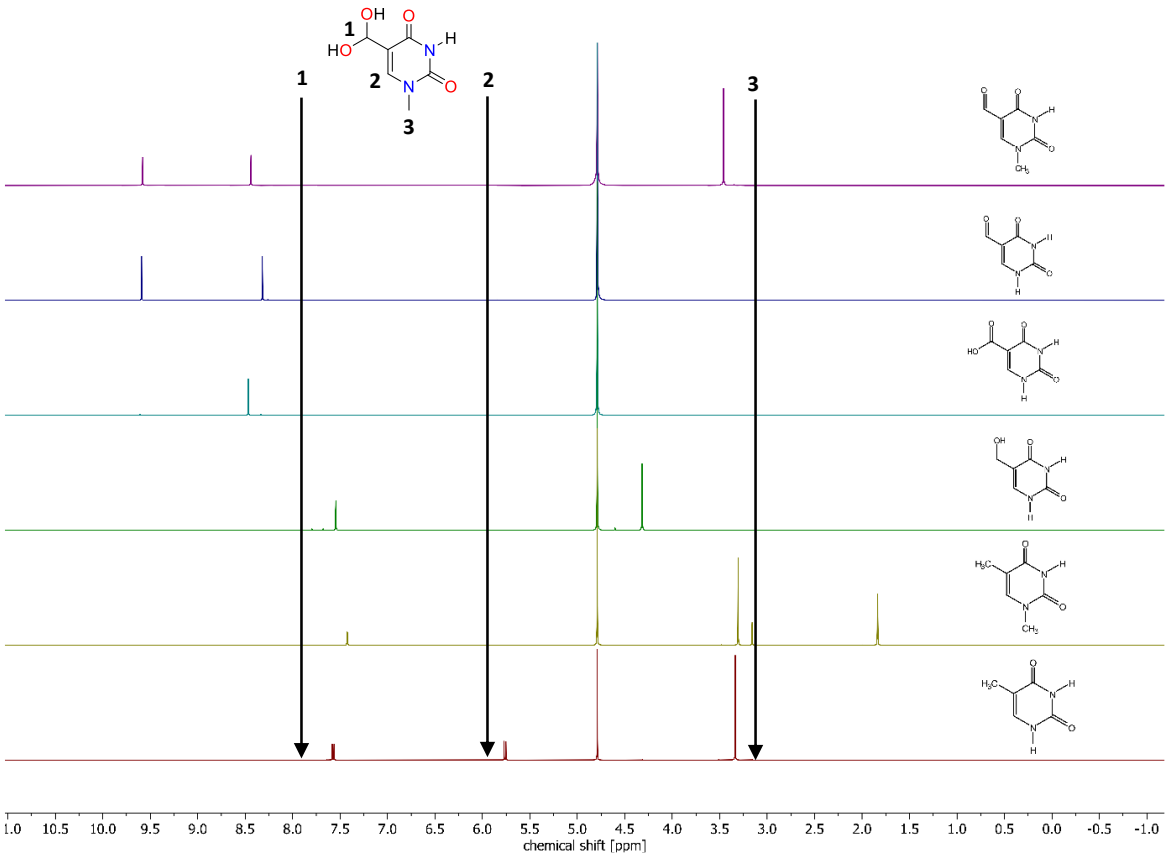


Figure S13: Stacked ^{1}H-NMR spectra in D_2O of important precursors and possible contaminations of **13b_hyd**.

3.4.4 5-Formylcytosine Hydrate (4a_hyd)

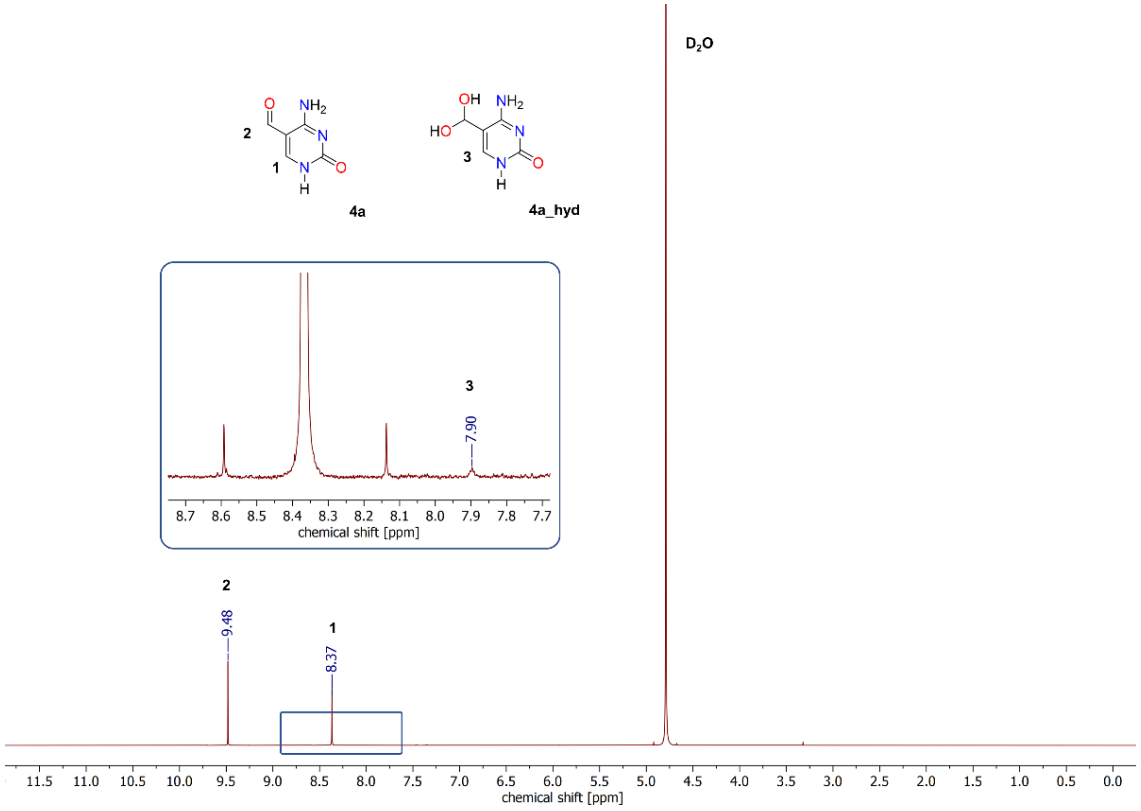


Figure S14: ^{1}H NMR spectra of **4a** in D_2O at pH = 6.7, assumed signal of **4a_hyd** below limit of quantification (SNR = 6, T = 295K).

The hydrated form of **4a (4a_hyd)** was not quantifiable under neutral *pH* conditions. Only a small signal, presumably, for the proton at the C-6 position of **4a_hyd** with a S/N ratio of 6 was detected (see Figure S14). A signal for the proton at the geminal diol position could not be assigned.

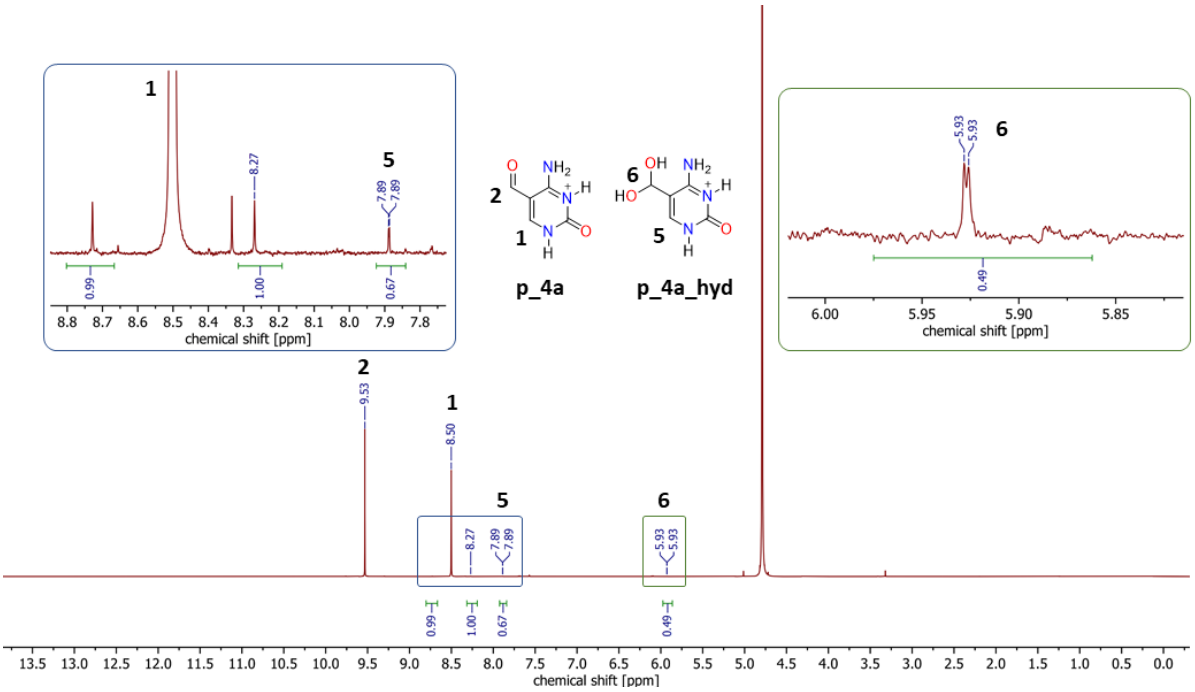


Figure S15: ^{1}H NMR spectra of protonated **p_4a** and its hydrate form **p_4a_hyd** in D_2O together with the relevant signal assignments (T = 295K, *pH* = 3.3).

Upon acidification of the NMR sample of neutral **4a** with 0.1 M DCL in D_2O, a hydrate signal for **p_4a_hyd** could be assigned (see Figure S15, δ (**p_4a_hyd**) = 7.89 (d, 1H, *J* = 1.0 Hz, 1H, 1H), 5.93 (d, 1H, *J* = 1.0 Hz)). These signals were above the limit of quantification LOQ (SNR = 16).

By calculating the LOD and LOQ for the acidified spectrum of **p_4a** (Figure S15, see section 3.5) the limits of the detectable, but not quantifiable hydrate signal (**4a_hyd**) of neutral **4a** for the free energy of hydration ΔG_{hyd} can be postulated between +24.8 – +27.8 kJ mol^{-1}, which is in agreement with the theoretically approximated value of +26.2 (see manuscript and section 3.5).

3.4.5 1-N-Methyl-5-formylcytosin Hydrate (4b_hyd)

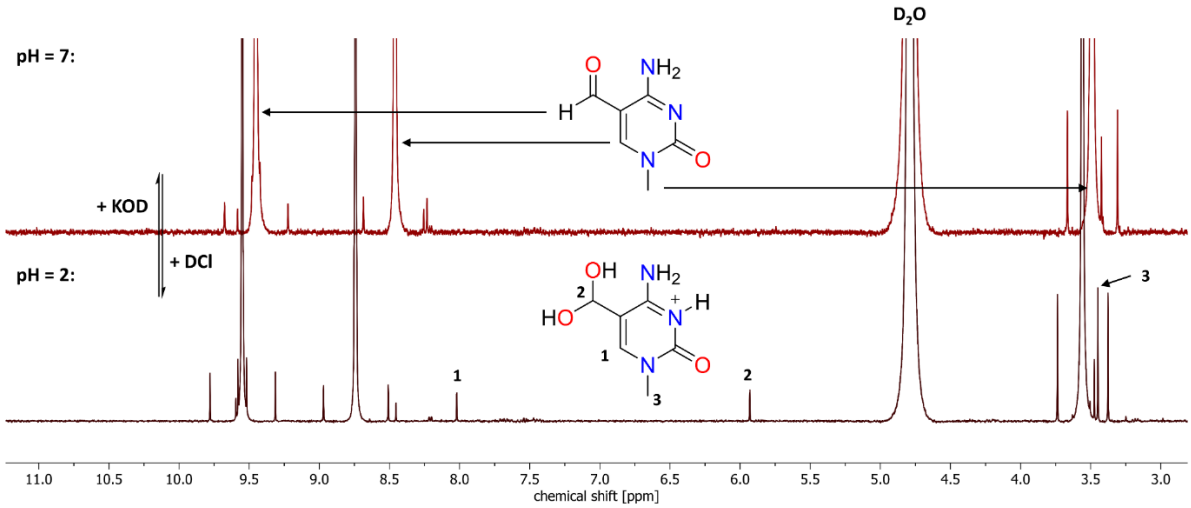


Figure S16: ^{1}H-NMR of 1-N-methyl-5-formyl cytosine **4b** in D_2O (c = 1.63x10^{-3} mol l^{-1}) shows the reversible and increased formation of the hydrated form **4b_hyd** under acidic compared to neutral conditions (pH adjusted by small aliquots of 0.1 M DCl in D_2O, T = 295 K).

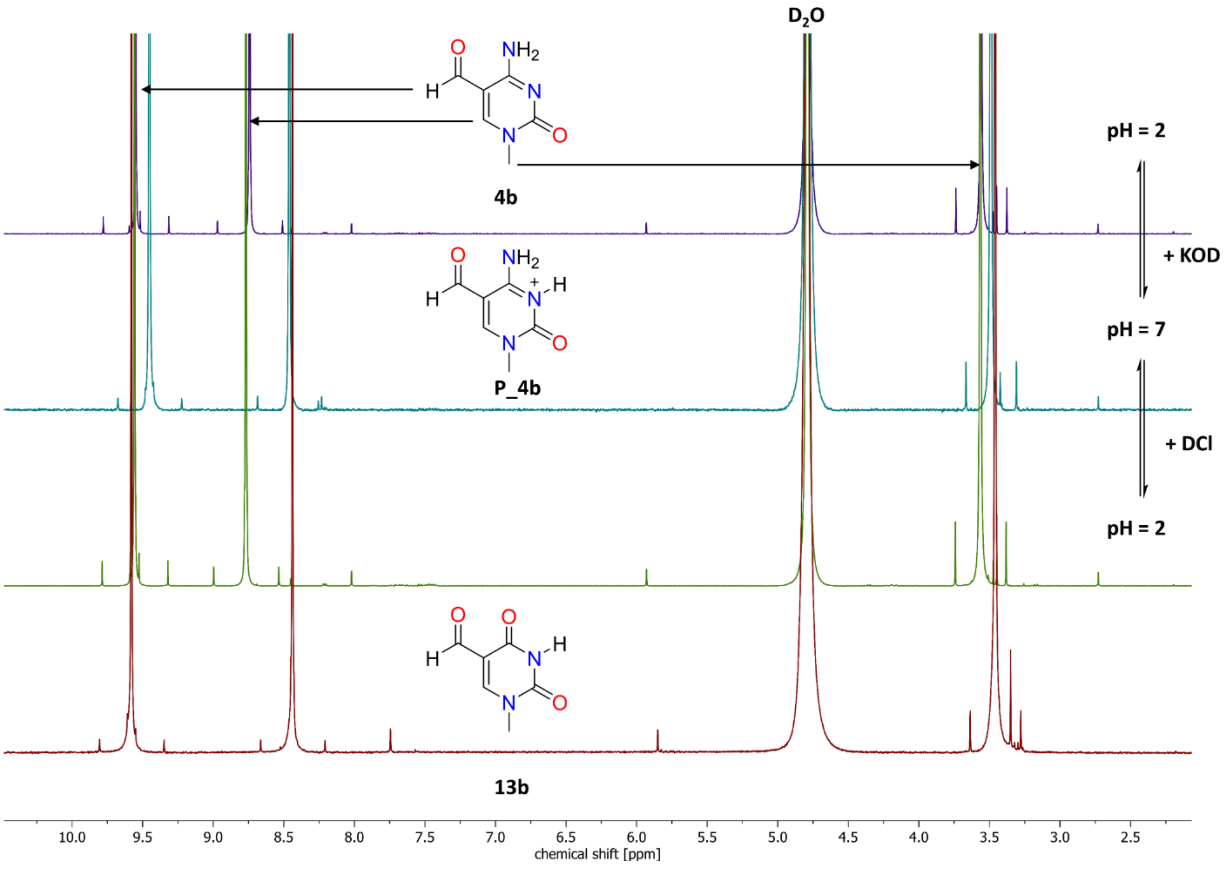


Figure S17: 1 H-NMR of 1-N-methyl-5-formyl cytosine **4b** in D_2O shows the reversible and increased formation of the hydrated form **4b_hyd** under acidic compared to neutral conditions (pH adjusted by small aliquots of 0.1 M DCl in D_2O, T = 295 K)); **13b** shown in comparison.

Under neutral conditions no hydrate form can be detected via ^{1}H-NMR spectroscopy (see Figure S16, top), although the ^{18}O-isotopic exchange experiment (see 3.3.2 1-N-Methyl-5-formylcytosine (**4b**)) indicates fast oxygen exchange at the formyl group through the hydrated form. This indicates that the

hydrate **4b_hyd** is in fact present in small amounts in neutral aqueous solutions, but below the detection level of NMR spectroscopy.

When acidifying the NMR sample of **4b** with small aliquots of 0.1 M DCl in D_2O (see Figure S16; bottom), three new NMR signals can be observed. These signals can be identified as **4b_hyd** by signal integration (see Figure S18), NMR shift calculations (see section 3.6) and by excluding reactant and side product contamination (see Figure S19). This process is reversible upon neutralization with 0.1 M KOD in D_2O with the hydrate signal disappearing (see Figure S16).

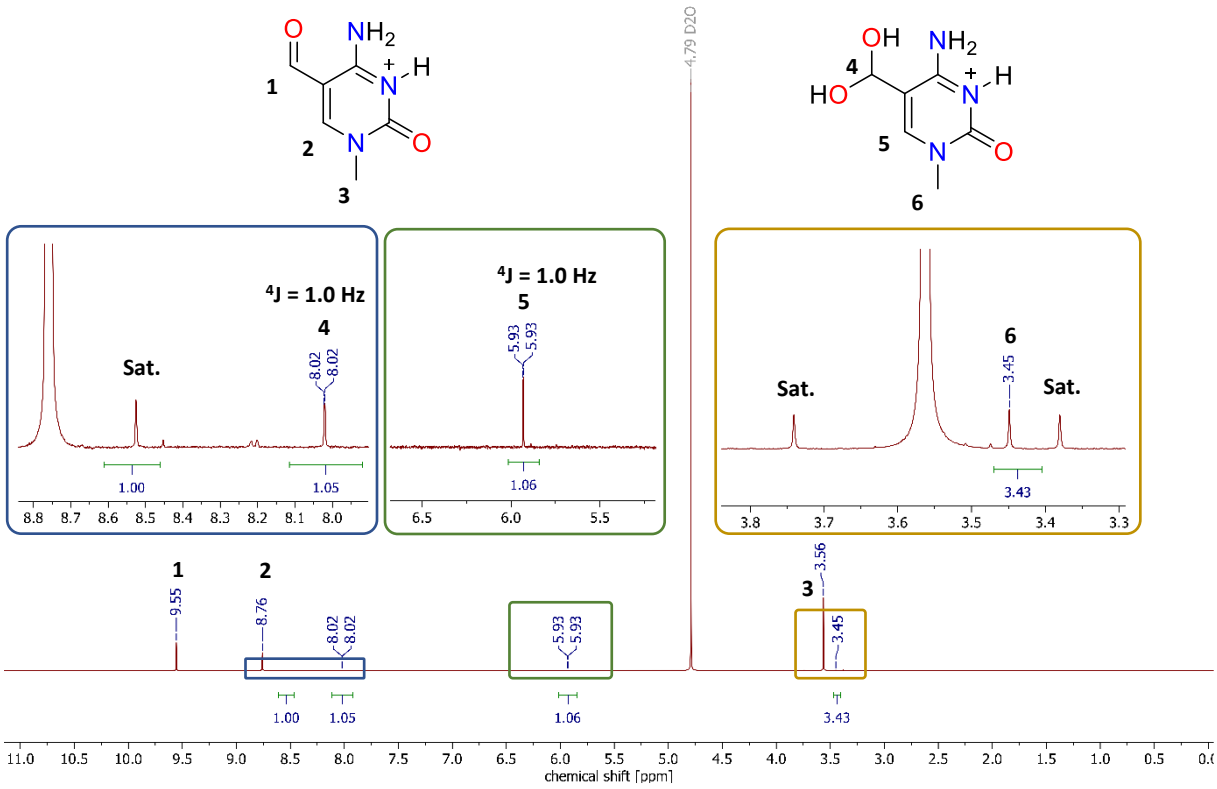


Figure S18: ^{1}H NMR spectra of **4b** in D_2O at a concentration of 2.6x10^{-3} mol l^{-1} together with the relevant signal assignments (pH = 2, T = 295 K)).

In comparison, the literature ^{1}H NMR chemical shifts for 5-formyl-2'-deoxycytidine in D_2O are 9.49 and 8.84 ppm for the formyl and the H-6 proton.[5]

Possible reactant contaminations of the preceding synthesis steps may include cytosine, 1-N-methyl-5-hydroxymethylcytosine, as well as 1-N-methyl-5-carboxycytosine as the result of overoxidation and 1-N-methylcytosine as the result of a deformylation processes. The ^{1}H NMR of 1-N-methylcytosine in D_2O is known in the literature[6] to resonate at a chemical shift of 7.39 (d, $^3J_{H6,H5}$ = 7.42 Hz, 1H, H6), 5.78 (d, $^3J_{H5,H6}$ = 7.23 Hz, 1H, H5) and 3.20 (s, 3H, NCH_3) ppm. Neither the chemical shifts nor the coupling constant match with our findings (see Figure S19). This compound was also synthesized and measured at low concentrations to exclude concentration effects on the chemical shift.

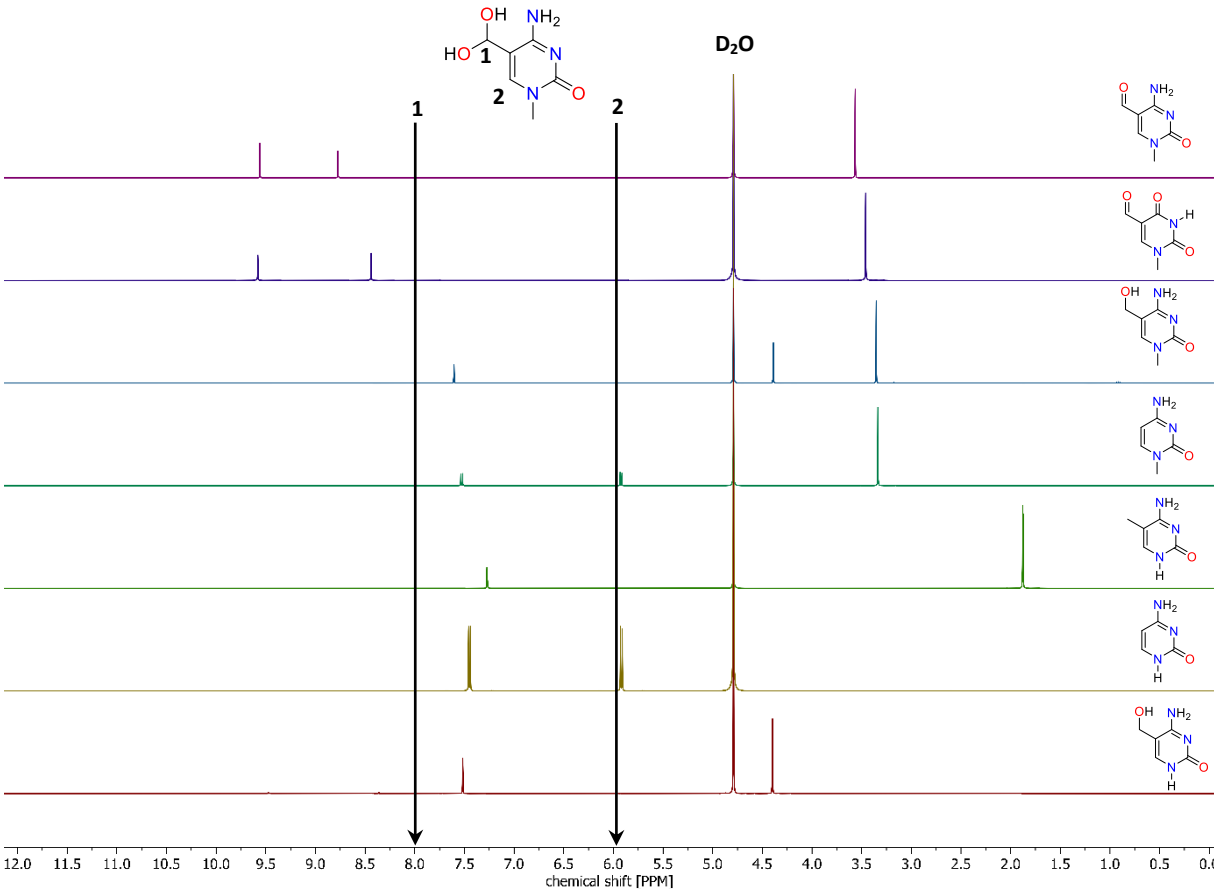


Figure S19: Stacked ^{1}H-NMR spectra in D_2O of important precursors and possible contaminations of **4b_hyd** (black arrow indicates signal of hydrate form).

3.4.6 5-Formyl-2'-deoxycytidine Hydrate (4_hyd)

A sample of 5-formyl-2'-deoxycytidine **4c** was obtained to compare the results of the 1-N-methylated nucleobase **4b** with the 2-desoxyribose as the substituent at the N-1 position.

Spectroscopic data of 5-formyl-2'-deoxycytidine **4c**:

^{1}H NMR (400 MHz, D_2O): δ 9.49 (s, 1H), 8.79 (s, 1H), 6.17 (t, *J* = 6.1 Hz, 1H), 4.51 – 4.33 (m, 1H), 4.11 (d, *J* = 3.4 Hz, 1H), 3.89 (dd, *J* = 12.6, 3.3 Hz, 1H), 3.82 – 3.67 (m, 2H), 3.62 (dd, *J* = 11.7, 4.3 Hz, 1H), 3.52 (dd, *J* = 11.7, 6.5 Hz, 1H), 2.53 (dd, *J* = 6.5, 4.9 Hz, 1H), 2.42 – 2.25 (m, 1H).

^{13}C-NMR (101 MHz, D_2O): δ = 190.29, 162.63, 155.27, 154.59, 105.65, 87.40, 87.12, 71.98, 69.70, 62.39, 60.59, 40.01 (including deoxy-β-L-ribose conformer).

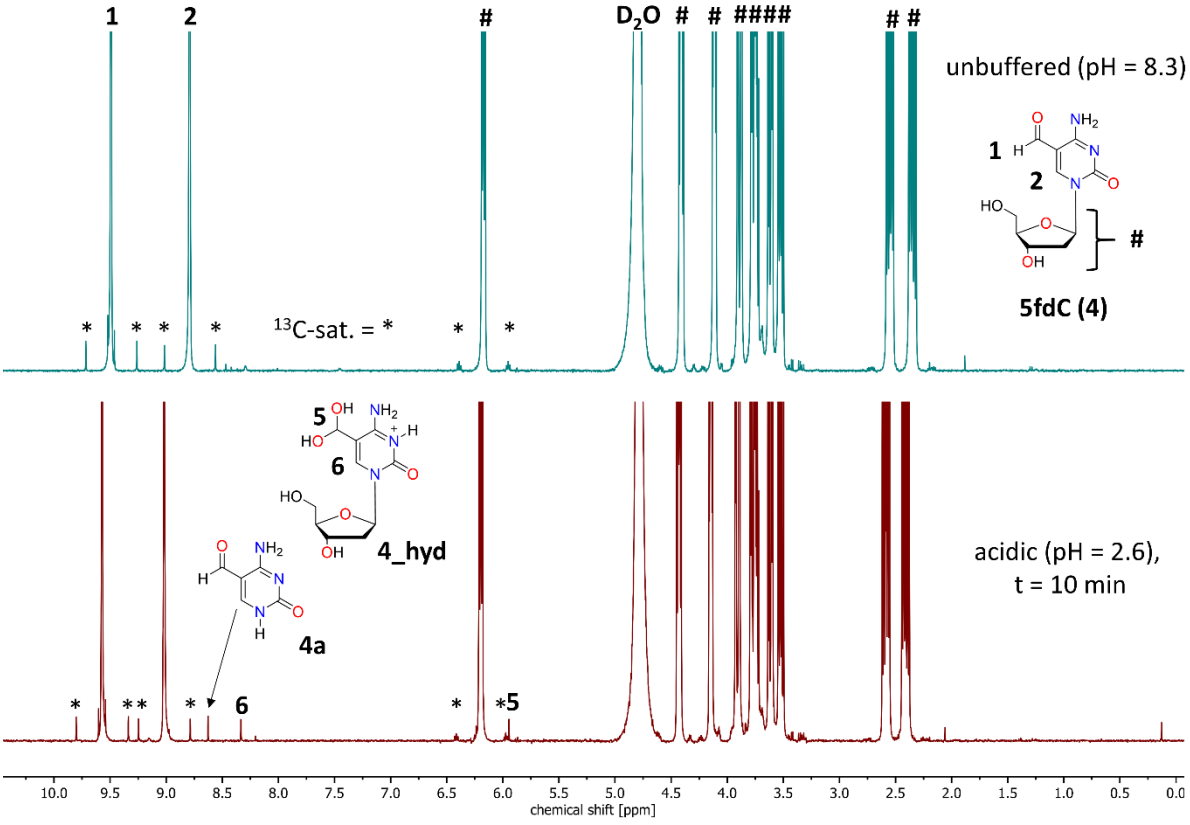


Figure S20: Stacked ^{1}H NMR spectra of 5fdC **(4c)** under neutral, unbuffered conditions vs. acidic conditions (c =1.3x10^{-2} mol l^{-1}). Signal of hydrated species appear only under acidic conditions. The nucleobase 5-formylcytosine **4a** is generated due to acidic cleavage of the glycosidic bond.

A ^{1}H NMR was measured under unbuffered conditions as well as after acidification. In Figure S20, a stacked spectra of 5fdC (**4c**) in D_2O under both unbuffered and acidic conditions is depicted. It can clearly be seen that the nucleotide **4c** behaves in the same way as the N-methylated nucleobase **4b** to form a hydrate **4c_hyd**. Under unbuffered conditions hydrate signals can only be seen as non-quantifiable signals (SNR = 4.34). Upon acidification with 0.1 M DCL in D_2O, distinct signals for the hydrate form arise (Figure S20, signals 5 and 6). The ^{1}H NMR shifts of the hydrate of the nucleotide **4c_hyd** (δ = 8.33 (d, *J* = 0.9 Hz, 1H), 5.95 (d, *J* = 0.9 Hz, 1H)) are very close to the NMR shifts of its 1-N-methylated form **4b_hyd** (δ = 8.02 (d, *J* = 1.0 Hz, 1H), 5.93 (d, *J* = 1.0 Hz, 1H), 3.45 (s, 3H).

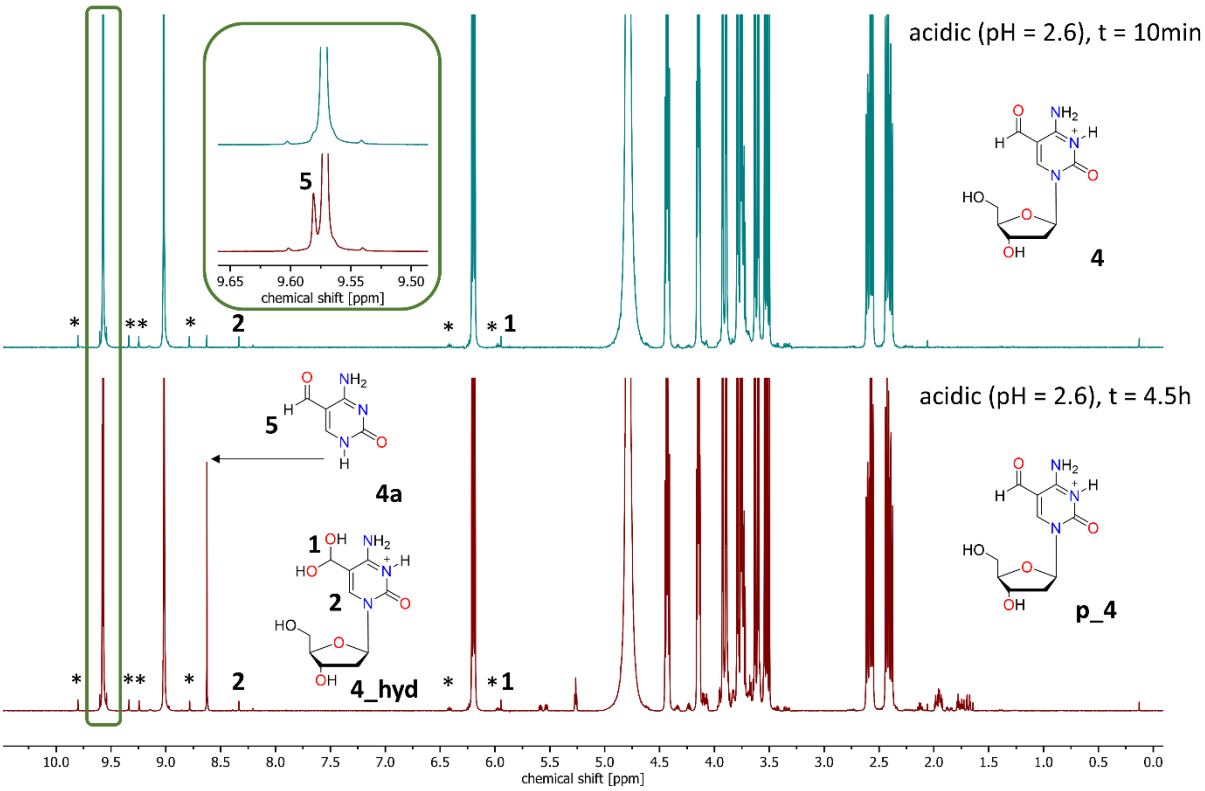


Figure S21: Stacked ^{1}H-NMR spectra of 5fdC (**4c**) in D_2O under acidic conditions after 10 min and 4.5h. The free nucleobase 5-formylcytosine **4a** is generated slowly due to acidic cleavage of the glycosidic bond.

After 4.5 h it can be seen that under acidic conditions the glycosidic bond of nucleotide **4** slowly cleaved to form 5-formylcytosine **4a** and deoxyribose (Figure S21). The ratio between 5fdC (**4**) and its hydrated form (**4_hyd**) of 0.68 % is likely to be not impacted by this decomposition, leading to a free energy of hydration of ΔG_{hyd} = 22.08 kJ/mol.

3.4.7 Benzaldehyde Hydrate (8_hyd)

A sample of freshly distilled benzaldehyde (**8**) in D_2O was measured at a concentration of $3.0x10^{-3}$ mol l^{-1}, to compare literature values with measured values at the concentration range at which hydrates are quantified in this study (see Table S2).

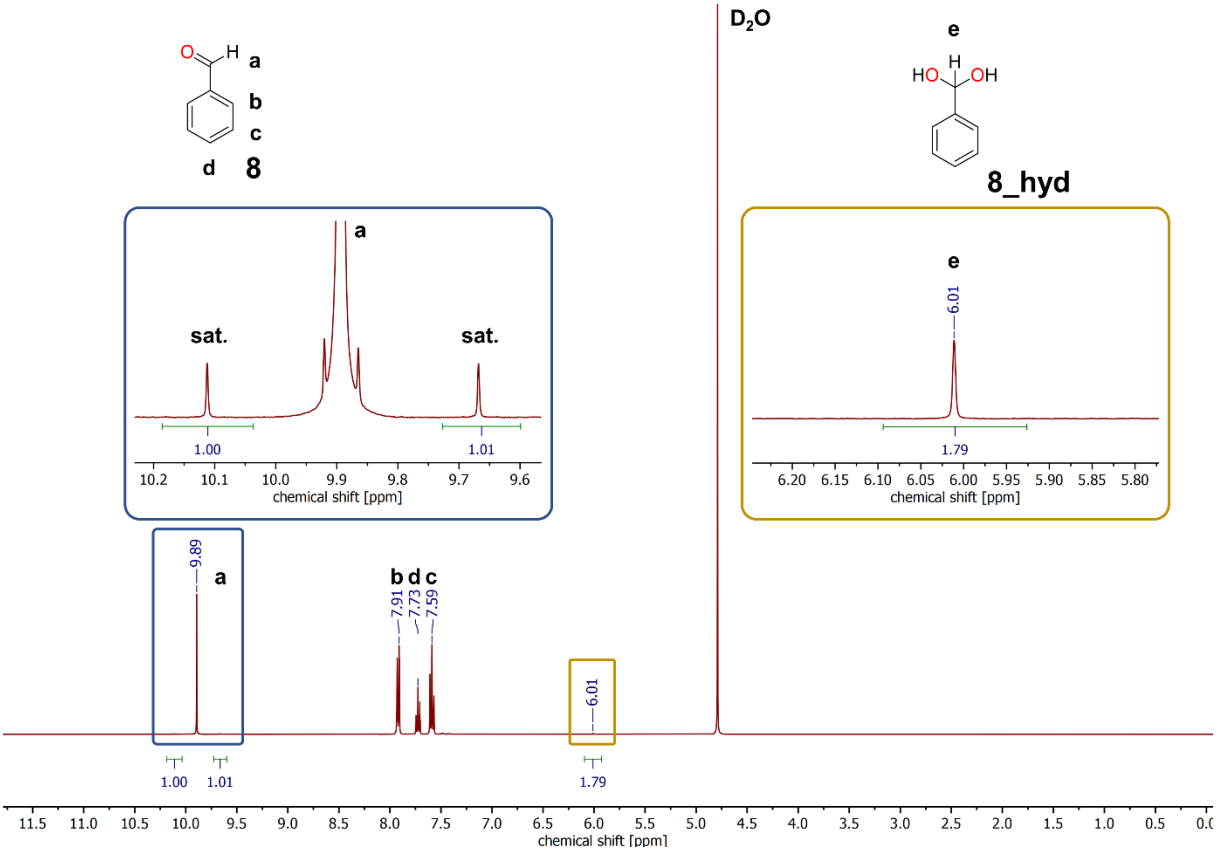
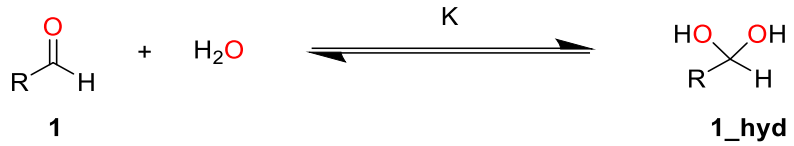


Figure22: ^1H NMR spectra of **8** in D_2O at a concentration of $3.0x10^{-3}$ mol l^{-1} together with the relevant signal assignments (T = 295 K).

The experimental free energy of hydration ($\Delta G_{295K}(exp)$ = 21.31 kJ mol^{-1}, see Table S2) is in good agreement with literature values (ΔG_{298K}(lit.) = 21.1 kJ mol^{-1}, see Manuscript).

3.5 Determination of Equilibrium Constant and Free Energy of Hydration

The reaction of the aldehyde **1** in water yields the respective hydrate **1_hyd** according to equation (1):



1 **1_hyd**

The equilibrium is described by the equilibrium constant *K*, which is defined by the quotient of the reaction concentration of the hydration reaction according to eq. (1):

$$K = \frac{[\mathbf{1\_hyd}]}{[\mathbf{1}][\mathrm{H_2O}]} \tag{1}$$

In dilute solutions it is practical to consider the concentration of water as a constant with $[H_2O]$ = 55.5 mol l^{-1} to form a new equilibrium constant in water K_w

$$K \times [\mathrm{H_2O}] = K_w = \frac{[\mathbf{1}_{\mathbf{hyd}}]}{[\mathbf{1}]} \tag{2}$$

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. (3):

$$\Delta G = -\mathrm{R}T \ \ln(K) \tag{3}$$

The equilibrium constant (1) can be determined by the integrals of the ^{1}H NMR experiments of the formyl compound and its hydrated form. In this study, the signal of the ^{13}C satellite of the excess formyl compound was integrated against the hydrate signals. This strategy increases the accuracy of the integration due to similar signal size (see section 3.4). The spectra were recorded with 256 scans to achieve a high signal to noise ratio (SNR or S/N ratio) above 100 and therefore enable the integration of residual signal in NMR samples with low sample concentration (see Figure S23 and Table S1).[7] These low sample concentrations were chosen for solubility reasons, as well as to exclude the possibility of base pairing events, that can alter the hydration equilibrium.

All signals were integrated after automatic phase and baseline correction by the MestreNova program. All integrals were taken up- and downfield at an area being 20-fold of the “full width at half maximum” (FWHM) in Hertz from the center of the signal. In case of contaminations in this area of integration a slightly shorter integration space was used as exception.

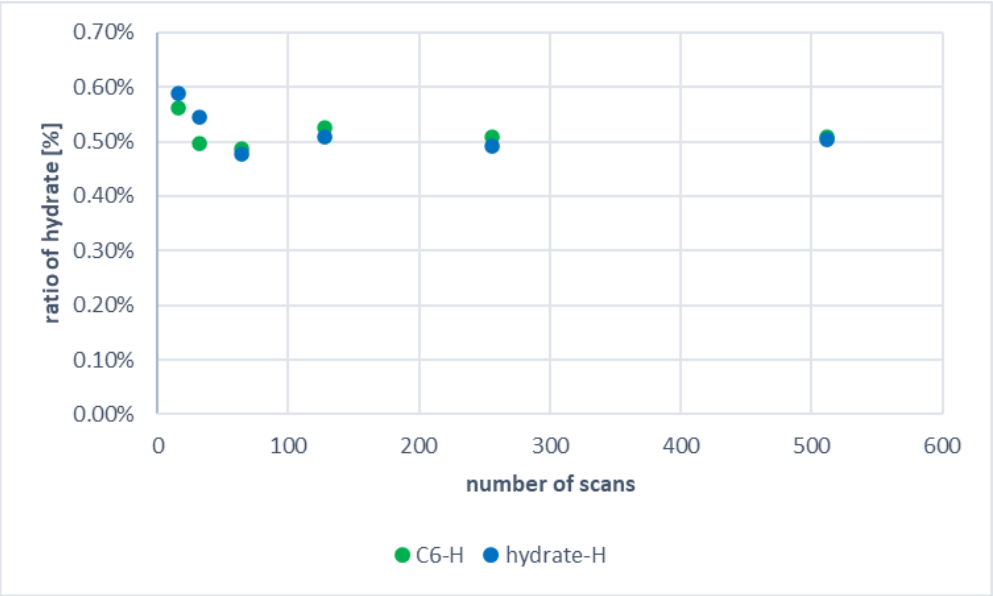


Figure S23: Abundance of hydrate **4b_hyd** determined by ^{1}H NMR measurements for **4b** in D_2O versus the number of scans.

In Figure S18 a plot of the observed **4b_hyd** ratio that can be observed in ^{1}H NMR experiments versus the number of scans can be seen. Since the accuracy of NMR integration above 200 scans stabilizes, all NMR experiments were performed with 256 scans to achieve sufficient S/N ratios and therefore reliable NMR integral values.

The concentration dependence of the hydrate formation for **4b_hyd** can be seen in Table S1. NMR dilution experiments were performed and no significant concentration dependent change in free energies of hydration ΔG was observed. The average free energy of hydration ΔG for 1-N-methyl-5-formyl cytosine **4b** is determined as 22.7 ± 0.1 kJ/mol.

Table S1: Determination of equilibrium constants K, equilibrium constants in water K_w and free energy of hydration ΔG for p_4b_hyd (pH = 2 in D_2O).

conc. 4b [mol/l]	hydrate ratio[a,b]	S/N ratio[b,c]	K[b]	K_w[b]	ΔG (H-6) [kJ/mol]	ΔG (hyd-H) [kJ/mol]	ΔG[c] [kJ/mol]
5.88E-03	0.57/100	123	1.03E-04	5.70E-03	22.5	22.56	22.5
5.34E-03	0.57/100	107	1.02E-04	5.67E-03	22.5	22.5	22.5
4.90E-03	0.52/100	94	9.40E-05	5.22E-03	22.7	22.8	22.7
2.61E-03	0.56/100	51	1.02E-04	5.64E-03	22.6	22.5	22.6
1.87E-03	0.50/100	40	9.01E-05	5.00E-03	22.8	22.9	22.9
1.63E-03	0.51/100	36	9.25E-05	5.14E-03	22.8	22.8	22.8
	Avg.		**Avg.**	**Avg.**		**Avg.**	**22.7**
	0.54/100		9.72E-05	5.40E-03		**Std.Dev.**	**0.1**

[a]: hydrate to aldehyde peak ratio; [b]: average of H-6 and hyd-H NMR integration; [c]: determined by MestreNovas built in S/N ratio tool.

All compounds with identifiable and quantifiable hydrated forms are listed under Table S2. After the accuracy of the hydrate quantification was developed in the example (Table S1, **p_4b**), all following calculations of the free energy of hydration were done by integrating a single sample at the conditions determined above.

Table S2: Summary of hydrate determination.

compound	c [mol l^{-1}]	hydrate ratio[a,b]	K	K_w	Δ*G(exp)* [kJ/mol]
8	$3.0x10^{-3}$	0.96%[c]	$1.74x10^{-4}$	$9.67x10^{-3}$	21.3
p_4a	$3.7x10^{-3}$	0.31%[c]	$5.60x10^{-5}$	$3.11x10^{-3}$	24.1
p_4b	$5.8x10^{-3}$	0.54%	$9.72x10^{-5}$	$5.40x10^{-3}$	22.7
p_4c	$1.3x10^{-2}$	0.68%[c]	$1.23 x10^{-4}$	$6.89x10^{-3}$	22.1
13a	$4.6x10^{-3}$	1.62%[c]	$2.94 x10^{-4}$	$1.63x10^{-2}$	20.0
13b	$3.1x10^{-3}$	1.36%[c]	$2.42 x10^{-4}$	$1.34x10^{-2}$	20.5

[a]: hydrate to aldehyde peak ratio; [b]: average of H-6 and hyd-H NMR integration; [c]: measurement of single NMR sample.

A signal to noise ratio (SNR or S/N ratio) of 3 is generally considered as the limit of detection in analytical procedures, while the limit of quantification (LOQ) is reached at a SNR higher than 10/1.[8] As in this work, the SNR ratios depend on the concentration of the sample and the number of scans in proton NMR measurements, the LOD is calculated individually.

As an example, the limit of detection of **4b_hyd** is derived from the ^{1}H NMR under acidic conditions of the lowest sample concentration (**p_4b_hyd**, c = 1.63 x 10^{-3} mol l^{-1}, 256 scans, Table S3). With an assumed limit of detection for SNR_{LOD} = 3, the hydrate ratio at the limit of detection for compound **p_4b_hyd** can be determined by (SNR values are listed under Table 3):

$$\text{hydrate ratio}_{\text{LOD}} = \left(\frac{[\text{4b\_hyd}]}{[\text{4b}]}\right)_{\text{LOD}} = \frac{[\text{p\_4b\_hyd}]}{[\text{p\_4b}]} \times \frac{\text{SNR}_{\text{LOD}}}{\text{SNR}_{\text{p\_4b\_hyd}}} = 0.0054 \times \frac{3}{35.5} = 0.046\% \quad (4)$$

This limit of detection derived from **p_4b_hyd** applies to **4b_hyd**, if the same NMR sample has been used for acidification. With the above hydrate ratio_{LOD} [%] individual K_{LOD}, $K_{w,LOD}$ and $\Delta G_{hyd,LOD}$ at the limit of detection can be determined by using eq. 1-3:

$$K_{\text{w,LOD}} = \text{hydrate ratio}_{\text{LOD}} = 0.00046$$

$$K_{\text{LOD}} = \frac{K_{\text{w,LOD}}}{[\text{H}_2\text{O}]} = 8.22 \times 10^{-6}$$

$$\Delta G_{\text{hyd,LOD}} = -\text{R}T \ln(K) = 28.7 \text{ kJ mol}^{-1}$$

Similar to equation (4) the hydrate ratio at the limit of quantification (LOQ) for compound **p_4b_hyd** and therefor **4b_hyd** as well as K_{LOQ}, $K_{w,LOQ}$ and $\Delta G_{hyd,LOQ}$ can be determined by:

$$\text{hydrate ratio}_{\text{LOQ}} = \left(\frac{[\text{4b\_hyd}]}{[\text{4b}]}\right)_{\text{LOQ}} = \frac{[\text{p\_4b\_hyd}]}{[\text{p\_4b}]} \times \frac{\text{SNR}_{\text{LOQ}}}{\text{SNR}_{\text{p\_4b\_hyd}}} = 0.0054 \times \frac{10}{35.5} = 0.152\% \quad (5)$$

$$K_{\text{w,LOQ}} = \text{hydrate ratio}_{\text{LOQ}} = 0.00152$$

$$K_{\text{LOQ}} = \frac{K_{\text{w,LOQ}}}{[\text{H}_2\text{O}]} = 2.74 \times 10^{-5}$$

$$\Delta G_{\text{hyd,LOQ}} = -\text{R}T \ln(K) = 25.8 \text{ kJ mol}^{-1}$$

Table S3: Data for the estimation of the limit of detection (LOD) and limit of quantification (LOQ) for hydrate equilibria in proton NMR measurements (256 scans).

compound	condition	C [mol/l]	SNR(^{13}C-sat.)	SNR(H-6)	SNR(H-hyd)	SNR_{avg}(hyd)	hydrate ratio [%]	$hydrate\ ratio_{LOD}$ [%]
p_4a_hyd	acidic	$3.7x10^{-3}$	32	16	16	16	0.36	0.068
								0.225[a]
p_4b_hyd	acidic	$1.6x10^{-3}$	44	34	37	35.5	0.51	0.046
								0.152[a]
p_4c_hyd	acidic	$1.3x10^{-2}$	44	46	n.d.[b]	46	0.68	0.044
								0.148[a]
[a]: LOQ ; [b]: due to signal overlapping.								

The $\Delta G_{hyd,LOD}$ and $\Delta G_{hyd,LOQ}$ values for **4a_hyd** and **4c_hyd** can be calculated in an analog manner to obtain Table S4.

Table S4: Estimation of the free energy of hydration $\mathbf{\Delta G_{hyd}}$ at the limit of detection (LOD) and limit of quantification (LOQ) for compound **4a**, **4b** and **4c** (derived from its acidified parent spectra of **p_4a**, **p_4b** and **p_4c**).

compound	$hydrate\ ratio_{LOD}$ [%]	$K_{w,LOD}$	K_{LOD}	$\Delta G_{hyd,LOD}$ [kJ/mol]
4a	0.068	$6.75\ x10^{-4}$	$1.22\ x10^{-5}$	27.9
4b	0.046	$4.30x10^{-4}$	$8.22x10^{-6}$	28.7
4c	0.044	$4.40x10^{-4}$	$7.93x10^{-6}$	28.8
compound	**$hydrate\ ratio_{LOQ}$ [%]**	**$K_{w,LOQ}$**	**K_{LOQ}**	**$\Delta G_{hyd,LOQ}$ [kJ/mol]**
4a	0.225	$2.25\ x10^{-3}$	$4.05\ x10^{-5}$	24.8
4b	0.152	$1.52\ x10^{-3}$	$2.74\ x10^{-5}$	25.8
4c	0.148	$1.48\ x10^{-3}$	$2.67\ x10^{-5}$	25.8

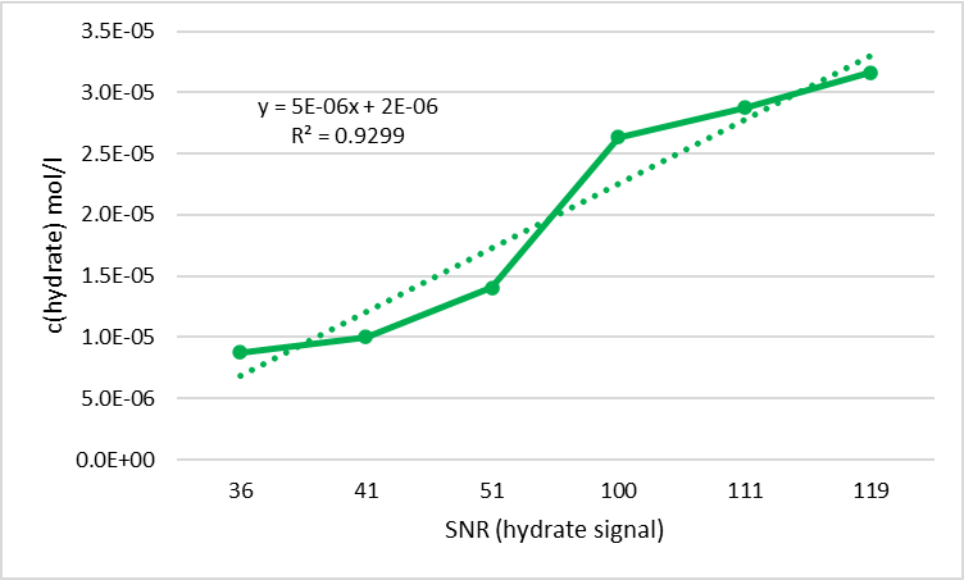


Figure S24: Linear correlation between hydrate concentration c of **p_4b_hydrate** and SNR at low concentrations (hydrate concentration derived from hydrate ratio and sample concentration).

3.6 NMR CALCULATIONS

Geometry optimization was performed at the B3LYP/6-31G(d) level of theory in gas phase.[9] The solution state was modeled both through addition of explicit water molecules and through the implicit continuum solvation model (SMD).[10] For chemical shielding the specifically optimized pcS-3 basis set was used with B3LYP in an H_2O continuum.[11] The isotropic chemical shielding values were calculated at B3LYP/pcS-3 level of theory in gas and solution phase. The ^1H chemical shifts were referenced with respect to synthesized and chemically closely related compounds (5-hydroxymethyluracil, 1-N-methyl-5-hydroxymethylcytosine). Calculations were performed using Gaussian09, Revision D.01.61.[12] Positive Δ_{shift} values indicate higher shielding (and therefor upfield shift) with respect to the reference NMR signal, and vice versa.

5-formyl uracil (13a)

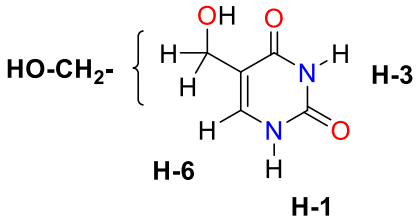


Table S5: Boltzmann averaged calculated chemical shielding of protons for 5-hydroxymethyluracil (5hmU) as reference compound.

		chemical shielding[a]			
	no. conf.	**HO-C*H*$_2$-[b]**	***H*-1**	***H*-3**	***H*-6**
5hmU	1	26.88	23.21	22.95	23.60

(a): SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p); [b]: averaged shielding over two protons

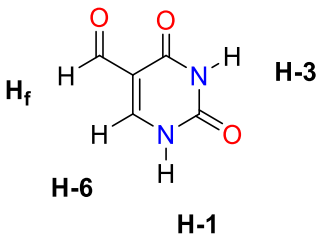


Table S6: Boltzmann averaged calculated chemical shielding of protons for 5-formyluracil (**13a**).

		chemical shielding[a]			
	no. conf.	**H_f**	***H*-1**	***H*-3**	***H*-6**
13a	2	21.32	22.64	22.91	22.80

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p)

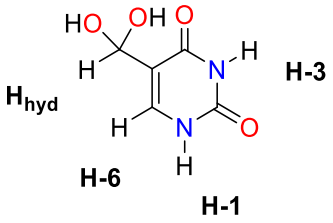


Table S7: Boltzmann averaged calculated chemical shielding of protons for geminal diol form of 5-formyluracil (**13a_hyd**).

		chemical shielding[a]			
	no. conf.	**Hhyd**	***H*-1**	***H*-3**	***H*-6**
13a_hyd	9	25.53	23.15	22.95	23.53

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p)

Table S8: Summary of calculated differences in chemical shift [Δ_{shift}] as well as chemicals shift with 5-hydroxymethyluracil as reference in comparison with observed experimental chemicals shifts in 1H-NMR spectroscopy.

		chemical shift[a] [ppm]			
		h_f	***H*-1**	***H*-3**	***H*-6**
13a	**Δ_{shift}[c]**	-5.56	-0.56	-0.03	-0.79
	calc. shift	**9.98**	**/ [B]**	**/ [B]**	**8.38**
	exp. shift	**9.88**	**/ [B]**	**/ [B]**	**8.33**
		H_{hyd}	***H*-1**	***H*-3**	***H*-6**
13a_hyd	**Δ_{shift}[c]**	-1.35	-0.06	0.01	-0.07
	calc. shift	**5.74**	**/ [B]**	**/ [B]**	**7.59**
	exp. shift	**5.67**	**/ [B]**	**/ [B]**	**7.61**

[a]: relative to 5-hydroxymethyl uracil (5hmU); [B]: not observable in D_2O due to fast proton exchange;
[c]: ^{1}H-NMR of reference 5hmU in D_2O: 7.54 (s, 1H, *H*-6), 4.32 (s, 1H, CH_2-OH)

1-N-Methyl-5-formyl uracil (13b)

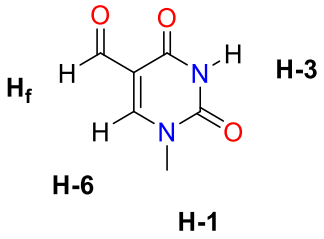


*Table S9: Boltzmann averaged calculated chemical shielding of protons for N-1-methyl-5-formyluracil (**13b**).*

		chemical shielding[a]			
	no. conf.	**H_f**	***H*-1**	***H*-3**	***H*-6**
5hmU	2	21.27	27.79	22.86	22.66

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p)

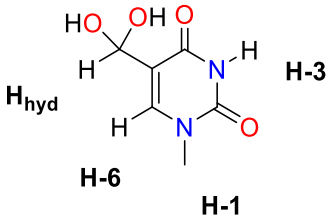


Table S10: Boltzmann averaged calculated chemical shielding of protons for geminal diol form of N-1-methyl-5-formyluracil (**13b_hyd**).

		chemical shielding[a]			
	no. conf.	**Hhyd**	***H*-1**	***H*-3**	***H*-6**
13a_hyd	4	25.79	27.95	22.89	23.50

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p)

Table S11: Summary of calculated differences in chemical shift [Δ_{shift}] as well as chemicals shift with 5-hydroxymethyluracil as reference in comparison with observed experimental chemicals shifts in ^{1}H NMR spectroscopy.

		chemical shift[a] [ppm]		
		h_f	*H*-3	*H*-6
13b	Δ_{shift}[c]	**-5.61**	**-0.09**	**-0.94**
	calc. shift	**9.93**	**/** [B]	**8.48**
	exp. shift	**9.58**	**/** [B]	**8.44**
		H_{hyd}	*H*-3	*H*-6
13b_hyd	Δ_{shift}[c]	**-1.08**	**-0.05**	**-0.09**
	calc. shift	**5.40**	**/** [B]	**7.63**
	exp. shift	**5.85**	**/** [B]	**7.74**

[a]: relative to 5-hydroxymethyl uracil (5hmU); [B]: not observable in D_2O due to fast proton exchange; [c]: ^{1}H-NMR of reference 5hmU in D_2O: 7.54 (s, 1H, *H*-6), 4.32 (s, 1H, C*H*$_2$-OH)

1-N-Methyl-5-formyl cytosine (4b)

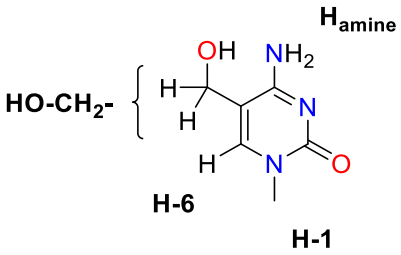


Table S12: Boltzmann averaged calculated chemical shielding of protons for 1-N-methyl-5-hydroxymethylcytosine (5hmC) as reference compound.

		chemical shielding[a]			
	no. conf.	**HO-C*H*$_2$-**[b]	***H*-1**	***H*$_{amine}$**[b]	***H*-6**
5hmC	4	26.79	28.01	25.62	23.59

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p); [b]: averaged shielding over two protons.

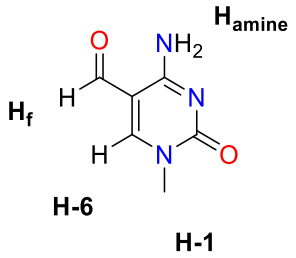


Table S13: Boltzmann averaged calculated chemical shielding of protons for 1-N-methyl-5-formylcytosine (**4b**).

		chemical shielding[a]			
	no. conf.	**H_f**	***H*-1**	**H_{amine}**	***H*-6**
4b	2	**21.61**	**27.84**	**24.02**	**22.88**

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p)

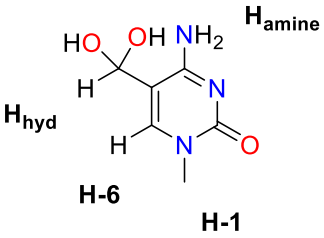


Table S14: Boltzmann averaged calculated chemical shielding of protons for geminal diol form of 1-N-Methyl-5-formyl-cytosine (**4b_hyd**).

		chemical shielding[a]			
	no. conf.	**H_{hyd}**	***H*-1**	**H_{amine}**	***H*-6**
4b_hyd	8	**25.23**	**28.00**	**25.35**	**23.47**

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p)

Table S15: Summary of calculated differences in chemical shift [Δ_{shift}] as well as chemicals shift with 5-hydroxymethylcytosine as reference in comparison with observed experimental chemicals shifts in 1H-NMR spectroscopy.

		chemical shift[a] [ppm]			
		h_f	*H*-1	H_{amine}	*H*-6
4b	Δ_{shift}[c]	**-5.18**	**-0.17**	**-1.60**	**-0.71**
	calc. shift	**9.57**	**3.52**	**/** [B]	**8.31**
	exp. shift	**9.55**	**3.56**	**/** [B]	**8.76**
		H_{hyd}	*H*-1	H_{amine}	*H*-6
4b_hyd	Δ_{shift}[c]	**-1.56**	**-0.01**	**-0.27**	**-0.12**
	calc. shift	**5.93**	**3.36**	**/** [B]	**7.72**
	exp. shift	**5.95**	**3.45**	**/** [B]	**8.02**

[a]: relative to 1-methyl-5-hydroxymethyl cytosine (1m5hmC); [B]: not observable in D_2O due to fast proton exchange; [c]: ^{1}H-NMR of reference 1m5hmC in D_2O: 7.60 (s, 1H, *H*-6), 4.39 (s, 1H, C*H*$_2$-OH), 3.35 (s, 3H, -C*H*$_3$).

Protonated 1-N-Methyl-5-formyl cytosine (p_4b)

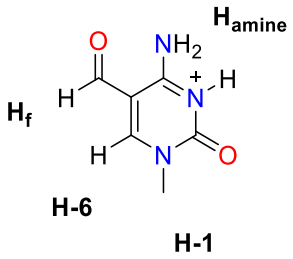


Table S16: Boltzmann averaged calculated chemical shielding of protons for protonated 1-N-methyl-5-formylcytosine (**p_4b**).

		chemical shielding[a]			
	no. conf.	H_f	*H*-1	H_{amine}	*H*-6
p_4b	1	**21.44**	**27.67**	**23.10**	**22.60**

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p).

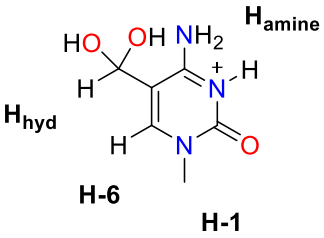


Table S17: Boltzmann averaged calculated chemical shielding of protons for protonated 1-N-methyl-5-formylcytosine hydrate (**p_4b_hyd**).

		chemical shielding[a]			
	no. conf.	**H_{hyd}**	***H*-1**	**H_{amine}**	***H*-6**
p_4b_hyd	7	**25.13**	**27.84**	**24.40**	**23.18**

[a]: SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p).

Table S18: Summary of calculated differences in chemical shift [Δ_{shift}] as well as chemicals shift with 5-hydroxymethylcytosine as reference in comparison with observed experimental chemicals shifts in 1H-NMR spectroscopy.

		chemical shift[a] [ppm]			
		h_f	***H*-1**	**H_{amine}**	***H*-6**
p_4b	**Δ_{shift}[c]**	**-5.35**	**-0.35**	**-2.52**	**-0.99**
	calc. shift	**9.74**	**3.70**	**/ [B]**	**8.59**
	exp. shift	**9.55**	**3.56**	**/ [B]**	**8.76**
		H_{hyd}	***H*-1**	**H_{amine}**	***H*-6**
p_4b_hyd	**Δ_{shift}[c]**	**-1.66**	**-0.18**	**-1.22**	**-0.41**
	calc. shift	**6.05**	**3.53**	**/ [B]**	**8.01**
	exp. shift	**5.95**	**3.45**	**/ [B]**	**8.02**

[a]: relative to 1-methyl-5-hydroxymethyl cytosine (1m5hmC); [B]: not observable in D_2O due to fast proton exchange; [c]: ^{1}H-NMR of reference 1m5hmC in D_2O: 7.60 (s, 1H, *H*-6), 4.39 (s, 1H, C*H*$_2$-OH), 3.35 (s, 3H, -C*H*$_3$)

Full List of Calculated Conformers

Table S19: Two main conformers and general numbering scheme of the reference 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) obtained for the calculation of NMR shieldings (see Table S20).

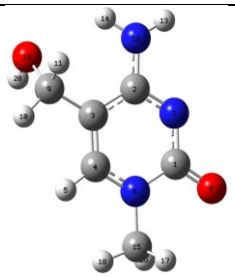
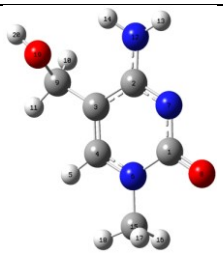
	
nb_5mOHC_smd_scanh	**nb_5mOHC_smd_scana**

Table S20: NMR shielding calculated for conformers of the reference 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hamino1	Hamino2	Hhm1	Hhm2	Hm	Hamino	Hhm	H6	Cm	C2	C4	C5	Ccy	C6
	Conf.	G298SOL	>0.02?	16	17	18	13	14	10	11	aver.	aver.	aver.	5	15	1	2	3	9	4
1	nb_5mOHC_smd_scanh	0.67	1	27.93	27.97	28.15	25.76	25.41	26.84	26.83	28.02	25.58	26.83	23.61	137.14	5.93	0.68	62.60	110.47	62.60
2	nb_5mOHC_smd_scana	0.21	1	27.97	27.93	28.14	25.78	25.36	26.49	26.84	28.01	25.57	26.66	23.56	137.10	5.88	0.43	64.37	110.99	64.37
3	nb_5mOHC_smd_scang	0.05	1	27.92	27.97	28.13	25.75	25.67	27.18	26.59	28.00	25.71	26.88	23.50	137.08	6.25	1.87	62.76	112.26	62.76
4	nb_5mOHC_smd_scanj	0.04	1	27.94	27.89	28.09	25.83	26.19	26.89	26.60	27.97	26.01	26.74	23.59	136.96	6.21	4.42	61.85	112.63	61.85
5	nb_5mOHC_smd_scann	0.03	1	27.94	27.88	28.10	25.83	26.22	26.61	26.61	27.97	26.03	26.61	23.59	136.95	6.37	4.41	64.36	113.48	64.36
	Boltzmann Avg.										28.01	25.62	26.79	23.59	137.11	5.96	0.96	63.00	110.85	63.00

Table S21: Two main conformers and general numbering scheme of the 1-methyl-5-formylcytosine (**4b**) obtained for the calculation of NMR shieldings (see Table S22).

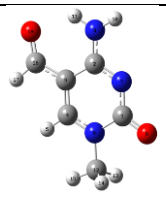
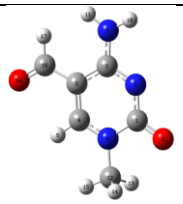
	
nb_5fCb	**nb_5fCa_a**

Table S22: NMR shielding calculated for conformers of 1-methyl-5formylcytosine (**4b**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hamino1	Hamino2	Hm	Hamino	Hf	H6	Cm	C2	C4	C5	Ccy	C6
	Conf.	G298SOL	>0.02?	13	14	15	10	11	aver.	aver.	17	5	12	1	2	3	16	4
1	nb_5fCb	1.00	1	27.78	27.78	27.95	142.48	25.25	27.84	83.87	21.61	22.88	135.70	8.37	3.67	63.41	-29.51	3.11
2	nb_5fCa_a	0.00	0	27.79	27.83	27.98	153.72	25.51	27.87	89.62	21.18	22.54	135.74	8.09	2.16	65.09	-25.13	12.13
	Boltzmann Avg.								27.84	83.89	21.61	22.88	135.70	8.37	3.66	63.41	-29.49	3.13

Table S23: Two main conformers and general numbering scheme of the 1-methyl-5-formylcytosine hydrate (**4b_hyd**) obtained for the calculation of NMR shieldings (see Table 24).

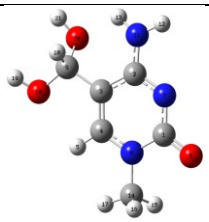
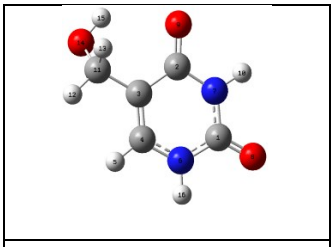
	
5fC_hydrat_scana_4	**5fC_hydrat_scanna**

Table S24: NMR shielding calculated for conformers of 1-methyl-5formylcytosine hydrate (**4b_hyd**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hamino1	Hamino2	Hm	Hamino	Hhy	H6	Cm	C2	C4	C5	Chy	C6
	Conf.	G298SOL	>0.02?	15	16	17	12	13	aver.	aver.	10	5	14	1	2	3	9	4
1	5fC_hydrat_scana_4	0.31	1	27.96	27.96	28.16	25.59	25.06	28.03	27.24	26.27	16.89	8.35	9.34	18.42	27.18	23.46	17.17
2	5fC_hydrat_scanna	0.18	1	27.88	27.94	28.09	25.81	25.16	27.97	27.28	26.35	16.99	8.39	9.32	18.42	27.20	23.54	17.24
3	5fC_hydrat_scanha	0.11	1	27.88	27.95	28.11	25.78	24.98	27.98	27.28	26.29	16.92	8.33	9.33	18.42	27.18	23.50	17.18
4	5fC_hydrat_scangb	0.11	1	27.96	27.94	28.14	25.73	24.58	28.01	27.27	26.15	16.77	8.19	9.34	18.43	27.14	23.40	17.04
5	5fC_hydrat_scanjb	0.10	1	27.92	27.91	28.10	25.78	24.91	27.98	27.26	26.26	16.90	8.30	9.33	18.41	27.17	23.47	17.16
6	5fC_hydrat_scanhb	0.07	1	27.97	27.95	28.16	25.59	25.10	28.03	27.23	26.28	16.89	8.37	9.34	18.42	27.18	23.47	17.18
7	5fC_hydrat_scan_10	0.04	1	27.95	27.87	28.09	25.80	25.34	27.97	27.25	26.41	17.04	8.45	9.32	18.41	27.21	23.57	17.30
8	5fC_hydrat_scan_7	0.03	1	27.91	27.92	28.11	25.74	24.83	27.98	27.26	26.23	16.86	8.28	9.33	18.41	27.15	23.45	17.12
	Boltzmann Avg.								28.00	27.26	26.28	16.90	8.33	9.33	18.42	27.18	23.48	17.17

Table S25: Main conformer and general numbering scheme of the reference 5-hydroxymethyluracil (**5hmU**) obtained for the calculation of NMR shieldings (see Table S26).

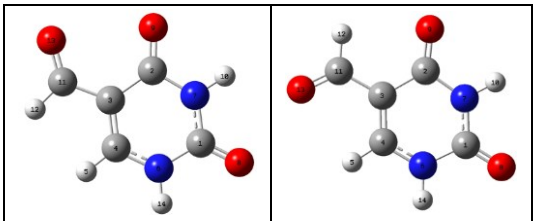


5hmUracil_14

Table S26: NMR shielding calculated for conformers of the reference 5-hydroxymethyuracil (**5hmU**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hhm1	Hhm2	H1	H3	Hhm	H6	C2	C4	C5	Cmh	C6
	Conf.	G298SOL	>0.02?	12	13	16	10	aver.	5	1	2	3	11	4
1	5hmUracil_14	1.00	1	27.28	26.48	23.21	22.95	26.88	23.60	14.07	-1.23	54.01	111.58	24.42
	Boltzmann Avg.					23.21	22.95	26.88	23.60	14.07	-1.23	54.01	111.58	24.42

Table S27: Two main conformers and general numbering scheme of the 5-formyluracil (**13a**) obtained for the calculation of NMR shieldings (see Table S28).

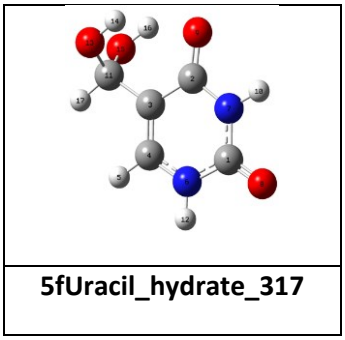


5fUracil_b3 **5fUracil_a2**

Table S28: NMR shielding calculated for conformers of 5-formyluracil (**13a**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				H1	H3	Hf	H6	C2	C4	C5	Cf	C6
		G298SOL	>0.02?	14	10	12	5	1	2	3	11	4
1	5fUracil_b3	0.56	1	22.62	22.94	21.54	22.91	16.08	3.60	56.45	5.15	5.15
2	5fUracil_a3	0.44	1	22.67	22.88	21.04	22.68	15.33	0.29	58.22	16.02	16.02
	Boltzmann Avg.			22.64	22.91	21.32	22.80	15.75	2.14	57.23	9.95	9.95

Table S29: Main conformer and general numbering scheme of 5-formyluracil hydrate (**13a_hyd**) obtained for the calculation of NMR shieldings (see Table S30).



5fUracil_hydrate_317

Table S30: NMR shielding calculated for conformers of 5-formyluracil hydrate (**13a_hyd**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

					Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				H1	H3	Hhy	H6	C2	C4	C5	Chy	C6
		G298SOL	>0.02?	12	10	17	5	1	2	3	11	4
1	5fUracil_hydrate_317	0.35	1	23.19	22.93	25.91	23.63	14.91	0.17	53.93	75.39	25.18
2	5fUracil_hydrate_312	0.25	1	23.14	22.96	25.53	23.61	14.91	-0.30	56.33	75.67	22.99
3	5fUracil_hydrate_3311	0.16	1	23.12	22.97	25.27	23.30	14.56	0.64	53.91	81.77	25.77
4	5fUracil_hydrate_349	0.05	1	23.10	22.97	24.90	23.31	14.48	0.65	55.22	82.24	24.88
5	5fUracil_hydrate_316	0.04	1	23.16	22.96	25.51	23.64	14.92	-0.36	55.00	75.52	23.99
6	5fUracil_hydrate_338	0.04	1	23.12	22.98	25.25	23.47	14.39	0.56	53.53	80.62	25.74
7	5fUracil_hydrate_314	0.03	1	23.20	23.00	25.18	23.58	14.58	1.46	55.94	75.65	22.25
8	5fUracil_hydrate_332	0.02	1	23.11	22.94	25.04	23.43	14.48	-0.25	53.82	81.09	26.72
9	5fUracil_hydrate_343	0.02	1	23.10	22.99	24.93	23.26	14.32	0.46	53.35	81.61	24.56
	Boltzmann Avg.			23.15	22.95	25.53	23.53	14.77	0.18	54.61	77.47	24.65

Table S31: Two main conformers and general numbering scheme of the 1-methyl-5-formyluracil (**13b**) obtained for the calculation of NMR shieldings (see Table S32).

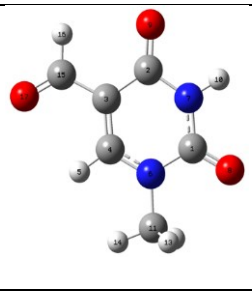
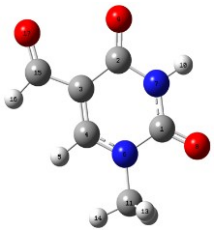
	
n5_formyl_uracil_tau3_a	**n5_formyl_uracil_tau3**

Table S32: NMR shielding calculated for conformers of 5-formyluracil (**13b**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hm	H3	Hf	H6	Cm	C2	C4	C5	Cf	C6
	Conf.	G298SOL	>0.02?	12	13	14	aver.	10	16	5	11	1	2	3	15	4
1	n5_formyl_uracil_tau3_a	0.62	1	27.79	27.73	27.88	27.80	22.83	21.07	22.55	136.47	14.00	0.31	58.69	-28.56	11.38
2	n5_formyl_uracil_tau3_	0.38	1	27.74	27.74	27.84	27.78	22.90	21.59	22.83	136.50	14.78	3.61	56.63	-29.53	0.53
	Boltzmann Avg.						27.79	22.86	21.27	22.66	136.48	14.29	1.58	57.90	-28.93	7.22

Table S33: Two main conformers and general numbering scheme of the 1-methyl-5-formyluracil hydrate (**13b**) obtained for the calculation of NMR shieldings (see Table S34).

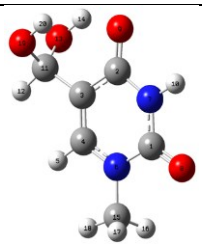
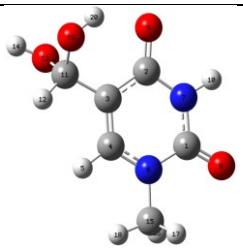
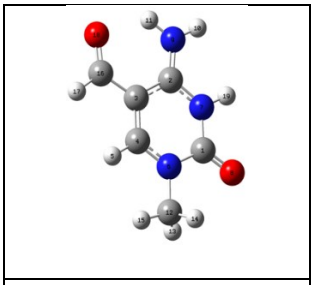
	
5fU_hydrate_a_10	**5fU_hydrate_a_5**

Table S34: NMR shielding calculated for conformers of 5-formyluracil hydrate (**13b_hyd**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hm	H3	Hhy	H6	Cm	C2	C4	C5	Chy	C6
	Conf.	G298SOL	>0.02?	16	17	18	aver.	10	12	5	15	1	2	3	11	4
1	5fU_hydrate_a_10	0.77	1	27.91	27.92	28.06	27.97	22.89	25.93	23.54	137.84	13.97	0.47	54.45	75.15	19.44
2	5fU_hydrate_a_5	0.10	1	27.93	27.87	28.05	27.95	22.91	25.52	23.51	137.76	14.06	-0.07	57.02	75.56	17.08
3	5fU_hydrate_a1_2	0.05	1	27.87	27.90	28.03	27.93	22.92	25.28	23.22	137.67	13.59	0.82	54.68	81.66	20.24
4	5fU_hydrate_a_9_b	0.03	1	27.92	27.89	28.06	27.96	22.91	25.52	23.56	137.80	14.07	-0.05	55.32	75.38	18.32
	Boltzmann Avg.						27.95	22.89	25.79	23.50	137.78	13.93	0.42	54.74	75.81	19.25



nb_5fC_prot_tau7

Table S35: NMR shielding calculated for conformers of protonated 5-formylcytosine (**p_4b**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hamino1	Hamino2	Hm	Hamino	Hf	H6	Cm	C2	C4	C5	Cf	C6
	Conf.	G298SOL	>0.02?	13	14	15	10	11	aver	aver	17	5	12	1	2	3	16	4
1	nb_5fC_prot_tau7	1.00	1	27.63	27.63	27.73	24.53	21.66	27.67	23.10	21.44	22.60	135.09	18.44	10.44	64.79	-30.13	1.29
2	nb_5fC_prot_tau1_a	0.00	0	27.63	27.69	27.75	24.80	23.95	27.69	24.38	21.14	22.32	135.11	17.95	9.02	66.12	-23.11	8.55
3	nb_5fC_prot_tau9	0.00	0	27.46	27.46	27.56	24.46	22.07	27.50	23.27	21.32	22.69	134.27	8.71	4.41	59.18	-31.89	3.35
12	Boltzmann Avg.								27.67	23.10	21.44	22.60	135.09	18.44	10.44	64.80	-30.11	1.31

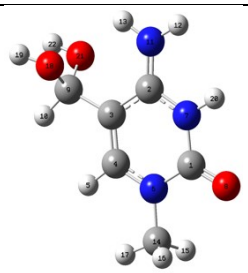
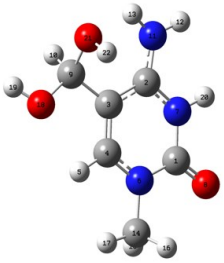
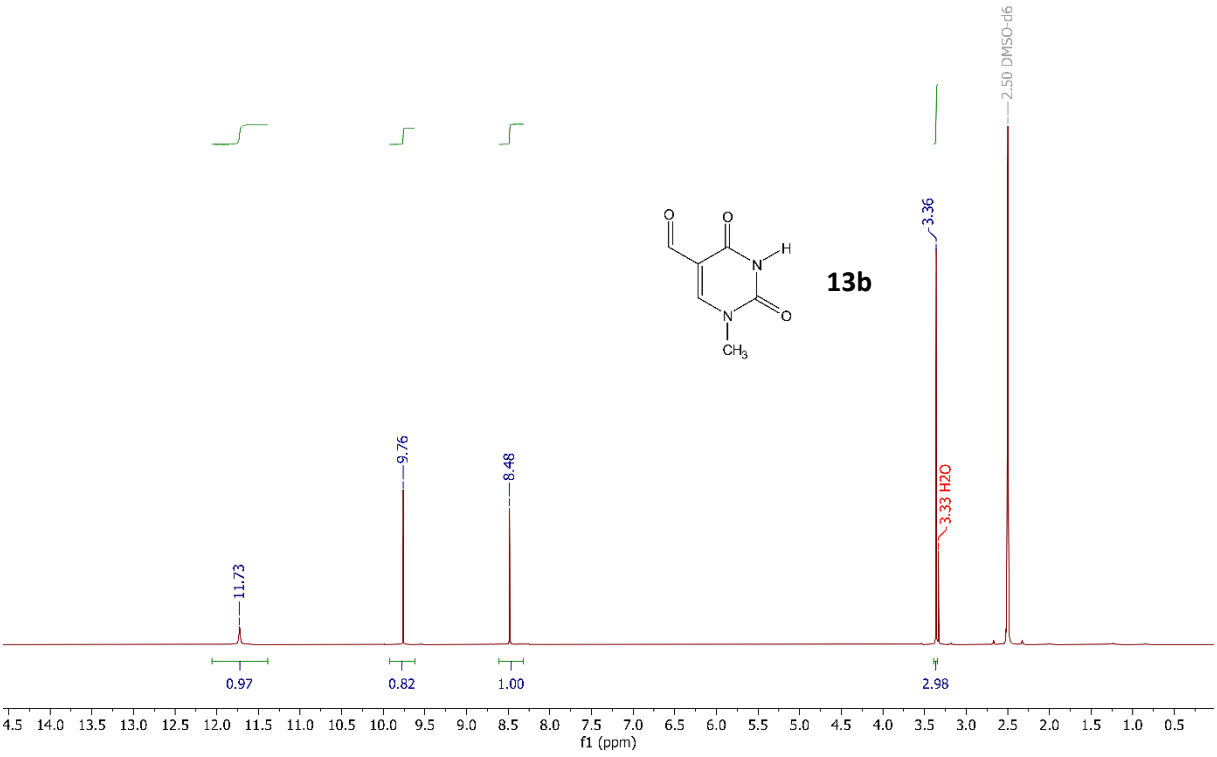
	
p5fC_hydrat_scanc_2	**p5fC_hydrat_scanc1_5**

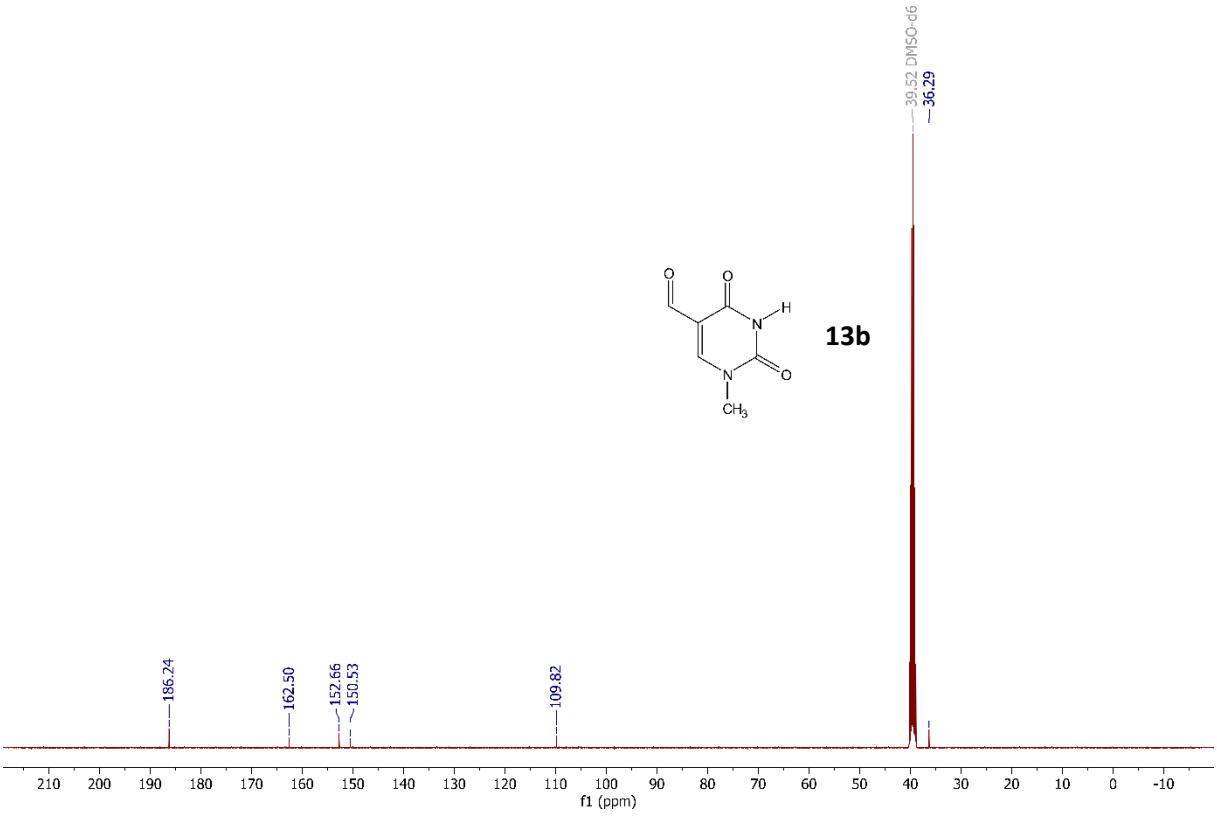
Table S36: NMR shielding calculated for conformers of protonated 5-formylcytosine hydrate (**p_4b_hyd**) at the SMD(H_2O)/B3LYP/PCS-3//SMD(H_2O)/B3LYP-D3/6-31+G(d,p) level of theory.

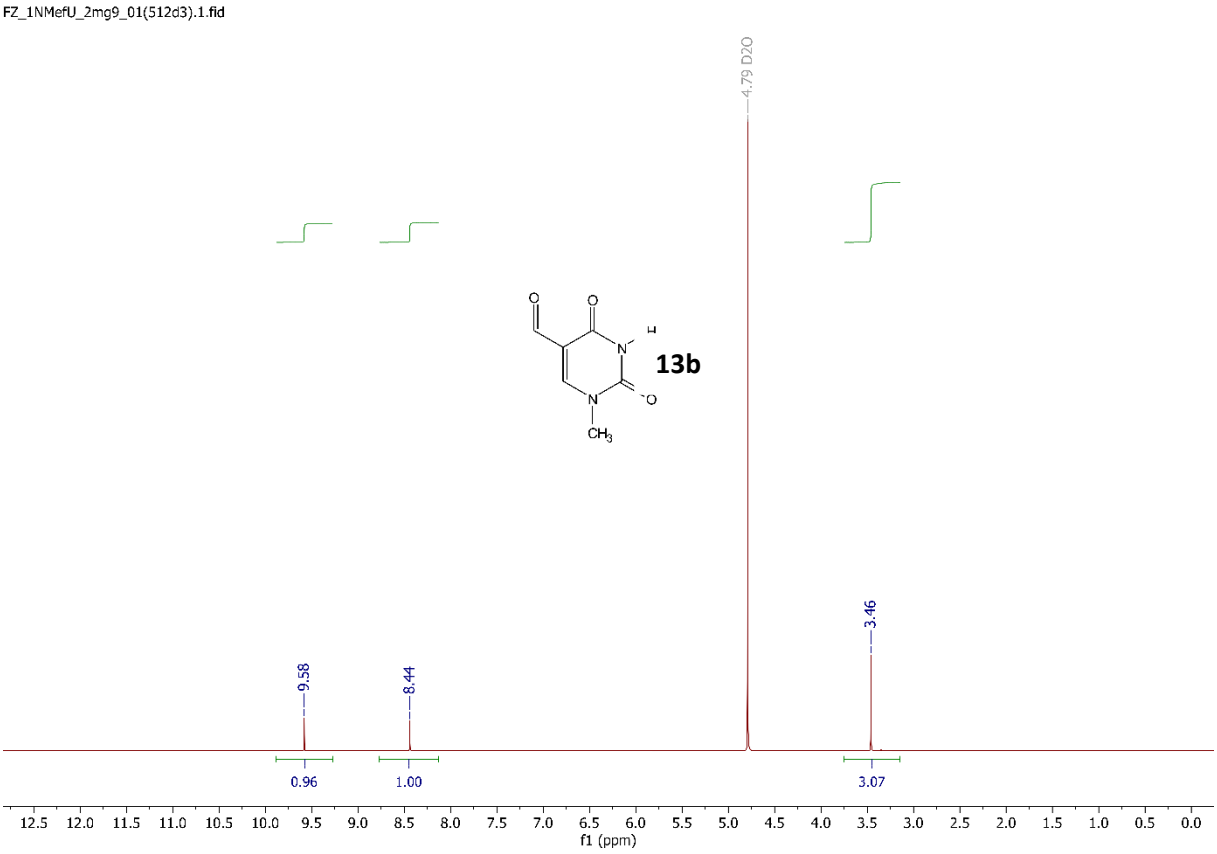
				Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield	Shield
				Hm1	Hm2	Hm3	Hamino1	Hamino2	Hm	Hamino	Hhy	H6	Cm	C2	C4	C5	Chy	C6
	Molecule2	G298	>0.02?	15	16	17	12	13	aver	aver	10	5	14	1	2	3	9	4
1	p5fC_hydrat_scanc_2	0.41	1	27.81	27.81	27.94	24.85	24.08	27.85	24.47	25.00	23.26	136.47	17.24	10.91	63.89	76.40	13.57
2	p5fC_hydrat_scanc1_5	0.18	1	27.80	27.75	27.90	24.92	23.81	27.82	24.37	25.32	23.03	136.21	17.02	10.01	63.24	81.45	15.39
3	p5fC_hydrat2a	0.09	1	27.79	27.83	27.94	24.86	24.06	27.85	24.46	25.07	23.30	136.53	17.19	11.05	62.57	76.14	14.25
4	p5fC_hydrat_scanc1_4	0.09	1	27.80	27.73	27.89	24.93	23.94	27.81	24.43	24.94	23.00	136.20	16.97	10.21	64.10	81.97	14.66
5	p5fC_hydrat_scanc_1	0.09	1	27.78	27.82	27.93	24.90	23.36	27.84	24.13	25.31	23.25	136.41	17.47	10.92	63.47	75.54	12.41
6	p5fC_hydrat_scanc1_7	0.09	1	27.77	27.77	27.90	24.91	23.81	27.81	24.36	25.28	23.10	136.17	16.92	10.41	62.14	80.31	14.96
7	p5fC_hydrat_scanc1_10	0.02	1	27.77	27.77	27.90	24.89	23.73	27.82	24.31	25.59	23.11	136.18	16.98	10.19	61.24	80.05	15.60
	Boltzmann Avg.								27.84	24.40	25.13	23.18	136.36	17.16	10.64	63.36	78.13	14.14

3.7 NMR SPECTRA

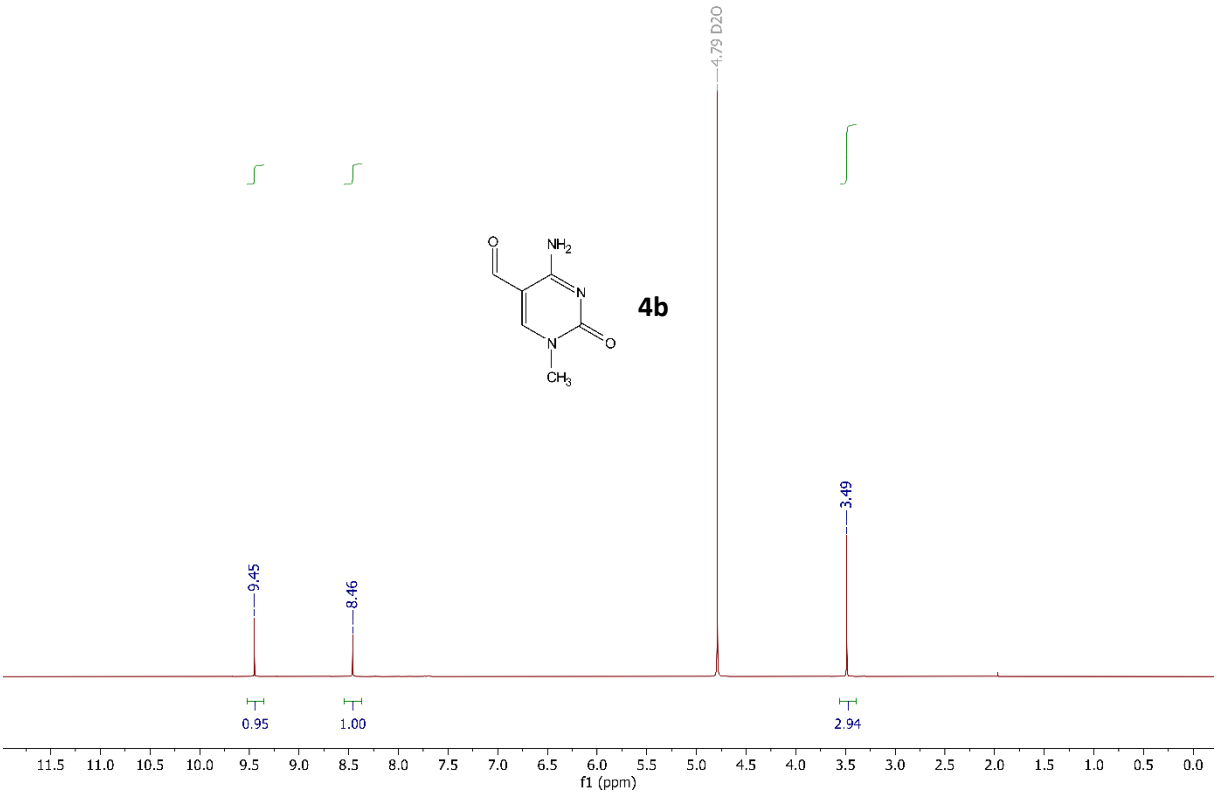
3.7.1 1-N-Methyl-5-formyluracil (13b)



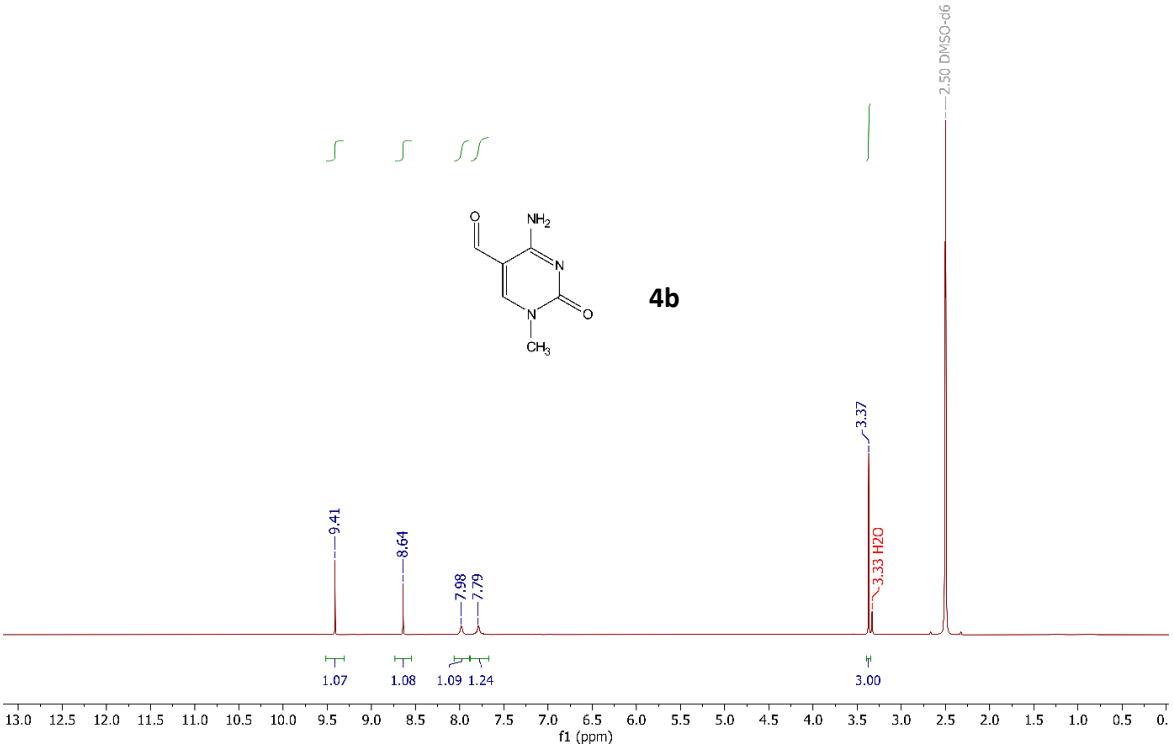




3.7.2 1-N-Methyl-5-formylcytosine (4b)

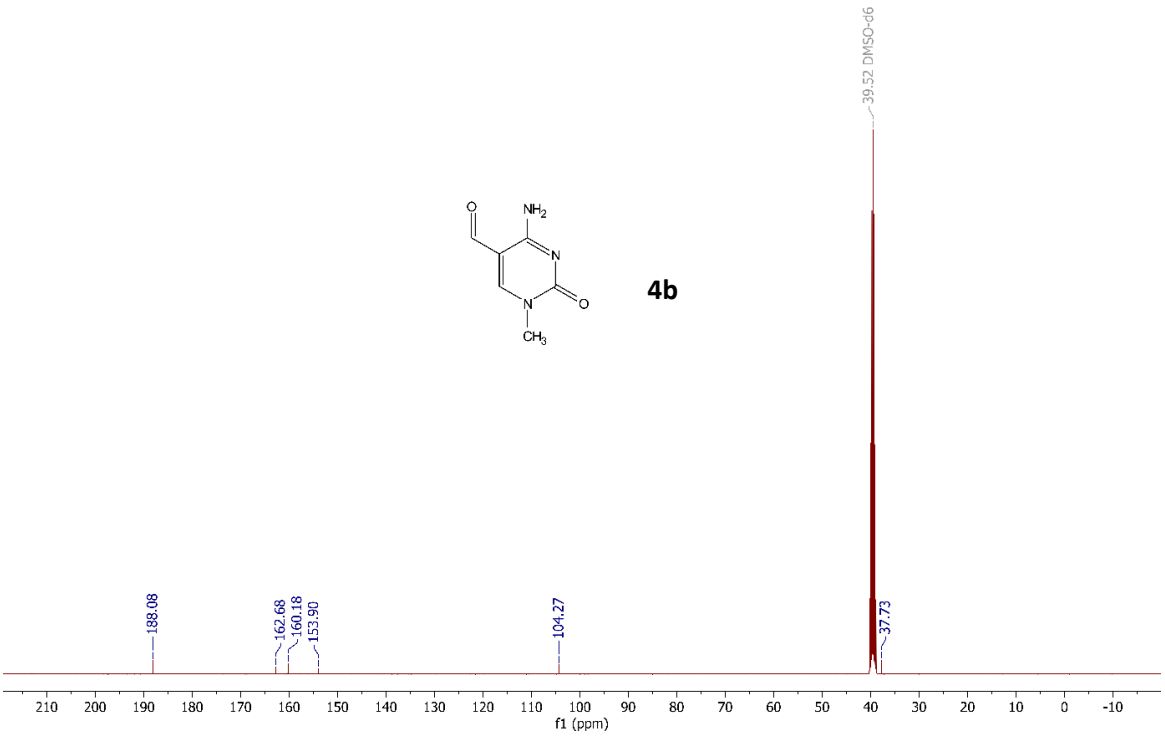


FZ_1m5fC_DMSO_dry_01(40064).1.fid



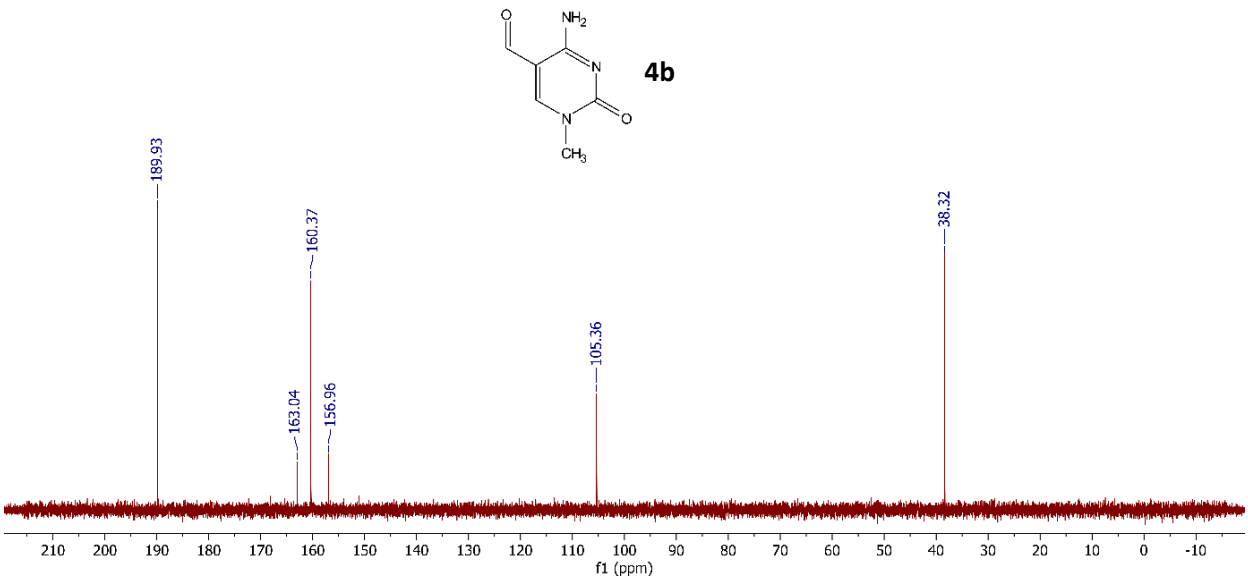
4b

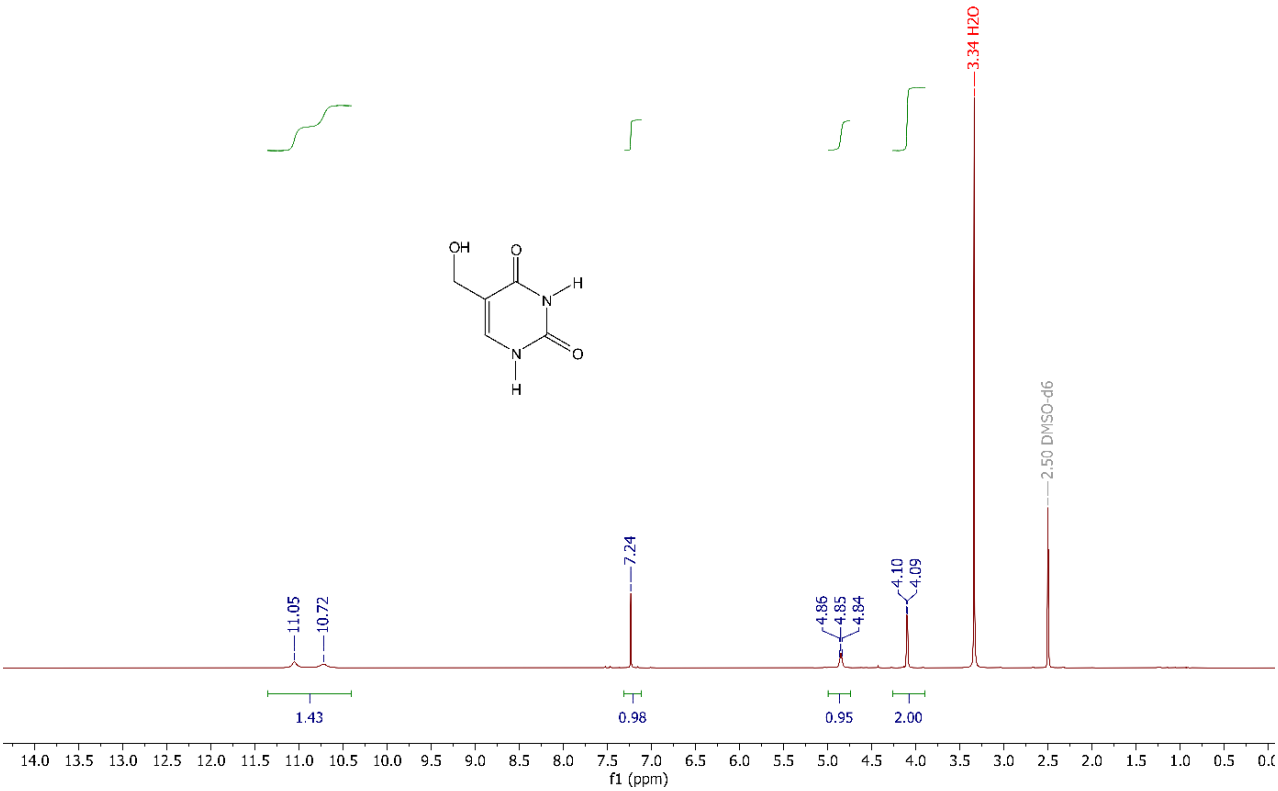
FZ_1m5fC_DMSO_dry_13(4001024).1.fid



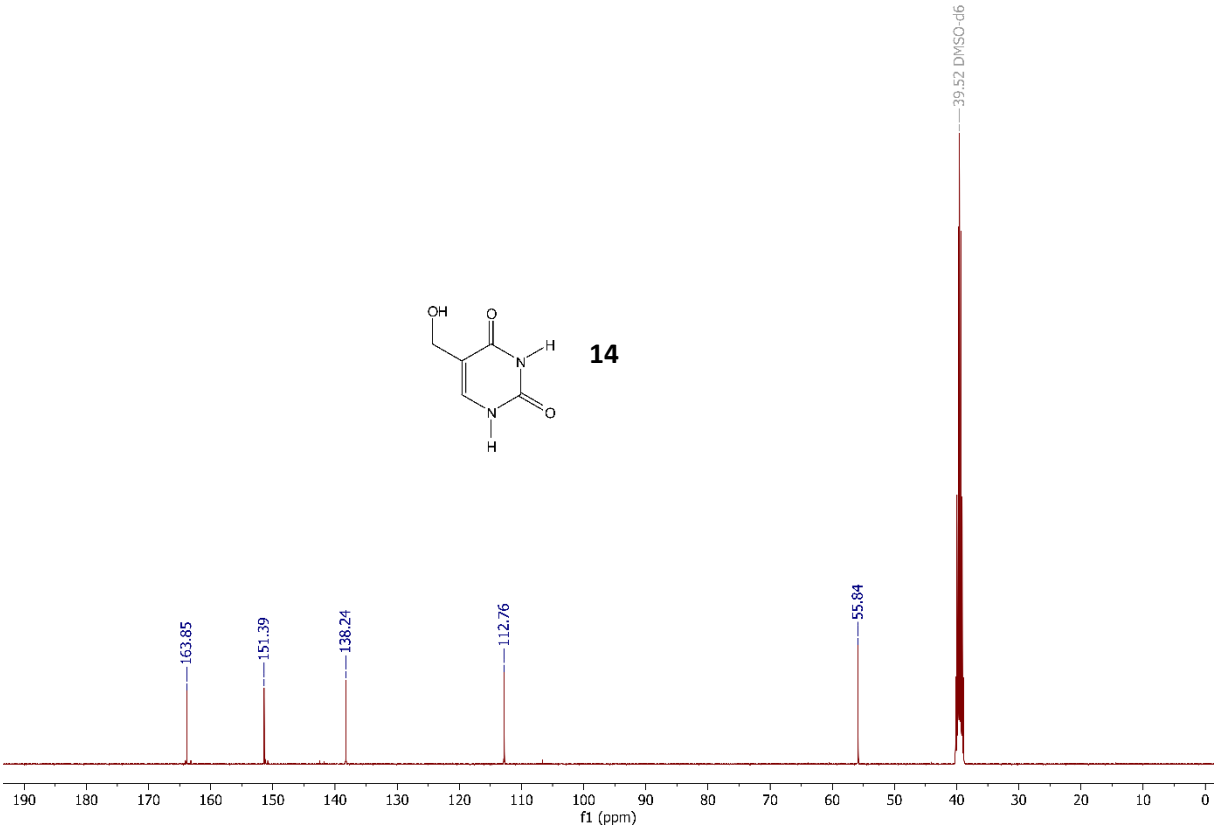
4b

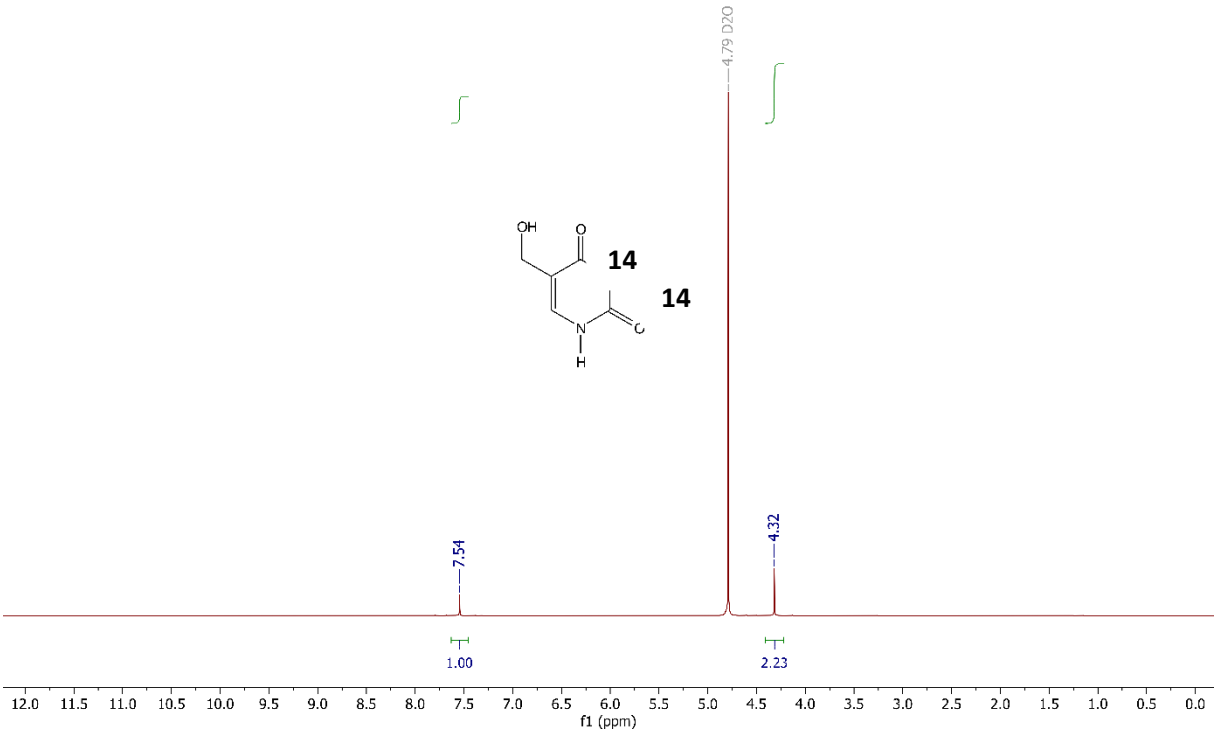
FZ_ep87_F1_13(400).2.fid

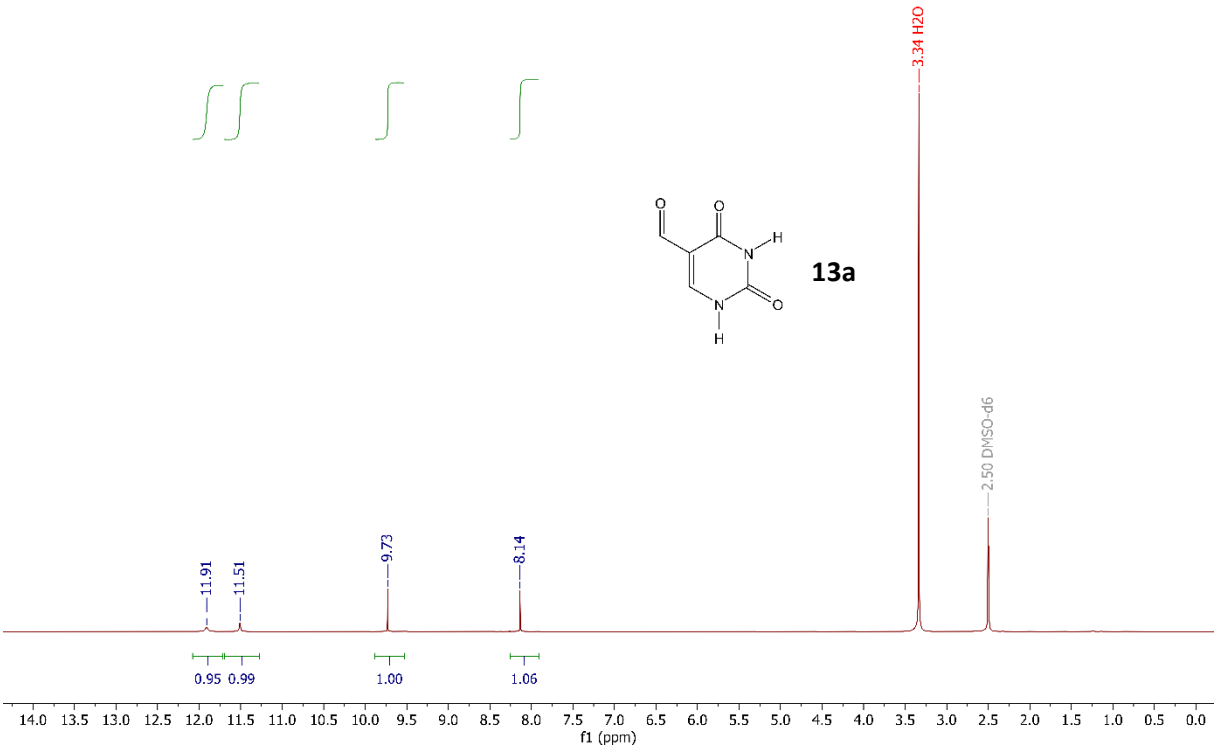




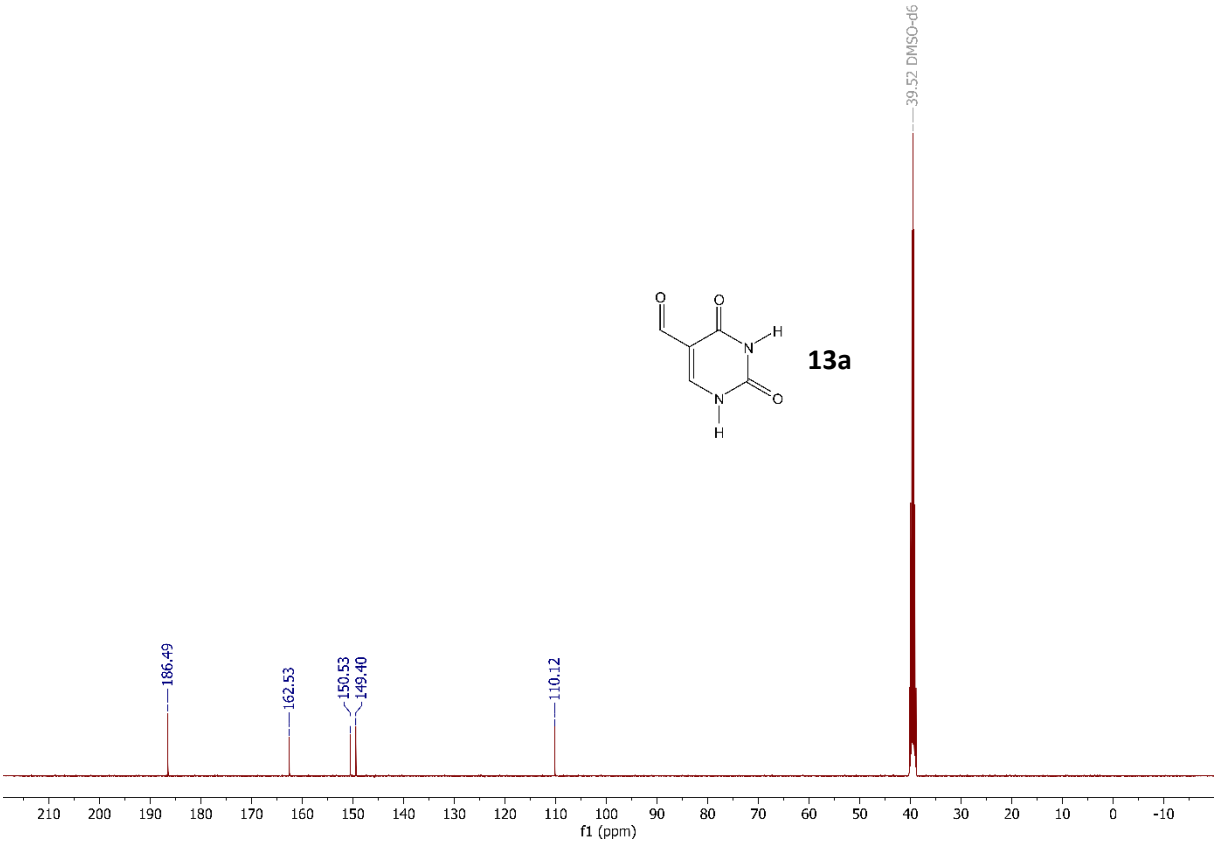
3.7.3 Other NMR Spectra



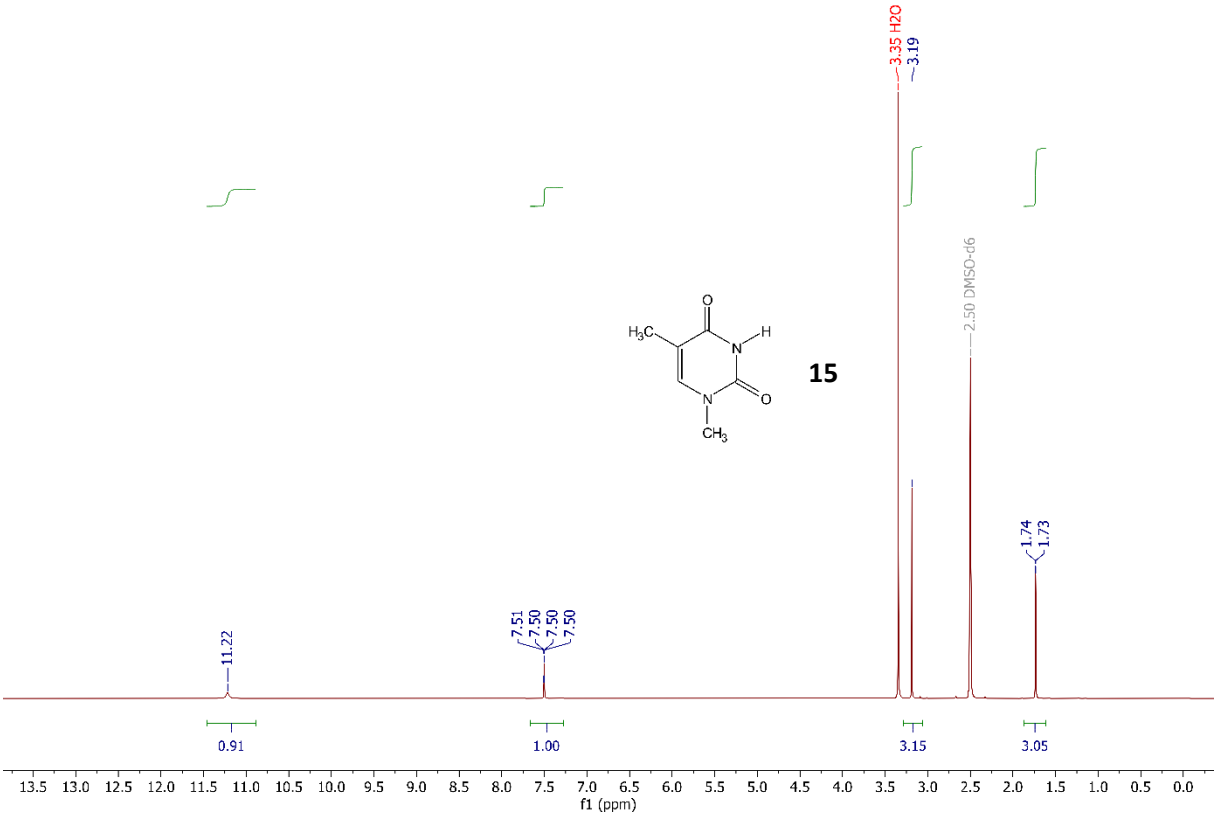



3.34 H2O
2.50 DMSO-d6
11.91
11.51
9.73
8.14
0.95
0.99
1.00
1.06
f1 (ppm)

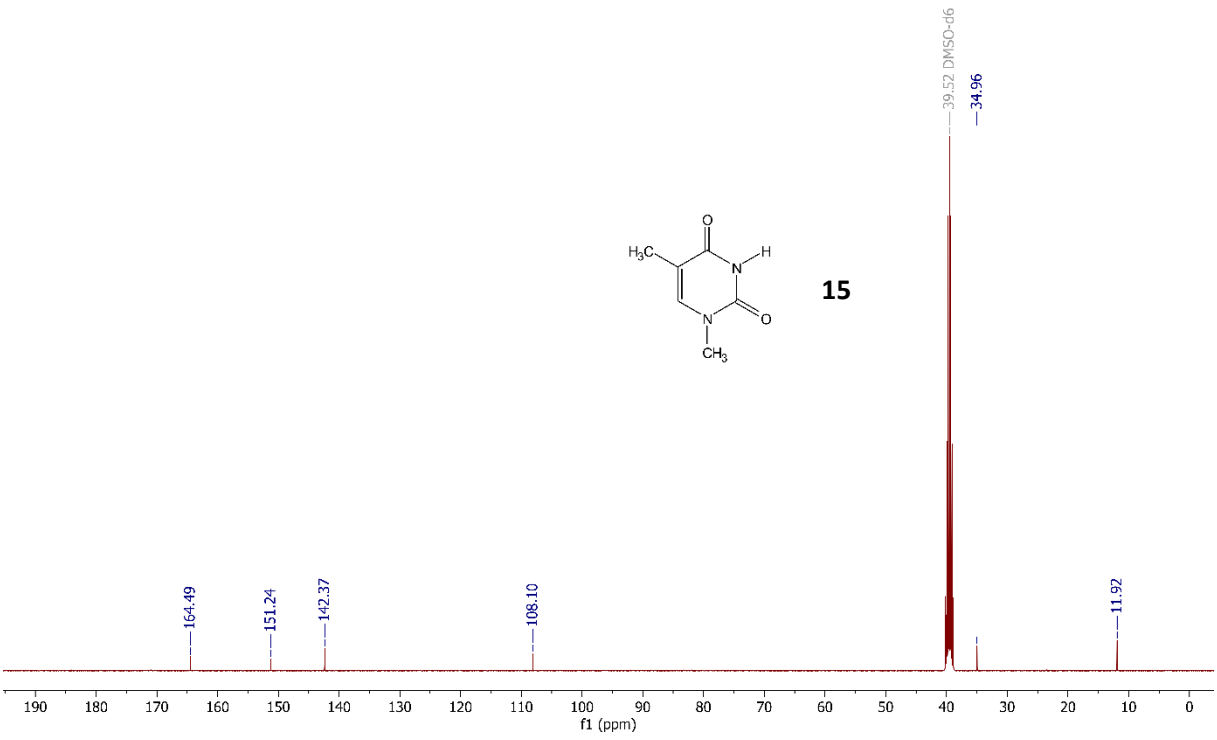
13a


39.52 DMSO-d6
186.49
162.53
150.53
149.40
110.12
f1 (ppm)

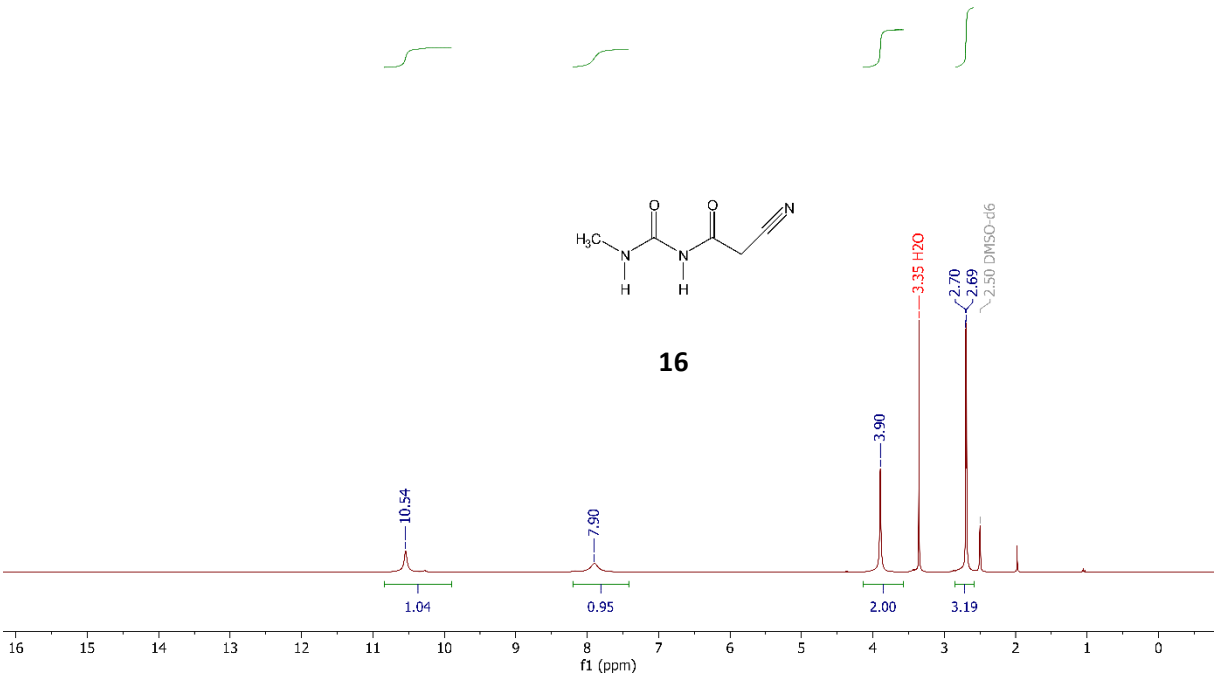
13a

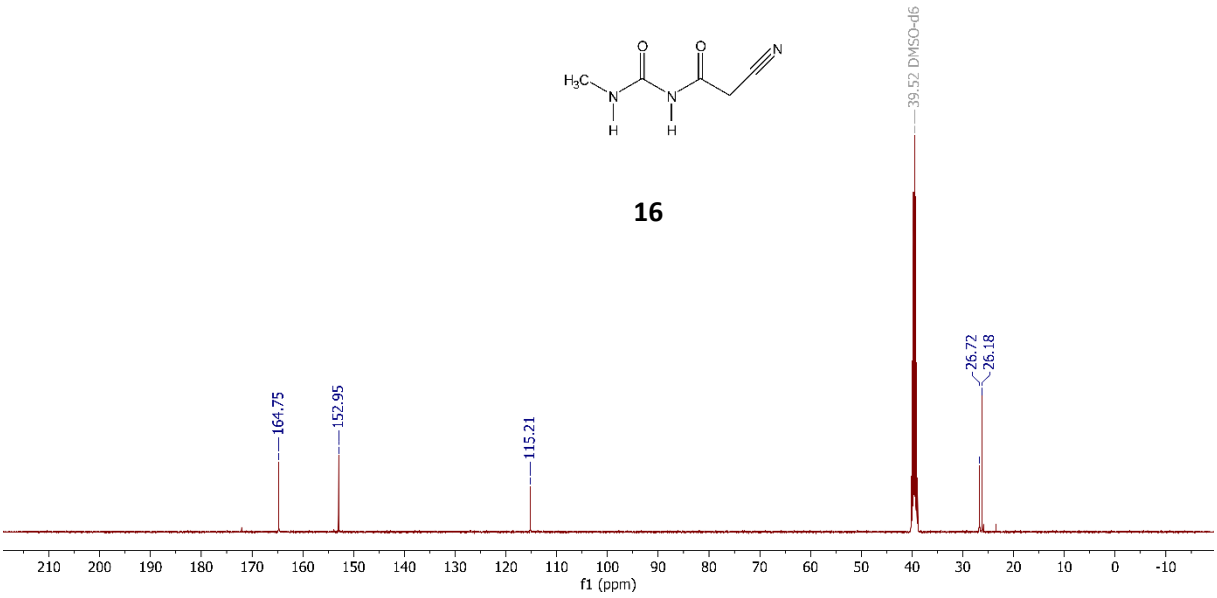

3.35 H2O
3.19
2.50 DMSO-d6
1.74
1.73
7.51
7.50
7.50
7.50
11.22
0.91
1.00
3.15
3.05
f1 (ppm)

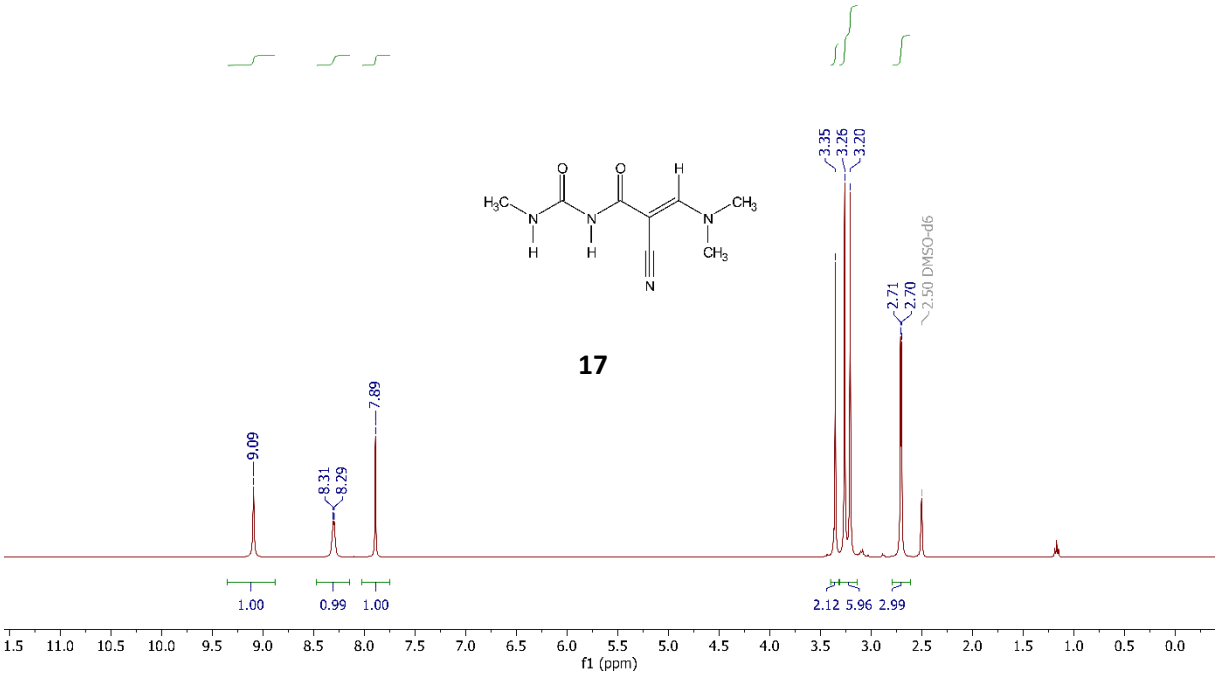
15

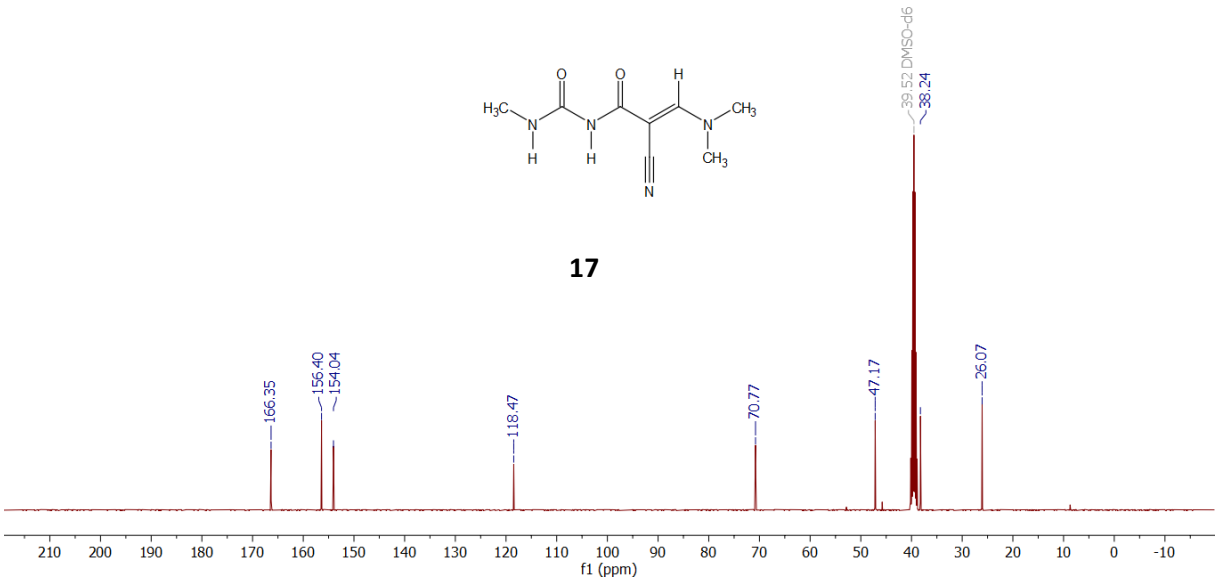

39.52 DMSO-d6
34.96
164.49
151.24
142.37
108.10
11.92
f1 (ppm)

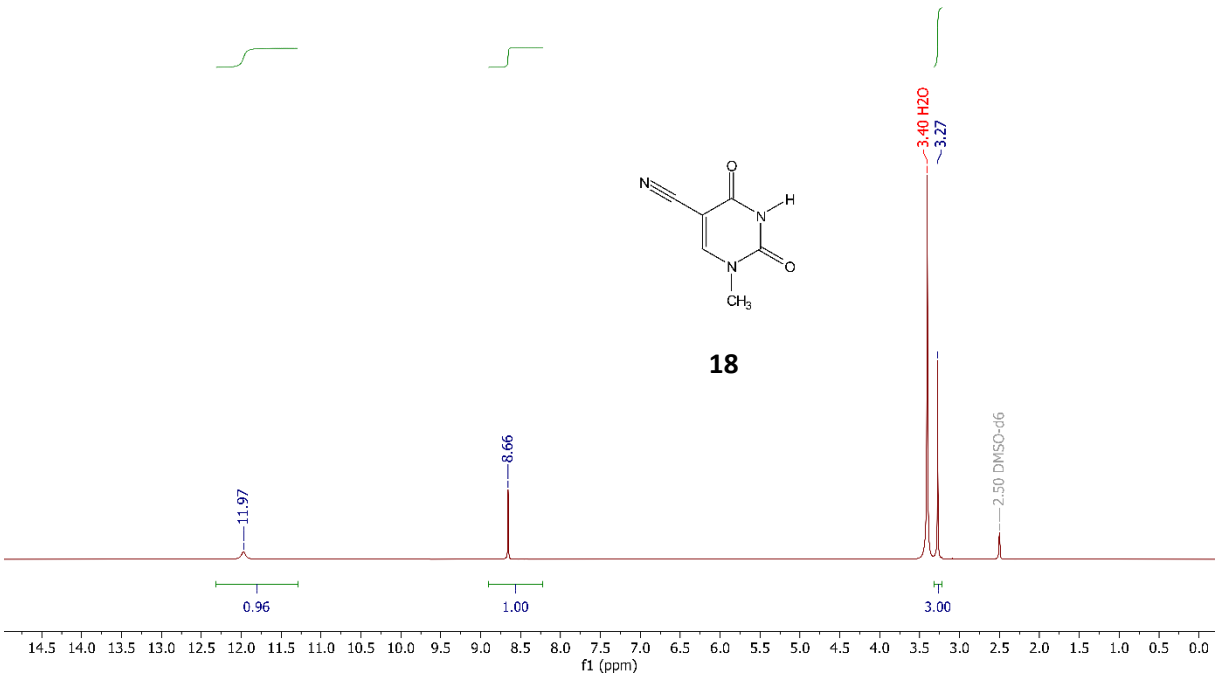
15


16
10.54
7.90
3.90
3.35 H2O
2.70
2.69
2.50 DMSO-d6
1.04
0.95
2.00
3.19
f1 (ppm)

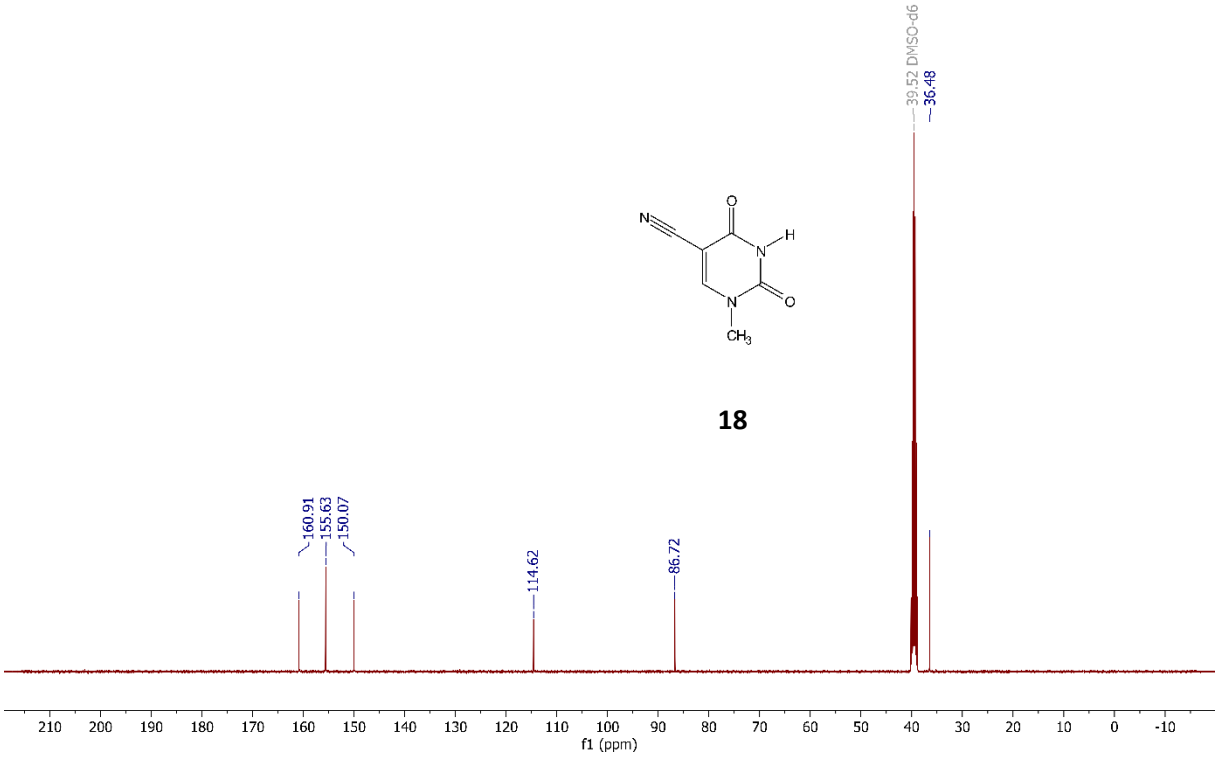

16
164.75
152.95
115.21
39.52 DMSO-d6
26.72
26.18
f1 (ppm)


17
9.09
8.31
8.29
7.89
3.35
3.26
3.20
2.71
2.70
2.50 DMSO-d6
1.00
0.99
1.00
2.12
5.96
2.99
f1 (ppm)


17
166.35
156.40
154.04
118.47
70.77
47.17
39.52 DMSO-d6
38.24
26.07
f1 (ppm)

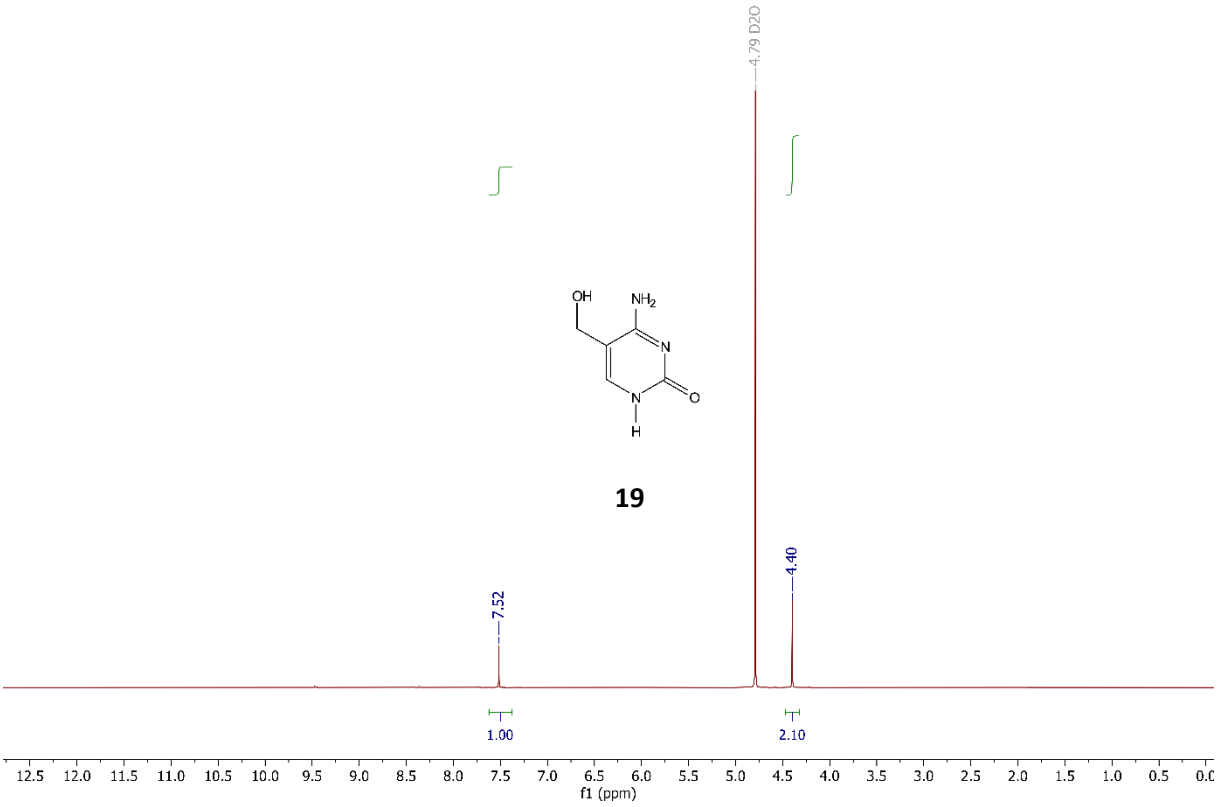

3.40 H2O
3.27
8.66
11.97
2.50 DMSO-d6
0.96
1.00
3.00
14.5 14.0 13.5 13.0 12.5 12.0 11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

18

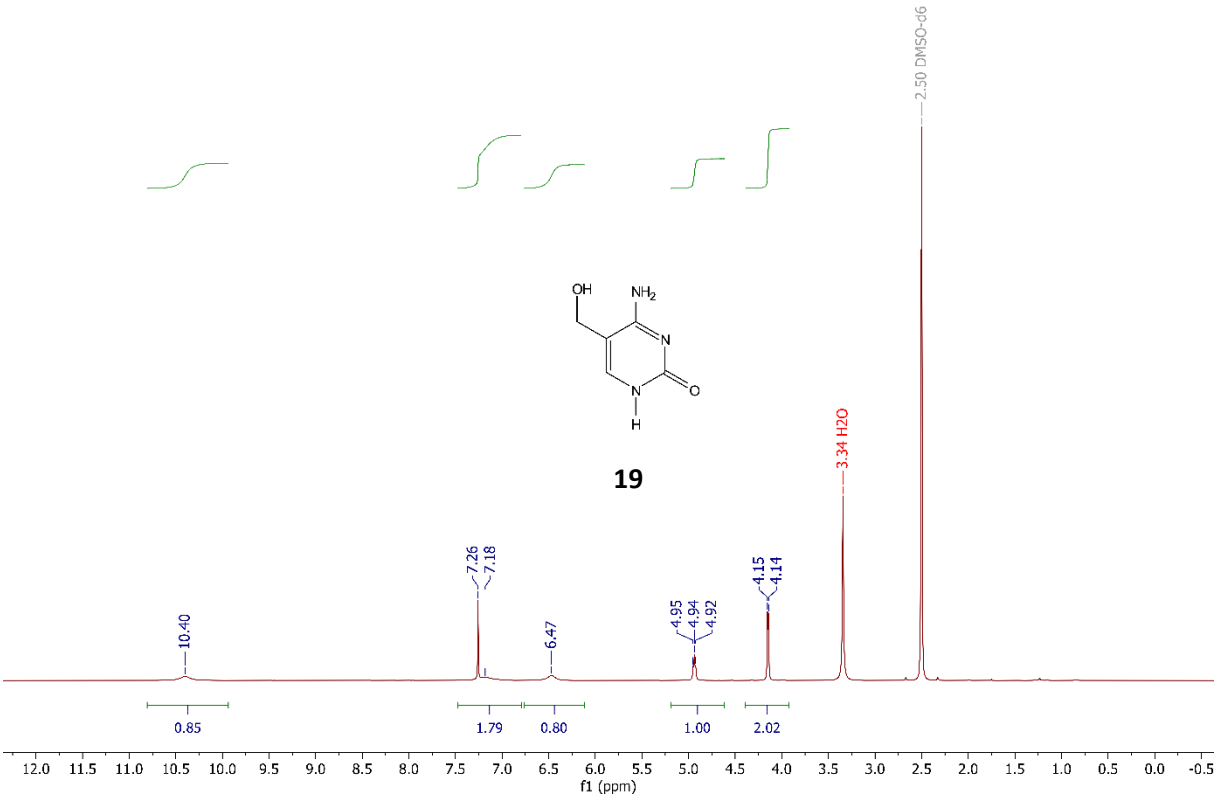

39.52 DMSO-d6
36.48
160.91
155.63
150.07
114.62
86.72
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)

18

FZ_ep82_F2_01(400).1.fid

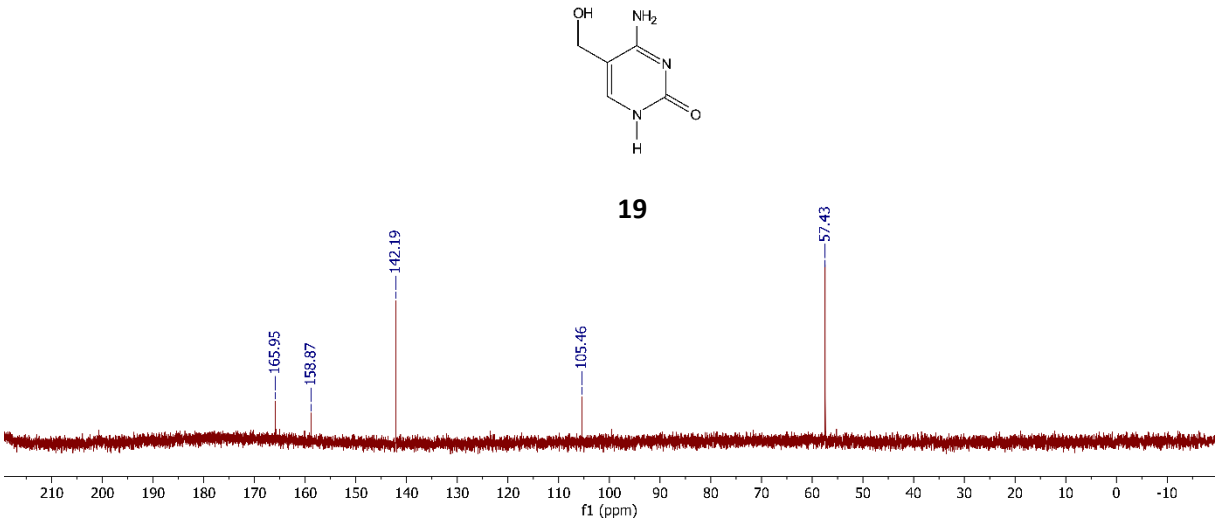


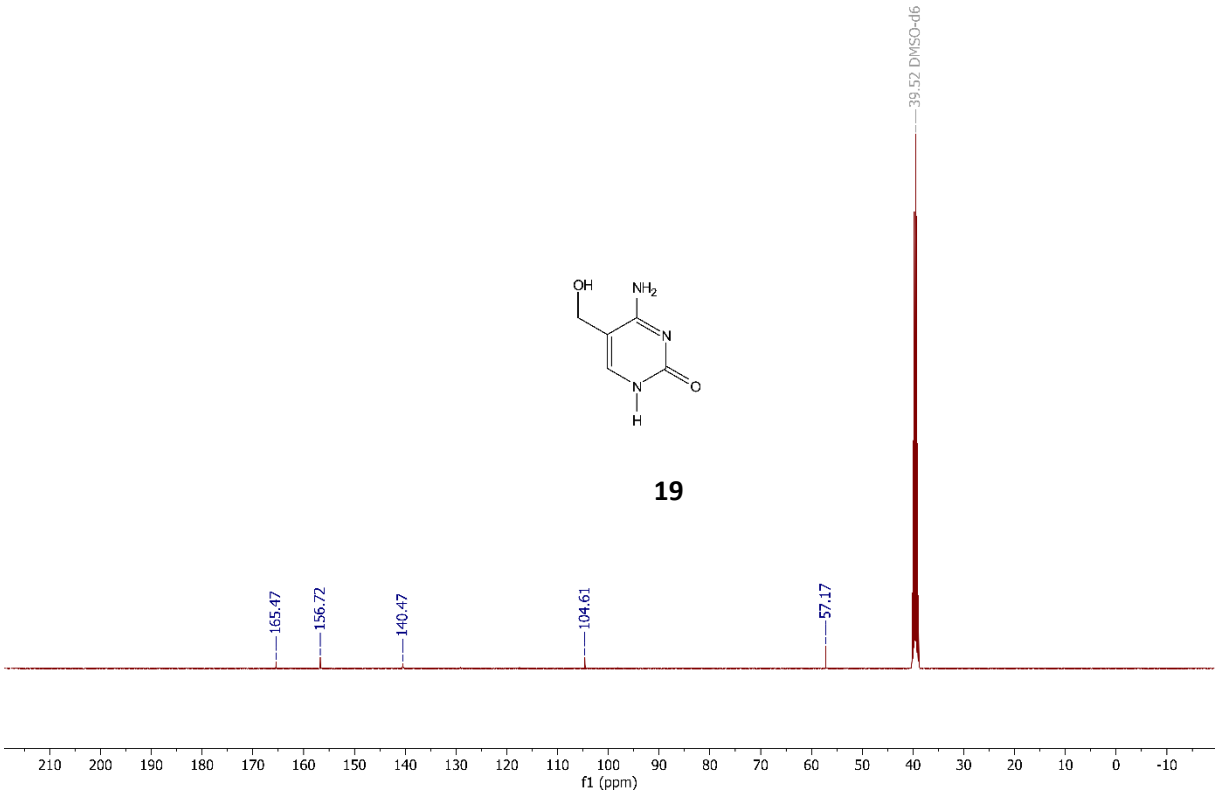
19

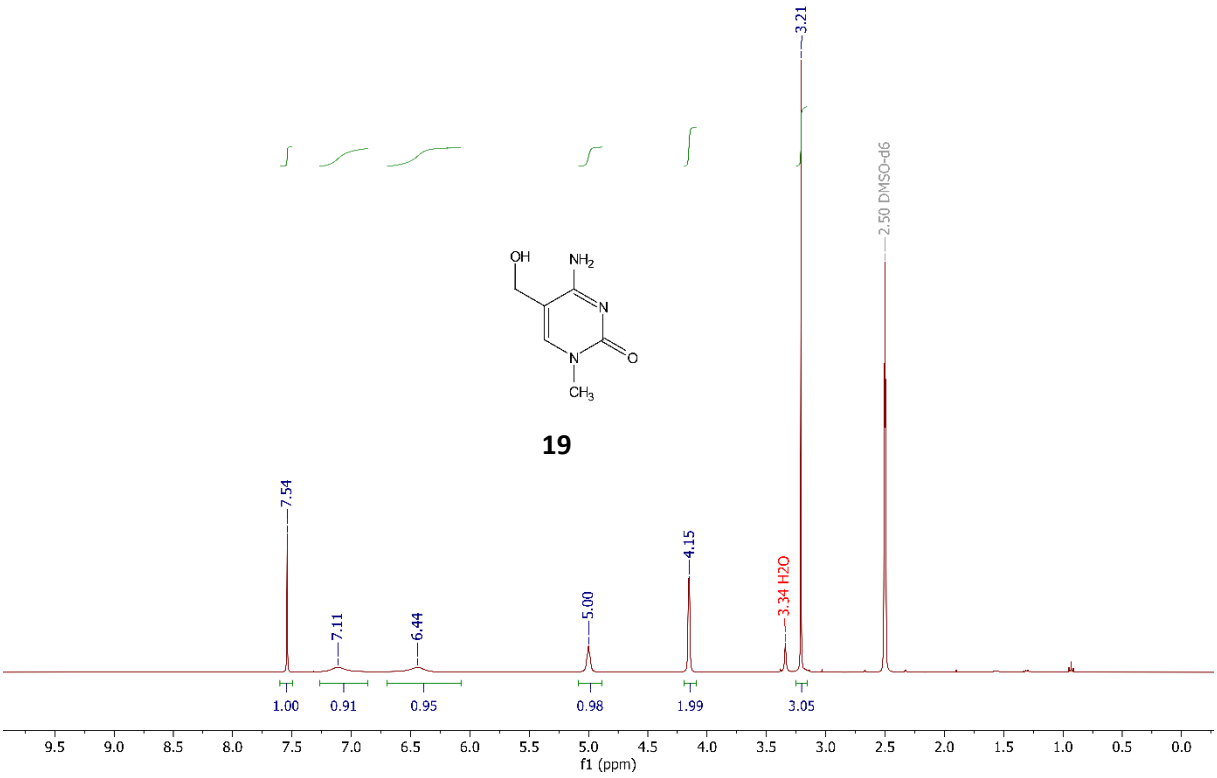


19

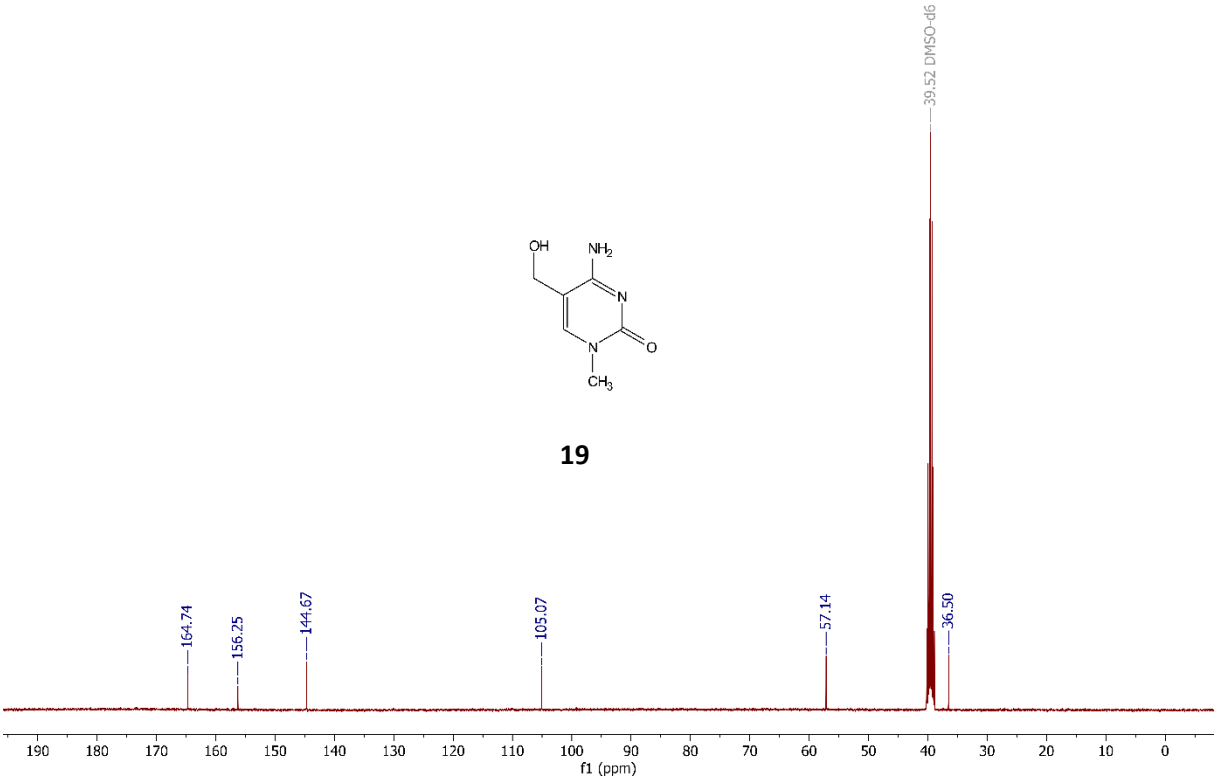
FZ_ep82_F2_13(400).1.fid



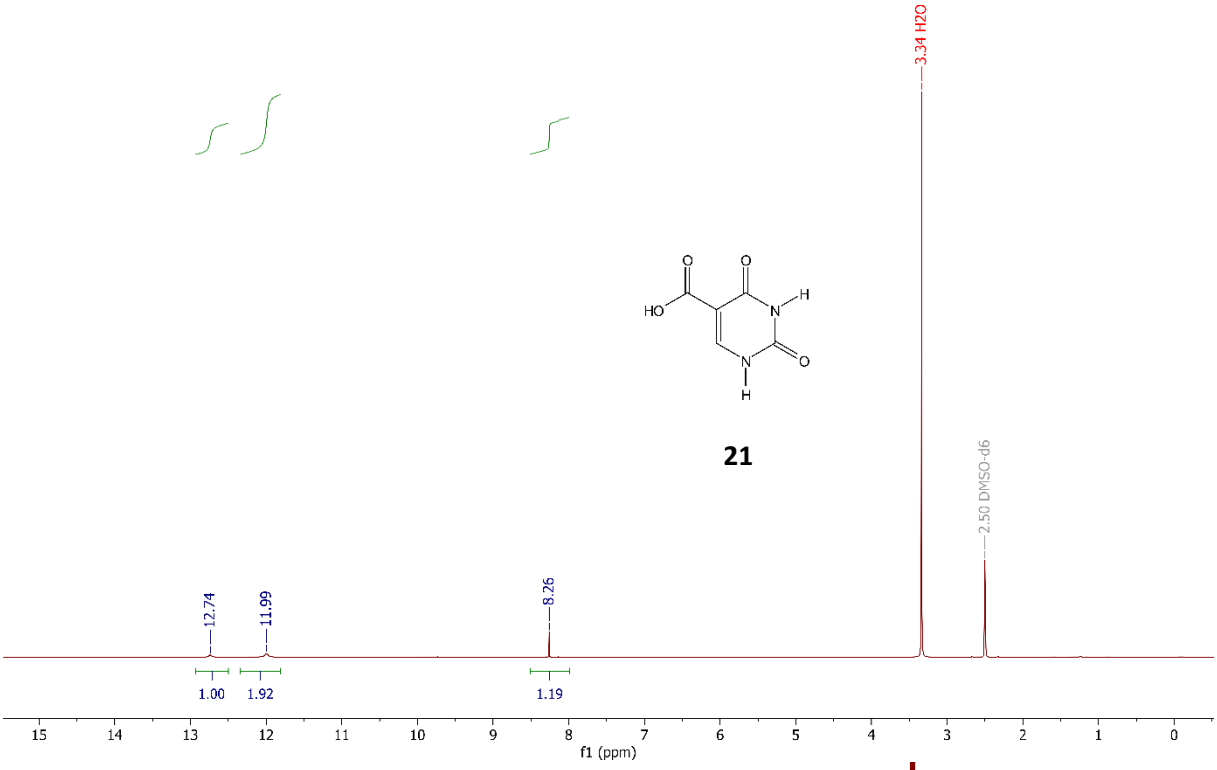



3.21
2.50 DMSO-d6
OH
NH2
N
O
N
CH3
7.54
7.11
6.44
5.00
4.15
3.34 H2O
1.00
0.91
0.95
0.98
1.99
3.05
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
f1 (ppm)

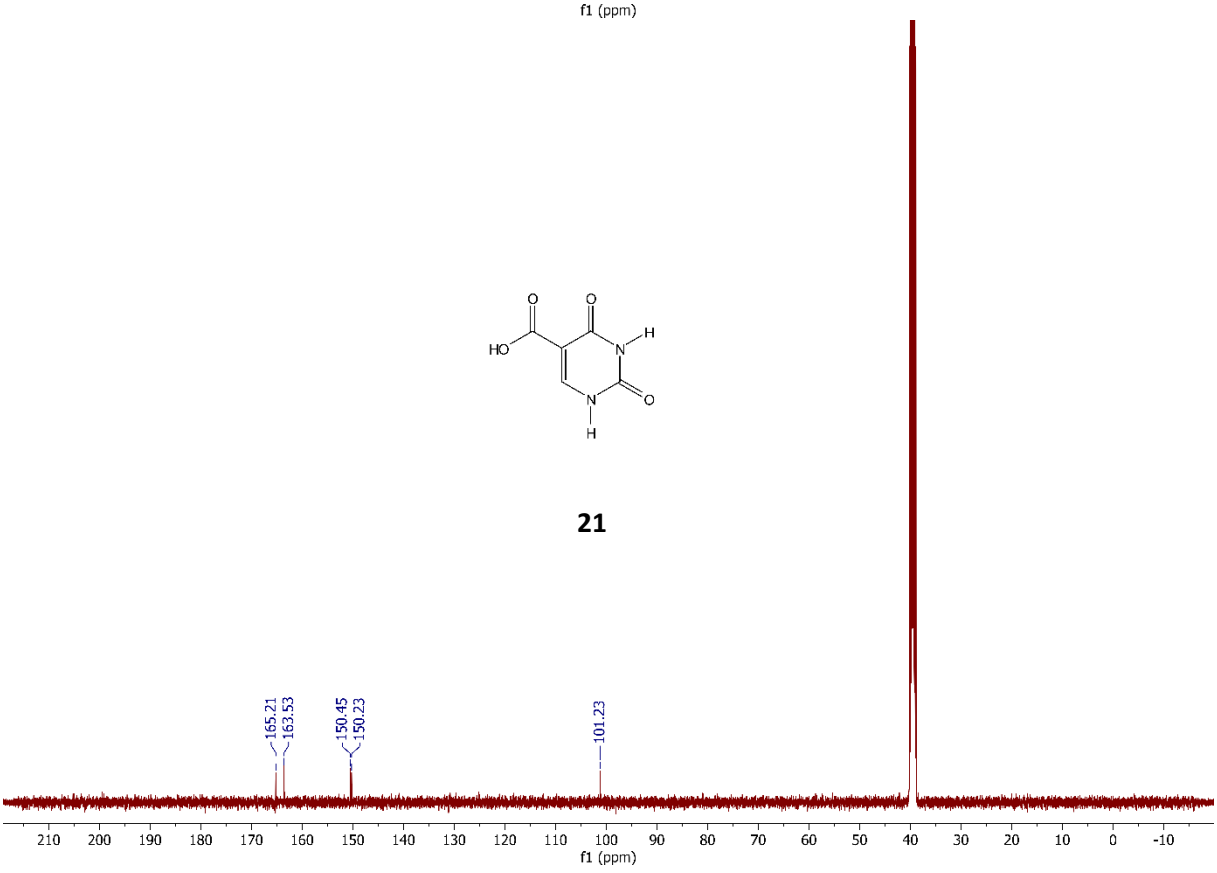
19


39.52 DMSO-d6
OH
NH2
N
O
N
CH3
164.74
156.25
144.67
105.07
57.14
36.50
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

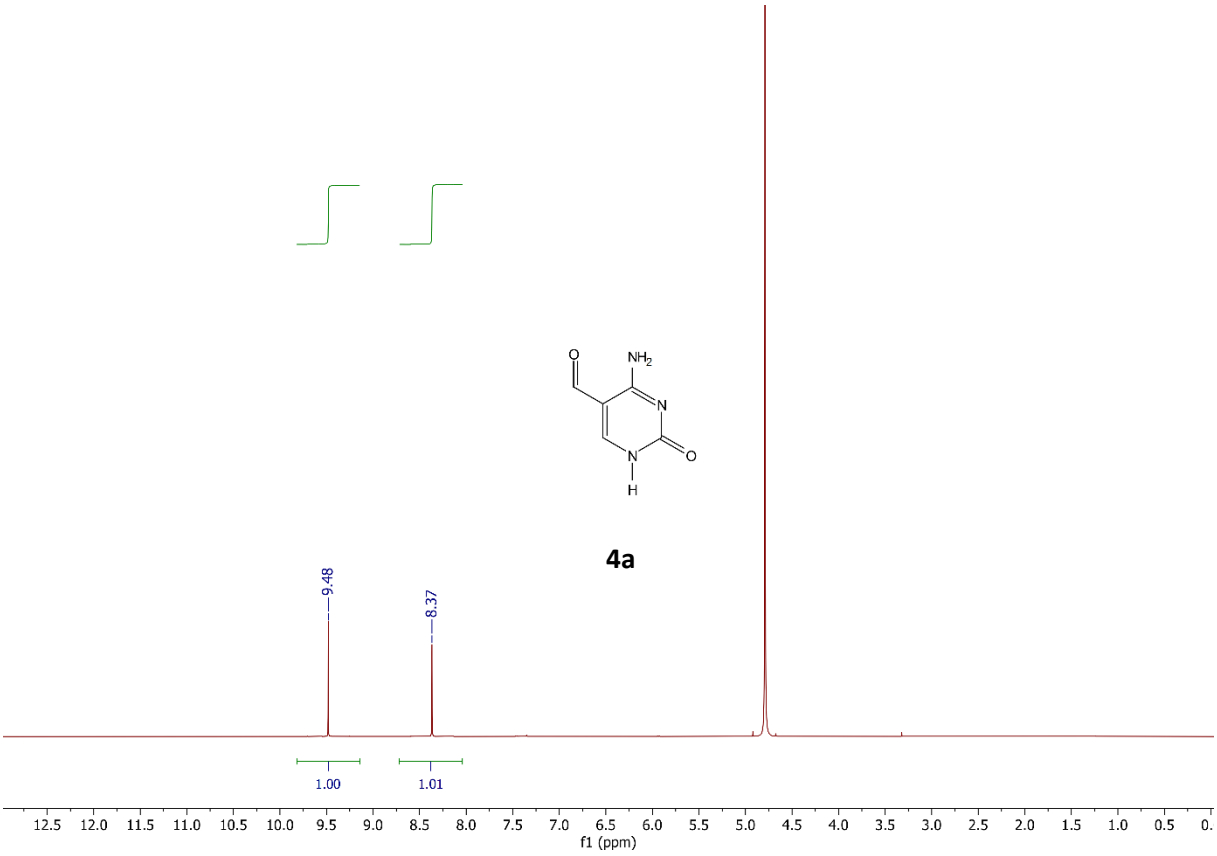
19

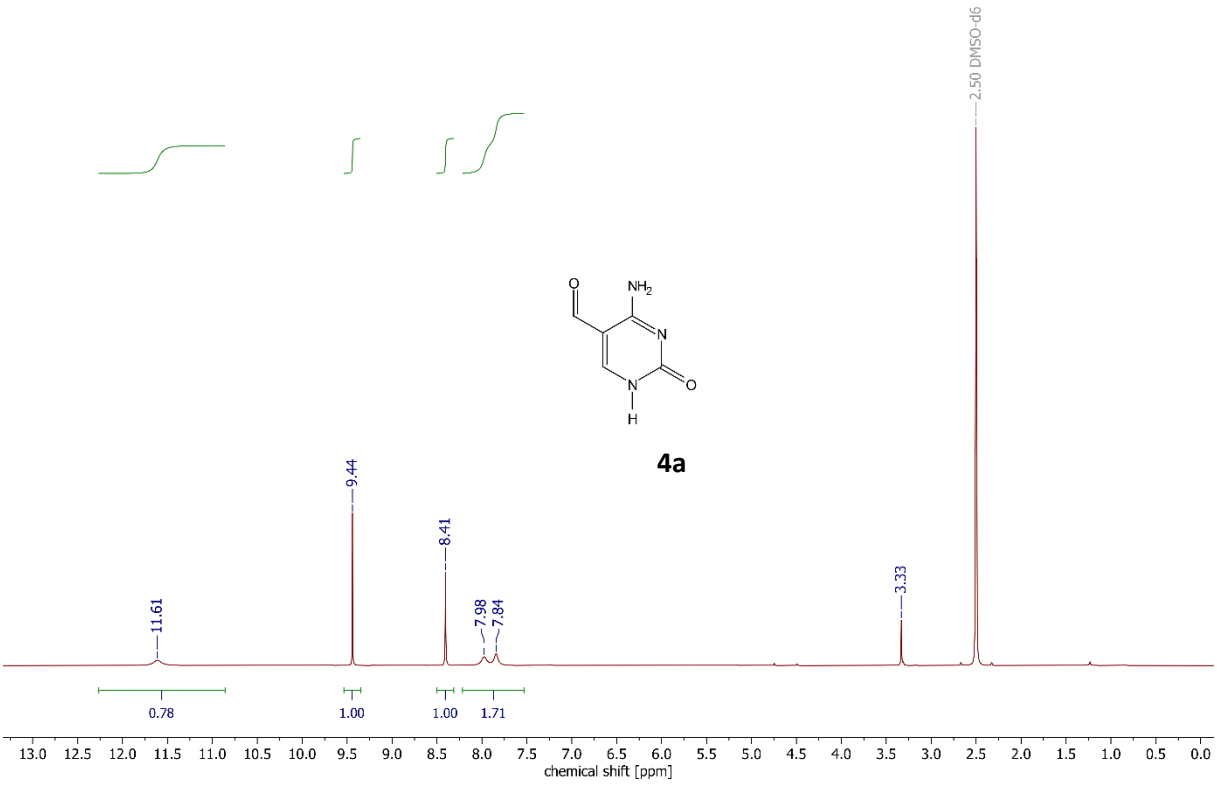

3.34 H2O
2.50 DMSO-d6
12.74
11.99
8.26
1.00
1.92
1.19
f1 (ppm)

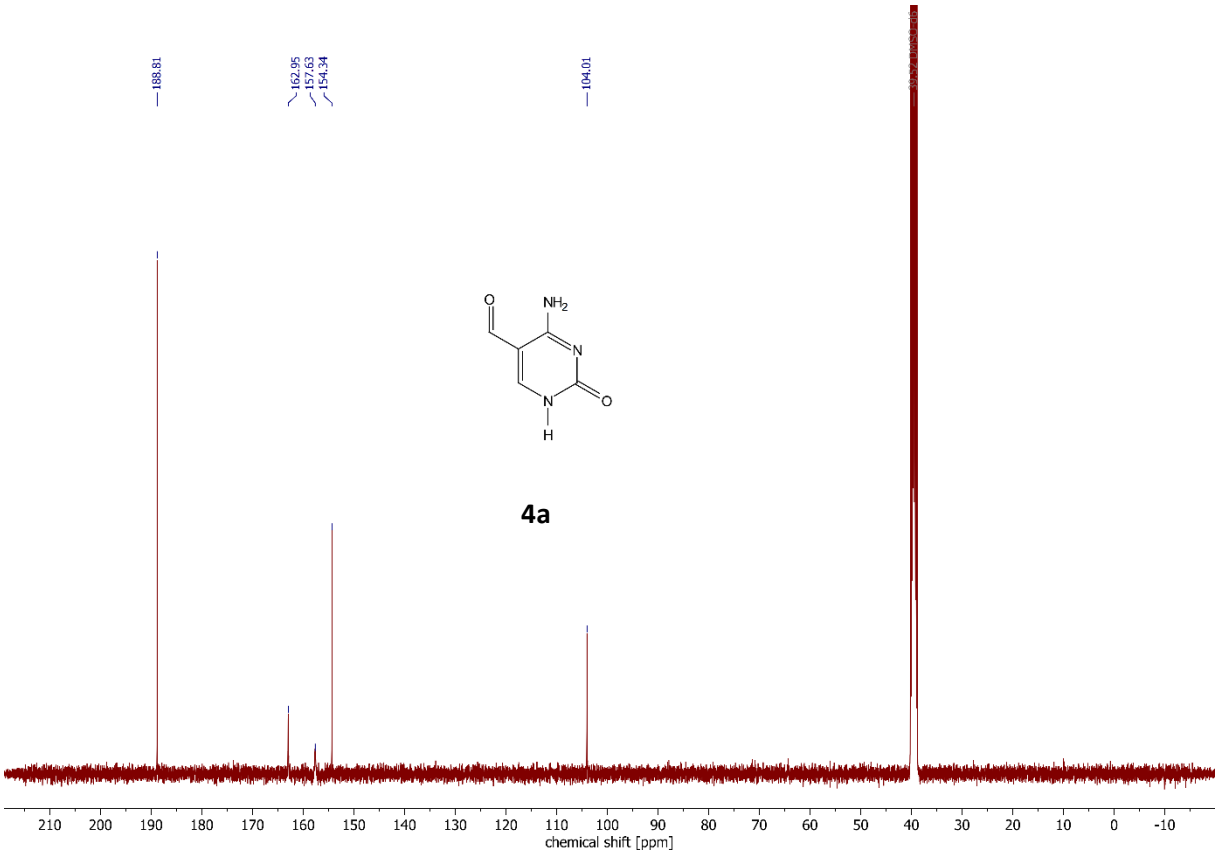
21


165.21
163.53
150.45
150.23
101.23
f1 (ppm)

21


O
NH2
N
N
O
H
4a
9.48
8.37
1.00
1.01
12.5 12.0 11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.
f1 (ppm)


2.50 DMSO-d6
O
NH2
N
N
O
H
4a
9.44
8.41
11.61
7.98
7.84
3.33
0.78
1.00
1.00
1.71
13.0 12.5 12.0 11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
chemical shift [ppm]


188.81
162.95
157.63
154.34
104.01
O
NH2
N
N
H
O
4a
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
chemical shift [ppm]

3.8 REFERENCES

[1] R. H. Hudson, Y. Liu and F. Wojciechowski, *Can. J. Chem.* **2007**, 85, 302-312.

[2] T. T. Sakai, A. L. Pogolotti Jr and D. V. Santi, *J. Heterocycl. Chem.* **1968**, 5, 849-851.

[3] S. Senda, K. Hirota and J. Notani, *Chem. Pharm. Bull.* **1972**, 20, 1389-1396.

[4] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld and T. Carell, *JACS* **2013**, 135, 14593-14599.

[5] X.-A. Zheng, H.-S. Huang, R. Kong, W.-J. Chen, S.-S. Gong and Q. Sun, *Tetrahedron* **2018**, 74, 7095-7101.

[6] A. Gaballa, H. Schmidt, C. Wagner and D. Steinborn, *Inorg. Chim. Acta.* **2008**, 361, 2070-2080.

[7] F. Malz and H. Jancke, J. Pharm. Biomed. Anal. **2005**, 38, 813-823.

[8] ICH Expert Working Group, presented in part at the International Conference on Harmonisation of Technical Requirements for Registration Of Pharmaceuticals for Human Use, Validation of Analytical Procedures: Text and Methodology Q2(R1), Geneva, CH, 2005.

[9] A. D. Becke, *J. Chem. Phys.* **1993**, 98, 5648-5652.

[10] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B* **2009**, 113, 6378-6396.

[11] F. Jensen, *J. Chem. Theory Comput.* **2008**, 4, 719-727.

[12] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, R. J. 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. Jr. Montgomery, 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, D. J. Fox, GAUSSIAN 09 (Revision D.01), Gaussian Inc., Wallingford CT, **2009**.

4 Acid-Base Properties of Cytosine and Uracil Derivatives

4.1 Introduction

In Chapter 3 we reported small amounts of the hydrate form of **5fC** (**5dhmC**) in aqueous, unbuffered solution. This compound is of particular interest due to its possible involvement in the TET-enzyme oxidation of **5fC** to **5caC** shown in Figure 1.

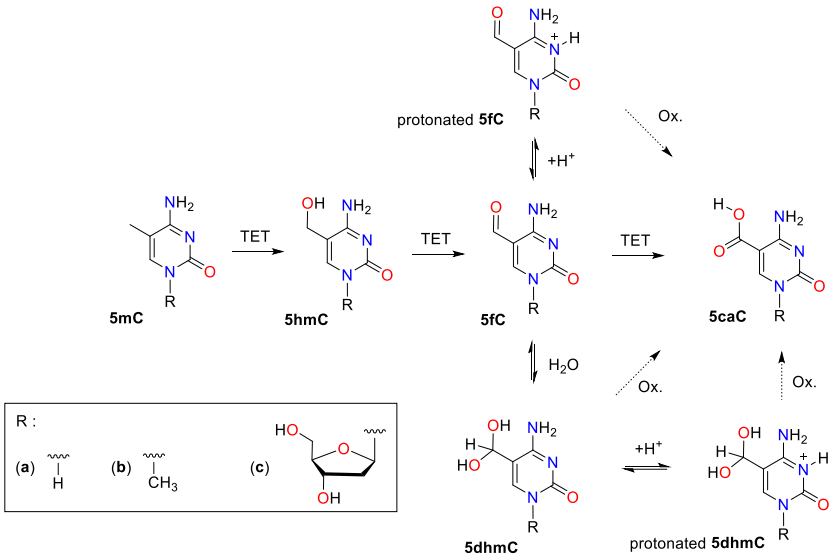


Figure 1: TET-oxidation of **5mC** with possible pathways via hydrated and protonated species of **5fC**.

The **5mC** can be iteratively oxidized to **5caC** via its intermediates **5hmC** and **5fC**.[1] The mechanism of the TET enzyme family is well described (see Chapter 2) and the hydrogen abstraction step of the catalytic cycle is believed to be the rate determining step.[2]

In a collaborative project with Vasily Korotenko, the bond dissociation energy (BDE) values as well as the pK_a values for important modified nucleobases, especially the hydrate form of 5-formylcytosine (**5dhmC**), have been calculated in order to investigate the nature of this oxidative sequence.

4.2 Theoretical Determination of Dissociation Constants

Dissociation constants can be determined by quantum chemical calculations with satisfactory accuracy at the molecular level.[3-7] Achieving chemical accuracy, however, is difficult, due to the fact that a computed difference in the free energy of 5.6 kJ mol^{-1}, which is state-of-the-art for gas-phase energies,[8-12] leads to an error in one unit of pK_a in the de-/protonation reaction. Different methods include absolute and relative pK_a calculations, where the absolute method (sometimes also called direct method) uses a proton-based thermodynamic cycle with the free energy of solvation of a proton $\Delta G_{H^+_{(solv)}}$ as an experimental reference (see Figure 2). This approach is the simplest, however the

limitations of this method are in the reliance of accurate solution phase calculations for all species as well as the inherent error that is introduced by using $\Delta G_{H^+_{(solv)}}$.[7]

Absolute Method:

$$AH^+_{(sol)} \xrightarrow{\Delta G^*(sol)} A_{(sol)} + H^+_{(sol)} \qquad pK_a = \frac{\Delta G^*_{sol}}{2.303RT}$$

Relative Method:

$$AH^+_{(sol)} + Ref_{(sol)} \xrightarrow{\Delta G^*(sol)} A_{(sol)} + RefH^+_{(sol)} \qquad \Delta pK_a = \frac{\Delta G^*_{sol}}{2.303RT}$$

Figure 2: Scheme for the calculation of dissociation constants pK_a with the absolute and relative method (thermodynamic cycles and computational details explained under sections 4.5.3 and 4.5.4, *R*: gas constant, *T*: temperature).

If the pK_a value of a chemically and structurally related molecule is known, the relative method, also called the proton-exchange method, can be employed. This thermodynamic cycle, described in detail in sections 4.5.3 and 4.5.4, is based on an isodesmic reaction in which the proton is shifted between an acid and a reference compound. This approach is advantageous due to the expected cancelation of errors between the solvation free energies of the charged species on both the reactant and product sides, while assuming close structural similarity between the investigated structure and its reference, as well as a similar pK_a values.[13]

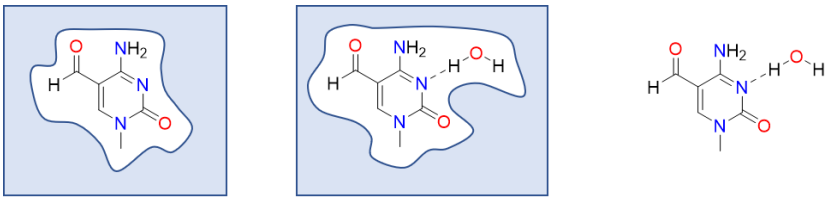


Figure 3: Symbolic representation of implicit, explicit and hybrid solvation models in computational chemistry (light blue space represents the reaction field of the continuous solvation model).

The standard SMD model is fairly inaccurate with mean unsigned errors (MUE) of 21.0 kJ mol^{-1} for anions, and 12.1 kJ mol^{-1} for cations.[14] Tuñón et al.[15] showed in early work that an explicit water solvent molecule plus the continuum solvent improves the solvation free energy of hydroxide anions (see Figure 3). In recent work it has been demonstrated that this implicit-explicit solvation model yields better approximations to experimental pK_a calculations.[3-7] The dissociation constants for methyl-substituted nucleobases have been improved in recent studies by using up to seven explicit water molecules at hydrogen-bonding positions.[16-17]

4.3 RESULTS

This section contains unpublished results. Section 4.3.1 contains bond dissociation energies (BDE) calculated by Vasily Korotenko, while the conformer search for all compounds was done jointly. The computational pK_a study in section 4.3.2 was conducted by Fabian Zott. These results will be published jointly.

4.3.1 BDE Calculations

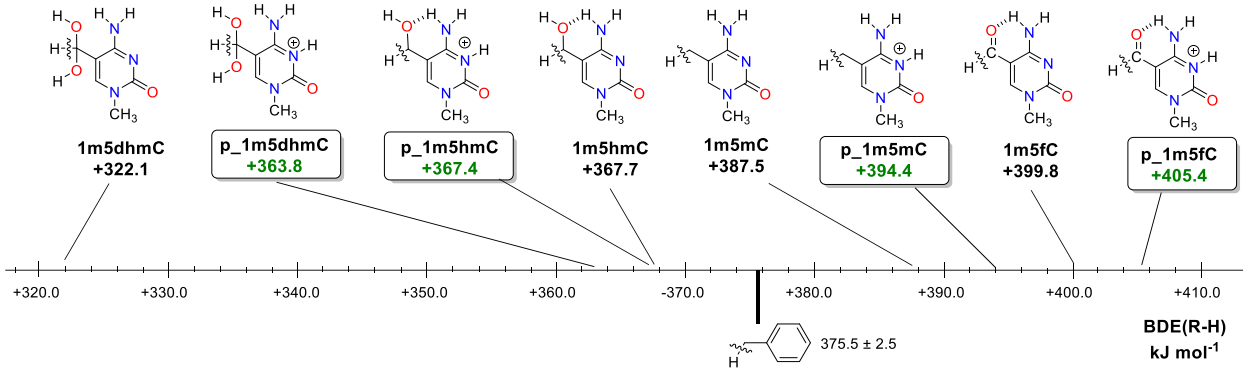


Figure 4: C-H bond dissociation energies (BDE(C-H)) of modified cytosine nucleobases at 298.15 K calculated at the SMD(H_2O)/DLPNO-CCSD(T)/CBS//SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) level of theory. BDE(C-H) values of protonated species are highlighted in boxes.[18]

In Figure 4 the C-H bond dissociation energies (BDE) for modified cytosine nucleobase model systems can be seen in their N^1-methylated form. The C-H bond strength for neutral and protonated cytosine bases increases in the following order **1m5dhmC** < **1m5hmC** < **1m5mC** < **1m5fC**. The BDE of the neutral, hydrated form of **5fC** (**1m5dhmC**) with 322.1 kJ mol^{-1} is by far the lowest value and increases significantly upon protonation (**p_1m5dhmC**, ΔBDE = +41.7 kJ mol^{-1}). While the BDE of **1m5hmC** does not change in case of protonation, the BDE values of both **1m5mC** and **1m5fC** increase by 6.9 and 5.6 kJ mol^{-1} respectively.

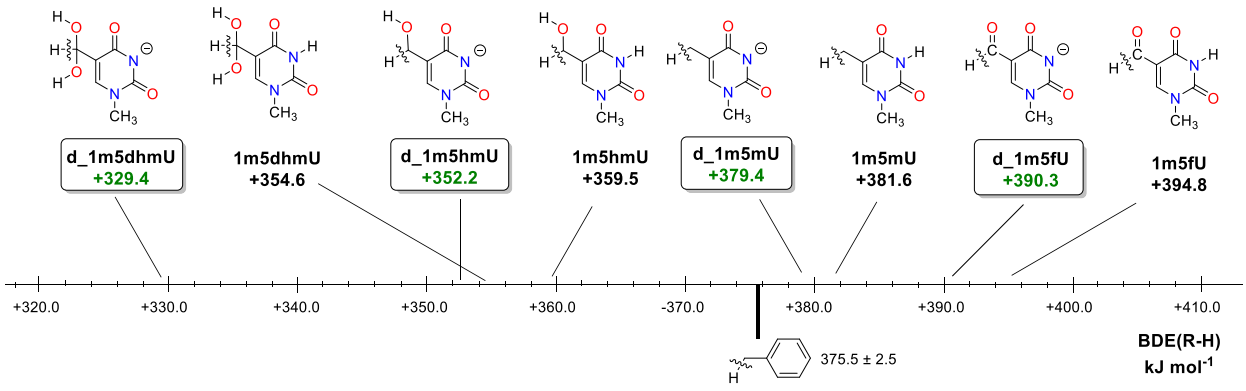


Figure 5: C-H bond dissociation energies (BDE(C-H)) of modified uracil nucleobases at 298.15 K calculated at the SMD(H_2O)/DLPNO-CCSD(T)/CBS//SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) level of theory. BDE(C-H) values of deprotonated species are highlighted in boxes.[18]

For the modified uracil system depicted in Figure 5, the CH bond strength for neutral and deprotonated bases increases in the following order **1m5dhmU** < **1m5hmU** < **1m5mU** < **1m5fU**. The BDE values of all uracil nucleobases decrease in case of deprotonation, with the biggest effect assigned to the **5fU** hydrate (**1m5dhmU,** ΔBDE = +25.2 kJ mol^{-1}).

4.3.2 pK_a Calculations

To determine the dissociation constants of the hydrate form of **1m5fC** as well as **1m5fU** and investigate the site specific, intrinsic acidity of **1m5caC**, a benchmark with known modified nucleobases and respective experimentally determined pK_a values has been conducted.

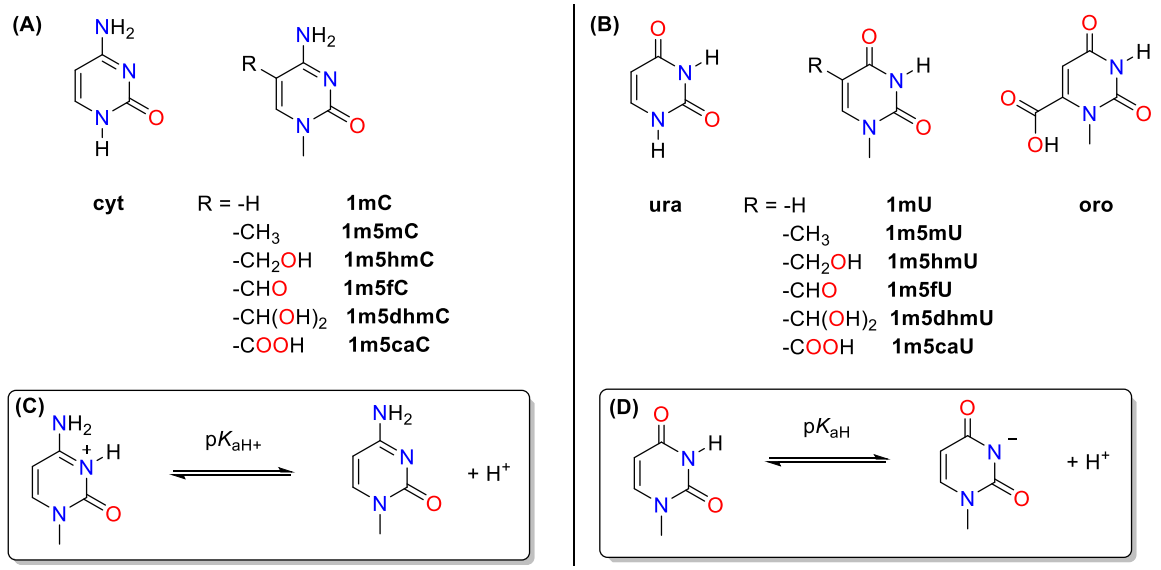


Figure 6: 1-N-Methylated pyrimidine nucleobases investigated in this section. Modified cytosine bases (**A**) and uracil bases (**B**) with its respective deprotonation at the N3 position (**C**) and (**D**). Nucleobases with known experimental pK_a values are collected in Table 1.

Table 1: Experimental pK_a values in water at 25 °C for modified cytosine and uracil nucleobases (averaged values over nucleobases, N-methylated nucleobases, ribonucleosides and deoxyribonucleosides from Table 4, paragraph 4.5.2).

molecule	pK_{aH+} (avg.)	molecule	pK_{aH} (avg.)
cyt	4.56	**ura**	9.46
1mC	4.48	**1mU**	9.49
1m5mC	4.52	**1m5U**	9.88
1m5hmC	3.74	**1m5hmU**	9.33
1m5fC	2.51	**1m5fU**	8.12
1m5caC	2.24 (COOH)	**1m5caU**	4.15 (COOH)
	4.25 (N3)		9.20 (N3)

The dissociation constants for many of these systems are experimentally well known for the respective nucleotides and deoxynucleotides (see Table 1, details: paragraph 4.5.2; Table 4). The N-methylated compounds have very similar chemical properties and are therefore used as model systems.[19-20] We

separately calculated the dissociation constants of the modified cytosine and uracil nucleobases due to their difference in deprotonation reactions. Cytosine bases are described as the cationic deprotonation at the N3 position pK_{aH+} to result in a neutral base (see Figure 6, (**C**)), while uracil bases undergo neutral deprotonation pK_{aH} at N3 to result in an anionic base (see Figure 6, (**D**)). These two classes are therefore described by separate references due to a different impact of absolute errors of the SMD solvation model.[14] Calculations have been performed at the (U)B3LYP-D3/6-31+G(d,p) as well as DLPNO-CCSD(T)/CBS level of theory (for details, see section 4.5.1). To minimize the effects of errors of very low frequency vibrational modes a scaling was applied by using the quasiharmonic oscillator approximation with an energy cut-off value of 30 and 100 cm^{-1}.[21]

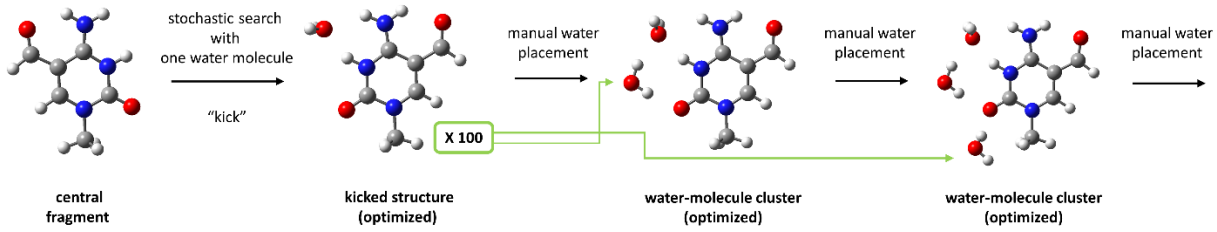


Figure 7: Searching for most stable water-complexed stationary points by a combination of stochastic "kick" protocol and Z-matrix manipulation.

To achieve accurate calculations of relevant dissociation constants a dual approach was taken. Firstly, relative calculations with cytosine and uracil as reference compounds were performed as well as secondly an increased number of explicit waters, to create water-molecule clusters, were tested. In contrast to other theoretical studies of pK_a calculations, these water molecule-clusters were not generated by chemical intuition, but following a strict protocol. To identify the most stable water-nucleobase cluster conformations a stochastic search procedure called "kick" was used which is based on work by Saunders et al.[22-23] This procedure, described in detail in 4.5.5, allows rapid screening of all possible configurations of water around a central molecule, which in this study is the nucleobase.[24] Explicit water molecules are placed in a random, predefined mode around this central molecule and after optimization of around 100 structures, all hydrogen-bonding sites are covered in the dataset.

In a second step water clusters are generated by manual placement of additional explicit water molecules. The ensemble of 100 conformers with one explicit water generated by the "kick" method is optimized and ranked by total energy E_{tot}. Now water-molecule clusters are generated by adding the lowest energy conformers to the existing clusters, followed by reoptimization (see Figure 7, indicated by green arrows). This leads to the coverage of all hydrogen bonding sites, beginning with the strongest. Also, care is taken to not crowd the hydrogen bonding sites by conformers often similar in energy which cover the same site.

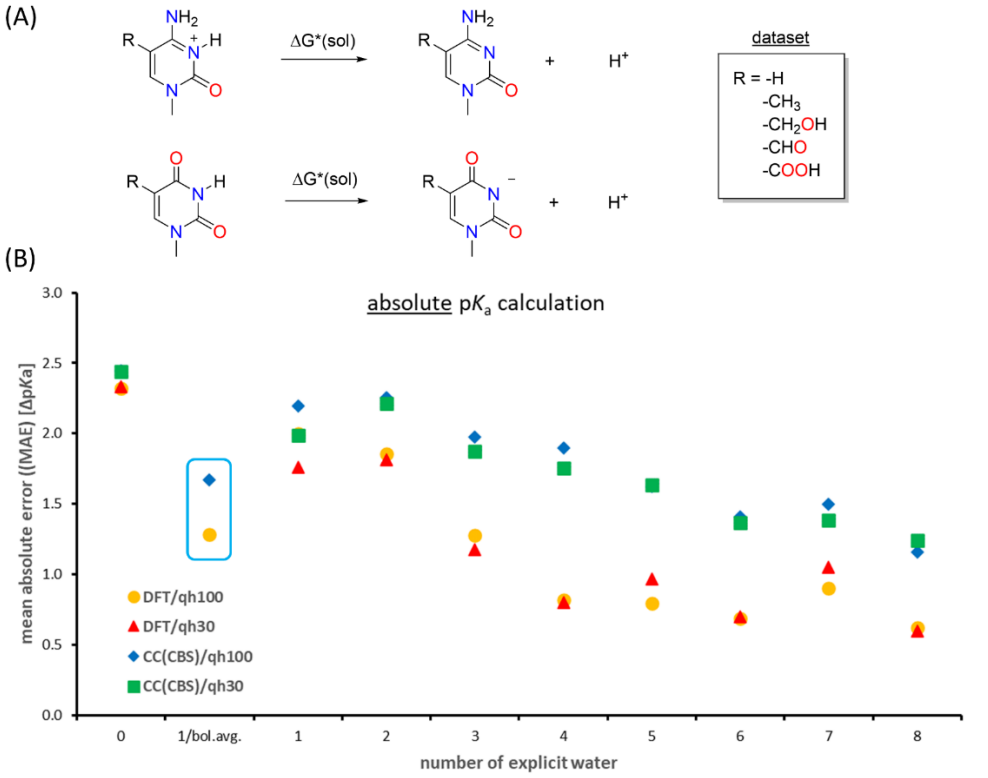


Figure 8: Mean absolute error (MAE) with increasing number of explicit waters (**B**) for absolute pK_a calculations (**A**) with ten systems investigated. Benchmarked compounds include: **1mC**, **1mU**, **1m5mC**, **1m5mU**, **1m5hmC**, **1m5hmU**, **1m5fC**, **1m5fU**, **1m5caC** and **1m5caU**. Using solution phase geometries SMD(H_2O)/UB3LYP-D3/6-31+G(d,p) and DLPNO-CCSD(T)/CBS single point calculations with the cc-pVTZ and cc-pVQZ basis set extrapolation; SMD solvation energies calculated at SMD(H2O)/B3LYP-D3/6-31+G(d,p) level; qh100 and qh30: quasi-harmonic approximation cut-off values [cm^{-1}].

The benchmark of absolute pK_a calculations for a test set of ten modified nucleobases known in literature yields results of an MAE value above one unit of pK_a (see Figure 8, experimental pK_a data: see 4.5.2, Table 4). The *N*-3 position of the uridine and cytosine modifications being the primary physiological site of deprotonation and protonation, respectively. The highest MAE are obtained for calculations without explicit water and deceasing upon addition of explicit water. After molecule-water clusters contain more than six water molecules, the values again become less accurate. An additional approach, using the best ten conformers with one "kicked" water and applying the Boltzmann averaged free energy, DFT as well as coupled-cluster methods do not match the experimental dissociation constants (see Fig. 8, (**B**), blue box; method explanation in paragraph 4.5.8). It can be noted, however, that for absolute pK_a calculations an ensemble of ten, Boltzmann-averaged conformers with one explicit water is superior to a single conformer with up to three explicit water molecules. Methods based on density functional theory (here: SMD(H2O)/UB3LYP-D3/6-31+G(d,p)) generally yield better results than the ones based on coupled cluster theory (here: DLPNO-CCSD(T)/CBS). The MAE value decreases with increased explicit solvent modelling with DFT methods between three and four explicit water and a quasi-harmonic cut-off value at 100 cm^{-1} (see Figure 8, "qh100"). Using the quasi-harmonic approximation with a cut-off value at 30 cm^{-1} gives usually slightly favourable errors (see Figure 7, "DFT/qh30" and "CC(CBS)/qh30").

In recent work of Schlegel *et al.*[17], a two-parameter adjustment for these pyrimidine modifications was used to reach good agreement with experimental values. The number of solvent molecules in these molecule-water clusters varied between 2-7 explicit water to generate a benchmark dataset of around 14 modified pyrimidines. The conformer search, on the contrary to our work, was not done by

a standardised, unbiased procedure, but chemical intuition, covering all hydrogen bonds by manual water placement.

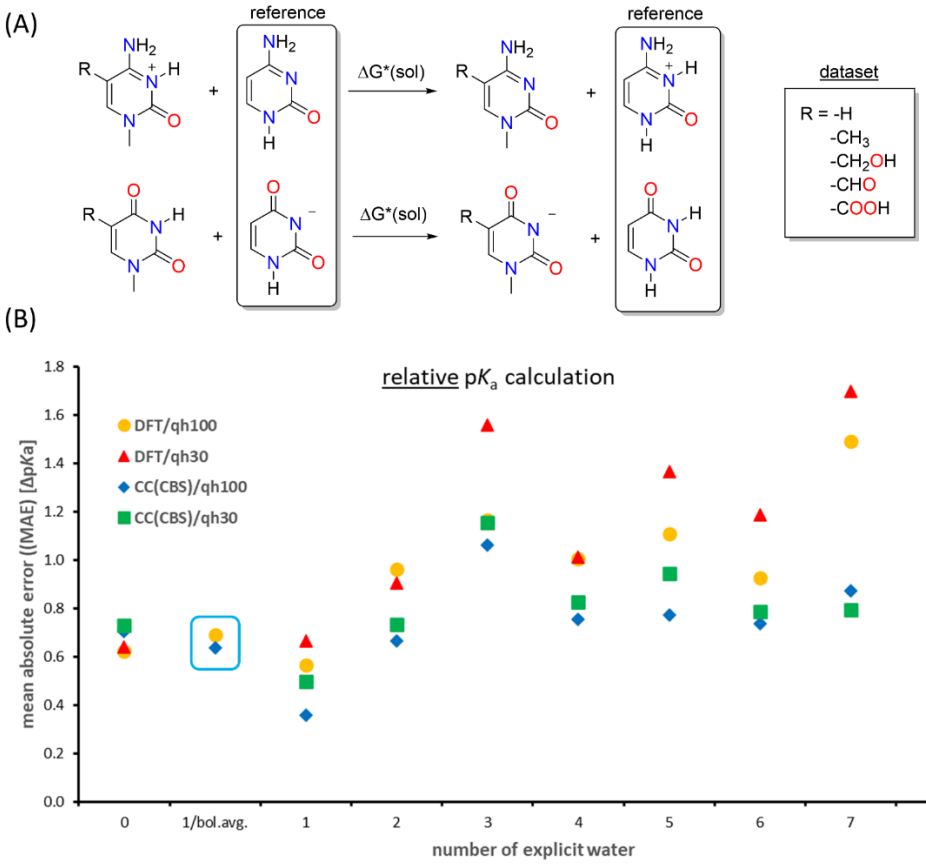


Figure 9: Mean absolute error (MAE) with increasing number of explicit waters (**B**) for relative pK_a calculations (**A**) with ten systems investigated. Benchmarked compounds include: **1mC**, **1mU**, **1m5mC**, **1m5mU**, **1m5hmC**, **1m5hmU**, **1m5fC**, **1m5fU**, **1m5caC** and **1m5caU**. Using solution phase geometries SMD(H_2O)/UB3LYP-D3/6-31+G(d,p) and DLPNO-CCSD(T)/CBS single point calculations with the cc-pVTZ and cc-pVQZ basis set extrapolation; SMD solvation energies calculated at SMD(H2O)/B3LYP-D3/6-31+G(d,p) level; qh100 and qh30: quasi-harmonic approximation cut-off values [cm^{-1}].

To yield better results, the generated water-molecule clusters were used for relative pK_a calculations, where the reference for modified pyrimidine bases are the well-known systems, cytosine (pK_{aH+}) and uracil (pK_a). In Figure 9 it can be seen, that relative pK_a calculations give better MAE values compared to absolute pK_a calculations (Figure 8). Not using any explicit water gives MAEs around 0.65 units of pK_a. When using the ten best Boltzmann-averaged conformers with one "kicked" water, similar results are obtained (see Fig. 9, (**B**), blue box). The reason why the results do not improve the same as for the absolute method (see Figure 8), is the cancellation of positive effects of the Boltzmann-averaging for charged and neutral species by having a structurally close reference on both sides of the reaction equation (see Figure 9, (**A**)). Upon increasing the size of water-molecule clusters above three explicit solvent molecules less favourable results are obtained. The best results are given with coupled cluster methods and one explicit water, with a MAE of 0.36 pK_a units for quasi-harmonic approximation cut-off values of 100 cm^{-1} and of 0.5 pK_a units for quasi-harmonic approximation cut-off values of 30 cm^{-1}. These MAE values indicate good agreement between experimental and theoretical values.

In Figure 10 and 11 linear correlations of experimental and theoretical pK_a values with different numbers of explicit water molecules can be seen. The theoretical values represent the best method determined before (SMD(H_2O)/DLPNO-CCSD(T)/CBS with quasi harmonic approximation by applying a cut-off value of 100 cm^{-1}) and the experimental values are averaged over ribonucleosides, deoxyribonucleosides and nucleobases to give a more robust reference base (see Table 1).

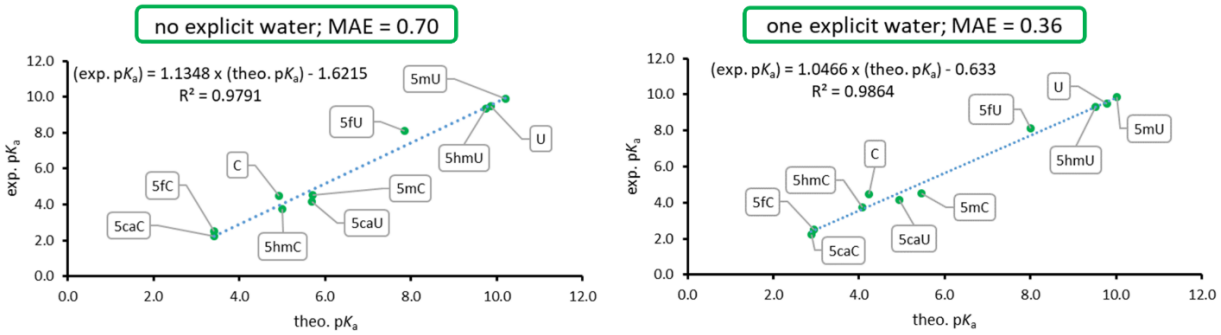


Figure 10: Correlation of experimental with theoretical pK_a values for the relative method without (left) and with one (right) explicit water (Method: SMD(H_2O)/DLPNO-CCSD(T)/CBS with quasi harmonic approximation (cut-off = 100 cm^{-1}), see Figure 9; theoretical pK_a vales for N-1 methylated compounds and experimental pK_a values as average over ribonucleosides and nucleobases, see Table 1).

When plotting the experimental versus the theoretically determined pK_a values for the best method (MAE = 0.36), we obtain a good linear fit with a small systematic shift of the theoretical data of - 0.633 units of pK_a and a very good coefficient of determination R^2 = 0.9864 (see Figure 10, right). The linear fits of the same method without explicit water also give good linear correlations but with higher systematic shift of – 1.6215 units of pK_a (see Figure 10, left).

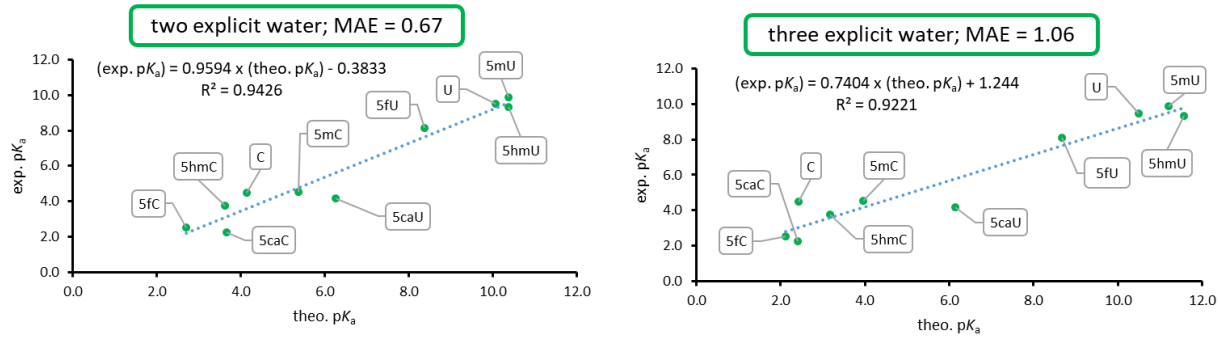


Figure 11: Correlation of experimental with theoretical pK_a values for the relative method with two (left) and three (right) explicit water (Method: SMD(H_2O)/DLPNO-CCSD(T)/CBS with quasi harmonic approximation (cut-off = 100 cm^{-1}), see Figure 9; theoretical pK_a vales for N-1 methylated compounds and experimental pK_a values as average over ribonucleosides and nucleobases, see Table 1).

Using solution phase calculations with the addition of two water leads to high quality correlation (R^2 = 0.9426) with the lowest systematic shift of -0.3833 units of pK_a, but higher MAE value of 0.67 (Figure 11, left). The addition of three explicit water yields a slightly less favourable correlation with a positive pK_a shift of +1.244 units and a MAE value of 1.06 (Figure 11, right). When looking at the

respective plot with two explicit waters (Figure 11, left), a good linear correlation can be found with **5caC** and **5caU** being the only outliers in this respective dataset. The same can also be observed for calculations with three explicit waters with **5caU** and **C** being the outliers (Figure 11, right).

All these findings lead to the conclusion that, relative yield better results than absolute pK_a calculations, with the best method being coupled cluster methods (SMD(H_2O)/DLPNO-CCSD(T)/CBS) using quasi-harmonic enthalpy corrections (cut-off = 100 cm^{-1}) and one or two explicit water molecules.

Table 2: Experimental and theoretical pK_a values for modified pyrimidine bases.

compound[a]		acidic site /form[b]	comp. pK_a	adj. pK_a[c]	exp. pK_a[d]	Schlegel *et al.* [17]
C		N3/P	4.24	3.80	4.48	4.6
5mC		N3/P	5.46	5.08	4.52	4.9
5hmC		N3/P	4.08	3.64	3.74	3.2
		OH/N	16.78	16.93	-	-
5dhmC		N3/P	4.82	4.41	-	-
		OH/P	16.30	16.43	-	-
		OH/N	16.76	16.88	-	-
5fC		N3/P	2.96	2.46	2.51	1.5
5caC	**cation**	COOH/P	3.50	3.03	2.24	-
		N3/P	2.89	2.39		-
	neutral	COOH/N	5.17[e]	5.07	4.25	-
		N3/N	5.51[e]	5.32		-
U		N3/N	9.78	9.61	9.49	9.6
5mU (T)		N3/N	10.02	9.86	9.88	9.2
5hmU		N3/N	9.53	9.34	9.33	-
5dhmU		N3/N	8.08	7.82	-	-
		OH/N	16.09	16.21	-	-
5fU		N3/N	8.01	7.75	8.12	-
5caU		COOH/N	4.94	4.45	4.15	-
		N3/NEG	9.27[f]	9.07	9.20	-

a: N^1-methylated; b: cationic (P), neutral (N) and anionic (NEG) form; c: parameters according to Figure 10 and 11; d: averaged over ribonucleoside and nucleobase values (see 4.3.2, Table 1 and 4.5.2, Table 4); e: three explicit water; f: reference is orotic acid (for better charge balance).

In Table 2 a summary of all calculated and experimentally determined pK_a values is shown. The term "adjusted pK_a" is used in accordance with work by Schlegel *et al.*[17], meaning a extrapolated value using equations represented in Figures 10 and 11. We support the experimental dissociation constants for modified pyrimidine nucleobases reported in the literature (see Table 4.3.2, Table 1). We do also match and sometimes even improve the theoretical pK_a values computed earlier by Schlegel *et al.*[17] For **1m5hmC**, **1m5fC** and **1m5mU** (thymine) the accuracy of computed values is improved by up to one unit of pK_a. For modified uracil bases, methylation (**1m5mU**) as well hydroxy methylation (**1m5hmU**, Figure 12, **A**) at the 5 position does not significantly change the pK_a values compared to the unmodified uracil system. 5-Formyluracil (**1m5fU**; Figure 12, **C**) on the other hand shows a two pK_a units lower value compared to uracil, which is also in line with the results for 5-formylcytosine (**1m5fC**). The pK_a of **1m5hmU** is only lowered by 0.52 units when compared to its 5-C-methylated nucleobase (**1m5mU**), while for **1m5hmC** this effect amounts to 1.44 units. This can be associated to the stabilizing effects of

a hydrogen bridge formed by explicit water between the carbonyl group at C-4 and the hydroxymethyl group at C-5 positions in contrast to **1m5hmC** (see Figure 12**, A** and **B**). The hydrate form of 5-formylcytosine (**1m5dhmC**) has a significantly higher pK_a value than its non-hydrated form **1m5fC** (ΔpK_a = +1.95), while the same cannot be observed for the hydration of 5-formyluracil hydrate (**1m5dhmU**, ΔpK_a = +0.07). The pK_a value of the hydroxy group of **1m5dhmC** is 16.43 for the protonated (cationic) and 16.88 for the unprotonated (neutral) form, resembling typical benzylic OH acidity (pK_a(benzylic alcohol) = 15.40).[25]

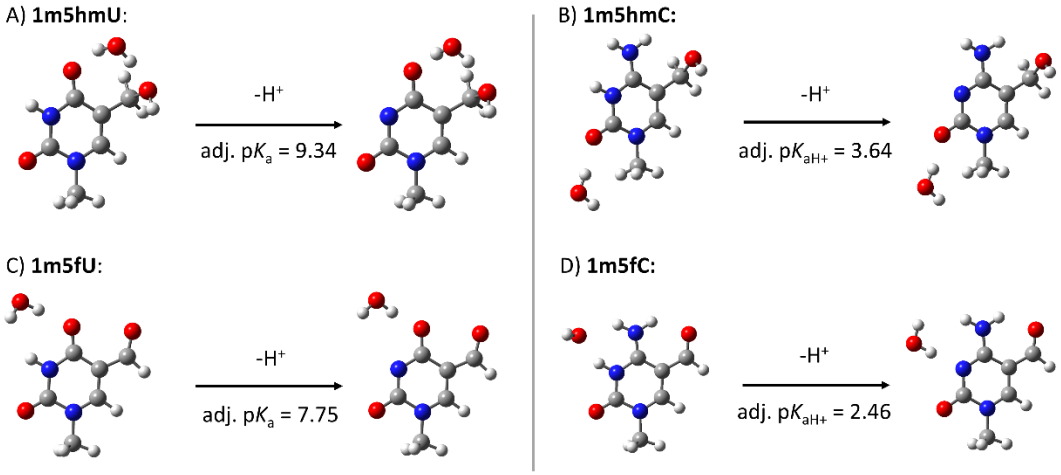


Figure 12: Lowest energy conformers with one explicit water for the deprotonation of A) 5-hydroxymethyluracil B) 5-hydroxymethylcytosine C) 5-formyluracil and D) 5-formylcytosine as well as its theoretical adjusted pK_a values (see Table 2).

The experimentally determined pK_a values for **1m5caC** are matched theoretically with good accuracy. The first deprotonation at the N3 position of the cationic form of **1m5caC** has a theoretical, adjusted pK_{aH+} of 2.39 (Figure 13), which is close to the experimental value of 2.24. The competing carboxy function has an adjusted pK_{aH+} value of 3.03 and is therefore 0.64 units of pK_a higher. This difference is bigger than the mean absolute error for this method (MAE = 0.36, see Figure 10). The resulting neutral form of **1m5caC** is in a dynamic, aqueous equilibrium between a neutral and a zwitterionic form (see Figure 13, **B** and **C**). The Boltzmann distribution between this neutral and zwitterionic species is determined to be 80:20 by weighting the best conformers with one explicit water of each tautomer with our best method (SMD(H_2O)/DLPNO-CCSD(T)/CBS).

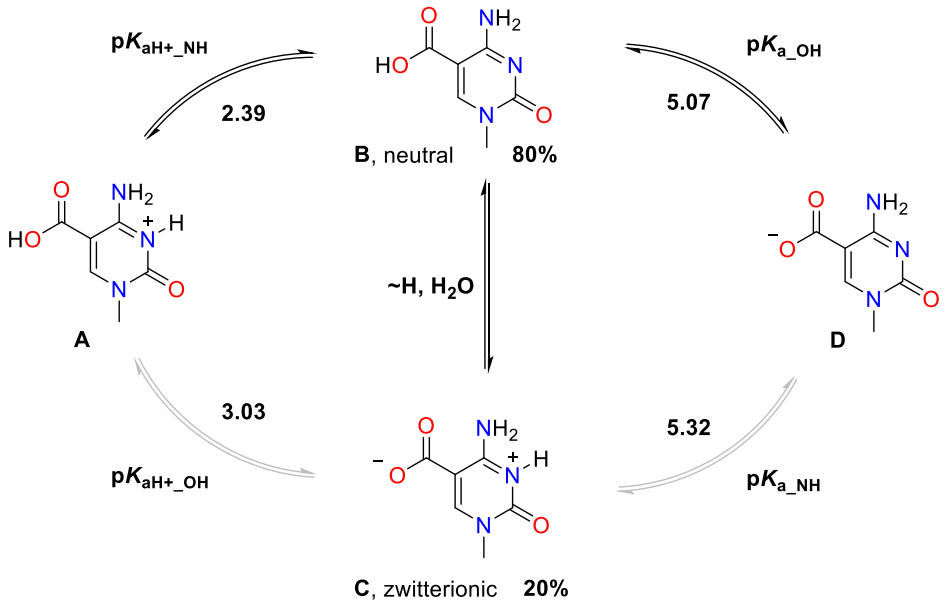


Figure 13: Adjusted theoretical, intrinsic, and averaged, pK_a values for **1m5caC** with Boltzmann distribution in H_2O (pK_{aH+} calculated with one explicit water and pK_a with three water molecules at SMD(H_2O)/DLPNO-CCSD(T)/CBS level of theory).

The adjusted, theoretical N3-acidity of the zwitterionic form is slightly stronger (Figure 13, **C**; pK_{a_OH} = 5.32) compared to the OH-acidity of its tautomer in water (Figure 13, **C**; pK_{a_NH} = 5.07). The experimental pK_a value differs by 0.82 units from the computed one, lying within the MAE of 1.06 for this method (see Figure 11). The theoretical determination of intrinsic pK_a values of the neutral and zwitterionic form of **1m5caC** (Figure 13, **B** and **C**) required the addition of three explicit water molecules (see paragraph 4.5.7). In a computational study by Maiti *et al.* the intrinsic acidities of **B** and **C** (Figure 13) as free energies were reported.[26] Although not determining the exact intrinsic pK_a values, they state that the zwitterionic form **C** is slightly less acidic than the tautomer **B**.

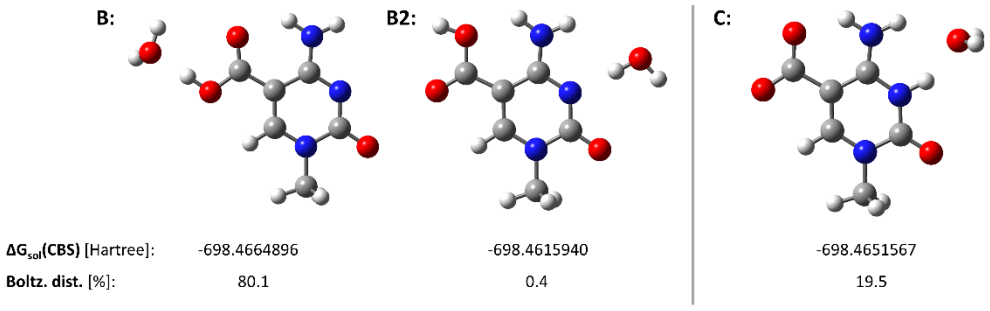


Figure 14: Comparison of lowest energy conformers and tautomers for **1m5caC** with <u>one</u> explicit water at the SMD(H2O)/DLPNO-CCSD(T)/CBS level of theory and its Boltzmann distribution.

In Figure 14 for **1m5caC**, the influence of conformers on the Boltzmann distribution in aqueous phase is shown. The lowest energy conformer **A** is weighted with 80.1%, while the next lowest energy conformer **B2** with 0.4% having no practical influence on the pK_a value of **1m5caC** and can therefore be neglected. The zwitterionic tautomer is shown in **C** with a Boltzmann weighting of 19.5% (Figure 14). This emphasises the selection of only two best conformer and tautomers being sufficient for the

theoretical determination of pK_a values in this study. Thus, for all other compounds, having only one de-/protonable site, one conformer is adequate.

The first dissociation constant pK_a of 5-carboxyuracil **1m5caU** with 4.45 is close to the experimental value of 4.15. For the second deprotonation step of the anionic form of **1m5caU**, the pK_{a^-} of a more suitable reference structure was used, which is structurally closer and has a better charge balance than uracil. With orotic acid as reference compound, the adjusted computed pK_{a^-} of **1m5caU** with one explicit water is 9.07, with 9.20 as the experimental value.

4.4 Discussion

In this chapter a new method to generate molecule water clusters for the theoretical determination of pK_a values has been developed and, after a benchmark experiment, has been employed in the prediction of dissociation constants of modified pyrimidine nucleobases in aqueous solution.

The hydrate form of 5-formylcytosine (**1m5dhmC**) is of particular interest due to its discovery and quantification in aqueous solutions.[20,27] Burrows *et al.* suggested that **5fC** exists in equilibrium with its hydrate form **5dhmC** when measuring base-flipping kinetics.[28] While the hydrate form was not reported in two recent NMR studies on the melting kinetics of dsDNA, the transient hydrate formation in the TET2-mediated oxidation of **5fC** cannot be excluded.[29-30] In 2015 Hu *et al.*[31] investigated the kinetics of the TET2-enzyme, determining kinetic efficiency k_{cat}/K_M values experimentally using the Michaelis-Menten equation. From these specificity constants k_{cat}/K_M the activation free energy $\Delta G^{\ddagger}$ can be determined using a modified form of the Eyring equation (see Table 3; $\Delta G^{\ddagger}$(exp); R: gas constant; h: Planck's constant; k_b: Boltzmann's constant; T: temperature):

$$\Delta G^{\ddagger} = -\mathrm{R}T \ln\left(\frac{k_{\mathrm{cat}}\,\mathrm{h}}{\mathrm{k_b}T}\right)$$

This shows that there is no significant difference in $\Delta G^{\ddagger}$ and thus the reaction rates of the epigenetic substrates **5mC**, **5hmC** and **5fC** do not differ significantly.

Table 3: Overview of theoretical and experimental free energy barriers $\Delta G^{\ddagger}$ for the TET2 enzyme.[2,31]

Substrate	K_M [µM]	k_{cat} [10^{-3} s^{-1}]	k_{cat}/K_M [M^{-1} s^{-1}]	$\Delta G^{\ddagger}$(exp)[a] [kJ mol^{-1}]	$\Delta G^{\ddagger}$(theo)[b] [kJ mol^{-1}]
5mC	0.48	2.12	4.42×10^{-3}	71.2	67.0
5hmC	0.90	0.63	0.70×10^{-3}	74.2	77.2
5fC	1.30	0.46	0.35×10^{-3}	75.0	113.8

a: determined by Eyring equation from k_{cat} and K_M; b: for the rate determining hydrogen atom abstraction (HAT) step.

When comparing these results with an extended computational investigation on the substrate preference of the TET2 enzyme,[2] a significant difference in the free energy barrier $\Delta G^{\ddagger}$ for the substrate **5fC** can be seen (see Table 3; $\Delta G^{\ddagger}$(theo)). Our BDE calculations in chapter 4.3.1 suggest that the ΔBDE between **5dhmC** and **5fC** is +77.7 kJ mol^{-1} (Figure 4) for the neutral and +41.6 kJ mol^{-1} for the protonated form. Even low concentrations of **5dhmC** in the enzyme active site can thus be rate

determining for the oxidation of **5fC** towards **5caC**. The increased pK_a value of 4.41 for **5dhmC** compared to 2.46 for **5fC** can have profound consequences for the enzymatic oxidation of **5fC** by facilitating an easier protonation step when incorporating the substrate into the active site. The hydration process is more favourable under acidic conditions,[20] after incorporation into the active site of the TET2 enzyme protonation state dictates a +41.7 kJ mol^{-1} difference in BDE for the HAT step of the oxidative cycle (Figure 4).

Another area of interest is the site-specific acidity of **5caC**, with early work assigning the first deprotonation step pK_{aH+} = 2.50 (Figure 13; **A**) to the carboxylic group and the second deprotonation pK_a = 4.28 to the N3 site of the pyrimidine heterocycle (Figure 13; **C**).[26,32] Later work by Tokmakoff *et al.*[33] concluded an opposite site assignment for the two **5caC** pK_a values, attributing the more acidic pK_{aH+} value to the N3 position by *pD*-dependent NMR and FTIR studies. Our quantum chemical calculations show a clear assignment for the first protonation step pK_{a_NH+} = 2.39 to the N3 position of **1m5caC** as well as the competing intrinsic at the carboxylic group pK_{a_OH+} = 3.03, which is not prevalent in *pH*-titration experiments (Figure 13; **A**). Since these two intrinsic pK_{aH+} values are very close to each other, depending on the experimental method, they can appear as an averaged pK_{aH+} of 2.52.

We also propose an equilibrium in aqueous phase between the two possible neutral species (Figure 13; **B** + **C**) that cannot be distinguished by slow experimental methods like NMR spectroscopy. The second deprotonation step show similar intrinsic pK_{a_NH} and pK_{a_OH} values (5.32 and 5.07), which leads to the conclusion that the neutral **1m5caC** is deprotonated almost unbiased at both the carboxy and the N3 sites. This results in an averaged pK_a value for the second deprotonation step of **5caC** of 5.12, being very close to experimental investigations (4.25), and also explaining the observed big ppm shifts for the second deprotonation step observed for the *pD*-depended NMR titration by Tokmakoff *et al.*[33] A recent study on the DNA duplex stability of **5caC** by the same authors reaffirms their pK_a assignment, with the pK_a of **5caC** nucleobase within duplex DNA between 4.1 – 4.8 at the carboxy site.[34] Since in a DNA duplex the N3 site is covered by a hydrogen bond and therefore not protonated, this value is close to the intrinsic pK_{a_OH} of neutral **1m5caC** of 5.07 at the carboxyl group (Figure 13; **B**).

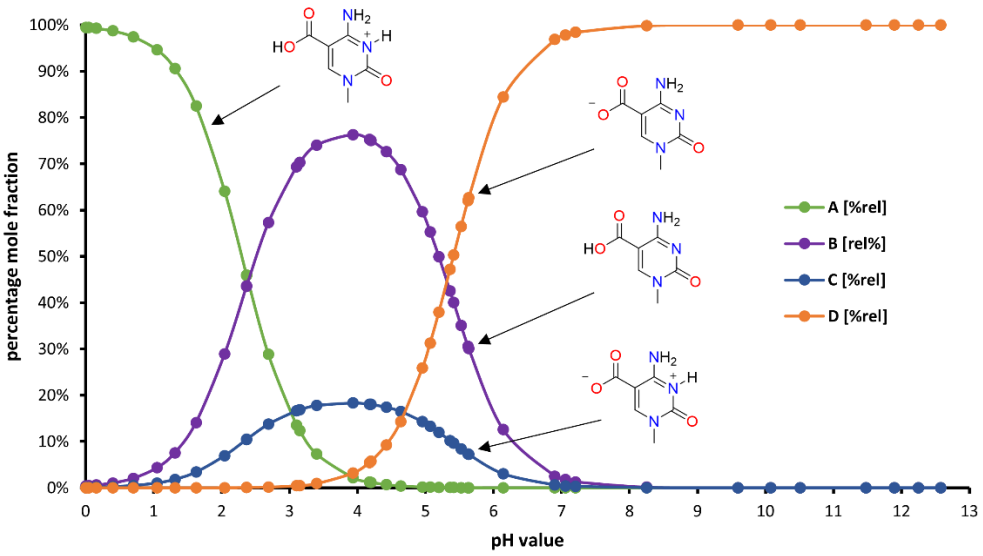


Figure 15: Steady-state analysis of the percentage mole fraction of all **1m5caC** species at different pH values in unbuffered water performed by the CoPaSi program (see Figure 12 for dissociation constants; CoPaSi[34] simulations described under paragraph 4.5.8).

In Figure 15 the percentage mol fractions of all protonated and deprotonated species of **1m5caC** are simulated at different pH values in unbuffered water by the CoPaSi program.[34] This simulation takes into account all intrinsic dissociation processes as well as the equilibrium between the neutral and zwitterionic form of **1m5caC** (see Figure 13 and 15 **B** + **C**, for details see paragraph 4.5.8). It can be seen that at pH values between 1.5 and 6.5 a mixture of zwitterionic and neutral species prevail, making a clear assignment of signals difficult for experimental measurements (see Figure 16, **A**).[33] The maximum simulated distribution of zwitterionic species **C** and neutral species **B** is 18.3% and 76.7% at pH = 3.94 with the remainder being fully protonates **A** and deprotonated **D** species (see paragraph 4.5.8, Table 5). This result is very close to the predicted equilibrium of **C** and **B** in water with a Boltzmann distribution of 80% to 20% (see Figure 13). From this data pK_a titration curves can be constructed either by using the simulated molar fractions obtained from CoPaSi and the ^{13}C-chemical shifts from Figure 16 **A** (see Figure 16 **B**) or by using theoretically determined absolute isotropic shielding values σ_{iso} calculated at the B3LYP/pcS-3 level of theory (see Figure 16 **C**). It can be seen, that these simulated curves resemble the experimentally determines titration curve very well (compare Figure 16 **A** with **B** and **C**). While graph **B** (Fig. 16) does not use exact shift values for the neutral species and therefore only the assumption that ^{13}C-chemical shifts are the same for the anionic and the neutral carboxylic groups, these differences are fully accounted for in graph **C**.

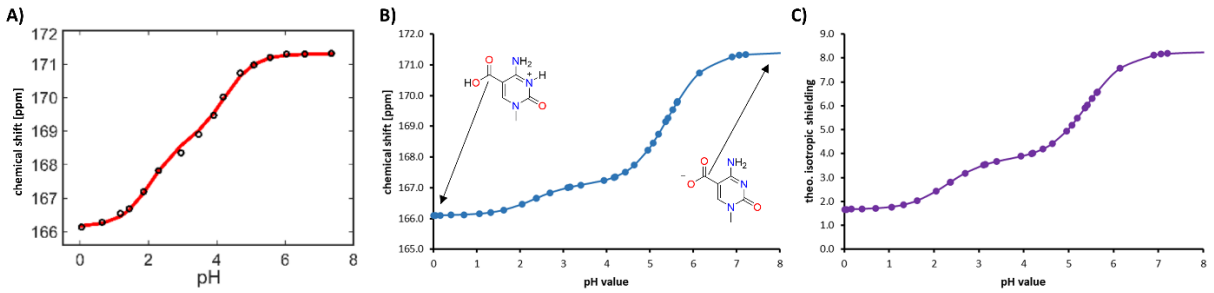


Figure 16: A) Experimentally determined pK_a titration curve for ^{13}C NMR chemical shifts of ^{13}C-labeled carboxyl group of 5-caboxycytidine (**5caC**) by Tokmakoff *et al.*[33]; B) simulated pK_a titration curve using chemical shift data from A) and molar fractions from the CoPaSi simulation; C) simulated pK_a titration curve by using theoretically determined isotropic shielding values σ_{iso} and molar fractions from the CoPaSi simulation. For details see*:* paragraph 4.5.8. **A)**: Reprinted and adapted with permission from [33]. Copyright 2023 American Chemical Society.

The tautomeric equilibrium between **B** and **C** (Figure 13 and Figure 15) and its intrinsic dissociation constants in aqueous solution also explain the limited TDG activity towards **5caC** under neutral conditions compared to **5fC**. Due to easier protonation at the N3 site, the glycosidic bond can be cleaved more facile by the TDG enzyme in a BER pathway.[26,33] Figure 17 represents the important tautomers during glycosidic bond cleavage for **5caC** and **5fC**, with N-3 protonated substrates being more likely to react.

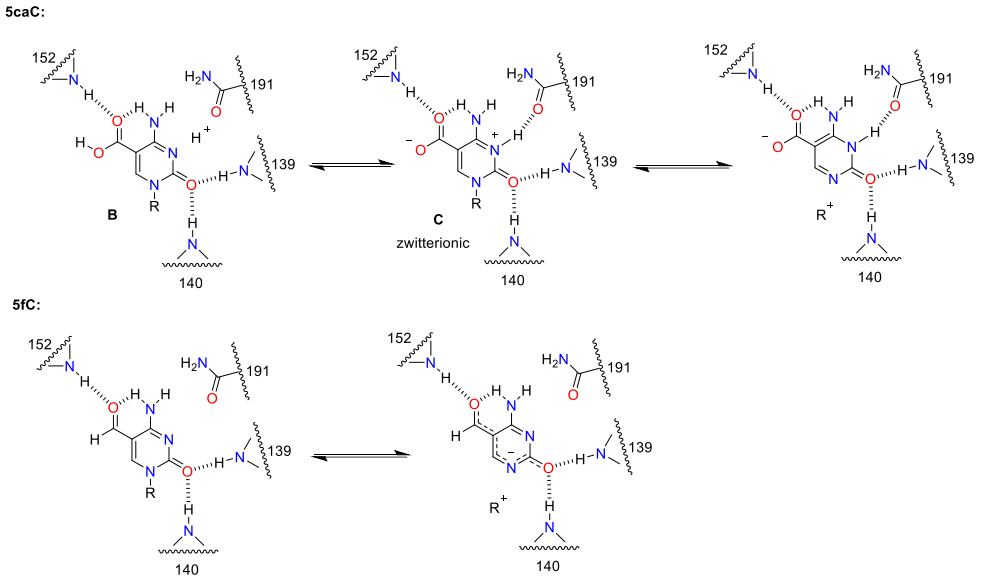


Figure 17: Potential differences in TDG excision of **5caC** (top) and **5fC** (bottom) due to different tautomerism and protonation state (structure for active site derived from previous studies).[26,36]

4.5 SUPPLEMENTARY MATERIAL

4.5.1 General Considerations

In this study geometry optimizations have been performed with a combination of the (U)B3LYP hybrid functional[37] complemented by the D3 dispersion correction[38] and the 6-31+G(d,p) basis set[39-40] in solution phase with Truhlar's SMD solvation model.[41] Thermochemical corrections (corr. Δ*H* & Δ*G*) to 298.15 K have been calculated at the same level of theory using the rigid rotor/harmonic oscillator model. In order to identify the effects of low-energy torsions, all vibrational frequencies were scaled by a scaling factor of 30 and 100 cm^{-1} introduced by Truhlar *et al.* using the Grimme quasi-harmonic oscillator approximation for the calculation of zero-point energies, thermal corrections, and entropies.[42-44] This quasi-harmonic frequency correction was implemented by using the "GoodVibes" program.[45] Enthalpies (ΔH_{298}) and Gibbs energies (ΔG_{298}) at (U)B3LYP-D3/6-31+G(d,p) level have been obtained through addition of corr. Δ*G* and corr. Δ*H* to ΔE_{tot}, respectively. All B3LYP calculations were performed using Gaussian 09, rev. D.01.[46]

Single point energies have subsequently been calculated with the DLPNO-CCSD(T) method[47-48] as implemented in ORCA 4.2.1[49] in combination with the cc-pVTZ and cc-pVQZ basis sets, followed by extrapolation to the complete basis set (CBS) limit to DLPNO-CCSD(T)/CBS total energies with the following equations:[50]

$$E_{HF}^{CBS} = \frac{E_{HF}^{(n)} \times e^{-\alpha\sqrt{m}} - E_{HF}^{(m)} \times e^{-\alpha\sqrt{n}}}{e^{-\alpha\sqrt{m}} - e^{-\alpha\sqrt{n}}} \quad (1)$$

$$E_{corr}^{CBS} = \frac{n^{\beta} E_{corr}^{(n)} - m^{\beta} E_{corr}^{(m)}}{n^{\beta} - m^{\beta}} \quad (2)$$

$$E_{DLPNO}^{CBS} = E_{HF}^{CBS} + E_{corr}^{CBS} \quad (3)$$

For the cc-pVTZ and cc-pVQZ basis sets the 3/4 extrapolation scheme with n=3 and m=4 and the constants α = 5.46, and β = 3.05 were used.[47,51] Enthalpies at 298.15 K have subsequently been calculated through combination of the DLPNO-CCSD(T)/CBS total energies with thermochemical corrections (Δ*H*) calculated at the (U)B3LYP-D3/6-31+G(d,p) level before. In order to estimate the impact of aqueous solvation, single point energies have been calculated using the SMD/H_2O//(U)B3LYP-D3/6-31+G(d,p) level using the solution phase-optimized geometries obtained before.[41] Solvation free energies in water (ΔG_{solv}) have then been obtained as the difference of total energies obtained from SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) and (U)B3LYP-D3/6-31+G(d,p) calculations. Combination of the solvation energies with gas phase enthalpies then yield enthalpy values in water. All tautomers and conformers with a Boltzmann averaged free energy population of >2% were considered for pK_a calculations.

4.5.2 Experimental pK_a- Values

Table 4: Experimental pK_a values in water at 25 °C for modified cytosine and uracil nucleobases.

molecule	acidic site	pK_a (nucleobase)	pK_a (N-1-methylated)	pK_a (nucleoside)	pK_a (desoxy-nuc.)	avg.
C	N3	4.63[52]/4.60[53]/ 4.45[54] (avg: **4.56**)	4.55[55]		4.31[56]/4.3[57]/ 4.5[33]	**4.48**
	N1	12.25[52]/12.1[53]/ 12.2[54]				**12.2**
5mC	N3	4.6[54]	4.76[58]	4.28[58]	4.40[58]4.4[59]/ 4.7[57]	**4.52**
	N1	12.4[54]				**12.4**
5hmC	N3			4.0[33]	3.81[56]/3.55[60]/ 3.6[61]	**3.74**
5fC	N3			2.4[33]	2.62[56]	**2.51**
5caC	-COOH			4.5/4.2[33]	2.45[26]/2.0[32]	**2.24**
	N3			2.4/2.1[33]	4.28[26]/4.0[32]	**4.25**
U	N3	9.36[52]/9.45[53]/ 9.51[62]/9.5[54] (avg: **9.46**)	9.99[53]/9.77[62]/ 9.75[54]	9.17[53]/9.25[62]	9.28[56]/9.3[62]	**9.49**
	N1	13.49[52]				**13.49**
5mU	N3	9.86[52]/9.9[54]	9.94[53]	9.8[62]		**9.88**
5hmU	N3				9.33[56]	**9.33**
5fU	N3				8.12[56]	**8.12**
5caU	COOH	4.2/4.25[63]/4.16[64]			4.0[63]	**4.15**
	N3	8.90/9.20[63]/8.89[64]			9.8[63]	**9.20**
orotic acid	-COOH	~2.8[54]/2.40[65]/2.07[64]				**2.42**
	N3	9.45[54]/9.50[65]/9.45[64]				**9.47**

4.5.3 pK_a- Calculations: Direct Method

To calculate theoretical pK_a values relative to the free energy of the proton in solvation, the difference in free energy for the deprotonation reaction must be calculated by eqs. (4) and (5):

$$\mathrm{p}K_{\mathrm{AH}} = \frac{\Delta G^*_{\mathrm{sol}}}{2.303RT} \tag{4}$$

$$\Delta G^*_{\mathrm{sol}} = G_{\mathrm{deprot}} + G_{\mathrm{H^+_{(aq)}}} - G_{\mathrm{prot}} \tag{5}$$

where R is the gas constant (0.008314511 kJ/K*mol) and T is the temperature (298.15 K). The free energy of the proton in water $G_{\mathrm{H^+_{(aq)}}}$ is obtained by the following equation (6)

$$G_{\mathrm{H^+_{(aq)}}} = G_{\mathrm{H^+_{(g)}}} + \Delta G^{1\mathrm{atm}\rightarrow 1\mathrm{M}} + \Delta G_{\mathrm{H^+_{(solv)}}} = -1130.92265\ \mathrm{kJ/mol} \tag{6}$$

with the gasphase free energy, $G_{\mathrm{H^+_{(g)}}} = -26.3048\,\mathrm{kJ/mol}$, the standard state correction from atm to 1 M, $\Delta G^{1\mathrm{atm}\rightarrow 1\mathrm{M}} = 7.9108\ \mathrm{kJ/mol}$, and the free energy of solvation of a proton, $\Delta G_{\mathrm{H^+_{(solv)}}} = -1112.5256\ \mathrm{kJ/mol}$.[66] An example for the calculation of pK_a values by the proton-based thermodynamic cycle is depicted under Figure 18.

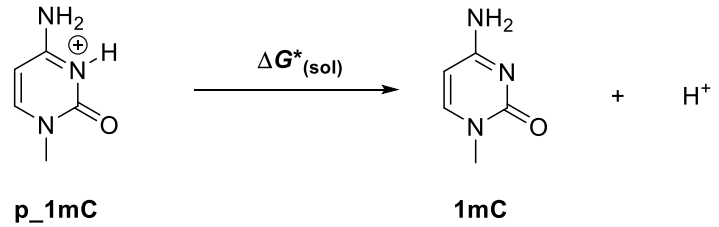


Figure 18: The reaction of deprotonation of N1-methylated cytosine described by the free energy of the deprotonation reaction in solution $\Delta G^*_{\mathrm{sol}}$ (solvent: water).

To obtain the free energy of the deprotonation reaction in solution $\Delta G^*_{\mathrm{sol}}$ for a given reaction a proton-based thermodynamic cycle is used (see Figure 19).

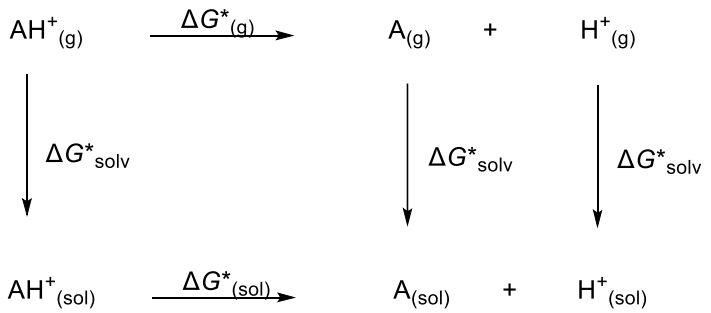


Figure 19: Proton-based thermodynamic cycle for the deprotonation reaction of AH^+.

For calculation at the SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) level of theory the free energies of solvation are corrected using quasi-harmonic approximation introduced by Cramer and Thrular with the "GoodVibes" program.[45] The original cutoff value of 100 cm^{-1} is used, to systematically scale up low energy frequencies to an consistent value.[42-45] All free energies must be corrected to the standard state by a conversion factor ($\Delta G^{1atm\rightarrow 1M}$) as mentioned above. The single point energies were calculated for the optimized geometries by using the high-precision method DLPNO-CCSD(T). 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, according to equation (7):

$$G^*_{\mathrm{sol}}(\mathrm{DLPNO}) = E^0_{\mathrm{tot}}[\mathrm{DLPNO}] + \Delta G_{\mathrm{solv,SMD}} + E_{\mathrm{cor,term}} \quad (7)$$

The total electronic energy E^0_{tot} is directly used from the ORCA output file, while the solvation free energy $\Delta G_{\mathrm{solv,SMD}}$ and thermal corrections $E_{\mathrm{cor,term}}$ are determined by eq. (8) and (9):

$$\Delta G_{\mathrm{solv,SMD}} = E^0_{\mathrm{tot}}[\mathrm{SMD/B3LYP-D3/6-31+G(d,p)}] - E^0_{\mathrm{tot}}[\mathrm{gasphase/B3LYP-D3/6-31+G(d,p)}] \quad (8)$$

$$E_{\mathrm{cor,term}} = E^0_{\mathrm{tot}}[\mathrm{SMD/B3LYP-D3/6-31+G(d,p)}] - (G_{298}\,[\mathrm{SMD/B3LYP-D3/6-31+G(d,p)} + \Delta G^{1\mathrm{atm}\rightarrow 1\mathrm{M}}) \quad (9)$$

4.5.4 pK_a- Calculations: Relative Method (Proton-Exchange Method)

In the proton exchange method, a proton is transferred between a reference molecule and the substrate. To minimize systematic errors in pK_a predictions a structurally similar reference system with a pK_a value close to the system under investigation is chosen. The reference molecule should also match the size, flexibility and structural strain of the of the molecule to determine.[21]

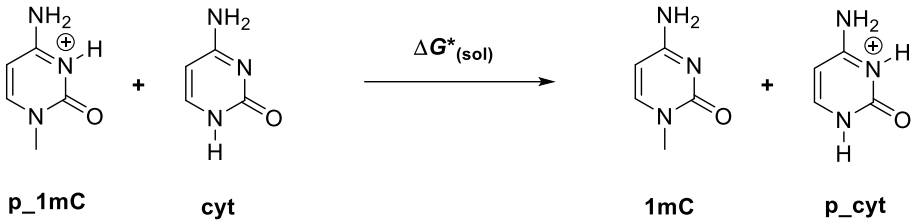


Figure 20: Deprotonation of N1-methylated cytosine (**1mC**) described by the free energy of the deprotonation reaction in solution $\Delta G^*_{\mathrm{sol}}$ (solvent: water, reference: cytosine).

In the example reaction depicted in Figure 20 cytosine is employed as reference molecule and the pK_a value is calculated according to the direct method (eq. (4)), but with a different equation to calculate the free energy for the deprotonation reaction $\Delta G^*_{\mathrm{sol}}$ leading to $\Delta \mathrm{p}K_{\mathrm{AH}}$ (eq. 11) to calculate the $\mathrm{p}K_{\mathrm{AH}}$ value from the reference $\mathrm{p}K_{\mathrm{AH}}(\mathrm{REF})$ (eq 11):

$$\Delta G^*_{\mathrm{sol}} = G_{\mathrm{A}} + G_{\mathrm{RefH+}} - G_{\mathrm{AH+}} + G_{\mathrm{Ref}} \tag{10}$$

$$\Delta \mathrm{p}K_{\mathrm{AH}} = \frac{\Delta G^*_{\mathrm{sol}}}{2.303RT} \tag{11}$$

$$\mathrm{p}K_{\mathrm{AH}} = \mathrm{p}K_{\mathrm{AH}}(\mathrm{REF}) + \Delta \mathrm{p}K_{\mathrm{AH}} \tag{11}$$

The general thermodynamic cycle for these reactions is:

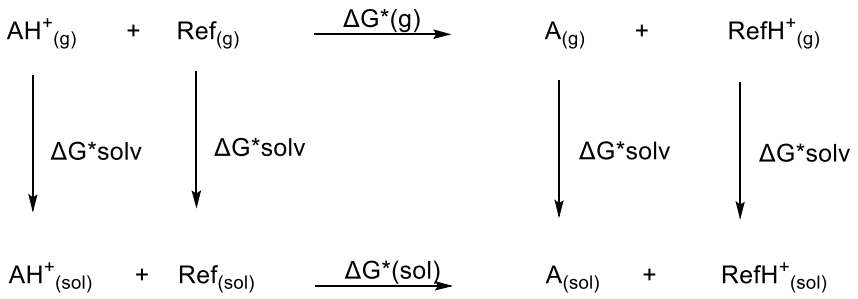


Figure 21: Thermodynamic cycle for the deprotonation reaction of AH+ with cytosine as reference molecule.

In this study we use cytosine and uracil as reference molecules due to their well-known experimental pK_a values (see Table 3) as well as structural similarity towards the investigated systems. An averaged

pK_a value was used for the reference, e.g. for cytosine (pK_a(AH^+) = 4.56; pK_a(A) = 12.20) and uracil (pK_a(A) = 9.46).

4.5.5 Explicit Water and the "kick" Procedure

To place explicit water molecules statistically around a central fragment (here the nucleobase) and search the conformational space for the best water-nucleobase conformer, we use Saunder's stochastic conformational search algorithm called "kick".[23] This algorithm was modified and implemented in a web version by D. Šakić.[22] This stochastic search procedure can locate the most stable water-complex by generating a high number of starting geometries for water-nucleobase complexes which are then optimized to stationary point using the SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) level of theory.

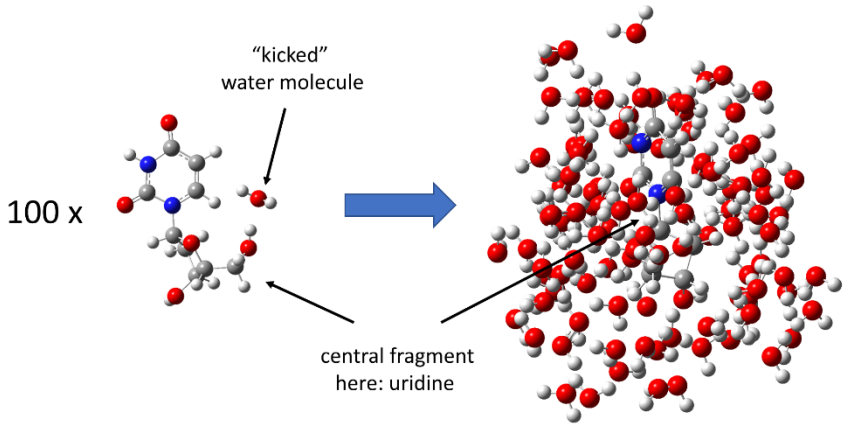


Figure 22: Example for "kick" procedure: one water was kicked around a central fragment 100 times and the individual water coordinates (before optimization) to visualize their spatial distribution.

Figure 22 visualizes how the conformational space of water-complexes is covered by statistically placing 100 water molecules around a central fragment (here: uridine) before optimization. After geometry optimization a Python script was used to extract thermodynamic data from Gaussian .log-files as well as free energies corrected by quasi-harmonic oscillator approximation with the "GoodVibes" program. The frequency cut-off for both entropy and enthalpy is used at its default value of 100 cm^{-1} as well as 30 cm^{-1}. All water-complexes were sorted by their quasi-harmonic corrected free energy G_{qh} and compared to the next lowest energy conformer by a geometric centroid analysis to exclude duplicate structures. As a last step, an energy cut-off of 0.026255 kJ/mol was used to also exclude mirror images. The best ten optimized structures were used for more expensive DLPNO calculations. The overall procedure is described under Figure 23 as a flowchart. These energies were used to calculate Boltzmann weighted dissociation constants with one explicit water.

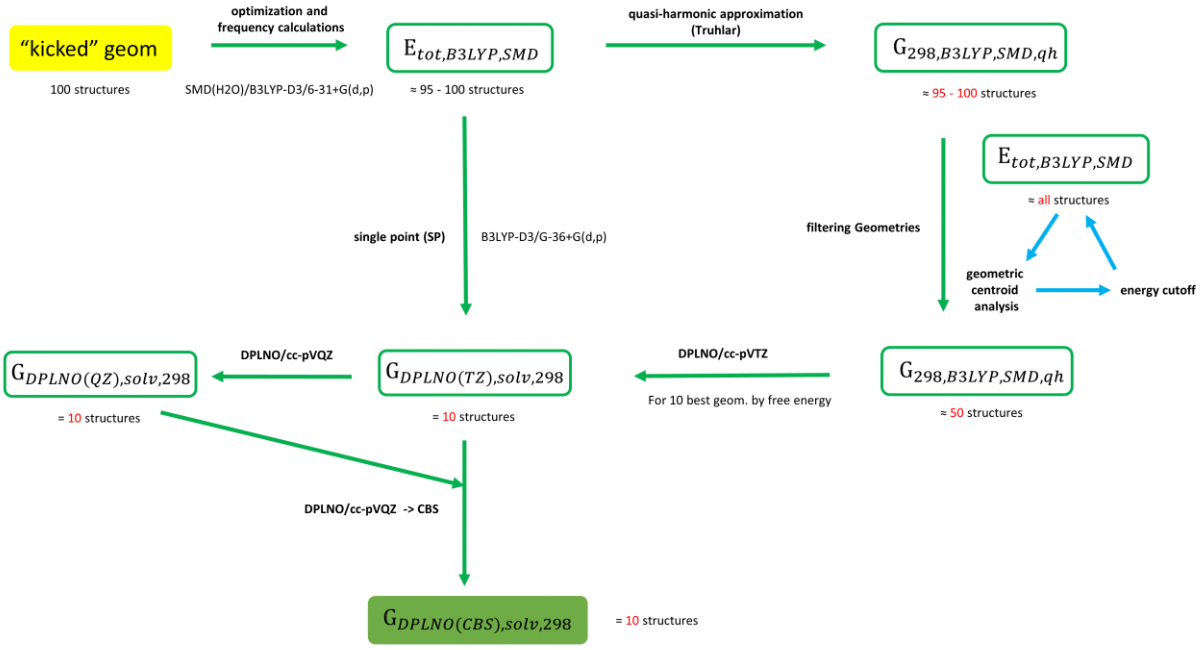


Figure 23: Flowchart for the search of most stable water-complexed stationary points for all relevant tautomers and conformers.

4.5.6 Generating Molecule-Water Clusters

The approximately one hundred "kicked" conformers with one explicit water were used to construct molecule-water cluster of up to nine explicit water molecules.

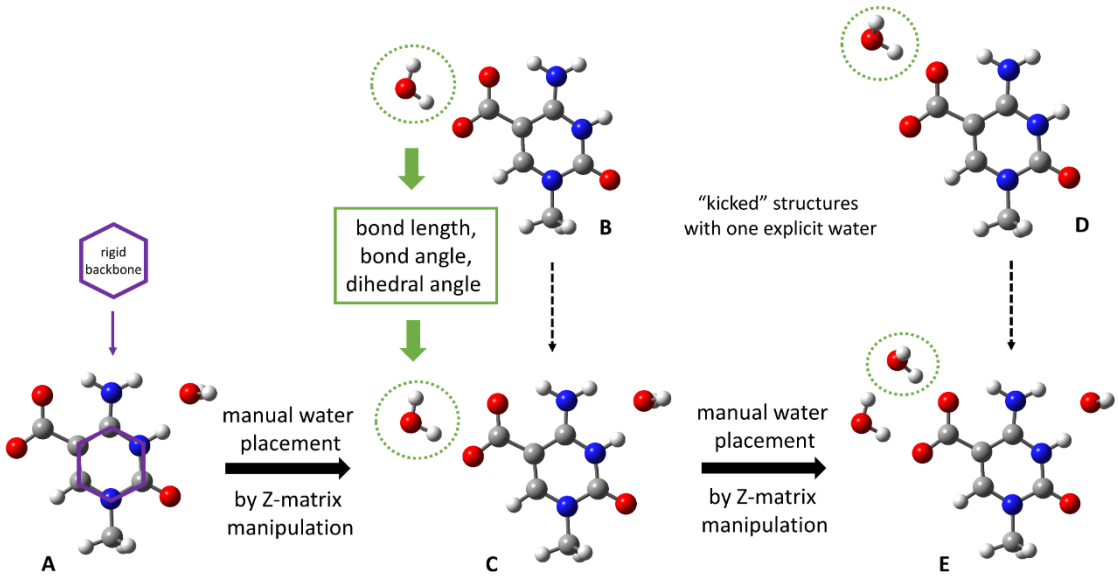


Figure 24: Basic steps for the construction of explicit molecule-water clusters by manual water placement using z-matrix manipulation from preoptimized structures with one explicit water generated by the "kick" procedure.

In Figure 24 a representation of the workflow for the zwitterionic form of neutral **1m5caC** is depicted. The "kicked" conformers were sorted by total energy E_{tot} and the structure of the lowest energy conformer was selected (Figure 24, **A**). Then the next structure (Figure 24, **B)** was selected and placed around structure **A** by z-matrix manipulation with the "GaussView 5" program.[46] By using the rigid backbone to define the common z-matrix that all nucleobases inherit, an exact placement around the central fragment can be achieved. The resulting structure with two explicit water molecules **C** was then reoptimized to account for water-water interactions. Then the same procedure was repeated with structure **D** to achieve a cluster with three explicit water **E**. This procedure can be repeated indefinitely, but is restricted by hydrogen bond site crowding, as well as increased interactions between water molecules leading to water clustering.

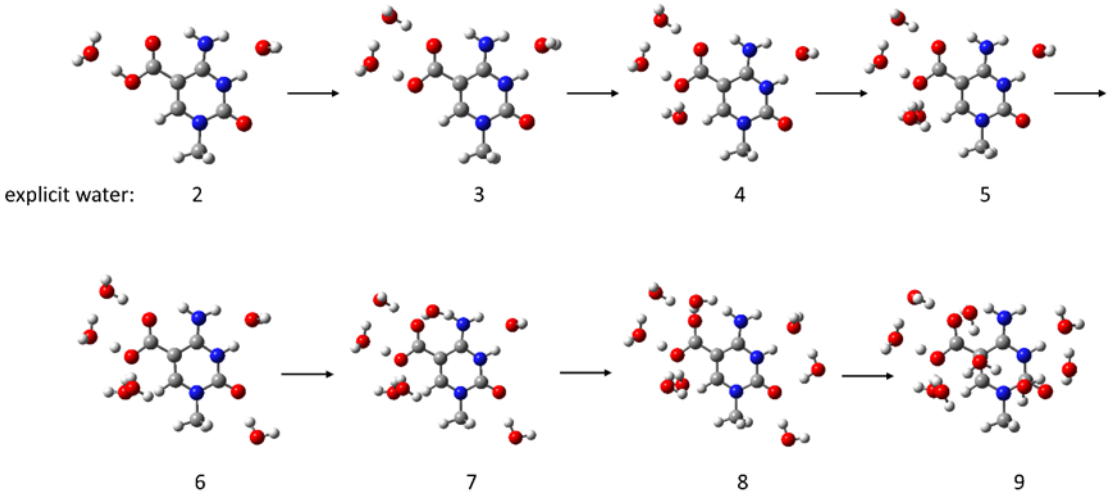


Figure 25: Example for the construction of a water-molecule cluster (here: protonated **1m5caC**) of up to nine explicit water molecules generated by manual placement of water with z-matrix manipulation.

In Figure 25 after the addition of seven explicit water, distinct water-water cluster are being formed leading to a decreased and/or distorted hydrogen bond site coverage. After reoptimization of neutral structures, the protonation (or deprotonation) is facilitated by the closest hydrogen bonding, explicit water to the acidic (or basic) site (Figure 26).[17]

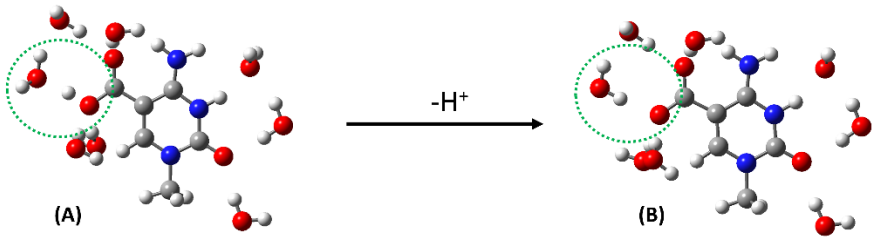


Figure 26: Example water-substrate cluster for **A** protonated **1m5caC** and **B** neutral **1m5caC** (zwitterionic tautomer) with eight explicit water molecules.

4.5.7 Investigations into the Quality of pK_a by Comparing 1m5caC and 1m5hmC

An example why for the deprotonation of neutral **1m5caC** three instead of one explicit water were deemed reasonable, is shown in Figure 27.

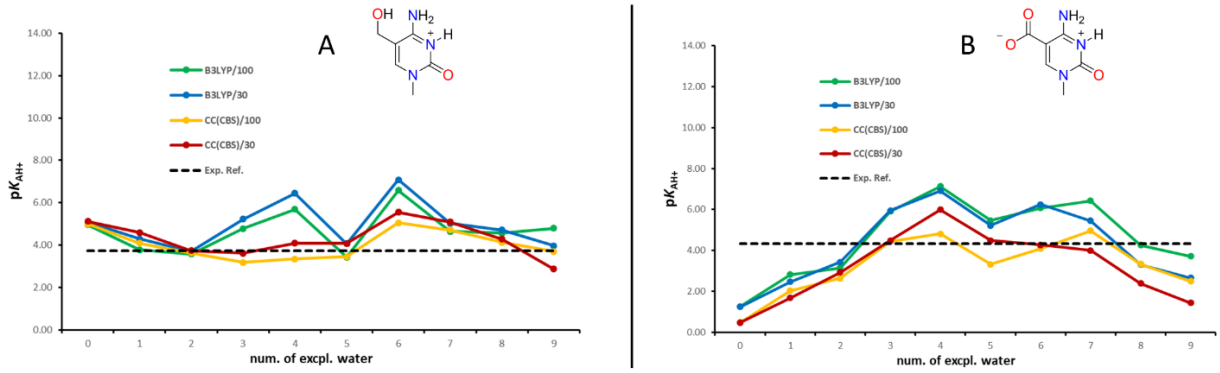


Figure 27: Relative pK_a values for **1mhmC** (**A**) and zwitterionic **1m5caC** (**B**) with increased numbers of explicit water molecules is shown as well as its experimental reference (black dotted line).

For **1m5hmC** (Figure 27, **A**) the error of the theoretical pK_a value is decreasing rapidly with the addition of one and two explicit solvent molecules. The addition of further explicit water molecules keeps the pK_a value somewhat stable but increased size of water cluster lead to a higher error and less predictable results. In case of the zwitterionic, neutral form of **1m5caC** (Figure 27, **B**) a minimum two water molecules is needed to cover the important hydrogen bonding sites due to a higher number of potential bonding sites available. The calculates pK_a therefore stabilizes around three water molecules to then have a higher spread due to bigger sized water clusters. In the following Figure 28 the calculations for **1m5dhmC** are shown as an example for a system with unknown experimental pK_a value. The pK_a calculations are stable within up to 4 explicit solvent molecules, suggesting a good quality of the resulting theoretical pK_a value.

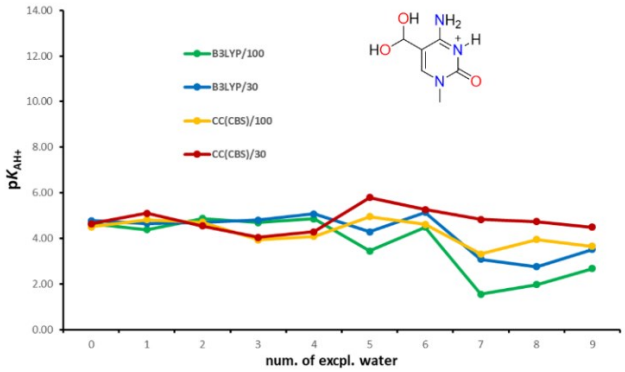


Figure 28: Relative calculated pK_a values for **1m5dhmC** with increased numbers of explicit water molecules.

4.5.8 Steady-State Simulation of 1m5caC Equilibrium with CoPaSi

With the help of the CoPaSi program[34] a steady-state analysis of the **1m5caC** system at different pH values was performed. Therefore, several chemical equilibria were defined (see Figure 13):

1. 1m5caC-H^+ (**A**) + H_2O $\overset{K_1}{\rightleftharpoons}$ 1m5caC (**B**) + H_3O^+
2. 1m5caC-H^+ (**A**) + H_2O $\overset{K_2}{\rightleftharpoons}$ 1m5caC_zw (**C**) + H_3O^+
3. 1m5caC (**B**) + H_2O $\overset{K_3}{\rightleftharpoons}$ 1m5caC$^-$ (**D**) + H_3O^+
4. 1m5caC_zw (**C**) + H_2O $\overset{K_4}{\rightleftharpoons}$ 1m5caC$^-$ (**D**) + H_3O^+
5. $2H_2O$ $\overset{K_w}{\rightleftharpoons}$ H_3O^+ + OH^-
6. 1m5caC (**B**) $\overset{K_{eq}}{\rightleftharpoons}$ 1m5caC_zw (**C**)

The equilibrium constants K_1, K_2, K_3 and K_4, are calculated from the intrinsic pK_a values according to Table 2 by following equation:

$$\mathrm{p}K_\mathrm{a} = -\log(K_{1-4} \times [\mathrm{H_2O}])$$

The reaction rates for the dissociation k_D and its back-reaction k_R can be defined as follows (with $k_R = 1$):

$$K_{1-4} = \frac{k_\mathrm{D}}{k_\mathrm{R}}$$

The self-dissociation in water K_w and is defined as:

$$K_\mathrm{w} = \frac{k_\mathrm{D}}{k_\mathrm{R}} = \frac{[\mathrm{H_3O^+}][\mathrm{OH^-}]}{[\mathrm{H_2O}][\mathrm{H_2O}]} = 10^{-14}$$

The chemical equilibrium between the neutral and the zwitterionic **1m5caC** species can be calculated by its free energy of reaction:

$$\Delta G = -R\mathrm{T}\ \ln K_\mathrm{eq}$$

To simulate the steady-state of this equilibria a total substrate concentration of 1.6×10^{-3} mol l^{-1} was used with varying concentrations of H_3O^+ and OH^--ions. The simulation of the equilibrated system was only used if fully converged. The resulting concentrations are shown in Table 5 as well as a simulated titration curve (see Figure 16, **B**) using the assumption that protonated and deprotonated carboxy groups have the same ^{13}C chemical shifts in water (166.1 and 171.4 ppm). Also, Table 5 shows in its last column a simulated titration curve with the theoretical determined isotropic shielding values σ_{iso} of the carbonyl C-atom for all four species defined for the CoPaSi simulation (also see Figure 13, simulated titration curve: see Figure 16, **C**). The calculated isotropic shielding values σ_{iso} were obtained according to paragraph 3.6:

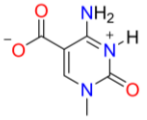
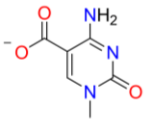
structure: **(as in Fig. 12)**	 1m5caC-H⁺ (A)	 1m5caC (B)	 1m5caC_zw (C)	 1m5caC⁻ (D)
isotropic shielding σ_{iso}(carboxyl-C):	-1.6572	-3.6081	-4.5353	-8.2601

Table 5: Steady-state concentrations for the simulation of the 1m5caC system by CoPaSi program[34] used to generate Figures 15 and 16.

pH	c(1m5caC) [mol/l]	c(1m5caC-H⁺) [mol/l]	c(5caC⁻) [mol/l]	c(1m5caC_zw) [mol/l]	c(H_3O^+) [mol/l]	c(OH^-) [mol/l]	simulatedchemical shift[a,b] [ppm]	absolute simulated shielding change[c]
0.01	6.57E-06	1.59E-03	3.22E-11	1.54E-06	9.83E-01	1.06E-14	166.11	1.668
0.05	7.18E-06	1.59E-03	3.86E-11	1.69E-06	9.01E-01	1.28E-14	166.11	1.669
0.16	9.31E-06	1.59E-03	6.50E-11	2.20E-06	6.98E-01	1.33E-14	166.11	1.673
0.40	1.64E-05	1.58E-03	2.03E-10	3.89E-06	3.98E-01	2.31E-14	166.11	1.684
0.70	3.27E-05	1.56E-03	8.18E-10	7.80E-06	1.98E-01	5.09E-14	166.13	1.711
1.05	6.92E-05	1.51E-03	3.78E-09	1.66E-05	8.85E-02	1.13E-13	166.15	1.771
1.31	1.21E-04	1.45E-03	1.20E-08	2.90E-05	4.85E-02	8.56E-14	166.20	1.857
1.63	2.26E-04	1.32E-03	4.62E-08	5.41E-05	2.37E-02	4.72E-13	166.28	2.030
2.05	4.63E-04	1.03E-03	2.50E-07	1.11E-04	8.97E-03	1.13E-12	166.47	2.423
2.37	6.97E-04	7.34E-04	7.92E-07	1.67E-04	4.27E-03	2.44E-12	166.66	2.812
2.37	0.0006965	7.36E-04	7.89E-07	0.000167	4.27E-03	2.43E-12	166.66	2.810
2.69	9.16E-04	4.61E-04	2.18E-06	2.20E-04	2.04E-03	5.31E-12	166.84	3.179
3.10	1.11E-03	2.17E-04	6.81E-06	2.66E-04	7.90E-04	1.46E-11	167.01	3.518
3.15	1.12E-03	1.98E-04	7.65E-06	2.70E-04	7.10E-04	1.62E-11	167.02	3.546
3.40	1.18E-03	1.16E-04	1.44E-05	2.84E-04	3.98E-04	3.01E-11	167.09	3.673
3.94	1.22E-03	3.50E-05	5.10E-05	2.93E-04	1.16E-04	1.08E-10	167.24	3.884
4.17	0.0012041	2.00E-05	8.69E-05	2.89E-04	6.69E-05	1.87E-10	167.35	4.004
4.20	0.0012011	1.88E-05	9.19E-05	0.000288	6.31E-05	1.99E-10	167.36	4.019
4.42	0.0011617	1.09E-05	0.000149	0.000279	3.78E-05	3.34E-10	167.52	4.189
4.64	1.10E-03	6.31E-06	2.29E-04	2.64E-04	2.32E-05	5.46E-10	167.73	4.421
4.95	9.55E-04	2.63E-06	4.14E-04	2.29E-04	1.11E-05	1.14E-09	168.23	4.941
5.07	8.85E-04	1.87E-06	5.00E-04	2.12E-04	8.54E-06	1.48E-09	168.46	5.184
5.20	7.99E-04	1.26E-06	6.08E-04	1.92E-04	6.35E-06	2.00E-09	168.75	5.484
5.36	6.81E-04	7.34E-07	7.55E-04	1.63E-04	4.36E-06	2.92E-09	169.14	5.897
5.41	6.41E-04	6.11E-07	8.04E-04	1.54E-04	3.85E-06	3.30E-09	169.27	6.035
5.52	5.61E-04	4.17E-07	9.03E-04	1.35E-04	3.00E-06	4.23E-09	169.54	6.312
5.62	4.90E-04	2.89E-07	9.93E-04	1.17E-04	2.38E-06	5.34E-09	169.78	6.562
5.63	0.0004816	2.77E-07	0.001003	0.000116	2.32E-06	5.48E-09	169.80	6.590
6.14	2.01E-04	3.58E-08	1.35E-03	4.82E-05	7.19E-07	1.77E-08	170.73	7.563
6.90	4.03E-05	1.25E-09	1.55E-03	9.67E-06	1.26E-07	1.01E-07	171.27	8.120
7.06	2.83E-05	6.11E-10	1.56E-03	6.79E-06	8.73E-08	1.46E-07	171.31	8.162
7.21	2.03E-05	3.12E-10	1.57E-03	4.87E-06	6.22E-08	2.05E-07	171.33	8.190
8.25	1.84E-06	2.53E-12	1.60E-03	4.42E-07	5.57E-09	2.29E-06	171.39	8.254
9.60	8.41E-08	6.03E-15	1.60E-03	2.02E-08	2.53E-10	5.01E-05	171.40	8.260

10.08	2.81E-08	1.64E-15	1.60E-03	6.74E-09	8.40E-11	1.50E-04	171.40	8.260
10.51	1.05E-08	1.07E-16	1.60E-03	2.53E-09	3.10E-11	4.00E-04	171.40	8.260
11.48	1.24E-09	2.18E-16	1.60E-03	2.99E-10	3.32E-12	3.40E-03	171.40	8.260
11.89	5.00E-10	-1.68E-15	1.60E-03	1.22E-10	1.28E-12	8.40E-03	171.40	8.260
12.25	1.47E-10	1.06E-15	1.60E-03	3.75E-11	5.63E-13	2.84E-02	171.40	8.260
12.58	1.08E-10	-7.94E-16	1.60E-03	2.82E-11	2.65E-13	3.84E-02	171.40	8.260

a: reference caboxyl ^{13}C shifts for species taken from Tokmakoff *et al.* with assumption that **1m5caC_zw** has same shift as **1m5caC**$^{-}$ and **1m5caC** same as **1m5caC-H**$^{+}$; b: calculated from percentage mol fraction of steady-state analysis at respective pH value; c: calculated from percentage mol fraction of steady-state analysis at respective pH value and theoretically determined isotropic shielding values for all species in Figure 12 at the B3LYP/pcS-3 level of theory.

4.5.9 pK_a-Values from Relative and Absolute Methods

In the following section the overall results of relative pK_a calculations with different numbers of explicit water are given for the cytosine and uracil systems with known experimental dissociation constants stated in the last column if available.

Table 6: Relative and absolute pK_a calculations for 1-N-metylcytosine (**1mC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1mC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	4.97		4.64		2.99		2.55		4.48
0	5.10	5.10	4.92	4.92	4.71	4.71	3.73	3.73	4.48
1	5.09	5.44	4.24	4.58	5.40	5.35	3.89	3.83	4.48
2	4.61	4.33	4.14	3.87	4.43	4.04	3.95	3.57	4.48
3	3.38	3.56	2.42	2.60	3.36	2.99	3.06	2.69	4.48
4	4.54	5.02	3.17	3.65	3.71	3.60	3.20	3.09	4.48
5	1.33	1.58	2.83	3.09	2.55	1.80	2.69	1.95	4.48
6	3.71	4.37	3.93	4.59	3.60	3.66	4.00	4.06	4.48
7	1.58	1.94	3.40	3.76	3.25	2.25	3.72	2.73	4.48
8	2.86	3.90	3.81	4.84	3.89	3.68	3.75	3.54	4.48
9	4.31	4.24	3.69	3.62	4.41	3.61	3.89	3.08	4.48

[a]: see Figure 20;[b]: see Figure 18;[c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

The difference between the methods with “1” explicit water and “1:boltz” in column one in Table 5 is the number of conformers to determine the dissociation constants. While “1” means that only the best conformer was taken into consideration (as for all other molecule-water clusters in column one), “1:boltz” means that the ten best structures by E_{tot} where Boltzmann weighted to result in a Boltzmann averaged free energy. This method does not yield better results for the theoretical determination of pK_a values by the relative method, that can be seen in Figure 9, paragraph 4.3.2. It does, however, slightly improve the pK_a calculations for the absolute method (see Figure 8).

Table 7: Relative and absolute pK_a calculations for 1,5-dimethylcytosine (**1m5mC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1m5mC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	5.74		5.31		3.77		3.22		4.52
0	5.97	6.11	5.72	5.86	5.58	5.71	4.53	4.67	4.52
1	5.56	5.65	5.46	5.55	5.87	5.56	5.11	4.80	4.52
2	6.11	5.94	5.37	5.20	5.93	5.65	5.18	4.90	4.52
3	5.25	5.88	3.95	4.58	5.23	5.30	4.59	4.67	4.52
4	5.62	5.20	3.17	2.75	4.79	3.78	3.19	2.19	4.52
5	3.42	4.08	4.79	5.46	4.64	4.30	4.65	4.32	4.52
6	4.86	5.31	4.49	4.95	4.75	4.61	4.56	4.42	4.52
7	2.80	4.07	4.60	5.88	4.46	4.39	4.93	4.85	4.52
8	3.65	5.06	5.27	6.68	4.68	4.84	5.21	5.38	4.52
9	4.58	4.82	4.47	4.71	4.68	4.18	4.67	4.17	4.52

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 8: Relative and absolute pK_a calculations for 1-N-metyl-5-hydroxymethylcytosine (**1m5hmC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1m5hmC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	5.49		5.40		3.51		3.31		3.74
0	4.95	5.08	4.99	5.12	4.56	4.69	3.80	3.93	3.74
1	3.78	4.28	4.08	4.59	4.09	4.20	3.73	3.84	3.74
2	3.57	3.69	3.62	3.74	3.39	3.41	3.43	3.44	3.74
3	4.78	5.22	3.18	3.62	4.76	4.65	3.82	3.71	3.74
4	5.69	6.44	3.34	4.09	4.86	5.02	3.36	3.53	3.74
5	3.43	4.06	3.45	4.08	4.65	4.28	3.31	2.94	3.74
6	6.58	7.08	5.05	5.55	6.47	6.37	5.12	5.02	3.74
7	4.64	5.03	4.71	5.09	6.31	5.34	5.03	4.07	3.74
8	4.56	4.71	4.13	4.28	5.58	4.50	4.07	2.98	3.74
9	4.79	3.97	3.70	2.87	4.89	3.33	3.89	2.33	3.74

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 9: Relative and absolute pK_a calculations for 1-N-metyl-5-formylcytosine (**1m5fC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**$pK_{a(AH+)}$**								
1m5fC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	3.52		3.76		1.55		1.67		2.51
0	2.58	2.52	3.42	3.36	2.18	2.12	2.22	2.17	2.51
1	2.98	3.18	2.96	3.15	3.30	3.09	2.61	2.40	2.51
2	2.44	2.37	2.70	2.64	2.26	2.08	2.51	2.33	2.51
3	2.49	3.25	2.12	2.88	2.46	2.67	2.76	2.96	2.51
4	2.79	3.51	2.26	2.98	1.96	2.09	2.29	2.42	2.51
5	0.84	1.40	2.45	3.01	2.05	1.62	2.31	1.88	2.51
6	2.02	2.53	3.23	3.73	1.92	1.82	3.30	3.20	2.51
7	-1.18	0.46	0.90	2.53	0.48	0.77	1.22	1.51	2.51
8	1.50	1.63	2.51	2.64	2.53	1.42	2.45	1.34	2.51
9	5.58	5.78	4.04	4.24	5.68	5.14	4.23	3.69	2.51

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 10: Relative and absolute calculations of intrinsic p$K_{a\ OH+}$ for protonated 1-N-metyl-5-carboxycytosine (**1m5caC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**$pK_{a(AH+)}$**								
1m5caC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	2.67		3.79		0.69		1.70		2.24
0	2.31	2.31	5.25	5.26	1.91	1.92	4.06	4.07	2.24
1	2.11	2.55	3.50	3.95	2.42	2.46	3.15	3.20	2.24
2	4.10	4.48	3.77	4.15	3.92	4.20	3.57	3.85	2.24
3	3.26	3.55	2.07	2.35	3.24	2.97	2.71	2.44	2.24
4	3.34	3.61	2.48	2.74	2.51	2.19	2.50	2.18	2.24
5	0.63	0.24	1.98	1.58	1.85	0.46	1.83	0.44	2.24
6	2.75	2.55	1.46	1.26	2.65	1.85	1.52	0.72	2.24
7	0.13	0.65	0.38	0.90	1.79	0.96	0.71	-0.12	2.24
8	-0.47	-0.51	-0.11	-0.14	0.55	-0.72	-0.17	-1.44	2.24
9	1.01	1.87	0.54	1.40	1.12	1.23	0.74	0.86	2.24

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 11: Relative and absolute calculations of intrinsic $pK_{a\ NH+}$ for protonated 1-N-metyl-5-carboxycytosine (**1m5caC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**$pK_{a(AH+)}$**								
1m5caC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									2.24
0	2.87	2.83	3.41	3.37	2.48	2.43	2.22	2.18	2.24
1	2.67	2.22	2.89	3.25	2.99	2.95	2.54	2.50	2.24
2	4.70	4.07	3.65	4.02	4.52	4.77	3.46	3.72	2.24
3	3.14	1.56	2.41	2.42	3.12	2.59	3.05	2.51	2.24
4	3.60	2.00	3.07	3.25	2.77	2.36	3.10	2.69	2.24
5	2.06	-0.40	3.27	3.13	3.28	2.14	3.12	1.99	2.24
6	2.95	-0.19	2.64	2.20	2.85	1.80	2.71	1.67	2.24
7	1.17	-2.44	2.75	2.65	2.84	1.39	3.08	1.63	2.24
8	2.06	-0.94	3.58	4.04	3.09	2.30	3.52	2.74	2.24
9	1.93	-1.93	3.58	4.27	2.03	1.98	3.78	3.73	2.24

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 12: Relative and absolute calculations of intrinsic $pK_{a\ OH}$ for neutral 1-N-metyl-5-carboxycytosine (**1m5caC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**$pK_{a(AH+)}$**								
1m5caC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									4.25
0	0.69	0.73	0.94	0.98	5.40	5.43	6.48	6.52	4.25
1	1.97	1.63	2.18	1.84	5.61	5.45	7.03	6.88	4.25
2	2.27	1.97	2.89	2.58	5.34	4.87	6.95	6.48	4.25
3	5.08	5.01	5.17	5.10	5.28	5.04	6.99	6.75	4.25
4	6.79	6.83	5.36	6.79	5.35	5.39	7.13	7.17	4.25
5	5.13	5.00	3.72	4.99	5.06	4.77	6.90	6.61	4.25
6	5.85	6.08	4.07	4.29	4.83	4.53	6.43	6.14	4.25
7	5.46	4.81	3.76	3.11	3.76	3.79	5.08	5.11	4.25
8	3.20	2.28	1.91	1.00	2.56	2.25	3.91	3.60	4.25
9	2.76	1.74	0.81	-0.21	2.38	2.08	3.76	3.47	4.25

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 13: Relative and absolute calculations of intrinsic p$K_{a\ NH}$ for neutral, zwitterionic 1-N-metyl-5-carboxycytosine (**1m5caC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1m5caC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									4.25
0	1.26	1.24	-0.89	-0.91	5.96	5.95	4.65	4.63	4.25
1	2.54	2.12	1.57	1.15	6.18	5.94	6.42	6.18	4.25
2	2.87	2.54	2.78	2.45	5.93	5.45	6.84	6.35	4.25
3	4.96	4.62	5.51	5.17	5.16	4.65	7.32	6.82	4.25
4	7.04	7.00	5.95	7.30	5.61	5.56	7.72	7.68	4.25
5	6.55	6.68	5.01	6.53	6.49	6.45	8.19	8.15	4.25
6	6.05	6.04	5.25	5.24	5.03	4.49	7.62	7.08	4.25
7	6.50	5.23	6.13	4.86	4.80	4.21	7.45	6.86	4.25
8	5.73	5.31	5.60	5.17	5.09	5.27	7.60	7.78	4.25
9	3.67	2.49	3.85	2.66	3.29	2.83	6.80	6.34	4.25

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 14: Relative and absolute pK_a calculations for 1-N-metyl-5-dihydroxymethylcytosine (**1m5dhmC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1m5dhmC	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									
0	4.64	4.78	4.50	4.64	4.25	4.39	3.31	3.45	
1	4.38	4.65	4.82	5.09	4.69	4.56	4.47	4.35	
2	4.86	4.70	4.70	4.54	4.68	4.41	4.51	4.24	
3	4.69	4.81	3.93	4.04	4.67	4.23	4.57	4.13	
4	4.86	5.07	4.09	4.29	4.03	3.65	4.12	3.73	
5	3.44	4.29	4.95	5.79	4.66	4.51	4.81	4.65	
6	4.49	5.14	4.61	5.26	4.38	4.43	4.68	4.73	
7	1.55	3.08	3.31	4.83	3.22	3.39	3.63	3.81	
8	1.98	2.76	3.95	4.73	3.00	2.55	3.88	3.43	
9	2.68	3.52	3.65	4.49	2.78	2.88	3.84	3.94	

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 15: Relative and absolute calculations of intrinsic p$K_{a\ OH+}$ for protonated 1-N-metyl-5-dihydroxymethylcytosine (**1m5dhmC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1m5dhmC	**Relative Method[a,f]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	14.34	14.42	16.95	17.04	13.94	15.76	14.03	15.84	
0	13.77	14.57	16.30	17.10	14.08	15.95	14.48	16.35	
1	14.50	14.67	16.57	16.74	14.32	16.38	14.38	16.44	
2	12.10	12.26	13.30	13.46	12.08	13.94	11.68	13.55	
3	11.48	11.60	12.50	12.62	10.65	12.52	10.18	12.06	
4	9.55	10.69	12.84	13.97	10.77	12.70	10.91	12.83	
5	12.51	13.54	14.00	15.04	12.40	14.07	12.84	14.50	
6	10.18	11.74	13.18	14.74	11.84	13.50	12.05	13.71	
7	10.30	11.34	13.29	14.33	11.33	13.23	11.13	13.03	
8	10.76	11.48	12.65	13.36	10.86	12.84	10.84	12.82	
9	14.34	14.42	16.95	17.04	13.94	15.76	14.03	15.84	

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}); [f]: reference cytosine.

Table 16: Relative and absolute calculations of intrinsic p$K_{a\ OH}$ for neutral 1-N-metyl-5-dihydroxymethylcytosine (**1m5dhmC**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH+)}$**								
1m5dhmC	**Relative Method[a,f]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	12.61	12.83	15.84	16.05	17.31	17.53	21.38	21.59	
0	12.77	12.63	14.39	14.25	16.41	16.45	19.24	19.28	
1	13.74	14.14	15.65	16.05	16.81	17.05	19.71	19.95	
2	13.72	13.54	14.72	14.54	13.91	13.57	16.53	16.19	
3	13.69	13.44	13.18	14.31	12.26	12.00	14.95	14.69	
4	11.55	11.75	11.23	12.82	11.48	11.52	14.40	14.45	
5	14.70	15.39	13.56	14.25	13.67	13.85	15.92	16.10	
6	15.06	13.92	14.36	13.22	13.36	12.90	15.67	15.21	
7	14.81	15.14	14.85	15.19	14.16	15.11	16.85	17.80	
8	13.30	12.13	12.59	11.41	12.92	12.47	15.54	15.09	
9	12.61	12.83	15.84	16.05	17.31	17.53	21.38	21.59	

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}); [f]: reference cytosine.

Table 17: Relative and absolute pK_a calculations for 1-methyluracil (**1mU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1mU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	10.08		9.82		11.47		12.09		9.49
0	10.13	10.08	9.85	9.81	14.06	14.01	14.06	14.02	9.49
1	10.08	9.88	9.78	9.58	12.84	12.24	13.74	13.14	9.49
2	10.29	10.36	10.06	10.13	12.08	11.90	13.28	13.10	9.49
3	10.60	11.08	10.50	10.97	10.83	10.68	12.58	12.43	9.49
4	10.09	9.94	9.09	8.94	10.00	9.50	11.86	11.36	9.49
5	8.49	8.20	8.12	7.84	9.30	8.93	11.36	10.99	9.49
6	8.44	8.33	8.60	8.50	9.65	9.48	11.02	10.85	9.49
7	8.34	8.22	8.71	8.59	9.19	8.90	10.84	10.55	9.49
8	8.41	8.65	8.74	8.98	8.61	8.19	10.59	10.16	9.49

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 18: Relative and absolute pK_a calculations for 1,5-dimethyluracil (**1m5mU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1m5mU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	10.70		10.41		12.09		12.69		9.88
0	10.78	10.75	10.20	10.16	14.71	14.68	14.41	14.37	9.88
1	11.27	11.61	10.02	10.37	14.03	13.98	13.98	13.93	9.88
2	11.00	10.94	10.37	10.32	12.78	12.47	13.59	13.28	9.88
3	11.15	11.27	11.21	11.33	11.38	10.87	13.30	12.78	9.88
4	10.54	10.61	9.76	9.84	10.45	10.17	12.54	12.26	9.88
5	8.74	8.24	8.82	8.32	9.55	8.96	12.06	11.47	9.88
6	8.89	8.67	8.59	8.37	10.10	9.82	11.01	10.72	9.88
7	9.34	9.46	9.29	9.42	10.19	10.14	11.43	11.38	9.88
8	9.59	9.96	9.44	9.81	9.79	9.49	11.29	10.99	9.88
9	10.88	10.27	10.09	9.48	9.64	9.57	11.32	11.25	9.88

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 19: Relative and absolute pK_a calculations for 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1m5hmU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	9.51		9.46		10.90		11.73		9.33
0	9.95	9.90	9.75	9.70	13.88	13.83	13.96	13.91	9.33
1	10.44	10.81	9.53	9.89	13.21	13.17	13.48	13.45	9.33
2	11.39	11.63	10.37	10.61	13.18	13.16	13.59	13.57	9.33
3	12.97	13.62	11.56	12.21	13.20	13.22	13.64	13.67	9.33
4	12.18	12.07	10.70	10.60	12.09	11.63	13.48	13.02	9.33
5	10.03	10.06	9.99	10.02	10.85	10.79	13.23	13.17	9.33
6	9.16	9.15	10.46	10.45	10.37	10.29	12.88	12.80	9.33
7	9.16	9.89	10.78	11.52	10.01	10.57	12.92	13.47	9.33
8	8.97	9.54	9.99	10.57	9.17	9.08	11.84	11.75	9.33
9	12.36	12.10	11.47	11.22	11.12	11.41	12.70	12.99	9.33

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 20: Relative and absolute pK_a calculations for 1-N-methyl-5-formyluracil (**1m5fU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1m5fU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]	7.60		8.06		8.99		10.34		
0	7.44	7.35	7.86	7.77	11.37	11.29	12.07	11.98	8.12
1	7.62	7.60	8.01	7.99	10.38	9.97	11.96	11.55	8.12
2	8.09	8.20	8.38	8.48	9.88	9.73	11.60	11.45	8.12
3	8.69	9.17	8.67	9.15	8.92	8.77	10.76	10.60	8.12
4	8.59	8.45	7.86	7.71	8.51	8.01	10.63	10.14	8.12
5	7.10	6.82	7.15	6.87	7.91	7.55	10.39	10.02	8.12
6	7.15	7.17	8.02	8.03	8.37	8.31	10.43	10.38	8.12
7	6.78	6.88	7.71	7.81	7.64	7.56	9.84	9.77	8.12
8	7.27	8.22	8.09	9.04	7.47	7.76	9.94	10.23	8.12
9	8.44	7.13	8.54	7.23	7.20	6.43	9.77	9.01	8.12

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 21: Relative and absolute pK_a calculations for 1-N-methyl-5carboxyuracil (**1m5caU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1m5caU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									
0	4.39	4.44	5.69	5.74	8.33	8.38	9.90	9.95	4.15
1	4.43	4.02	4.94	4.53	7.20	6.38	8.90	8.08	4.15
2	6.19	6.18	6.25	6.24	7.98	7.72	9.47	9.21	4.15
3	6.11	6.59	6.14	6.62	6.34	6.19	8.22	8.08	4.15
4	5.70	5.98	4.81	5.09	5.61	5.54	7.58	7.52	4.15
5	4.22	4.40	3.98	4.16	5.04	5.13	7.22	7.31	4.15
6	3.45	3.52	4.48	4.56	4.66	4.67	6.90	6.91	4.15
7	3.60	2.88	4.94	4.23	4.45	3.56	7.07	6.19	4.15
8	3.33	4.22	4.69	5.58	3.54	3.76	6.54	6.77	4.15
9	3.67	2.86	4.98	4.18	2.43	2.17	6.21	5.95	4.15

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 22: Relative and absolute pK_a calculations for 1-N-metyl-5-dihydroxymtehyluracil (**1m5dhmU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1m5dhmU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									
0	4.64	4.78	4.50	4.64	4.25	4.39	3.31	3.45	
1	4.38	4.65	4.82	5.09	4.69	4.56	4.47	4.35	
2	4.86	4.70	4.70	4.54	4.68	4.41	4.51	4.24	
3	4.69	4.81	3.93	4.04	4.67	4.23	4.57	4.13	
4	4.86	5.07	4.09	4.29	4.03	3.65	4.12	3.73	
5	3.44	4.29	4.95	5.79	4.66	4.51	4.81	4.65	
6	4.49	5.14	4.61	5.26	4.38	4.43	4.68	4.73	
7	1.55	3.08	3.31	4.83	3.22	3.39	3.63	3.81	
8	1.98	2.76	3.95	4.73	3.00	2.55	3.88	3.43	
9	2.68	3.52	3.65	4.49	2.78	2.88	3.84	3.94	

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 23: Relative and absolute calculations of intrinsic p$K_{a\ OH}$ for neutral 1-N-methyl-5-dihydroxymethyluracil (**1m5hdmU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing.

Molecule	**p$K_{a(AH)}$**								
1mdhmU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									
0	14.64	14.67	18.04	18.07	18.57	18.60	22.25	22.28	
1	12.78	13.17	16.09	15.10	15.54	15.53	20.05	18.66	
2	11.62	11.80	13.06	13.24	13.40	13.34	16.27	16.21	
3	13.14	14.18	14.51	15.56	13.37	13.78	16.60	17.02	
4	13.07	13.91	13.21	14.06	12.98	13.47	15.99	16.48	
5	11.38	11.75	12.10	12.47	12.19	12.48	15.34	15.63	
6	9.89	10.66	13.16	13.93	11.11	11.80	15.58	16.28	
7	9.80	10.08	12.55	12.84	10.65	10.76	14.69	14.80	
8	11.49	12.74	12.41	13.65	11.69	12.27	14.26	14.84	
9	13.24	13.13	13.45	13.34	12.00	12.43	14.68	15.11	

[a]: see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

Table 24: Relative and absolute pK_{a-} calculations for deprotonated 1-N-metyl-5-carboxyuracil (**1m5caU**) with increasing number of explicit water molecules, different quantum-chemical methods, and data processing. Refence compound is orotic acid.

Molecule	**p$K_{a(A-)}$**								
1m5caU	**Relative Method[a]**				**Absolute Method[b]**				**exp. Ref.**
Method:	**B3LYP[c]**		**DPLNO(CBS) [c]**		**B3LYP[c]**		**DPLNO(CBS) [c]**		
expl. H_2O	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	100[d]	30[d]	
1:boltz[e]									9.20
0	10.44	10.41	10.32	10.29	15.08	15.01	16.09	16.03	9.20
1	11.37	11.39	10.26	10.28	14.59	14.66	15.57	15.64	9.20
2	9.85	10.22	9.48	9.85	12.73	12.80	14.65	14.72	9.20
3	9.56	9.63	9.27	9.34	12.23	12.11	14.34	14.22	9.20
4	9.81	9.93	9.64	9.76	12.50	12.31	14.84	14.66	9.20
5	8.92	9.05	9.06	9.19	11.68	11.56	14.12	14.01	9.20
6	9.51	9.23	9.26	8.98	12.05	11.81	14.12	13.88	9.20
7	9.21	9.71	9.16	9.66	11.60	12.03	13.61	14.04	9.20
8	9.19	9.19	9.20	9.20	10.98	10.79	13.19	13.00	9.20
9	7.86	8.08	9.05	9.27	9.68	10.07	12.35	12.74	9.20

[a]: reference is orotic acid, see Figure 20; [b]: see Figure 18; [c]: using solution phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries (see 4.5.1); [d]: imaginary frequency cut-off value [cm^{-1}]; [e]: Boltzmann weighted ten best conformers (by free energy G_{qh}).

4.5.10 Thermodynamic Data from Explicit Water-Clusters

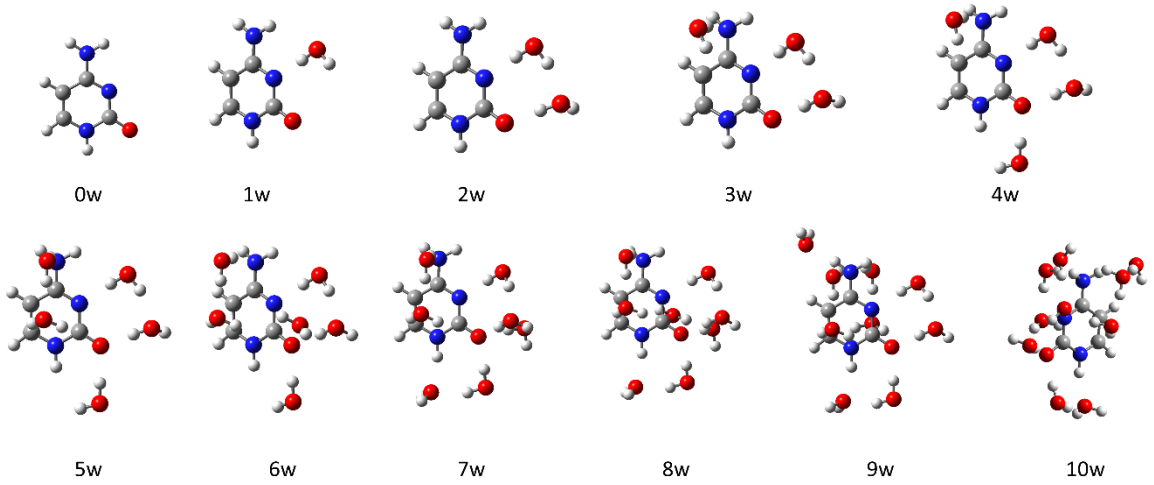


Figure 29: Molecule-water clusters for neutral cytosine (**cyt**) with relevant thermodynamic data provided in Table 25.

Table *25*: Gibbs free energies for solution phase optimized molecule-water clusters of neutral cytosine (**cyt**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
cytosine_0w_trans	0	-394.967638	-394.938447	-394.935434	-394.935434	-394.454469	-394.451456	-394.451456
cytosine_1w_trans	1	-471.419663	-471.377032	-471.372043	-471.372043	-470.831299	-470.826310	-470.826310
cytosine_2w_trans	2	-547.871288	-547.813995	-547.808707	-547.808707	-547.207546	-547.202259	-547.202259
cytosine_3w_trans	3	-624.311042	-624.247023	-624.240239	-624.240239	-623.580969	-623.574186	-623.574186
cytosine_4w_trans	4	-700.749683	-700.682250	-700.675194	-700.675194	-699.957379	-699.950324	-699.950324
cytosine_5w_trans	5	-777.191975	-777.119610	-777.111544	-777.111544	-776.334471	-776.326405	-776.326405
cytosine_6w_trans_2	6	-853.628816	-853.553546	-853.544388	-853.544388	-852.709117	-852.699959	-852.699959
cytosine_7w_trans_10	7	-930.087516	-929.996726	-929.985668	-929.985668	-929.088882	-929.077824	-929.077824
cytosine_8w_trans_11	8	-1006.537472	-1006.430876	-1006.420062	-1006.420062	-1005.462279	-1005.451465	-1005.451465
cytosine_9w_trans_4	9	-1082.977033	-1082.866155	-1082.852241	-1082.852241	-1081.837579	-1081.823666	-1081.823666
cytosine_10w_trans	10	-1159.456457	-1159.300242	-1159.290059	-1159.290059	-1158.206523	-1158.196340	-1158.196340

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

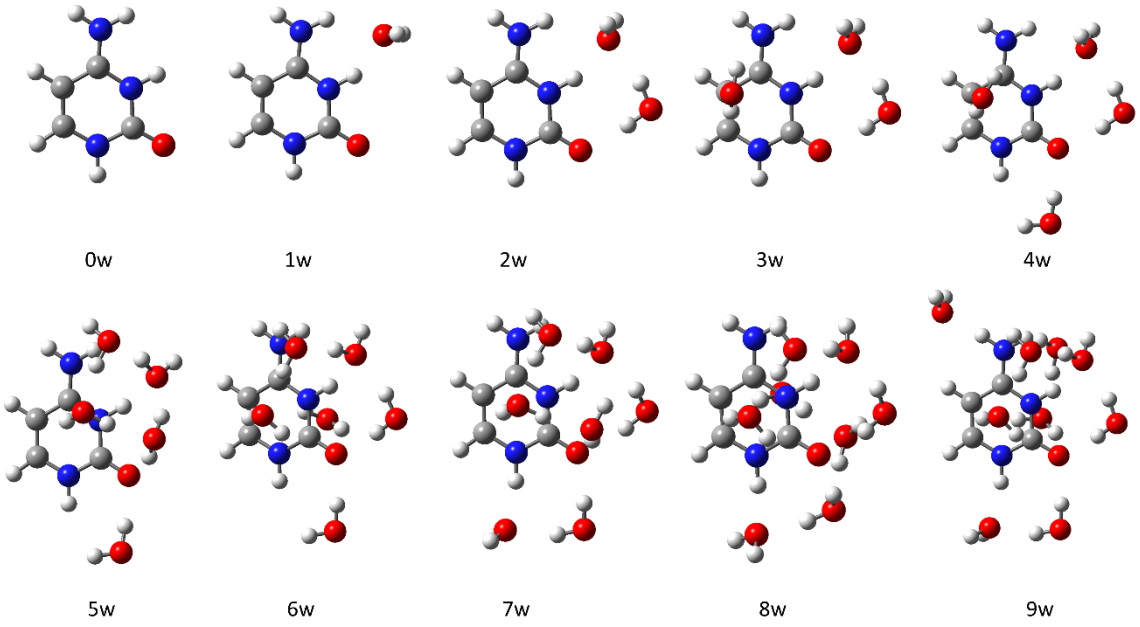


Figure 30: Molecule-water clusters for protonated cytosine (**p_cyt**) with relevant thermodynamic data provided in Table 26.

Table 26: Gibbs free energies for solution phase optimized molecule-water clusters of protonated cytosine (**p_cyt**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
p_cytosine_0w_trans_4	0	-395.345622	-395.378251	-395.375238	-395.375238	-394.892541	-394.889528	-394.889528
p_cytosine_1w_trans_4	1	-471.812241	-471.817303	-471.814290	-471.813382	-471.270132	-471.267120	-471.266212
p_cytosine_2w_2_trans_4	2	-548.239621	-548.253944	-548.250901	-548.248978	-547.647468	-547.644425	-547.642502
p_cytosine_3w_2_trans_4	3	-624.665370	-624.686153	-624.683140	-624.680843	-624.021544	-624.018532	-624.016235
p_cytosine_4w_2_trans_4	4	-701.097141	-701.119663	-701.116650	-701.114052	-700.396656	-700.393643	-700.391045
p_cytosine_5w_trans_4	5	-777.579451	-777.560744	-777.557731	-777.554851	-776.772646	-776.769634	-776.766754
p_cytosine_6w_trans_4	6	-853.986924	-853.992386	-853.989373	-853.984817	-853.148335	-853.145322	-853.140766
p_cytosine_7w_trans_4	7	-930.472877	-930.438241	-930.434660	-930.429949	-929.527486	-929.523906	-929.519195
p_cytosine_8w_3_trans_4	8	-1006.921014	-1006.870794	-1006.867780	-1006.862954	-1005.899833	-1005.896819	-1005.891993
p_cytosine_9w_trans_4	9	-1083.349952	-1083.305394	-1083.301412	-1083.293128	-1082.277024	-1082.273042	-1082.264758

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

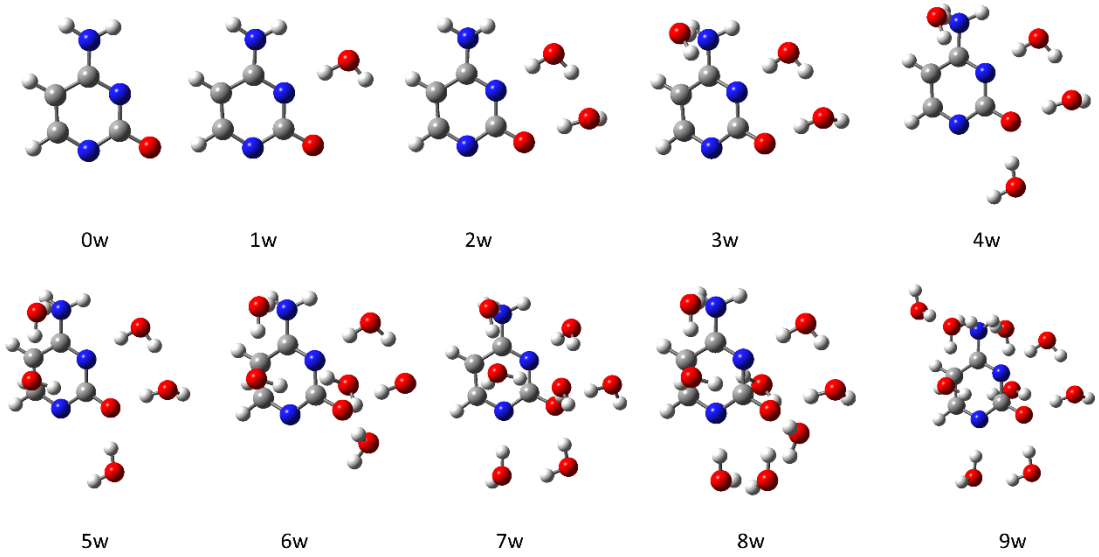


Figure 31: Molecule-water clusters for deprotonated cytosine (**d_cyt**) with relevant thermodynamic data provided in Table 27.

Table 27: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated cytosine (**d_cyt**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_cytosine_0w_trans	0	-394.400530	-394.470947	-394.467934	-394.467934	-393.985148	-393.982135	-393.982135
d_cytosine_1w_trans	1	-470.866190	-470.911440	-470.908233	-470.906849	-470.363076	-470.359869	-470.358485
d_cytosine_2w_trans	2	-547.331696	-547.350316	-547.347014	-547.345065	-546.741582	-546.738280	-546.736331
d_cytosine_3w_trans	3	-623.783512	-623.785576	-623.782563	-623.779405	-623.116410	-623.113397	-623.110239
d_cytosine_4w_trans	4	-700.240705	-700.225180	-700.221622	-700.216681	-699.495268	-699.491711	-699.486770
d_cytosine_5w_trans	5	-776.692641	-776.661855	-776.658568	-776.653085	-775.870990	-775.867703	-775.862220
d_cytosine_6w_trans	6	-853.144859	-853.094951	-853.091938	-853.087701	-852.243355	-852.240342	-852.236105
d_cytosine_7w_2_trans	7	-929.604471	-929.541534	-929.538262	-929.532092	-928.627131	-928.623860	-928.617690
d_cytosine_8w_2_trans	8	-1006.055536	-1005.973429	-1005.970384	-1005.964188	-1004.999083	-1004.996038	-1004.989842
d_cytosine_9w_3_trans	9	-1082.503439	-1082.407437	-1082.404117	-1082.395800	-1081.371606	-1081.368286	-1081.359969
d_cytosine_10w_2_trans	10	-1158.975298	-1158.848873	-1158.845681	-1158.836888	-1157.748497	-1157.745306	-1157.736513

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

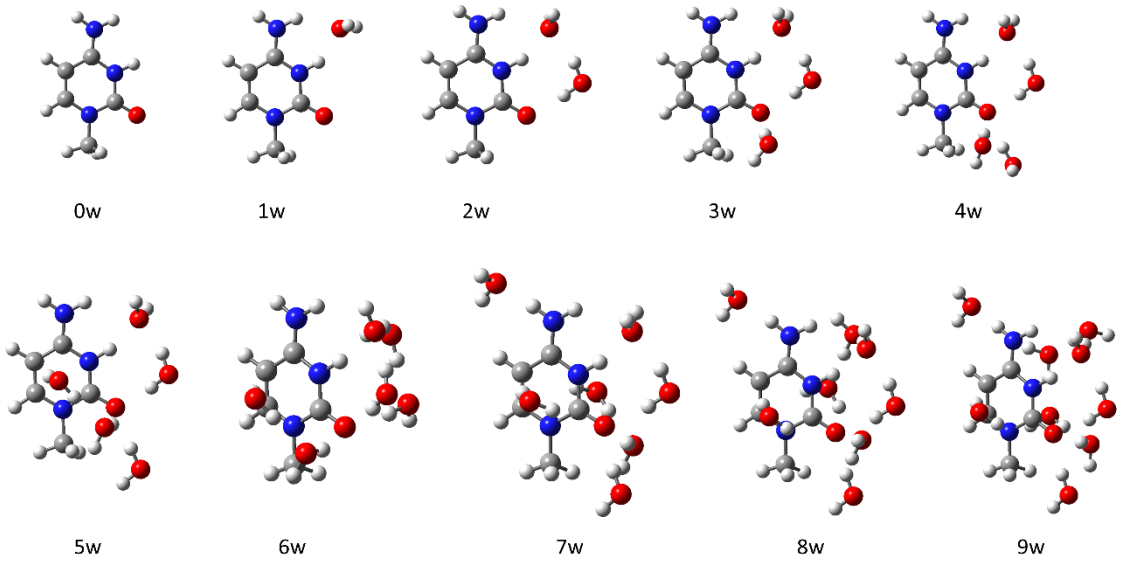


Figure 32: Molecule-water clusters for protonated 1-methylcytosine (**1mC**) with relevant thermodynamic data provided in Table 28.

Table 28: Gibbs free energies for solution phase optimized molecule-water clusters of protonated 1-methylcytosine (**1mC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
p_1mecy_0w	0	-434.666821	-434.665469	-434.662456	-434.662456	-434.117012	-434.113999	-434.113999
p_1mecy_1w	1	-511.132570	-511.104482	-511.101469	-511.100445	-510.494590	-510.494590	-510.493566
p_1mecy_2w	2	-587.560443	-587.539626	-587.536613	-587.535195	-586.870320	-586.870320	-586.868902
p_1mecy_3w	3	-663.992253	-663.974515	-663.971502	-663.968763	-663.246467	-663.246467	-663.243728
p_1mecy_4w	4	-740.432514	-740.411600	-740.407909	-740.403888	-739.623768	-739.623090	-739.619069
p_1mecy_5w	5	-816.870398	-816.845099	-816.842086	-816.838354	-815.997773	-815.997773	-815.994041
p_1mecy_6w	6	-893.330501	-893.283088	-893.280075	-893.275749	-892.373355	-892.373355	-892.369029
p_1mecy_7w_3	7	-969.753871	-969.715606	-969.712593	-969.707046	-968.749557	-968.749557	-968.744010
p_1mecy_8w	8	-1046.217974	-1046.156307	-1046.153239	-1046.146991	-1045.126846	-1045.126791	-1045.120543
p_1mecy_9w	9	-1122.679463	-1122.593108	-1122.590006	-1122.582934	-1121.500533	-1121.500444	-1121.493372

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

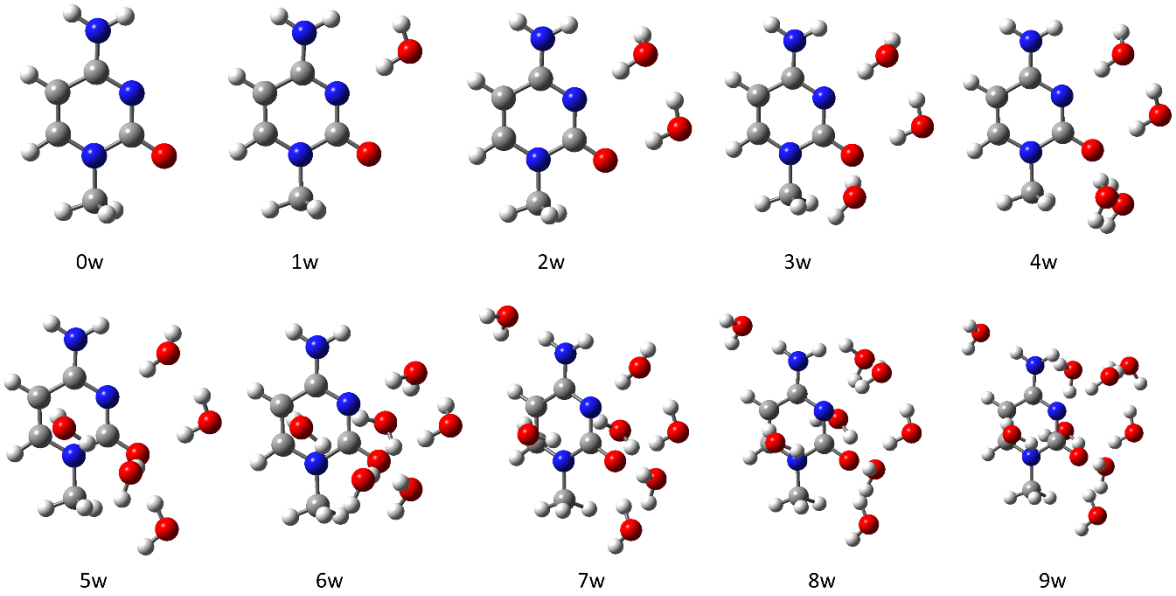


Figure 33: Molecule-water clusters for neutral 1-methylcytosine (**1mC**) with relevant thermodynamic data provided in Table 29.

Table 29: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methylcytosine (**1mC)** at the (U)B3LYP-D3\6-31G+(d,p) level of theory . (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
1mecy_0w	0	-434.283283	-434.224492	-434.221479	-434.221479	-433.678150	-433.675137	-433.675137
1mecy_1w	1	-510.726546	-510.662099	-510.659086	-510.657951	-510.055506	-510.055506	-510.054371
1mecy_2w	2	-587.181466	-587.100335	-587.097074	-587.094825	-586.432061	-586.431813	-586.429564
1mecy_3w	3	-663.622506	-663.537790	-663.534261	-663.530718	-662.810381	-662.809865	-662.806322
1mecy_4w	4	-740.059695	-739.972769	-739.969329	-739.965065	-739.186057	-739.185630	-739.181366
1mecy_5w	5	-816.510132	-816.411018	-816.407424	-816.402071	-815.563378	-815.562797	-815.557444
1mecy_6w_2_6	6	-892.964698	-892.844374	-892.841361	-892.837173	-891.933784	-891.933784	-891.929596
1mecy_7w	7	-969.393588	-969.280440	-969.276946	-969.269241	-968.313351	-968.312870	-968.305165
1mecy_8w	8	-1045.848397	-1045.717533	-1045.714482	-1045.707785	-1044.688379	-1044.688341	-1044.681644
1mecy_9w	9	-1122.303740	-1122.154587	-1122.151421	-1122.142594	-1121.063158	-1121.063005	-1121.054178

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

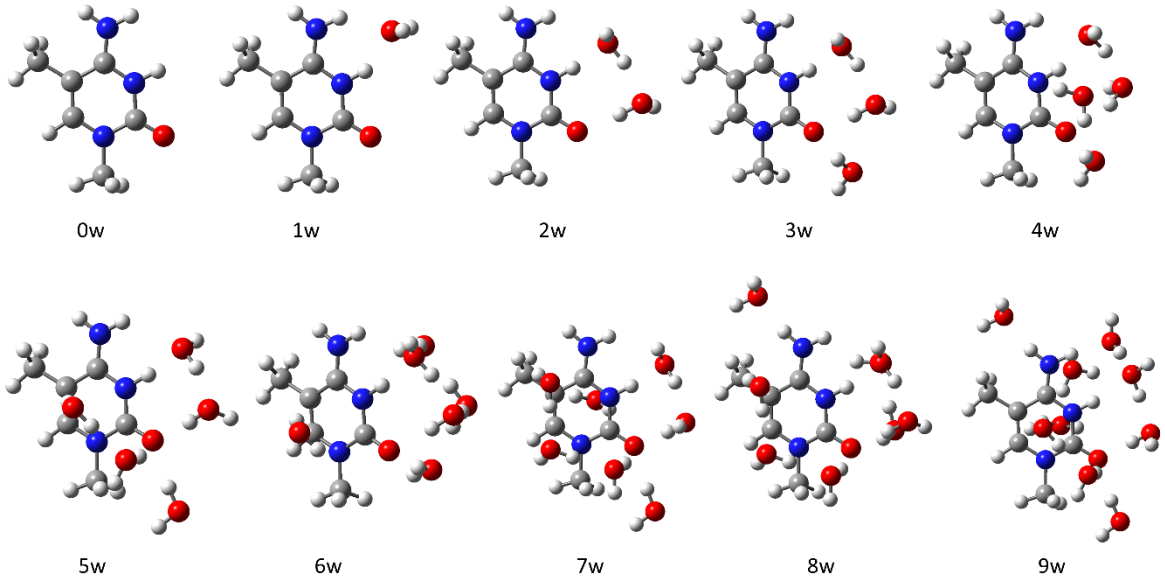


Figure 34: Molecule-water clusters for protonated 1,5-dimethylcytosine (**1m5mC**) with relevant thermodynamic data provided in Table 30.

Table 30: Gibbs free energies for solution phase optimized molecule-water clusters of protonated 1,5-dimethylcytosine (**1m5mC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
p_15mecy1ne1w_0w	0	-473.994000	-473.965732	-473.962719	-473.962234	-473.355416	-473.352404	-473.351919
p_15mecy1ne1w_1w	1	-550.459168	-550.403427	-550.400414	-550.399391	-549.731694	-549.731694	-549.730671
p_15mecy1ne1w_2w	2	-626.911658	-626.840730	-626.837555	-626.835829	-626.108142	-626.107980	-626.106254
p_15mecy1ne1w_3w	3	-703.348198	-703.276876	-703.273825	-703.270402	-702.484723	-702.484685	-702.481262
p_15mecy1ne1w_4w	4	-779.799121	-779.710646	-779.707633	-779.705343	-778.857049	-778.857049	-778.854759
p_15mecy1ne1w_5w	5	-856.223371	-856.148153	-856.144987	-856.140518	-855.237208	-855.237055	-855.232586
p_15mecy1ne1w_6w_2	6	-932.671766	-932.583254	-932.580223	-932.576128	-931.610000	-931.609982	-931.605887
p_15mecy1ne1w_7w_3	7	-1009.101617	-1009.022369	-1009.018422	-1009.010669	-1007.991712	-1007.990778	-1007.983025
p_15mecy1ne1w_8w	8	-1085.574292	-1085.457184	-1085.453885	-1085.447015	-1084.361920	-1084.361634	-1084.354764
p_15mecy1ne1w_9w_2	9	-1162.013394	-1161.893834	-1161.890781	-1161.882407	-1160.739555	-1160.739515	-1160.731141

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

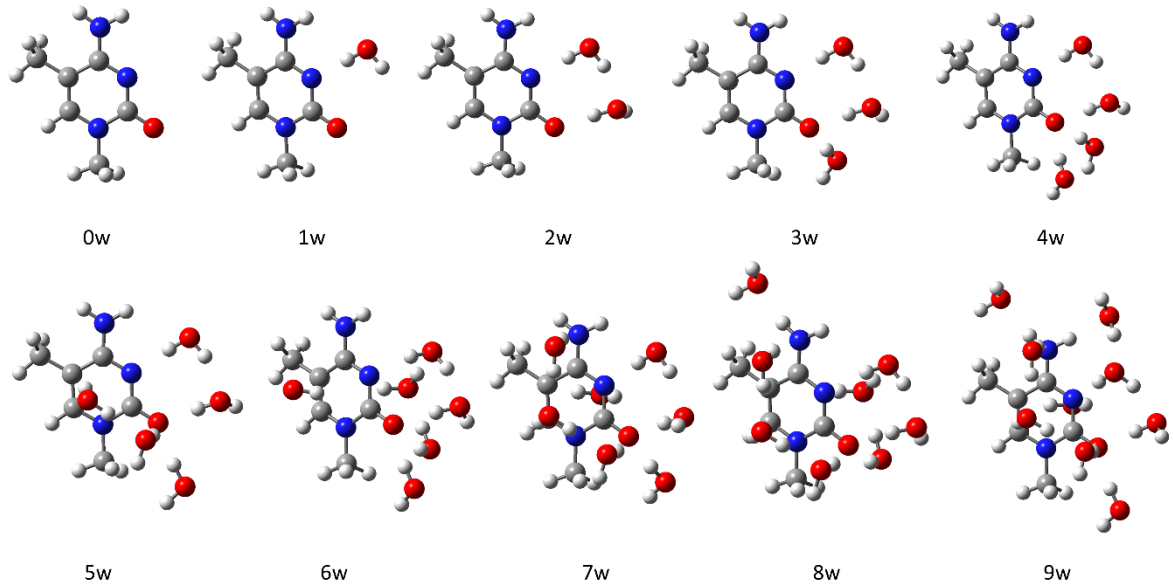


Figure 35: Molecule-water clusters for neutral 1,5-dimethylcytosine (**1m5mC**) with relevant thermodynamic data provided in Table 31.

Table 31: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1,5-dimethylcytosine (**1m5mC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion)

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
15mecy1ne1w_0w	0	-473.606828	-473.522559	-473.519546	-473.519364	-472.914526	-472.911513	-472.911331
15mecy1ne1w_1w	1	-550.059320	-549.960584	-549.957571	-549.955879	-549.290513	-549.290513	-549.288821
15mecy1ne1w_2w	2	-626.511899	-626.397534	-626.394521	-626.392185	-625.666588	-625.666588	-625.664252
15mecy1ne1w_3w	3	-702.955675	-702.834565	-702.831552	-702.828290	-702.043795	-702.043795	-702.040533
15mecy1ne1w_4w	4	-779.396977	-779.271917	-779.268657	-779.264189	-778.421783	-778.421536	-778.417068
15mecy1ne1w_5w	5	-855.843478	-855.708457	-855.704880	-855.699686	-854.797485	-854.796921	-854.791727
15mecy1ne1w_6w	6	-932.294996	-932.142475	-932.139462	-932.135053	-931.169631	-931.169631	-931.165222
15mecy1ne1w_7w	7	-1008.726732	-1008.582875	-1008.578134	-1008.570225	-1007.551205	-1007.549477	-1007.541568
15mecy1ne1w_8w_3_2	8	-1085.202238	-1085.015739	-1085.012607	-1085.006101	-1083.919315	-1083.919196	-1083.912690
15mecy1ne1w_9w	9	-1161.638308	-1161.454379	-1161.450949	-1161.441479	-1160.300126	-1160.299709	-1160.290239

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

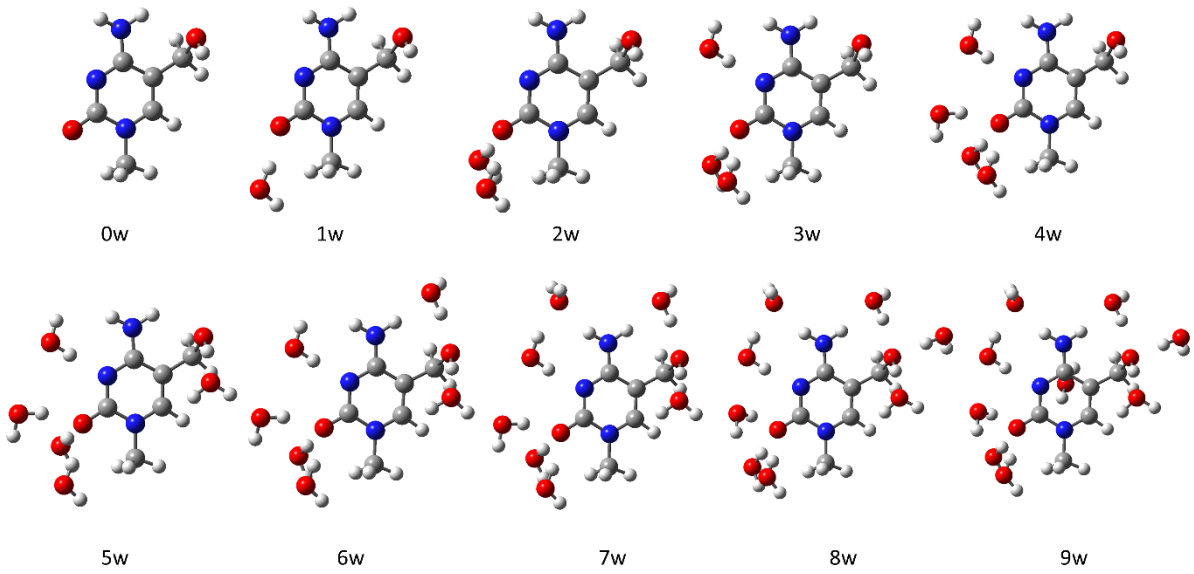


Figure 36: Molecule-water clusters for neutral 1-methyl-5-hydroxymetylcytosine (**1m5hmC**) with relevant thermodynamic data provided in Table 32.

Table 32: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5-hydroxymetylcytosine (**1m5hmC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5hm_mecy1ne_3_0w	0	-548.828074	-548.744559	-548.741546	-548.740621	-548.078610	-548.078610	-548.077685
5hm_mecy1ne_3_1w	1	-625.272258	-625.180567	-625.177512	-625.175848	-624.455597	-624.455555	-624.453891
5hm_mecy1ne_3_2w	2	-701.713597	-701.616396	-701.613383	-701.610949	-700.831951	-700.831951	-700.829517
5hm_mecy1ne_3_3w	3	-778.154383	-778.055465	-778.051448	-778.047054	-777.211049	-777.210045	-777.205651
5hm_mecy1ne_3_4w	4	-854.596101	-854.491538	-854.488377	-854.483636	-853.586903	-853.586755	-853.582014
5hm_mecy1ne_3_5w	5	-931.041208	-930.927254	-930.924084	-930.918101	-929.963079	-929.962922	-929.956939
5hm_mecy1ne_3_6w	6	-1007.491730	-1007.361912	-1007.358731	-1007.352137	-1006.336590	-1006.336422	-1006.329828
5hm_mecy1ne_3_7w	7	-1083.941539	-1083.798910	-1083.795657	-1083.787670	-1082.711609	-1082.711369	-1082.703382
5hm_mecy1ne_3_8w	8	-1160.386657	-1160.237241	-1160.233365	-1160.223232	-1159.088888	-1159.088025	-1159.077892
5hm_mecy1ne_3_9w	9	-1236.820343	-1236.672985	-1236.669080	-1236.656914	-1235.465519	-1235.464627	-1235.452461

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

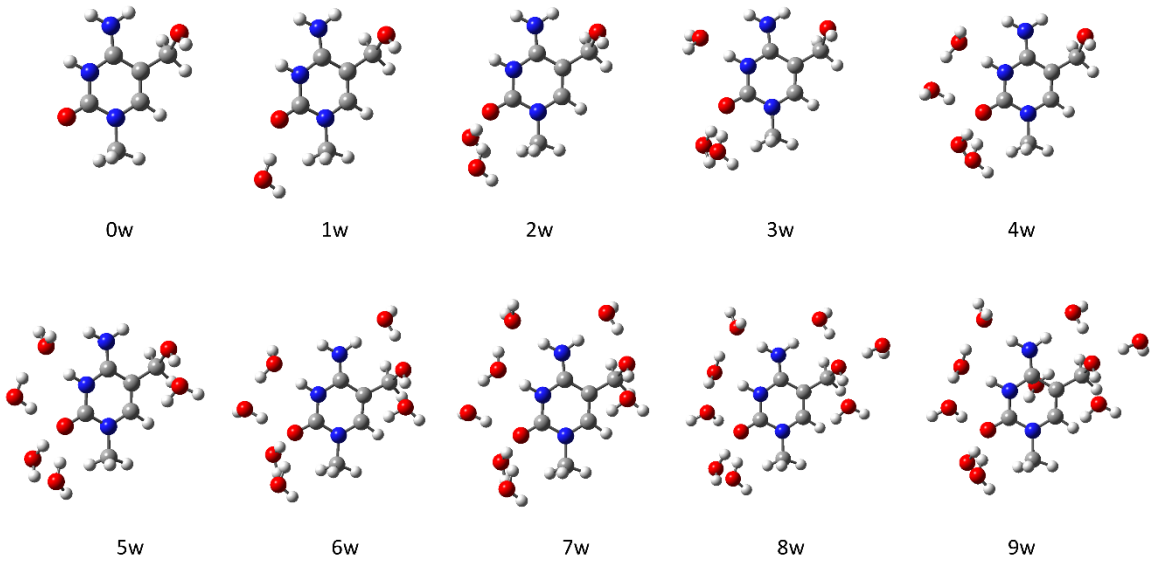


Figure 37: Molecule-water clusters for protonated 1-methyl-5-hydroxymetylcytosine (**1m5hmC**) with relevant thermodynamic data provided in Table 33.

Table 33: Gibbs free energies for solution phase optimized molecule-water clusters of protonated 1-methyl-5-hydroxymetylcytosine (**1m5hmC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
p_5hm_mecy1ne_3_0w	0	-549.211072	-549.185496	-549.182483	-549.181282	-548.517898	-548.517898	-548.516697
p_5hm_mecy1ne_3_1w	1	-625.644176	-625.620928	-625.617383	-625.615488	-624.895182	-624.894650	-624.892755
p_5hm_mecy1ne_3_2w	2	-702.078689	-702.054608	-702.051535	-702.049077	-701.270237	-701.270177	-701.267719
p_5hm_mecy1ne_3_3w	3	-778.546016	-778.495314	-778.492301	-778.488147	-777.648854	-777.648854	-777.644700
p_5hm_mecy1ne_3_4w	4	-854.996274	-854.933873	-854.930044	-854.924943	-854.025992	-854.025176	-854.020075
p_5hm_mecy1ne_3_5w	5	-931.423946	-931.367506	-931.364134	-931.358950	-930.400418	-930.400059	-930.394875
p_5hm_mecy1ne_3_6w	6	-1007.895688	-1007.806658	-1007.803328	-1007.796948	-1006.778398	-1006.778081	-1006.771701
p_5hm_mecy1ne_3_7w	7	-1084.354910	-1084.241199	-1084.238013	-1084.232128	-1083.151128	-1083.150955	-1083.145070
p_5hm_mecy1ne_3_8w	8	-1160.799182	-1160.676929	-1160.673896	-1160.666122	-1159.525274	-1159.525254	-1159.517480
p_5hm_mecy1ne_3_9w	9	-1237.223403	-1237.110142	-1237.107065	-1237.098296	-1235.900506	-1235.900442	-1235.891673

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

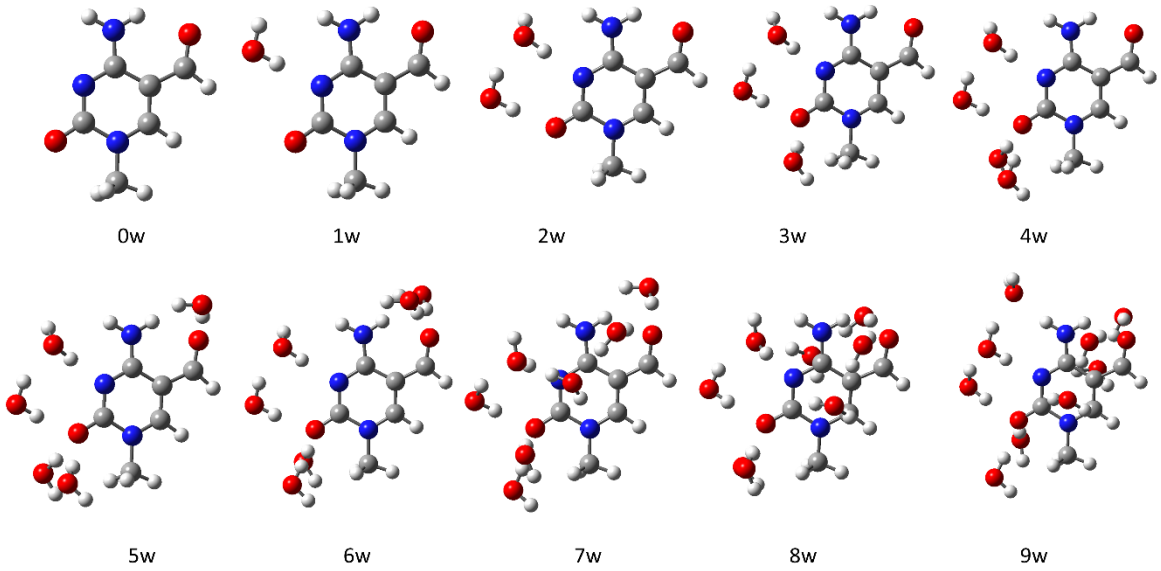


Figure 38: Molecule-water clusters for neutral 1-methyl-5formylcytosine (**1m5fC**) with relevant thermodynamic data provided in Table 34.

Table 34: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5formylcytosine (**1m5fC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5f_mecy1ne0w_2	0	-547.627046	-547.559227	-547.556214	-547.555904	-546.890827	-546.887814	-546.887504
5f_mecy1ne1w	1	-624.069242	-623.995864	-623.992851	-623.991482	-623.267463	-623.264450	-623.263081
5f_mecy1ne2w	2	-700.524026	-700.433987	-700.430974	-700.428521	-699.644216	-699.641204	-699.638751
5f_mecy1ne3w	3	-776.960509	-776.868494	-776.865163	-776.862200	-776.020049	-776.016719	-776.013756
5f_mecy1ne4w	4	-853.400616	-853.302620	-853.299607	-853.296172	-852.394932	-852.391919	-852.388484
5f_mecy1ne5w	5	-929.836308	-929.740152	-929.736754	-929.730528	-928.773436	-928.770038	-928.763812
5f_mecy1ne6w	6	-1006.273144	-1006.171969	-1006.168828	-1006.161681	-1005.146222	-1005.143081	-1005.135934
5f_mecy1ne7w	7	-1082.727168	-1082.610389	-1082.607376	-1082.601518	-1081.522100	-1081.519087	-1081.516242
5f_mecy1ne8w_6	8	-1159.166707	-1159.046588	-1159.043225	-1159.036239	-1157.897731	-1157.897381	-1157.890395
5f_mecy1ne9w_6	9	-1235.611314	-1235.482489	-1235.479315	-1235.471036	-1234.273094	-1234.272933	-1234.264654

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

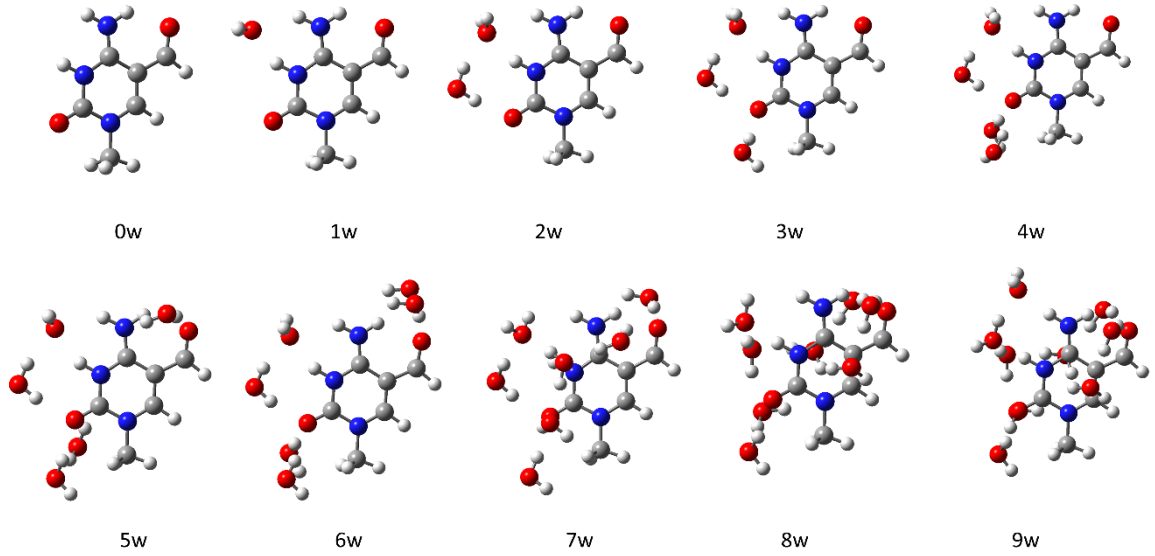


Figure 39: Molecule-water clusters for protonated 1-methyl-5formylcytosine (**1m5fC**) with relevant thermodynamic data provided in Table 35.

Table 35: Gibbs free energies for solution phase optimized molecule-water clusters of protonated 1-methyl-5formylcytosine (**1m5fC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
p_5f_mecy1ne0w	0	-547.998862	-547.994593	-547.991580	-547.991395	-547.326285	-547.323272	-547.323087
p_5f_mecy1ne1w	1	-624.464967	-624.433327	-624.430314	-624.429394	-623.703427	-623.700414	-623.699494
p_5f_mecy1ne2w	2	-700.892022	-700.869688	-700.866244	-700.864177	-700.080470	-700.077027	-700.074960
p_5f_mecy1ne3w	3	-777.326564	-777.305150	-777.301723	-777.298306	-776.457338	-776.453912	-776.450495
p_5f_mecy1ne4w	4	-853.759985	-853.737912	-853.734899	-853.731179	-852.830935	-852.827922	-852.824202
p_5f_mecy1ne5w	5	-930.202011	-930.174198	-930.171016	-930.165737	-929.208043	-929.204861	-929.199582
p_5f_mecy1ne6w	6	-1006.619308	-1006.606691	-1006.603527	-1006.596597	-1005.583942	-1005.580779	-1005.573849
p_5f_mecy1ne7w	7	-1083.076208	-1083.042970	-1083.039798	-1083.033317	-1081.956285	-1081.953114	-1081.949645
p_5f_mecy1ne8w_2	8	-1159.535127	-1159.480063	-1159.477050	-1159.472478	-1158.331042	-1158.331042	-1158.326470
p_5f_mecy1ne9w_2	9	-1236.006562	-1235.924376	-1235.921244	-1235.914137	-1234.711832	-1234.711713	-1234.704606

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

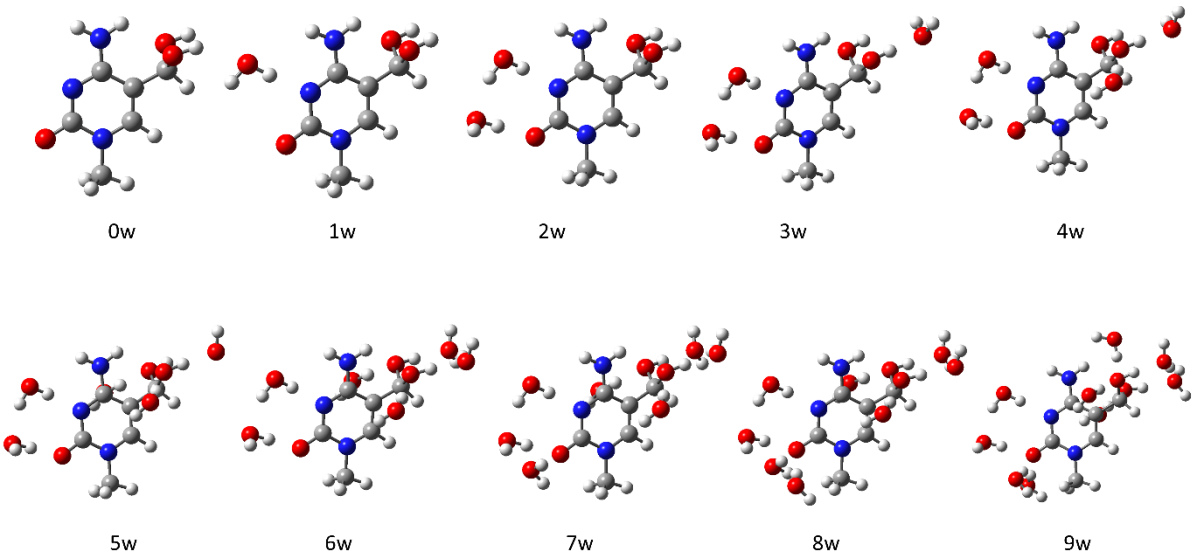


Figure 40: Molecule-water clusters for neutral 1-methyl-5-dihydroxymethylcytosine (**1m5dhmC**) with relevant thermodynamic data provided in Table 36.

Table 36: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5-dihydroxymethylcytosine (**1m5dhmC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5dhmCa_4_0w	0	-624.054468	-623.979414	-623.976401	-623.975338	-623.257537	-623.257537	-623.256474
5dhmCa_4_1w	1	-700.506726	-700.416964	-700.413951	-700.411548	-699.633313	-699.633313	-699.630910
5dhmCa_4_2w	2	-776.959381	-776.854554	-776.850828	-776.847683	-776.009932	-776.009219	-776.006074
5dhmCa_4_3w	3	-853.405423	-853.289600	-853.286484	-853.282377	-852.384519	-852.384416	-852.380309
5dhmCa_4_4w	4	-929.848543	-929.727410	-929.724155	-929.719453	-928.761473	-928.761231	-928.756529
5dhmCa_4_5w	5	-1006.290065	-1006.162967	-1006.159395	-1006.153705	-1005.136385	-1005.135826	-1005.130136
5dhmCa_4_6w	6	-1082.750365	-1082.603913	-1082.600144	-1082.593294	-1081.513905	-1081.513149	-1081.506299
5dhmCa_4_7w	7	-1159.190721	-1159.039742	-1159.035543	-1159.027752	-1157.890523	-1157.889337	-1157.881546
5dhmCa_4_8w	8	-1235.632155	-1235.477747	-1235.474124	-1235.463840	-1234.269017	-1234.268407	-1234.258123
5dhmCa_4_9w	9	-1312.077652	-1311.911323	-1311.908003	-1311.897975	-1310.641217	-1310.640910	-1310.630882

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

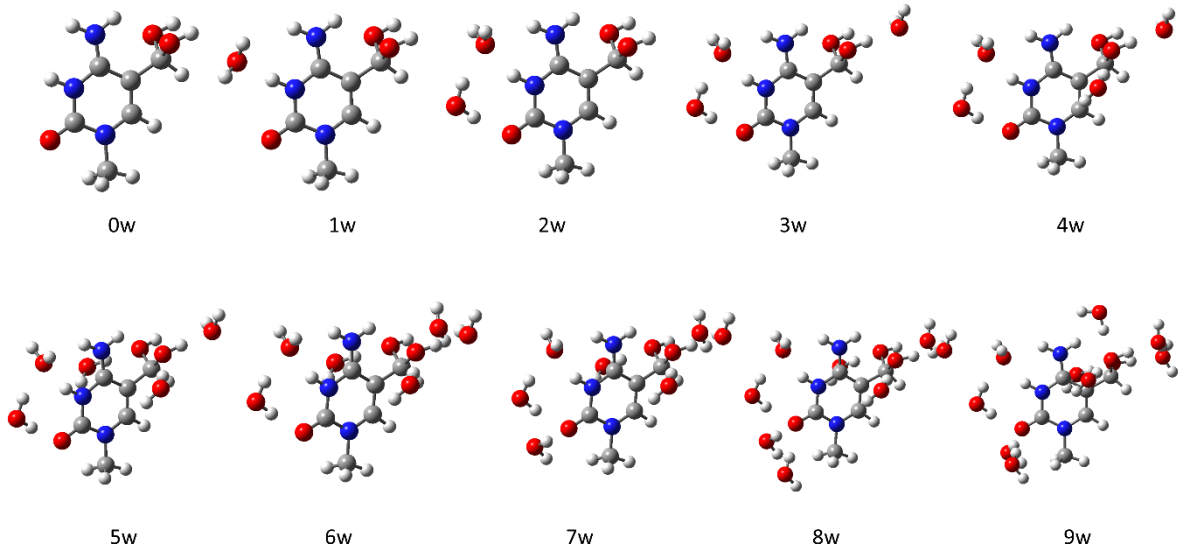


Figure 41: Molecule-water clusters for protonated 1-methyl-5-dihydroxymethylcytosine (**1m5dhmC**) with relevant thermodynamic data provided in Table 37.

Table 37: Gibbs free energies for solution phase optimized molecule-water clusters of protonated 1-methyl-5-dihydroxymethylcytosine (**1m5dhmC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
p_5dhmCa_4_0w	0	-624.445214	-624.419700	-624.416687	-624.415325	-623.695775	-623.695775	-623.694413
p_5dhmCa_4_1w	1	-700.899567	-700.857627	-700.854614	-700.852490	-700.073507	-700.073507	-700.071383
p_5dhmCa_4_2w	2	-777.341016	-777.294407	-777.291173	-777.288616	-776.449408	-776.449187	-776.446630
p_5dhmCa_4_3w	3	-853.789595	-853.729590	-853.726432	-853.723274	-852.824291	-852.824146	-852.820988
p_5dhmCa_4_4w	4	-930.227933	-930.166008	-930.162838	-930.158970	-929.200254	-929.200097	-929.196229
p_5dhmCa_4_5w	5	-1006.667739	-1006.602953	-1006.599940	-1006.594581	-1005.576694	-1005.576694	-1005.571335
p_5dhmCa_4_6w	6	-1083.131453	-1083.043712	-1083.040532	-1083.033564	-1081.954346	-1081.954179	-1081.947211
p_5dhmCa_4_7w	7	-1159.565596	-1159.476793	-1159.473670	-1159.465495	-1158.328477	-1158.328367	-1158.320192
p_5dhmCa_4_8w	8	-1235.997443	-1235.913771	-1235.910411	-1235.901115	-1234.706958	-1234.706611	-1234.697315
p_5dhmCa_4_9w	9	-1312.439769	-1312.348390	-1312.345019	-1312.334772	-1311.080591	-1311.080233	-1311.069986

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

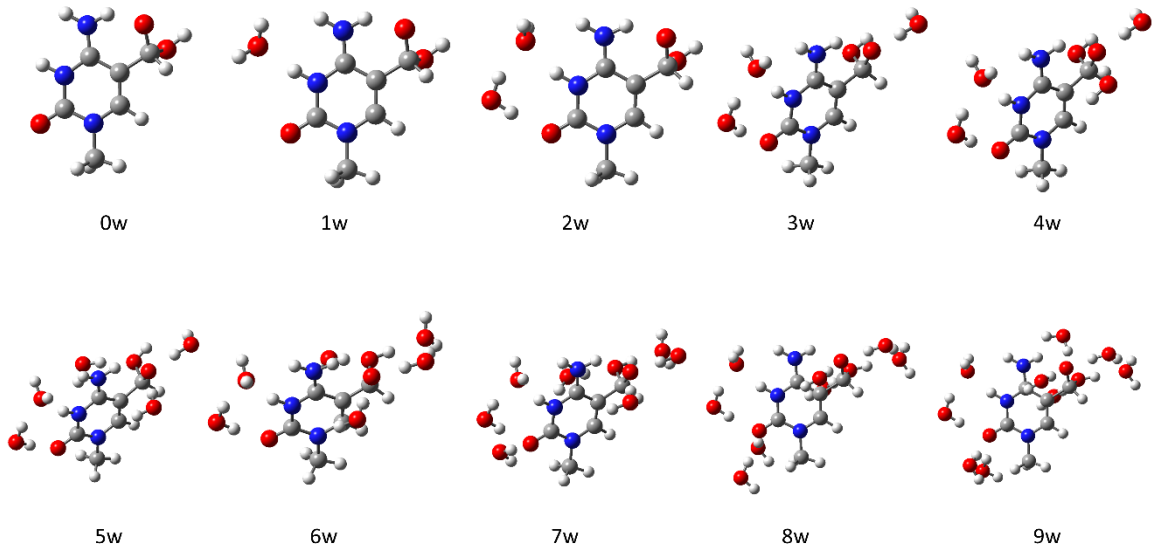


Figure 42: Molecule-water clusters for neutral, zwitterionic 1-methyl-5-dihydroxymethylcytosine (**1m5dhmC**) with relevant thermodynamic data provided in Table 38.

Table 38: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1-methyl-5-dihydroxymethylcytosine (**1m5dhmC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
zw_5dhmCa_4_0w	0	-624.023593	-623.958451	-623.955437	-623.954259	-623.230577	-623.230576	-623.229398
zw_5dhmCa_4_1w	1	-700.469764	-700.395392	-700.392379	-700.391122	-699.607204	-699.607204	-699.605947
zw_5dhmCa_4_2w	2	-776.914859	-776.832366	-776.829150	-776.826731	-775.982897	-775.982694	-775.980275
zw_5dhmCa_4_3w	3	-853.366654	-853.273311	-853.270284	-853.266268	-852.363960	-852.363946	-852.359930
zw_5dhmCa_4_4w	4	-929.816052	-929.713334	-929.709946	-929.705066	-928.743504	-928.743129	-928.738249
zw_5dhmCa_4_5w_2	5	-1006.261262	-1006.148492	-1006.145479	-1006.140416	-1005.118046	-1005.118046	-1005.112983
zw_5dhmCa_4_6w_2	6	-1082.707988	-1082.585381	-1082.581876	-1082.575848	-1081.492387	-1081.491895	-1081.485867
zw_5dhmCa_4_7w	7	-1159.143669	-1159.020054	-1159.016720	-1159.009003	-1157.868124	-1157.867803	-1157.860086
zw_5dhmCa_4_8w	8	-1235.580941	-1235.458881	-1235.455463	-1235.445735	-1234.247929	-1234.247524	-1234.237796
zw_5dhmCa_4_9w	9	-1312.023773	-1311.894019	-1311.890696	-1311.880403	-1310.621916	-1310.621606	-1310.611313

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

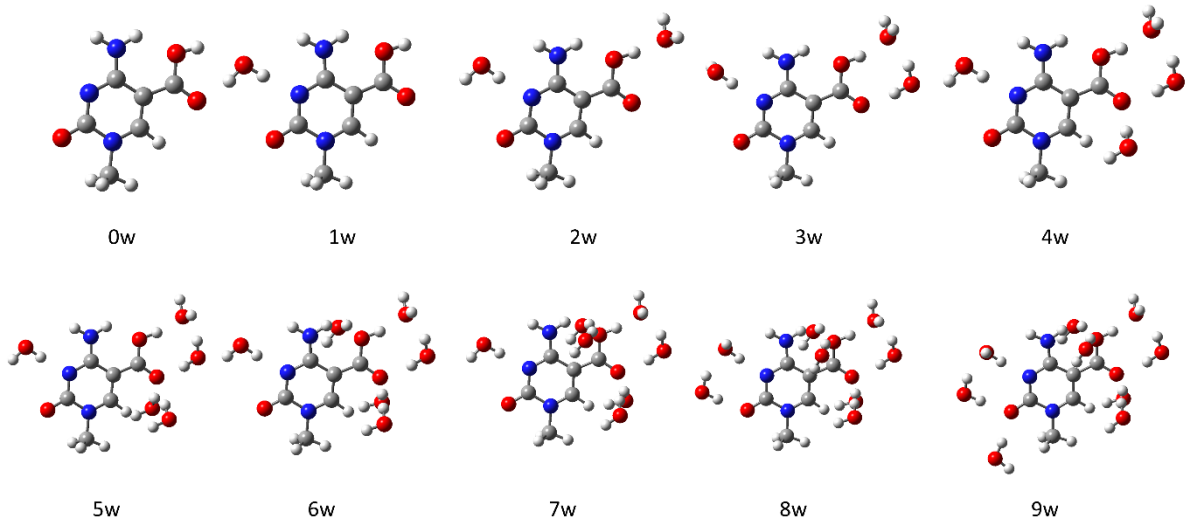


Figure 43: Molecule-water clusters for second best conformer (2ne_4) for neutral 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 39.

Table 39: Gibbs free energies for solution phase optimized molecule-water clusters for second best conformer (2ne_4) of neutral 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5caC_2ne_4_N_0w	0	-622.879905	-622.812131	-622.809118	-622.808156	-622.087434	-622.087434	-622.086472
5caC_2ne_4_N_1w	1	-699.331110	-699.251132	-699.247547	-699.244733	-698.464980	-698.464408	-698.461594
5caC_2ne_4_N_2w	2	-775.778542	-775.687172	-775.684159	-775.681354	-774.840602	-774.840602	-774.837797
5caC_2ne_4_N_3w_2	3	-852.221168	-852.123868	-852.120455	-852.115805	-851.216749	-851.216349	-851.211699
5caC_2ne_4_N_4w	4	-928.663647	-928.559622	-928.556523	-928.550508	-927.592027	-927.591941	-927.585926
5caC_2ne_4_N_5w	5	-1005.101015	-1004.994947	-1004.991934	-1004.985880	-1003.967862	-1003.967862	-1003.961808
5caC_2ne_4_N_6w	6	-1081.537253	-1081.429676	-1081.426255	-1081.418795	-1080.343254	-1080.342846	-1080.335386
5caC_2ne_4_N_7w	7	-1157.971938	-1157.862617	-1157.859295	-1157.851266	-1156.716602	-1156.716293	-1156.708264
5caC_2ne_4_N_8w	8	-1234.417531	-1234.299837	-1234.296772	-1234.287633	-1233.094116	-1233.094064	-1233.084925
5caC_2ne_4_N_9w	9	-1310.855885	-1310.734100	-1310.730543	-1310.720658	-1309.469362	-1309.468818	-1309.458933

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

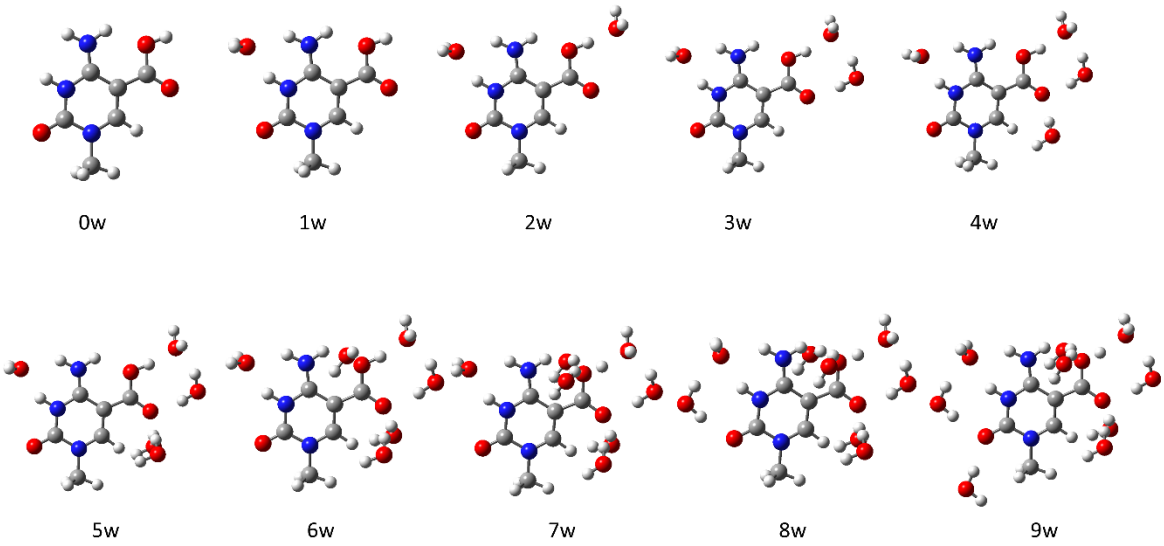


Figure 44: Molecule-water clusters for second best conformer (2ne_4) of protonated 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 40.

Table 40: Gibbs free energies for solution phase optimized molecule-water clusters for second best conformer (2ne_4) of protonated 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
p_5caC_2ne_4_N_0w	0	-623.254651	-623.247776	-623.244763	-623.244227	-622.522510	-622.522510	-622.521974
p_5caC_2ne_4_N_1w	1	-699.721081	-699.689125	-699.685908	-699.683427	-698.902243	-698.902039	-698.899558
p_5caC_2ne_4_N_2w	2	-776.176843	-776.127411	-776.124398	-776.121394	-775.279822	-775.279822	-775.276818
p_5caC_2ne_4_N_3w	3	-852.615319	-852.564626	-852.560941	-852.556440	-851.655878	-851.655206	-851.650705
p_5caC_2ne_4_N_4w	4	-929.048791	-928.998365	-928.995039	-928.989399	-928.031035	-928.030722	-928.025082
p_5caC_2ne_4_N_5w	5	-1005.480467	-1005.432411	-1005.429332	-1005.423893	-1004.406090	-1004.406024	-1004.400585
p_5caC_2ne_4_N_6w	6	-1081.911826	-1081.866988	-1081.863630	-1081.856634	-1080.781813	-1080.781468	-1080.774472
p_5caC_2ne_4_N_7w	7	-1158.341834	-1158.299378	-1158.296295	-1158.288522	-1157.154990	-1157.154920	-1157.147147
p_5caC_2ne_4_N_8w	8	-1234.771621	-1234.734784	-1234.731581	-1234.723464	-1233.531278	-1233.531088	-1233.522971
p_5caC_2ne_4_N_9w	9	-1311.202807	-1311.169208	-1311.165511	-1311.156101	-1309.907677	-1309.906993	-1309.897583

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

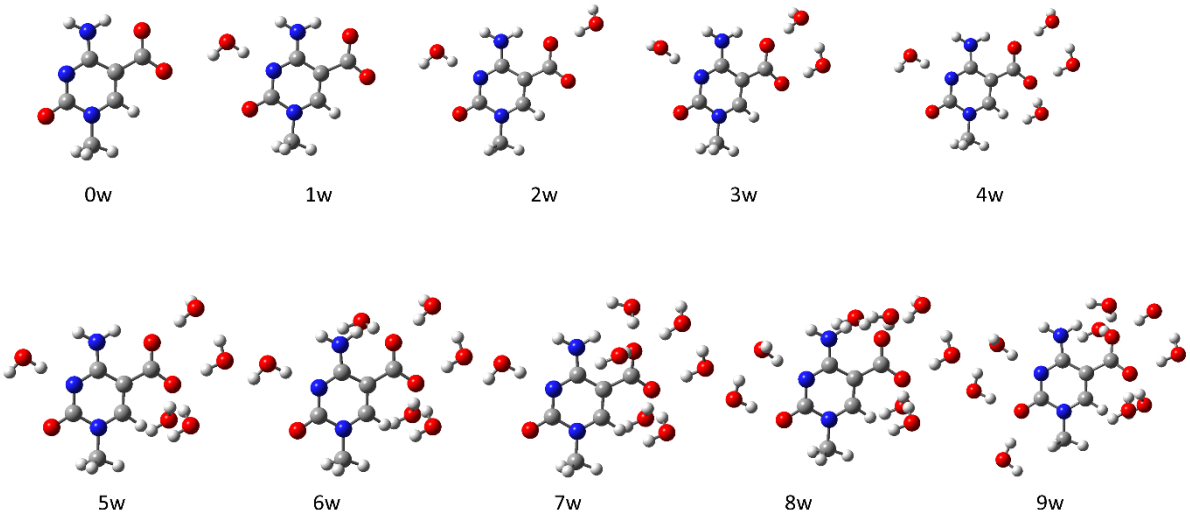


Figure 45: Molecule-water clusters for second best conformer (2ne_4) for deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 41.

Table 41: Gibbs free energies for solution phase optimized molecule-water clusters for second best conformer (2ne_4) of deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_5caC_2ne_4_N_0w	0	-622.350453	-622.371923	-622.368909	-622.367963	-621.645388	-621.645387	-621.644441
d_5caC_2ne_4_N_1w	1	-698.808565	-698.810051	-698.807037	-698.804787	-698.021679	-698.021678	-698.019428
d_5caC_2ne_4_N_2w	2	-775.263834	-775.247734	-775.244244	-775.240846	-774.398727	-774.398250	-774.394852
d_5caC_2ne_4_N_3w	3	-851.725998	-851.685098	-851.681386	-851.677399	-850.775028	-850.774329	-850.770342
d_5caC_2ne_4_N_4w	4	-928.178024	-928.124364	-928.120565	-928.114675	-927.153858	-927.153072	-927.147182
d_5caC_2ne_4_N_5w_2	5	-1004.625071	-1004.562441	-1004.558537	-1004.551174	-1003.532045	-1003.531154	-1003.523791
d_5caC_2ne_4_N_6w	6	-1081.073235	-1080.996384	-1080.993371	-1080.986089	-1079.905989	-1079.905989	-1079.898707
d_5caC_2ne_4_N_7w_2	7	-1157.516479	-1157.434548	-1157.431048	-1157.421768	-1156.284536	-1156.284049	-1156.274769
d_5caC_2ne_4_N_8w_2	8	-1233.969324	-1233.872491	-1233.868909	-1233.858639	-1232.662250	-1232.661681	-1232.651411
d_5caC_2ne_4_N_9w_2	9	-1310.410580	-1310.307321	-1310.303131	-1310.292073	-1309.038388	-1309.037211	-1309.026153

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

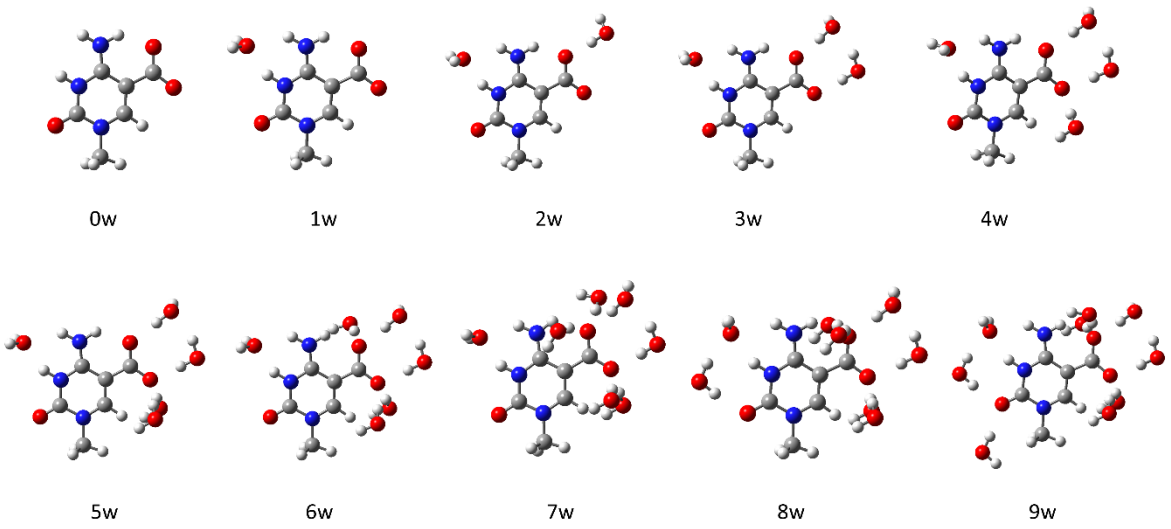


Figure 46: Molecule-water clusters for neutral, zwitterionic 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 42.

Table 42: Gibbs free energies for solution phase optimized molecule-water clusters of neutral, zwitterionic 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
zw_5caC_2ne_1_N_0w	0	-622.855955	-622.815601	-622.812588	-622.811680	-622.089212	-622.089212	-622.088304
zw_5caC_2ne_1_N_1w	1	-699.310649	-699.254481	-699.251468	-699.249598	-698.467027	-698.467027	-698.465157
zw_5caC_2ne_1_N_2w	2	-775.758950	-775.692186	-775.688778	-775.685115	-774.844238	-774.843843	-774.840180
zw_5caC_2ne_1_N_3w	3	-852.216336	-852.128584	-852.125147	-852.121416	-851.218844	-851.218420	-851.214689
zw_5caC_2ne_1_N_4w	4	-928.658543	-928.566593	-928.563213	-928.557791	-927.598037	-927.597670	-927.592248
zw_5caC_2ne_1_N_5w	5	-1005.099327	-1005.003786	-1005.000148	-1004.993644	-1003.975810	-1003.975185	-1003.968681
zw_5caC_2ne_1_N_6w	6	-1081.540190	-1081.437399	-1081.434336	-1081.427812	-1080.350084	-1080.350034	-1080.343510
zw_5caC_2ne_1_N_7w_2	7	-1157.978545	-1157.875025	-1157.871427	-1157.862773	-1156.728406	-1156.727821	-1156.719167
zw_5caC_2ne_1_N_8w	8	-1234.413059	-1234.310491	-1234.306762	-1234.297254	-1233.103998	-1233.103282	-1233.093774
zw_5caC_2ne_1_N_9w	9	-1310.849051	-1310.744928	-1310.740414	-1310.730066	-1309.480594	-1309.479093	-1309.468745

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

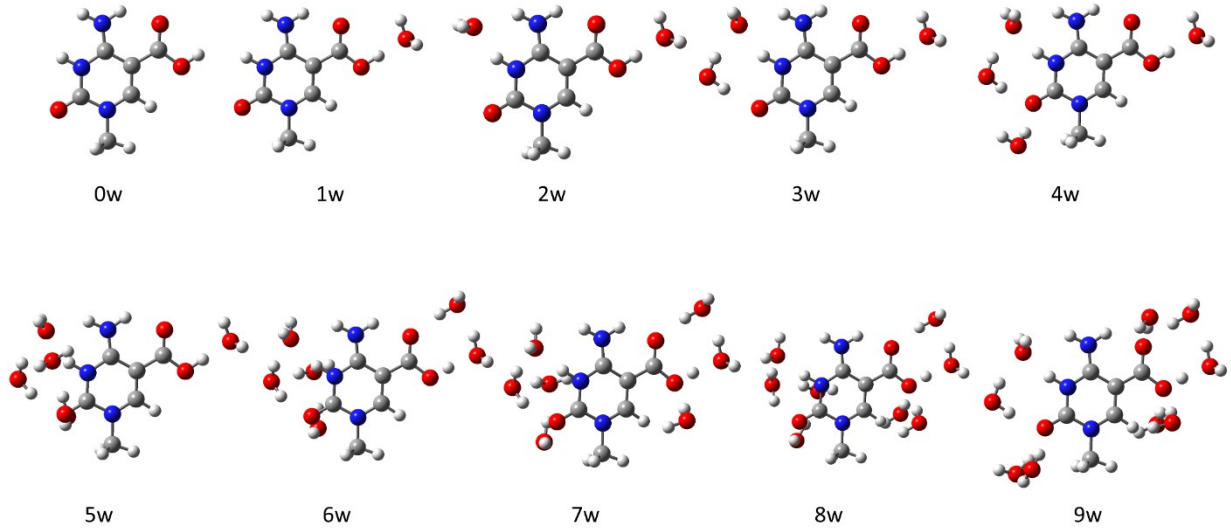


Figure 47: Molecule-water clusters for best conformer (2ne_1) of protonated 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 43.

Table 43: Gibbs free energies for solution phase optimized molecule-water clusters for best conformer (2ne_1) of protonated 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
p_5caC_2ne_1_0w	0	-623.260215	-623.250520	-623.247507	-623.246583	-622.525789	-622.525789	-622.524865
p_5caC_2ne_1_1w	1	-699.719032	-699.690645	-699.687566	-699.685600	-698.904786	-698.904720	-698.902754
p_5caC_2ne_1_2w	2	-776.183792	-776.130811	-776.127798	-776.124484	-775.283471	-775.283471	-775.280157
p_5caC_2ne_1_3w	3	-852.615860	-852.563796	-852.560783	-852.557937	-851.656898	-851.656898	-851.654052
p_5caC_2ne_1_4w	4	-929.053885	-928.999969	-928.996956	-928.992337	-928.032961	-928.032961	-928.028342
p_5caC_2ne_1_5w	5	-1005.504481	-1005.435309	-1005.432296	-1005.428769	-1004.407287	-1004.407287	-1004.403760
p_5caC_2ne_1_6w	6	-1081.953580	-1081.872193	-1081.869096	-1081.864313	-1080.782436	-1080.782352	-1080.777569
p_5caC_2ne_1_7w_2	7	-1158.385084	-1158.307338	-1158.303903	-1158.297881	-1157.158509	-1157.158087	-1157.152065
p_5caC_2ne_1_8w	8	-1234.817890	-1234.741151	-1234.737878	-1234.730971	-1233.533134	-1233.532874	-1233.525967
p_5caC_2ne_1_9w_2	9	-1311.240685	-1311.177183	-1311.173737	-1311.163481	-1309.911982	-1309.911549	-1309.901293

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

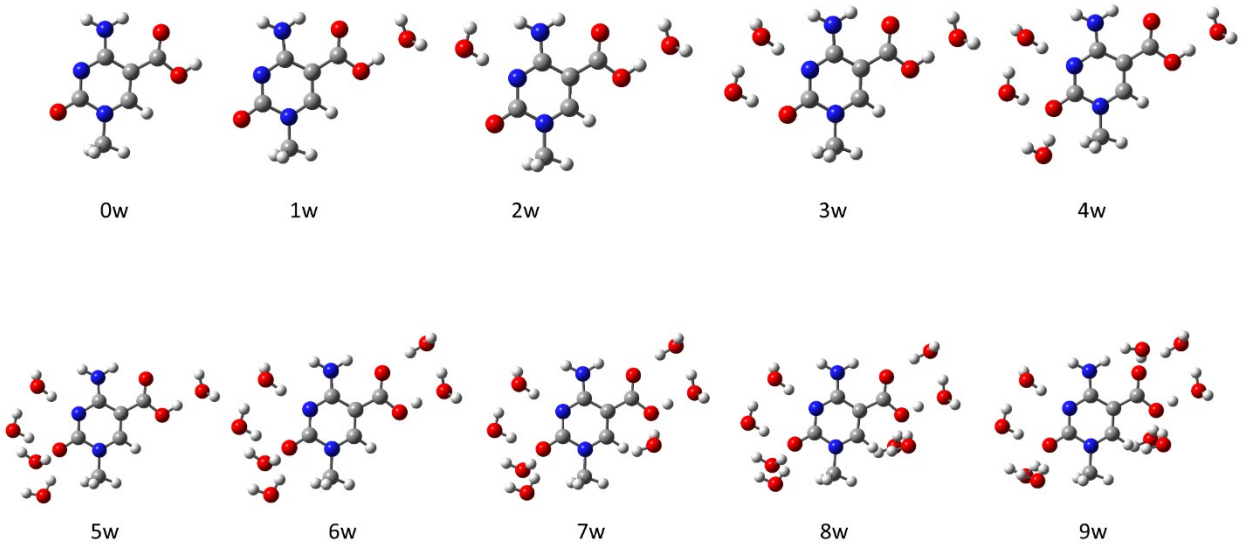


Figure 48: Molecule-water clusters for best conformer (2ne_1) for neutral 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 44.

Table 44: Gibbs free energies for solution phase optimized molecule-water clusters for best conformer (2ne_1) of neutral 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5caC_2ne_1_0w	0	-622.883851	-622.814480	-622.811467	-622.810444	-622.090307	-622.090307	-622.089284
5caC_2ne_1_1w	1	-699.334134	-699.253417	-699.250404	-699.248361	-698.468533	-698.468533	-698.466490
5caC_2ne_1_2w	2	-775.777380	-775.689688	-775.686675	-775.683912	-774.844647	-774.844647	-774.841884
5caC_2ne_1_3w	3	-852.233158	-852.127549	-852.124414	-852.120410	-851.220810	-851.220688	-851.216684
5caC_2ne_1_4w	4	-928.684217	-928.564232	-928.561071	-928.555567	-927.596506	-927.596358	-927.590854
5caC_2ne_1_5w	5	-1005.127241	-1005.000140	-1004.996893	-1004.990892	-1003.972457	-1003.972223	-1003.966222
5caC_2ne_1_6w	6	-1081.582779	-1081.438315	-1081.434429	-1081.427379	-1080.348855	-1080.347982	-1080.340932
5caC_2ne_1_7w	7	-1158.021510	-1157.873873	-1157.870139	-1157.860969	-1156.724522	-1156.723801	-1156.714631
5caC_2ne_1_8w	8	-1234.457889	-1234.305251	-1234.302126	-1234.293516	-1233.096291	-1233.096179	-1233.087569
5caC_2ne_1_9w	9	-1310.899359	-1310.742469	-1310.738687	-1310.728320	-1309.473468	-1309.472699	-1309.462332

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

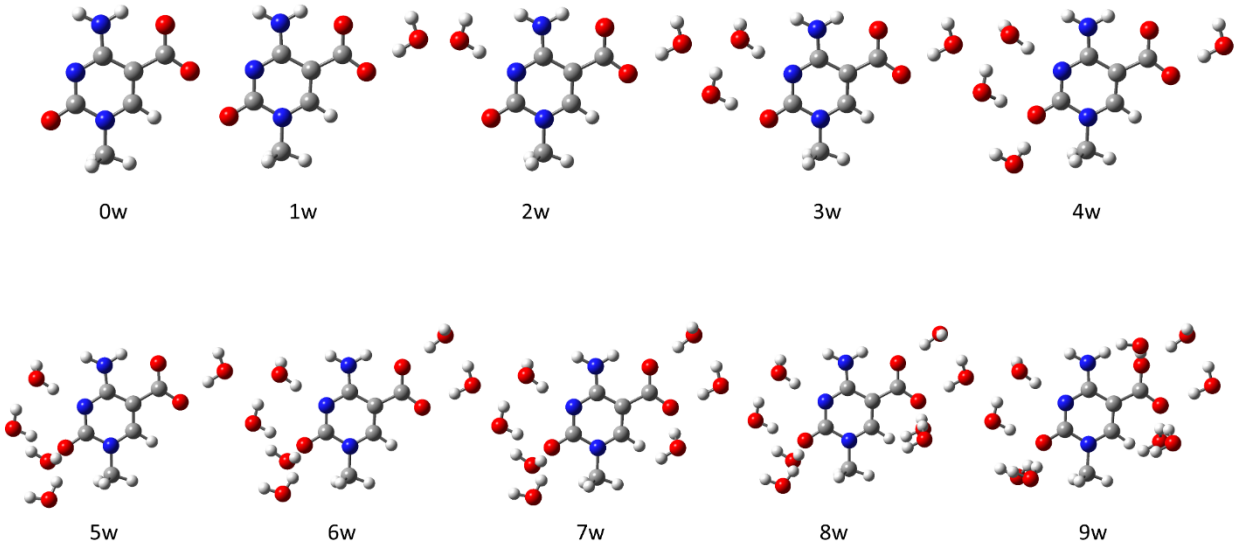


Figure 49: Molecule-water clusters for best conformer (2ne_1) for deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 45.

Table 45: Gibbs free energies for solution phase optimized molecule-water clusters for best conformer (2ne_1) of deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
d_5caC_2ne_1_0w	0	-622.350453	-622.371923	-622.368910	-622.367963	-621.645388	-621.645388	-621.644441
d_5caC_2ne_1_1w	1	-698.812412	-698.811102	-698.807802	-698.805420	-698.023119	-698.022832	-698.020450
d_5caC_2ne_1_2w	2	-775.262330	-775.248344	-775.245331	-775.241560	-774.399802	-774.399802	-774.396031
d_5caC_2ne_1_3w	3	-851.724445	-851.685916	-851.682711	-851.678193	-850.775459	-850.775267	-850.770749
d_5caC_2ne_1_4w	4	-928.179371	-928.122049	-928.118601	-928.113181	-927.150467	-927.150032	-927.144612
d_5caC_2ne_1_5w	5	-1004.627098	-1004.559840	-1004.555771	-1004.549143	-1003.528157	-1003.527101	-1003.520473
d_5caC_2ne_1_6w	6	-1081.084711	-1080.997571	-1080.993824	-1080.986129	-1079.904627	-1079.903893	-1079.896198
d_5caC_2ne_1_7w	7	-1157.532227	-1157.435031	-1157.431161	-1157.422048	-1156.282809	-1156.281952	-1156.272839
d_5caC_2ne_1_8w	8	-1233.978150	-1233.870077	-1233.866489	-1233.857212	-1232.658171	-1232.657596	-1232.648319
d_5caC_2ne_1_9w	9	-1310.423583	-1310.307407	-1310.303410	-1310.292409	-1309.035390	-1309.034406	-1309.023405

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

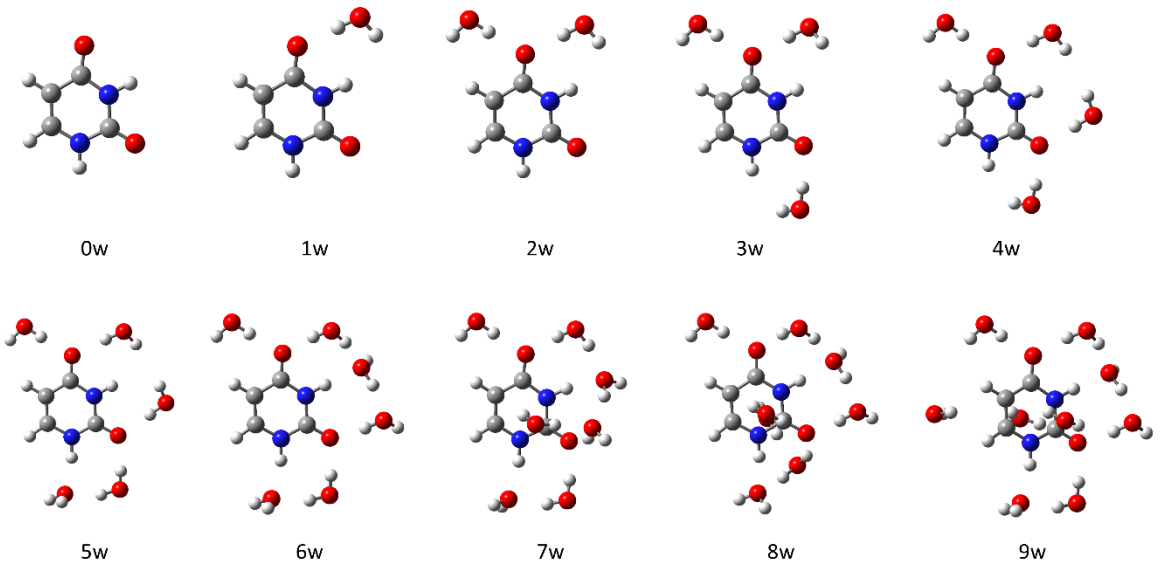


Figure 50: Molecule-water clusters for neutral uracil (**ura**) with relevant thermodynamic data provided in Table 46.

Table 46: Gibbs free energies for solution phase optimized molecule-water clusters of neutral uracil (**ura**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
ur1ne0w	0	-414.853148	-414.823824	-414.820810	-414.820810	-414.341605	-414.338591	-414.338591
ur1ne1w	1	-491.294057	-491.259556	-491.256543	-491.255896	-490.717960	-490.717960	-490.717313
ur1ne2w	2	-567.734593	-567.695925	-567.692911	-567.690932	-567.094664	-567.094663	-567.092684
ur1ne3w	3	-644.175206	-644.129394	-644.126381	-644.124347	-643.469584	-643.469584	-643.467550
ur1ne4w	4	-720.611031	-720.565114	-720.561919	-720.558362	-719.846700	-719.846518	-719.842961
ur1ne5w	5	-797.065695	-797.004795	-797.000919	-796.996114	-796.223410	-796.222547	-796.217742
ur1ne6w	6	-873.537215	-873.445878	-873.442189	-873.436212	-872.599190	-872.598514	-872.592537
ur1ne7w	7	-949.978380	-949.882650	-949.878318	-949.871993	-948.975231	-948.973912	-948.967587
ur1ne8w_2	8	-1026.418945	-1026.312374	-1026.309361	-1026.303210	-1025.346040	-1025.346040	-1025.339889
ur1ne9w	9	-1102.842829	-1102.745981	-1102.742422	-1102.732486	-1101.722694	-1101.722148	-1101.712212

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

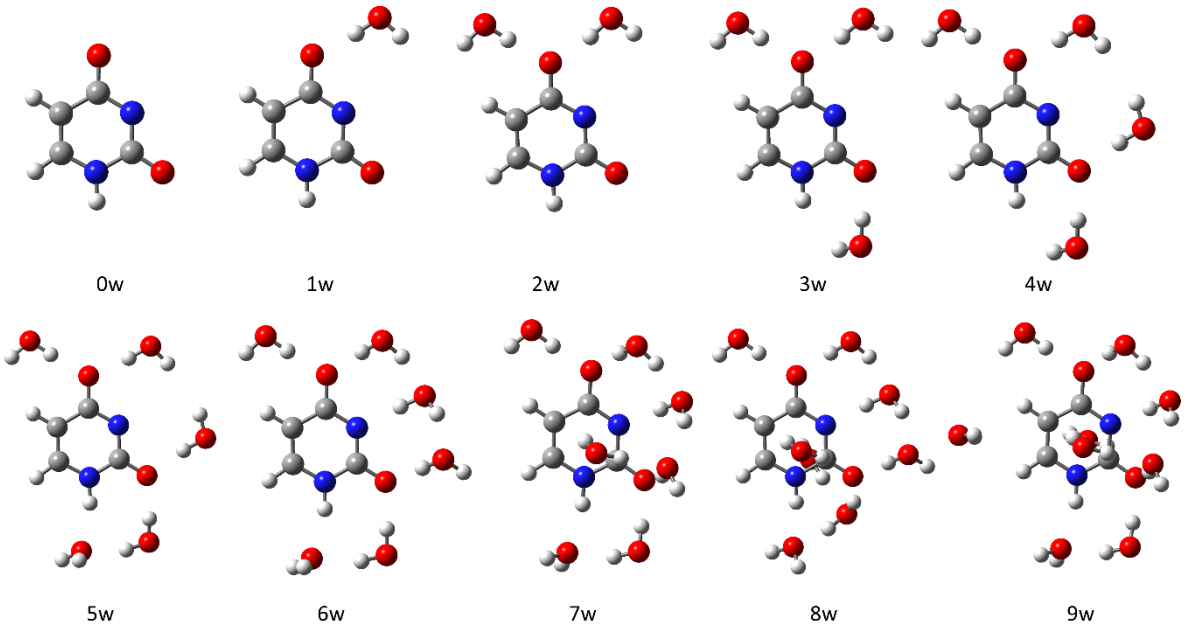


Figure 51: Molecule-water clusters for deprotonated uracil (**d_ura**) with relevant thermodynamic data provided in Table 47.

Table 47: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated uracil (**d_ura**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_ur1ne0w	0	-414.286895	-414.363953	-414.360940	-414.360940	-413.881133	-413.878120	-413.878120
d_ur1ne1w	1	-490.749271	-490.803223	-490.800088	-490.798573	-490.259030	-490.258908	-490.257393
d_ur1ne2w	2	-567.203369	-567.241292	-567.238262	-567.235728	-566.636918	-566.636901	-566.634367
d_ur1ne3w	3	-643.657889	-643.678945	-643.675932	-643.672530	-643.015098	-643.015098	-643.011696
d_ur1ne4w	4	-720.112699	-720.114782	-720.111559	-720.107237	-719.390139	-719.389929	-719.385607
d_ur1ne5w	5	-796.554246	-796.551293	-796.548018	-796.543026	-795.764632	-795.764370	-795.759378
d_ur1ne6w_2	6	-873.015275	-872.991548	-872.988380	-872.982258	-872.142242	-872.142087	-872.135965
d_ur1ne7w_2	7	-949.464166	-949.428745	-949.425520	-949.418819	-948.518549	-948.518337	-948.511636
d_ur1ne8w_2	8	-1025.910013	-1025.862486	-1025.859054	-1025.851456	-1024.892565	-1024.892146	-1024.884548
d_ur1ne9w_2	9	-1102.347278	-1102.296365	-1102.292619	-1102.283867	-1101.267707	-1101.266974	-1101.258222

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

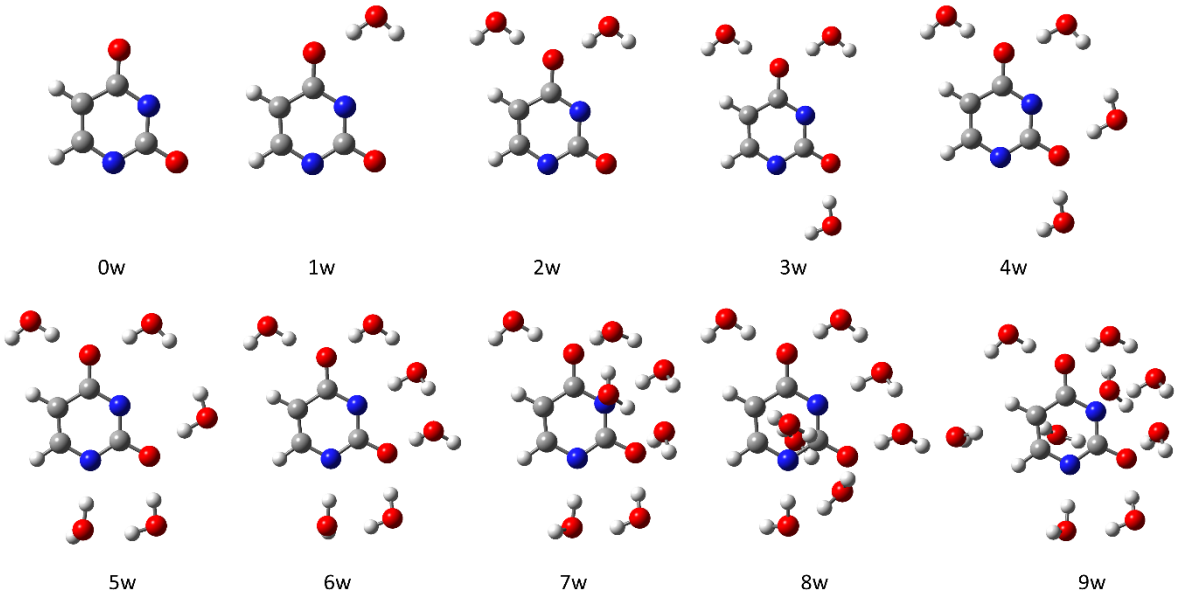


Figure 52: Molecule-water clusters for twice deprotonated uracil (**dd_ura**) with relevant thermodynamic data provided in Table 48.

Table 48: Gibbs free energies for solution phase optimized molecule-water clusters of twice deprotonated uracil (**dd_ura**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
dd_ur1ne0w	0	-413.570489	-413.887725	-413.884712	-413.884712	-413.398071	-413.395058	-413.395058
dd_ur1ne1w	1	-490.048920	-490.328471	-490.325414	-490.323996	-489.777704	-489.777660	-489.776242
dd_ur1ne2w	2	-566.517985	-566.768820	-566.765546	-566.763072	-566.157716	-566.157455	-566.154981
dd_ur1ne3w	3	-642.992386	-643.207978	-643.204913	-643.201409	-642.536311	-642.536259	-642.532755
dd_ur1ne4w	4	-719.461112	-719.644409	-719.641396	-719.637416	-718.912002	-718.912002	-718.908022
dd_ur1ne5w	5	-795.930766	-796.086432	-796.083150	-796.078193	-795.291623	-795.291354	-795.286397
dd_ur1ne6w_2	6	-872.398446	-872.529077	-872.525577	-872.519277	-871.671954	-871.671467	-871.665167
dd_ur1ne7w	7	-948.860043	-948.966946	-948.963494	-948.957722	-948.047956	-948.047517	-948.041745
dd_ur1ne8w	8	-1025.305977	-1025.400413	-1025.397138	-1025.389374	-1024.423256	-1024.422994	-1024.415230
dd_ur1ne9w	9	-1101.757801	-1101.837333	-1101.833889	-1101.824584	-1100.799148	-1100.798717	-1100.789412

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

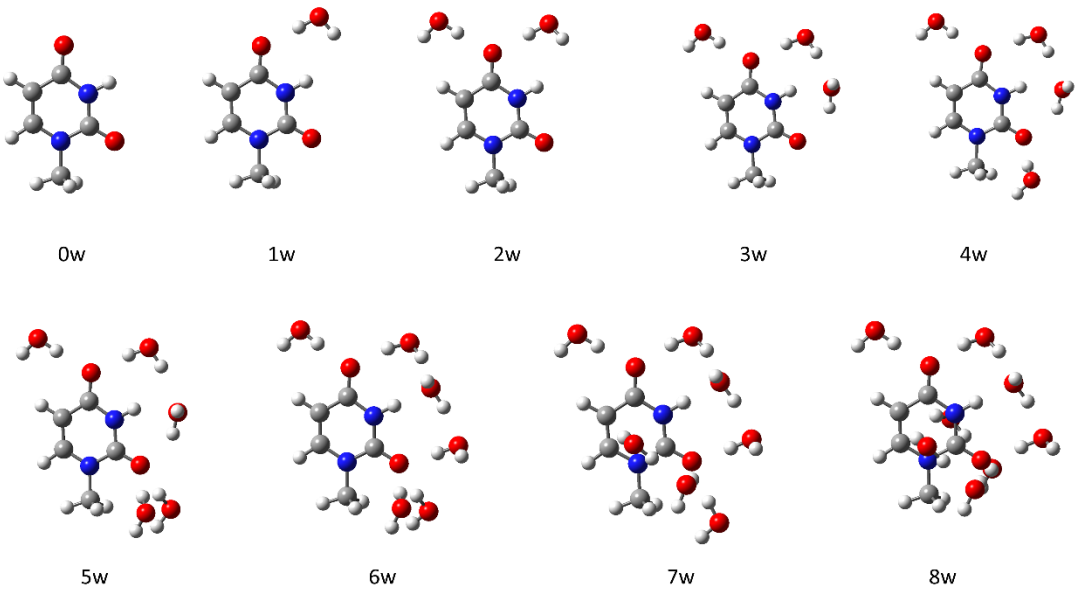


Figure 53: Molecule-water clusters for neutral 1-methyluracil (**1mU**) with relevant thermodynamic data provided in Table 49.

Table 49: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyluracil (**1mU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
meur1ne_0w	0	-454.168150	-454.110188	-454.107175	-454.107175	-453.565282	-453.565282	-453.565282
meur1ne_1w	1	-530.609757	-530.546818	-530.543805	-530.542653	-529.942393	-529.942393	-529.941241
meur1ne_2w	2	-607.050815	-606.983502	-606.980353	-606.977895	-606.319400	-606.319264	-606.316806
meur1ne_3w	3	-683.505590	-683.420097	-683.416798	-683.413668	-682.694767	-682.694481	-682.691351
meur1ne_4w	4	-759.946249	-759.855428	-759.851823	-759.847504	-759.071069	-759.070477	-759.066158
meur1ne_5w	5	-836.384177	-836.290220	-836.286978	-836.281519	-835.446831	-835.446602	-835.441143
meur1ne_6w	6	-912.840165	-912.730710	-912.727609	-912.720899	-911.824120	-911.824032	-911.817322
meur1ne_7w	7	-989.285030	-989.165992	-989.162670	-989.155906	-988.199043	-988.198734	-988.191970
meur1ne_8w	8	-1065.725438	-1065.600903	-1065.597591	-1065.590009	-1064.573538	-1064.573239	-1064.565657

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

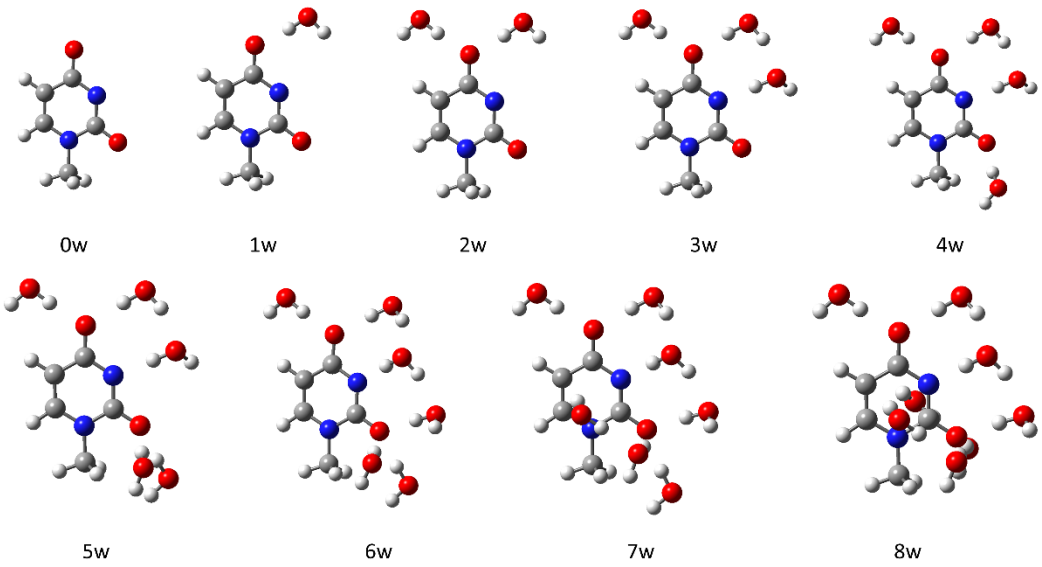


Figure 54: Molecule-water clusters for deprotonated 1-methyluracil (**1mU**) with relevant thermodynamic data provided in Table 50.

Table 50: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1-methyluracil (**1mU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
d_meur1ne_0w	0	-453.599507	-453.648968	-453.645955	-453.645858	-453.104056	-453.104056	-453.103959
d_meur1ne_1w	1	-530.062289	-530.089744	-530.086438	-530.083983	-529.483364	-529.483071	-529.480616
d_meur1ne_2w	2	-606.516999	-606.527104	-606.523740	-606.520884	-605.860392	-605.860041	-605.857185
d_meur1ne_3w	3	-682.978252	-682.965913	-682.962822	-682.959365	-682.236781	-682.236703	-682.233246
d_meur1ne_4w	4	-759.427943	-759.404277	-759.400414	-759.395006	-758.615876	-758.615026	-758.609618
d_meur1ne_5w	5	-835.876357	-835.840627	-835.836816	-835.830548	-834.992754	-834.991956	-834.985688
d_meur1ne_6w	6	-912.330175	-912.279735	-912.276247	-912.269166	-911.370174	-911.369699	-911.362618
d_meur1ne_7w	7	-988.780537	-988.716556	-988.712579	-988.705173	-987.746013	-987.745049	-987.737643
d_meur1ne_8w	8	-1065.228456	-1065.153182	-1065.149040	-1065.140532	-1064.121527	-1064.120398	-1064.111890

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

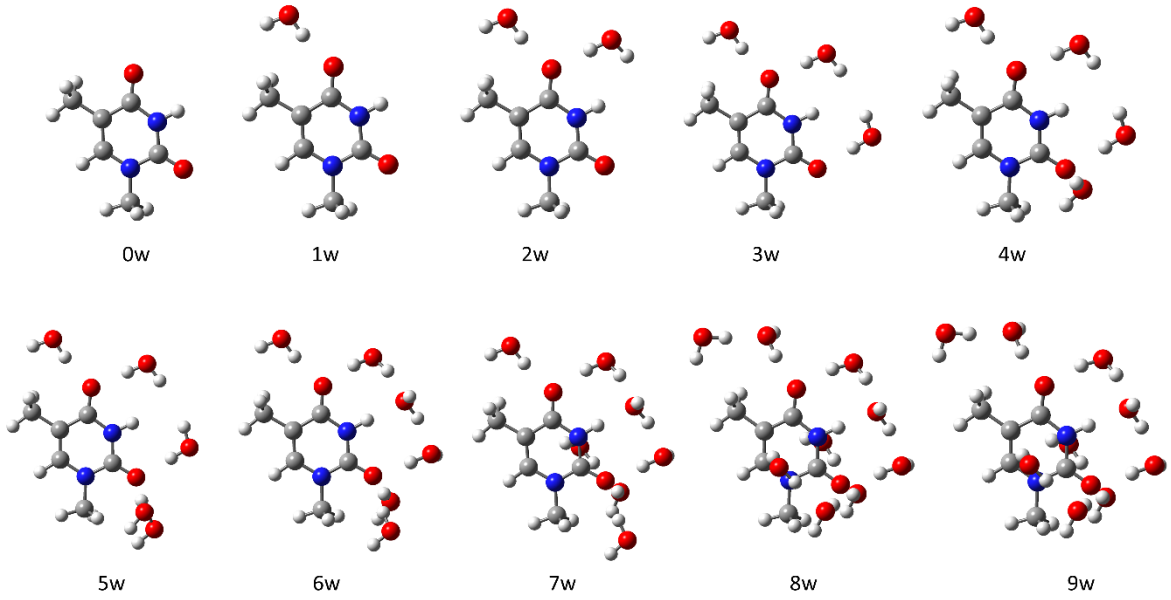


Figure 55: Molecule-water clusters for neutral 1,5-dimethyluracil or 1-methylthymine (**1m5mU**) with relevant thermodynamic data provided in Table 51.

Table 51: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1,5-dimethyluracil or 1-methylthymine (**1m5mU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
15meur1ne1w_0w	0	-493.493867	-493.409096	-493.406083	-493.405953	-492.802285	-492.799272	-492.799142
15meur1ne1w_1w	1	-569.937513	-569.846036	-569.842234	-569.841099	-569.179538	-569.178749	-569.177614
15meur1ne1w_2w	2	-646.377249	-646.282550	-646.279076	-646.276535	-645.556484	-645.556023	-645.553482
15meur1ne1w_3w	3	-722.815525	-722.715782	-722.712660	-722.709733	-721.930751	-721.930642	-721.927715
15meur1ne1w_4w_4	4	-799.253310	-799.151217	-799.147568	-799.143742	-798.307150	-798.306514	-798.302688
15meur1ne1w_5w	5	-875.690246	-875.586444	-875.583345	-875.578265	-874.683418	-874.683332	-874.678252
15meur1ne1w_6w	6	-952.168630	-952.027733	-952.024538	-952.018439	-951.058600	-951.058418	-951.052319
15meur1ne1w_7w	7	-1028.611917	-1028.464653	-1028.461123	-1028.454354	-1027.435328	-1027.434811	-1027.428042
15meur1ne1w_8w	8	-1105.057904	-1104.901323	-1104.897225	-1104.889566	-1103.810931	-1103.809846	-1103.802187
15meur1ne1w_9w	9	-1181.506565	-1181.336154	-1181.332900	-1181.324212	-1180.184545	-1180.184304	-1180.175616

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

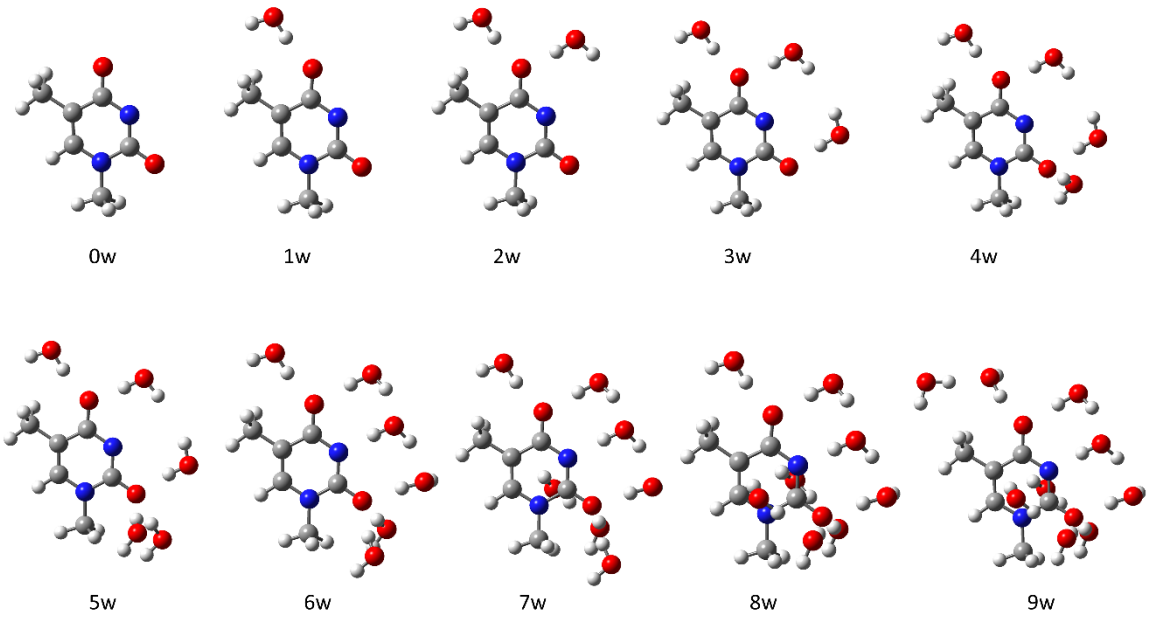


Figure 56: Molecule-water clusters for deprotonated 1,5-dimethyluracil or 1-methylthymine (**1m5mU**) with relevant thermodynamic data provided in Table 52.

Table 52: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1,5-dimethyluracil or 1-methylthymine (**1m5mU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
d_15meur1ne_0w	0	-492.924037	-492.946425	-492.943412	-492.943214	-492.340283	-492.337270	-492.337072
d_15meur1ne_1w	1	-569.381538	-569.384589	-569.381099	-569.379846	-568.718199	-568.717722	-568.716469
d_15meur1ne_2w	2	-645.839824	-645.824846	-645.821205	-645.817992	-645.097018	-645.096390	-645.093177
d_15meur1ne_3w	3	-722.298383	-722.261926	-722.258274	-722.254235	-721.472737	-721.472098	-721.468059
d_15meur1ne_4w	4	-798.745890	-798.698006	-798.694702	-798.690277	-797.849393	-797.849102	-797.844677
d_15meur1ne_5w	5	-875.193037	-875.136647	-875.133106	-875.126743	-874.228163	-874.227635	-874.221272
d_15meur1ne_6w	6	-951.654117	-951.575961	-951.572447	-951.565726	-950.604854	-950.604353	-950.597632
d_15meur1ne_7w	7	-1028.105452	-1028.011527	-1028.008321	-1028.001449	-1026.979515	-1026.979322	-1026.972450
d_15meur1ne_8w	8	-1104.555557	-1104.449855	-1104.445834	-1104.437535	-1103.356203	-1103.355195	-1103.346896
d_15meur1ne_9w	9	-1181.010349	-1180.885033	-1180.881340	-1180.872499	-1179.729775	-1179.729095	-1179.720254

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

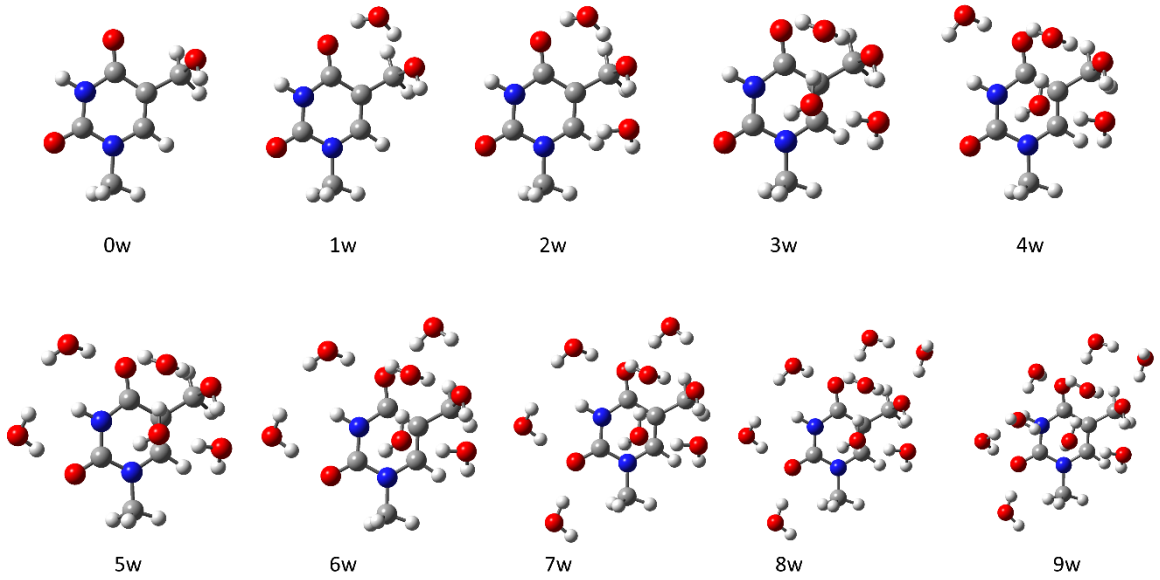


Figure 57: Molecule-water clusters for neutral 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with relevant thermodynamic data provided in Table 53.

Table 53: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5-hydroxymethyluracil (**1m5hmU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5hm_meur1ne_4_0w	0	-568.706520	-568.630894	-568.627881	-568.627135	-567.967419	-567.964407	-567.963661
5hm_meur1ne_4_1w	1	-645.154850	-645.068174	-645.064970	-645.063392	-644.342835	-644.342644	-644.341066
5hm_meur1ne_4_2w	2	-721.596787	-721.504430	-721.501022	-721.498223	-720.718981	-720.718586	-720.715787
5hm_meur1ne_4_3w	3	-798.064472	-797.942842	-797.939829	-797.938143	-797.093956	-797.093956	-797.092270
5hm_meur1ne_4_4w	4	-874.505003	-874.379075	-874.376062	-874.373574	-873.470703	-873.470703	-873.468215
5hm_meur1ne_4_5w	5	-950.945792	-950.814546	-950.811484	-950.807477	-949.847200	-949.847151	-949.843144
5hm_meur1ne_4_6w	6	-1027.388873	-1027.249831	-1027.246817	-1027.243361	-1026.221441	-1026.221440	-1026.217984
5hm_meur1ne_4_7w	7	-1103.825380	-1103.687105	-1103.683629	-1103.677994	-1102.600503	-1102.600040	-1102.594405
5hm_meur1ne_4_8w	8	-1180.273917	-1180.125199	-1180.121903	-1180.114638	-1178.977403	-1178.977120	-1178.969855
5hm_meur1ne_4_9w_2	9	-1256.740329	-1256.564987	-1256.561974	-1256.555277	-1255.352557	-1255.352557	-1255.345860

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

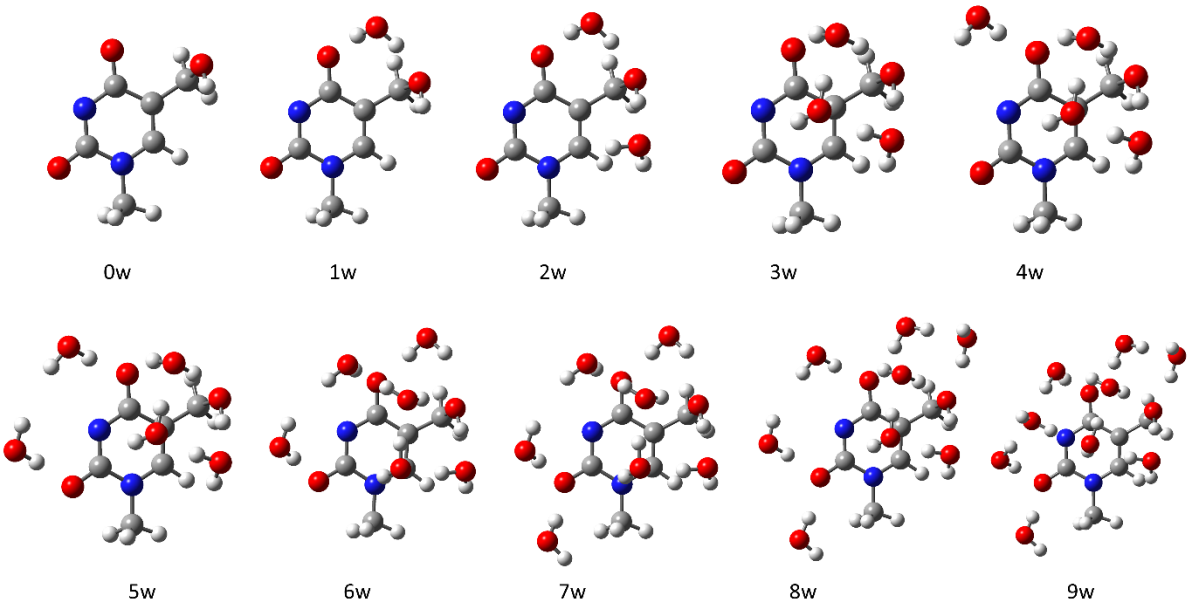


Figure 58: Molecule-water clusters for deprotonated 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with relevant thermodynamic data provided in Table 54.

Table 54: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1-methyl-5-hydroxymethyluracil (**1m5hmU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_5hm_meur1ne_4_0w	0	-568.142290	-568.170069	-568.167056	-568.166210	-567.506417	-567.503404	-567.502558
d_5hm_meur1ne_4_1w	1	-644.604707	-644.608716	-644.605588	-644.603932	-643.882769	-643.882654	-643.880998
d_5hm_meur1ne_4_2w	2	-721.048889	-721.044779	-721.041658	-721.038819	-720.258432	-720.258324	-720.255485
d_5hm_meur1ne_4_3w	3	-797.525807	-797.483347	-797.480334	-797.478703	-796.633489	-796.633489	-796.631858
d_5hm_meur1ne_4_4w	4	-873.983491	-873.923079	-873.920030	-873.916544	-873.011681	-873.011645	-873.008159
d_5hm_meur1ne_4_5w	5	-950.439439	-950.360596	-950.357276	-950.353146	-949.388074	-949.387767	-949.383637
d_5hm_meur1ne_4_6w	6	-1026.893589	-1026.796699	-1026.793686	-1026.790060	-1025.762866	-1025.762866	-1025.759240
d_5hm_meur1ne_4_7w	7	-1103.338694	-1103.233310	-1103.229903	-1103.225480	-1102.140390	-1102.139996	-1102.135573
d_5hm_meur1ne_4_8w	8	-1179.789219	-1179.675669	-1179.671416	-1179.663953	-1178.522064	-1178.520824	-1178.513361
d_5hm_meur1ne_4_9w	9	-1256.237461	-1256.109625	-1256.106428	-1256.100361	-1254.893746	-1254.893562	-1254.887495

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

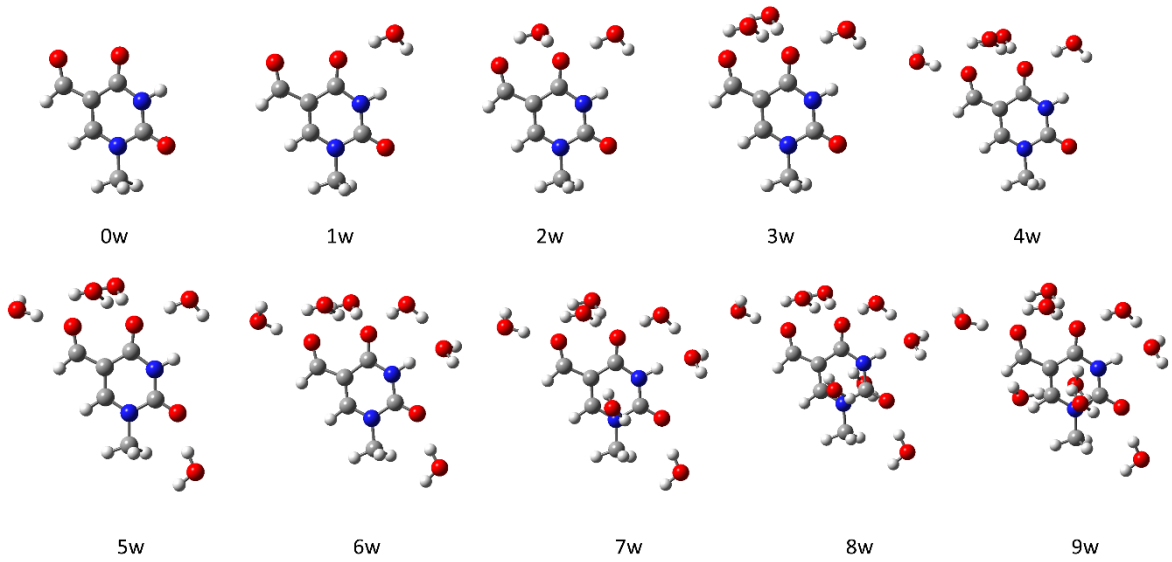


Figure 59: Molecule-water clusters for neutral 1-methyl-5-formyluracil (**1m5fU**) with relevant thermodynamic data provided in Table 55.

Table 55: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5-formyluracil (**1m5fU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5f_meur1ne1w_2_0w	0	-567.492056	-567.439896	-567.436883	-567.436579	-566.774994	-566.771981	-566.771677
5f_meur1ne1w_2_1w	1	-643.933195	-643.875901	-643.872888	-643.871473	-643.151892	-643.151892	-643.150477
5f_meur1ne1w_2_2w	2	-720.377739	-720.314187	-720.311031	-720.307561	-719.529436	-719.529293	-719.525823
5f_meur1ne1w_2_3w	3	-796.823342	-796.750792	-796.747357	-796.742952	-795.904627	-795.904205	-795.899800
5f_meur1ne1w_2_4w	4	-873.263974	-873.186843	-873.183463	-873.177366	-872.280668	-872.280301	-872.274204
5f_meur1ne1w_2_5w	5	-949.701132	-949.622927	-949.618844	-949.611389	-948.658757	-948.657687	-948.650232
5f_meur1ne1w_2_6w	6	-1026.155924	-1026.059015	-1026.055338	-1026.047844	-1025.031980	-1025.031316	-1025.023822
5f_meur1ne1w_2_7w	7	-1102.591088	-1102.494449	-1102.490639	-1102.480912	-1101.408766	-1101.407969	-1101.398242
5f_meur1ne1w_2_8w	8	-1179.024859	-1178.927398	-1178.923088	-1178.912578	-1177.782767	-1177.781470	-1177.770960
5f_meur1ne1w_2_9w_2	9	-1255.469207	-1255.361690	-1255.358338	-1255.347843	-1254.156492	-1254.156153	-1254.145658

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

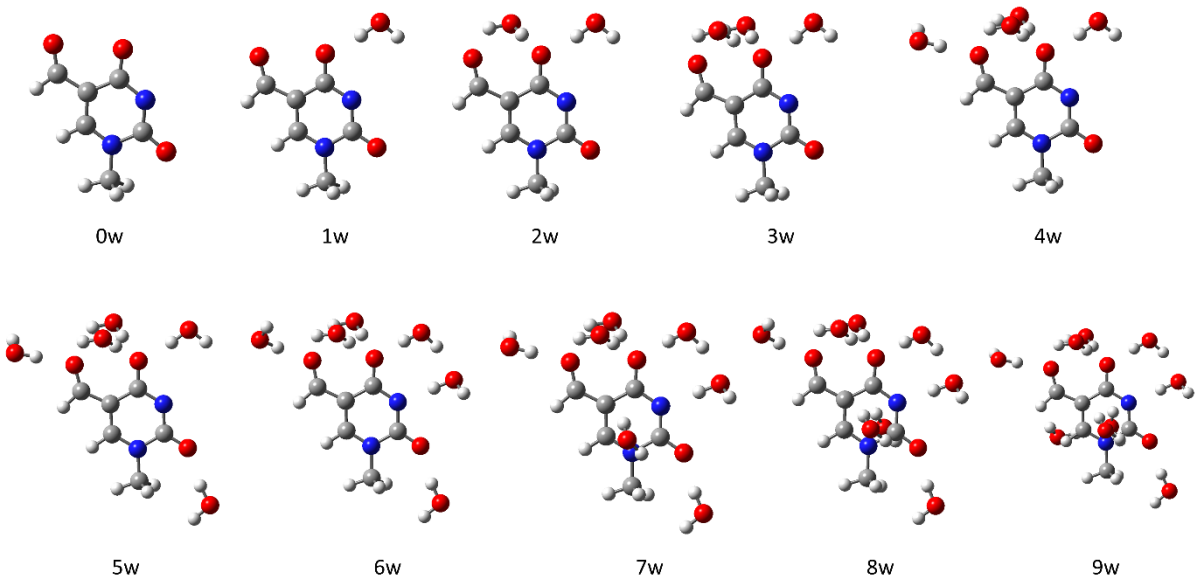


Figure 60: Molecule-water clusters for deprotonated 1-methyl-5-formyluracil (**1m5fU**) with relevant thermodynamic data provided in Table 56.

Table 56: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1-methyl-5-formyluracil (**1m5fU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_5f_meur1ne1w_2_0w	0	-566.939029	-566.984610	-566.981597	-566.981103	-566.318199	-566.315186	-566.314692
d_5f_meur1ne1w_2_1w	1	-643.398941	-643.423690	-643.420470	-643.418157	-642.696240	-642.696033	-642.693720
d_5f_meur1ne1w_2_2w	2	-719.854609	-719.862593	-719.859131	-719.855329	-719.074108	-719.073659	-719.069857
d_5f_meur1ne1w_2_3w	3	-796.309728	-796.300793	-796.297548	-796.292809	-795.450637	-795.450405	-795.445666
d_5f_meur1ne1w_2_4w	4	-872.759736	-872.739624	-872.735301	-872.728122	-871.828826	-871.827516	-871.820337
d_5f_meur1ne1w_2_5w	5	-949.207712	-949.175741	-949.171683	-949.163433	-948.206191	-948.205146	-948.196896
d_5f_meur1ne1w_2_6w	6	-1025.663136	-1025.610640	-1025.606519	-1025.598908	-1024.579106	-1024.577998	-1024.570387
d_5f_meur1ne1w_2_7w	7	-1102.105190	-1102.049111	-1102.043460	-1102.033561	-1100.958627	-1100.955989	-1100.946090
d_5f_meur1ne1w_2_8w	8	-1178.546075	-1178.479057	-1178.475474	-1178.465586	-1177.329055	-1177.328485	-1177.318597
d_5f_meur1ne1w_2_9w	9	-1254.995370	-1254.917793	-1254.913604	-1254.901450	-1253.706999	-1253.705823	-1253.693669

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

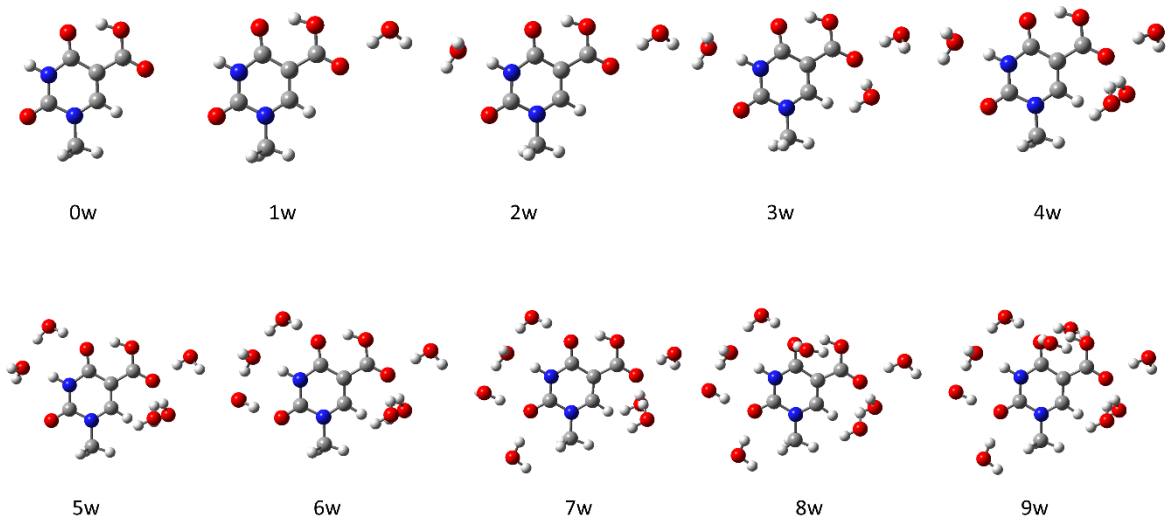


Figure 61: Molecule-water clusters for neutral 1-methyl-5-carboxyuracil (**1m5caU**) with relevant thermodynamic data provided in Table 57.

Table 57: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5-carboxyuracil (**1m5caU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
5ca_meur2ne_2_0w	0	-642.762212	-642.700763	-642.697750	-642.696906	-641.979722	-641.976709	-641.975865
5ca_meur2ne_2_1w	1	-719.205842	-719.135597	-719.132583	-719.130825	-718.354402	-718.354401	-718.352643
5ca_meur2ne_2_2w2	2	-795.654683	-795.574825	-795.571398	-795.567816	-794.733361	-794.732947	-794.729365
5ca_meur2ne_2_3w	3	-872.095732	-872.007799	-872.004729	-872.000554	-871.107027	-871.106970	-871.102795
5ca_meur2ne_2_4w	4	-948.533987	-948.442905	-948.439716	-948.434438	-947.482596	-947.482420	-947.477142
5ca_meur2ne_2_5w	5	-1024.977256	-1024.879490	-1024.876229	-1024.870191	-1023.858005	-1023.857757	-1023.851719
5ca_meur2ne_2_6w	6	-1101.434028	-1101.316450	-1101.312660	-1101.306299	-1100.233328	-1100.232551	-1100.226190
5ca_meur2ne_2_7w	7	-1177.872187	-1177.750795	-1177.747767	-1177.740640	-1176.609003	-1176.608988	-1176.601861
5ca_meur2ne_2_8w	8	-1254.311897	-1254.188361	-1254.183899	-1254.174642	-1252.986415	-1252.984966	-1252.975709
5ca_meur2ne_2_9w	9	-1330.748771	-1330.623043	-1330.619358	-1330.609344	-1329.361863	-1329.361191	-1329.351177

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

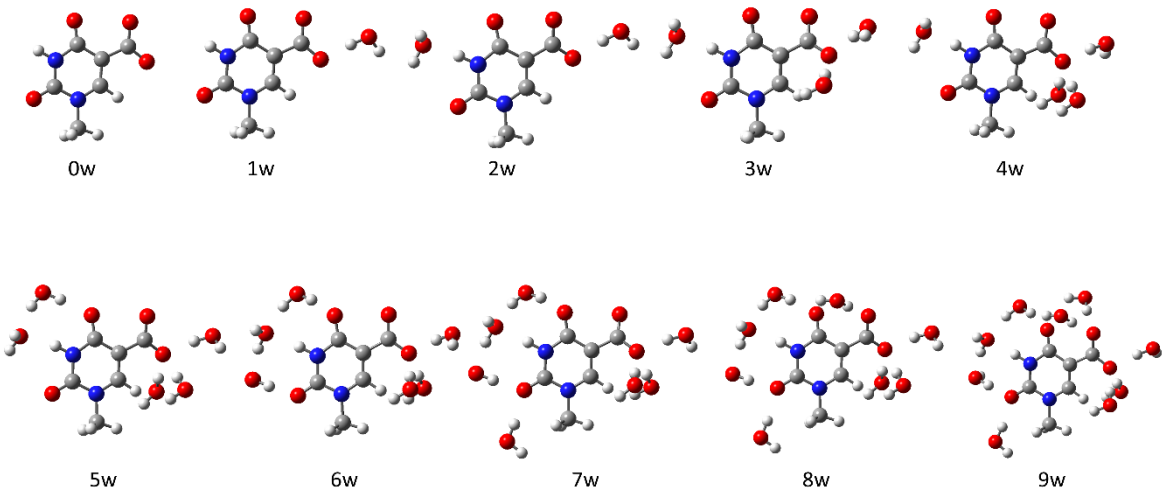


Figure 62: Molecule-water clusters for deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) with relevant thermodynamic data provided in Table 58.

Table 58: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_5ca_meur2ne_2_0w	0	-642.204108	-642.251799	-642.248786	-642.248057	-641.527340	-641.524328	-641.523599
d_5ca_meur2ne_2_1w	1	-718.661778	-718.691604	-718.687964	-718.684432	-717.906703	-717.906076	-717.902544
d_5ca_meur2ne_2_2w	2	-795.102920	-795.126887	-795.123874	-795.119719	-794.282179	-794.282179	-794.278024
d_5ca_meur2ne_2_3w	3	-871.560762	-871.563672	-871.560517	-871.556022	-870.658803	-870.658661	-870.654166
d_5ca_meur2ne_2_4w_2	4	-948.011142	-948.000115	-947.996923	-947.991496	-947.035511	-947.035332	-947.029905
d_5ca_meur2ne_2_5w	5	-1024.462317	-1024.437473	-1024.434325	-1024.428492	-1023.411244	-1023.411109	-1023.405276
d_5ca_meur2ne_2_6w	6	-1100.924248	-1100.875037	-1100.871768	-1100.865417	-1099.787045	-1099.786789	-1099.780438
d_5ca_meur2ne_2_7w	7	-1177.369484	-1177.313349	-1177.309271	-1177.300216	-1176.165856	-1176.164791	-1176.155736
d_5ca_meur2ne_2_8w	8	-1253.824500	-1253.748246	-1253.744976	-1253.736208	-1252.539757	-1252.539500	-1252.530732
d_5ca_meur2ne_2_9w_2	9	-1330.275553	-1330.188620	-1330.183904	-1330.173317	-1328.919209	-1328.917506	-1328.906919

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

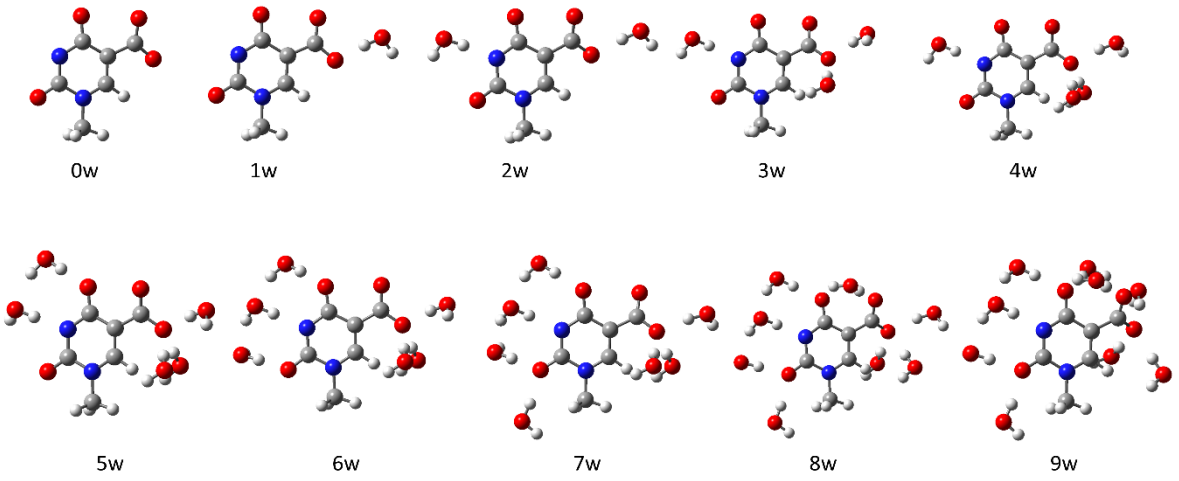


Figure 63: Molecule-water clusters for twice deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) with relevant thermodynamic data provided in Table 59.

Table 59: Gibbs free energies for solution phase optimized molecule-water clusters of twice deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
dd_5ca_meur2ne_2_0w_2	0	-641.521623	-641.788408	-641.785395	-641.784528	-641.061744	-641.058731	-641.057864
dd_5ca_meur2ne_2_1w	1	-717.987611	-718.228615	-718.225332	-718.221968	-717.441582	-717.441312	-717.437948
dd_5ca_meur2ne_2_2w	2	-794.455991	-794.668345	-794.665292	-794.661286	-793.819463	-793.819423	-793.815417
dd_5ca_meur2ne_2_3w	3	-870.921482	-871.106665	-871.103438	-871.098677	-870.197202	-870.196988	-870.192227
dd_5ca_meur2ne_2_4w	4	-947.379126	-947.542650	-947.539411	-947.533574	-946.572943	-946.572717	-946.566880
dd_5ca_meur2ne_2_5w	5	-1023.846981	-1023.981744	-1023.978435	-1023.972348	-1022.950202	-1022.949906	-1022.943819
dd_5ca_meur2ne_2_6w	6	-1100.309763	-1100.418602	-1100.415332	-1100.408473	-1099.326111	-1099.325854	-1099.318995
dd_5ca_meur2ne_2_7w	7	-1176.763389	-1176.855685	-1176.852375	-1176.844252	-1175.703807	-1175.703510	-1175.695387
dd_5ca_meur2ne_2_8w	8	-1253.227224	-1253.294355	-1253.290776	-1253.281592	-1252.081046	-1252.080480	-1252.071296
dd_5ca_meur2ne_2_9w_2	9	-1329.684897	-1329.735549	-1329.731263	-1329.721517	-1328.460338	-1328.459065	-1328.449319

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

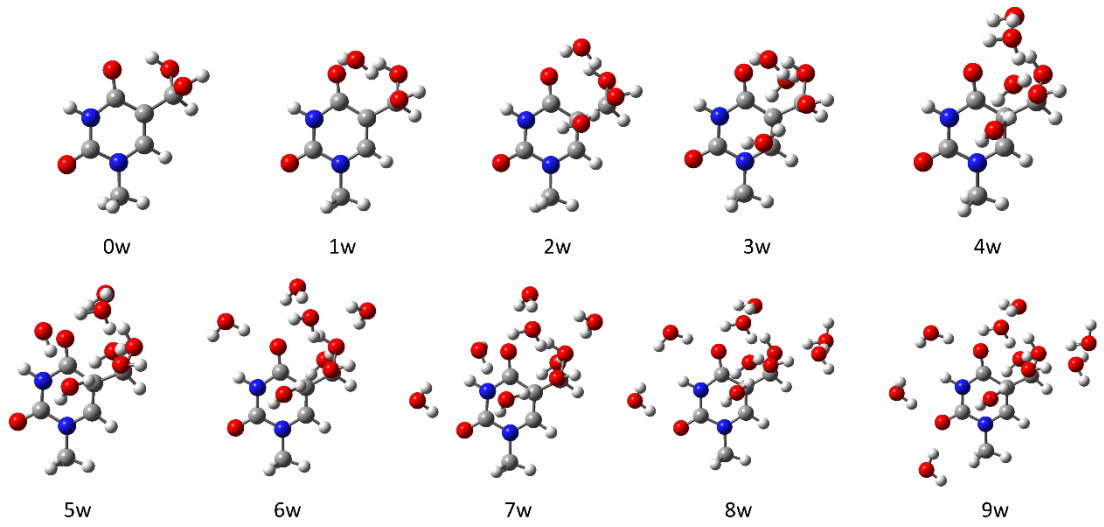


Figure 64: Molecule-water clusters for neutral 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) with relevant thermodynamic data provided in Table 60.

Table 60: Gibbs free energies for solution phase optimized molecule-water clusters of neutral 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
1m5dhmUa_5_0w	0	-643.946029	-643.866353	-643.863340	-643.862701	-643.145572	-643.142559	-643.141920
1m5dhmUa_5_1w	1	-720.393134	-720.303381	-720.300031	-720.298122	-719.520880	-719.520543	-719.518634
1m5dhmUa_5_2w	2	-796.834119	-796.738811	-796.735798	-796.733591	-795.895686	-795.895686	-795.893479
1m5dhmUa_5_3w	3	-873.275434	-873.176961	-873.173948	-873.170002	-872.273613	-872.273613	-872.269667
1m5dhmUa_5_4w	4	-949.733123	-949.613002	-949.609858	-949.605172	-948.647015	-948.646884	-948.642198
1m5dhmUa_5_5w_3	5	-1026.184910	-1026.049890	-1026.046701	-1026.042414	-1025.022478	-1025.022302	-1025.018015
1m5dhmUa_5_6w	6	-1102.607604	-1102.486831	-1102.483161	-1102.475622	-1101.402017	-1101.401360	-1101.393821
1m5dhmUa_5_7w_2	7	-1179.061813	-1178.920009	-1178.916948	-1178.910857	-1177.774205	-1177.774157	-1177.768066
1m5dhmUa_5_8w	8	-1255.524809	-1255.360189	-1255.356930	-1255.348450	-1254.151316	-1254.151070	-1254.142590
1m5dhmUa_5_9w	9	-1331.962548	-1331.797208	-1331.793075	-1331.782492	-1330.529820	-1330.528700	-1330.518117

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

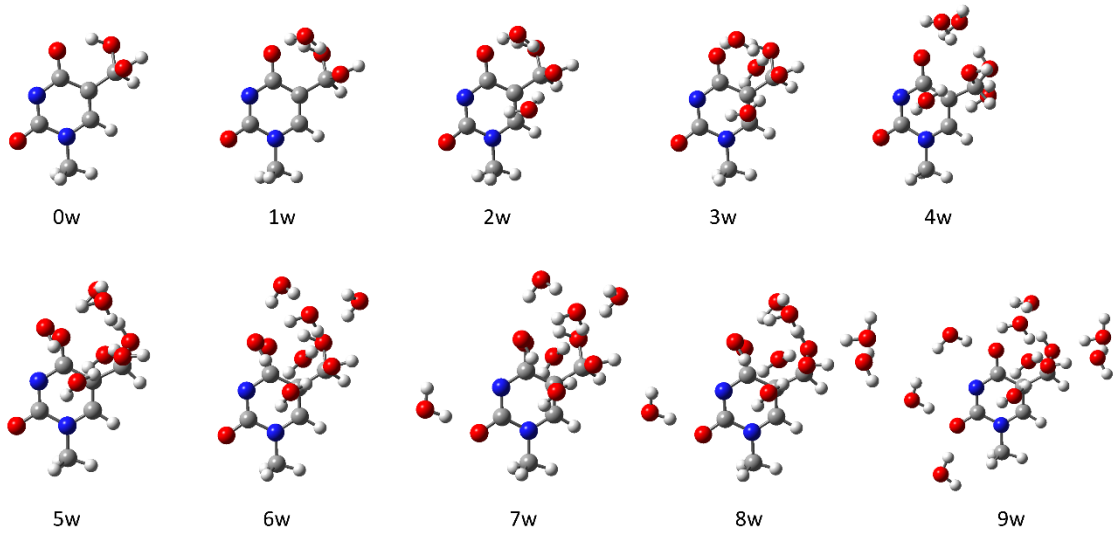


Figure 65: Molecule-water clusters for deprotonated 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) with relevant thermodynamic data provided in Table 61.

Table 61: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
d_1m5dhmUa_5_0w	0	-643.391606	-643.408632	-643.405619	-643.404772	-642.687833	-642.684820	-642.683973
d_1m5dhmUa_5_1w	1	-719.850845	-719.846082	-719.843069	-719.841664	-719.063127	-719.063127	-719.061722
d_1m5dhmUa_5_2w	2	-796.301091	-796.283163	-796.280150	-796.278042	-795.438937	-795.438937	-795.436829
d_1m5dhmUa_5_3w	3	-872.751918	-872.721896	-872.718751	-872.714634	-871.816722	-871.816590	-871.812473
d_1m5dhmUa_5_4w	4	-949.204081	-949.156662	-949.153121	-949.148177	-948.189310	-948.188782	-948.183838
d_1m5dhmUa_5_5w	5	-1025.666959	-1025.593715	-1025.590471	-1025.586249	-1024.564104	-1024.563873	-1024.559651
d_1m5dhmUa_5_6w	6	-1102.117993	-1102.034126	-1102.030963	-1102.025390	-1100.944466	-1100.944316	-1100.938743
d_1m5dhmUa_5_7w	7	-1178.572388	-1178.473883	-1178.469871	-1178.462071	-1177.323412	-1177.322413	-1177.314613
d_1m5dhmUa_5_8w_2	8	-1255.034488	-1254.909684	-1254.906147	-1254.897966	-1253.694904	-1253.694380	-1253.686199
d_1m5dhmUa_5_9w	9	-1331.473925	-1331.343567	-1331.339862	-1331.329626	-1330.069941	-1330.069249	-1330.059013

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

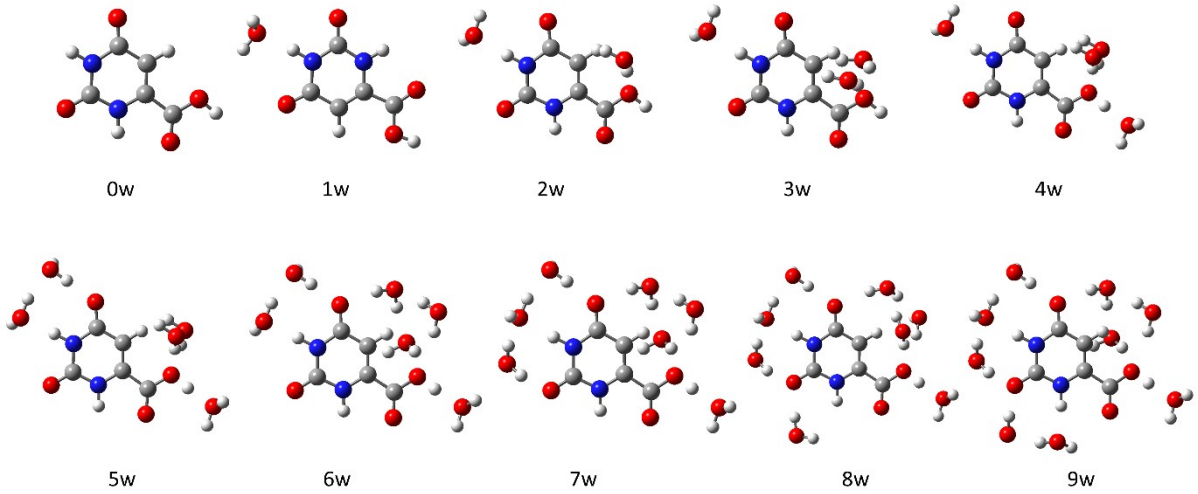


Figure 66: Molecule-water clusters for neutral orotic acid (**oro**) with relevant thermodynamic data provided in Table 62.

Table 62: Gibbs free energies for solution phase optimized molecule-water clusters of neutral orotic acid (**oro**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
oro5ne_0w	0	-603.440099	-603.404453	-603.401440	-603.400632	-602.746224	-602.743212	-602.742404
oro5ne_1w	1	-679.887160	-679.841500	-679.838487	-679.835747	-679.123125	-679.123125	-679.120385
oro5ne_2w	2	-756.325408	-756.275037	-756.271925	-756.268730	-755.497332	-755.497233	-755.494038
oro5ne_3w_2	3	-832.762210	-832.709562	-832.706455	-832.701468	-831.872165	-831.872071	-831.867084
oro5ne_4w	4	-909.219062	-909.150521	-909.147507	-909.142168	-908.251177	-908.251176	-908.245837
oro5ne_5w	5	-985.674671	-985.588600	-985.585544	-985.579095	-984.627673	-984.627630	-984.621181
oro5ne_6w	6	-1062.130696	-1062.026030	-1062.022350	-1062.016041	-1061.003212	-1061.002545	-1060.996236
oro5ne_7w	7	-1138.579817	-1138.465315	-1138.461278	-1138.453657	-1137.380651	-1137.379627	-1137.372006
oro5ne_8w	8	-1215.022185	-1214.901257	-1214.897488	-1214.888648	-1213.757230	-1213.756474	-1213.747634
oro5ne_9w	9	-1291.484700	-1291.337634	-1291.333584	-1291.324775	-1290.130729	-1290.129692	-1290.120883

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

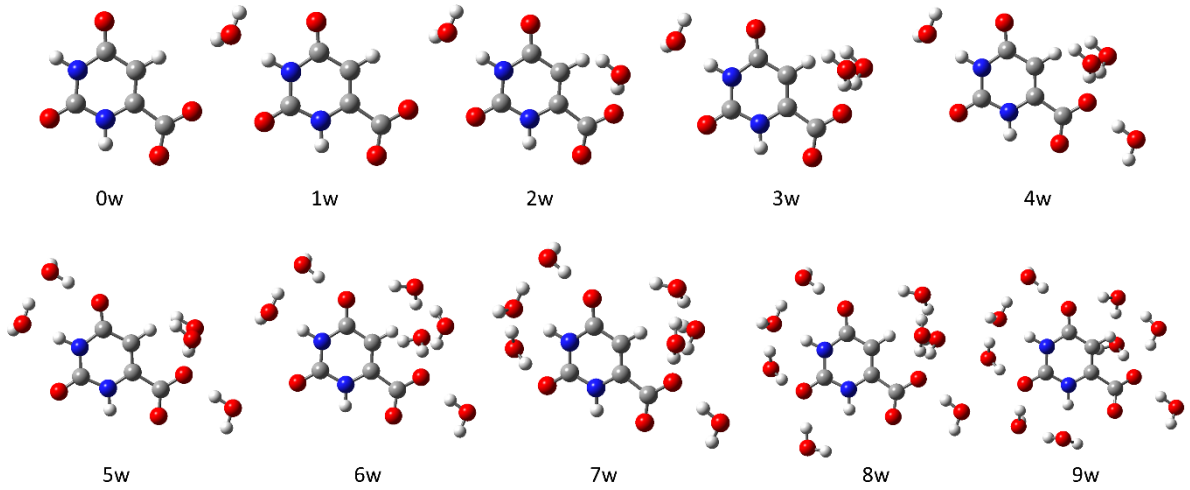


Figure 67: Molecule-water clusters for deprotonated (at COO*H*) orotic acid (**oro**) with relevant thermodynamic data provided in Table 63.

Table 63: Gibbs free energies for solution phase optimized molecule-water clusters of deprotonated (at COO*H*) orotic acid (**oro**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh-30}$[b]	$G_{DFT,qh-100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh-30}$[b,c,d]	$G_{CCSD(T)/CBS,qh-100}$[b,c,d]
d_COOH_oro5ne_0w	0	-602.926242	-602.971400	-602.968387	-602.967884	-602.308617	-602.305604	-602.305101
d_COOH_oro5ne_1w	1	-679.368149	-679.408384	-679.405371	-679.402782	-678.685653	-678.685653	-678.683064
d_COOH_oro5ne_2w	2	-755.823309	-755.844358	-755.841345	-755.838282	-755.062186	-755.062186	-755.059123
d_COOH_oro5ne_3w	3	-832.275206	-832.282650	-832.279598	-832.274949	-831.440574	-831.440535	-831.435886
d_COOH_oro5ne_4w	4	-908.728619	-908.721864	-908.717882	-908.711220	-907.818768	-907.817799	-907.811137
d_COOH_oro5ne_5w	5	-985.189783	-985.158841	-985.155286	-985.148239	-984.194181	-984.193639	-984.186592
d_COOH_oro5ne_6w	6	-1061.649181	-1061.595621	-1061.591379	-1061.584201	-1060.568383	-1060.567154	-1060.559976
d_COOH_oro5ne_7w	7	-1138.103452	-1138.034875	-1138.029571	-1138.021216	-1136.945542	-1136.943251	-1136.934896
d_COOH_oro5ne_8w	8	-1214.548001	-1214.469643	-1214.465558	-1214.455955	-1213.320769	-1213.319697	-1213.310094
d_COOH_oro5ne_9w	9	-1291.004989	-1290.907391	-1290.902833	-1290.892558	-1289.695625	-1289.694080	-1289.683805

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

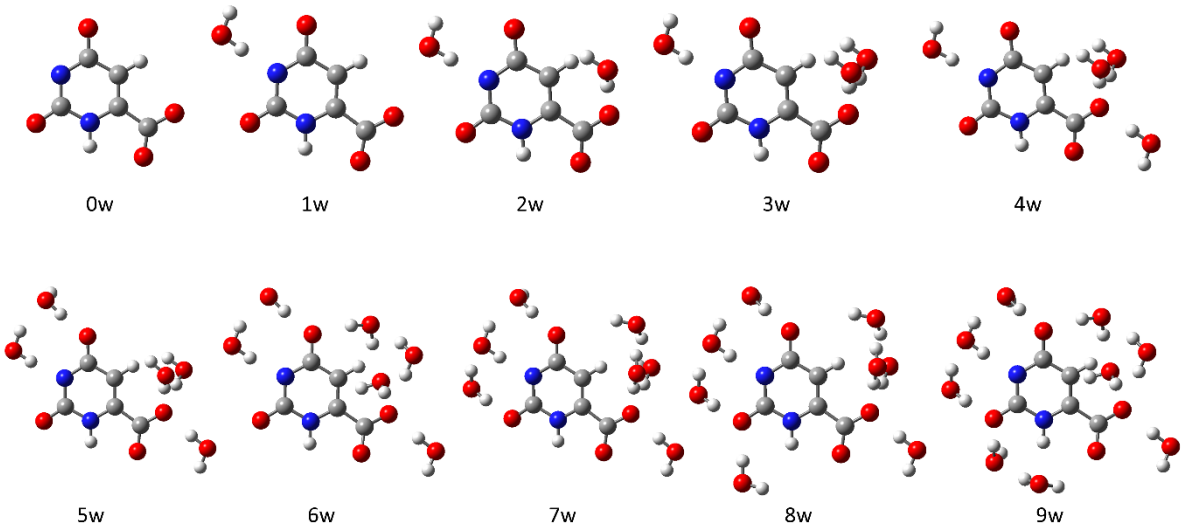


Figure 68: Molecule-water clusters for twice deprotonated orotic acid (**oro**) with relevant thermodynamic data provided in Table 64.

Table 64: Gibbs free energies for solution phase optimized molecule-water clusters of twice deprotonated orotic acid (**oro**) at the (U)B3LYP-D3\6-31G+(d,p) level of theory (p_ = cation, d_ = anion).

structure	no. expl. H_2O	$E_{tot,DFT}$[a,b]	G_{DFT}[b]	$G_{DFT,qh\text{-}30}$[b]	$G_{DFT,qh\text{-}100}$[b]	$G_{CCSD(T)/CBS}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}30}$[b,c,d]	$G_{CCSD(T)/CBS,qh\text{-}100}$[b,c,d]
dd_COOH_oro5ne_0w	0	-602.248645	-602.510475	-602.507462	-602.506883	-601.845212	-601.842199	-601.841620
dd_COOH_oro5ne_1w	1	-678.717790	-678.950623	-678.947328	-678.944864	-678.223350	-678.223068	-678.220604
dd_COOH_oro5ne_2w	2	-755.181601	-755.388210	-755.384813	-755.381098	-754.601050	-754.600666	-754.596951
dd_COOH_oro5ne_3w	3	-831.641225	-831.826830	-831.823283	-831.818206	-830.979526	-830.978992	-830.973915
dd_COOH_oro5ne_4w	4	-908.101642	-908.266521	-908.261775	-908.254456	-907.357951	-907.356218	-907.348899
dd_COOH_oro5ne_5w	5	-984.565256	-984.703062	-984.698901	-984.691323	-983.733378	-983.732230	-983.724652
dd_COOH_oro5ne_6w	6	-1061.031074	-1061.138855	-1061.134824	-1061.127759	-1060.106580	-1060.105562	-1060.098497
dd_COOH_oro5ne_7w	7	-1137.495538	-1137.577808	-1137.573610	-1137.565101	-1136.483989	-1136.482804	-1136.474295
dd_COOH_oro5ne_8w	8	-1213.949765	-1214.016259	-1214.011171	-1214.001137	-1212.862582	-1212.860507	-1212.850473
dd_COOH_oro5ne_9w	9	-1290.407365	-1290.451769	-1290.447588	-1290.437671	-1289.236785	-1289.235617	-1289.225700

[a]: Using solution phase geometries UB3LYP-D3/6-31+G(d,p); [b]: Using solution phase geometries SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) without standard state correction (1bar -> 1M); [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.

4.5.11 Thermodynamic Data of Tautomers and Conformers

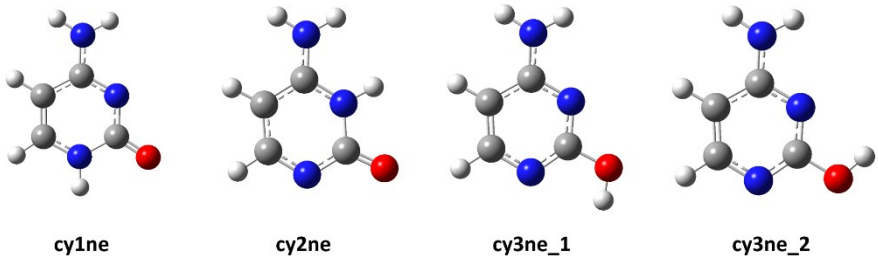


Figure 69: Tautomers for neutral cytosine (**cyt**) with relevant thermodynamic data provided in Table 65.

Table 65: QM properties for solution phase optimized tautomers/conformers of cytosine (**cyt**) shown in Figure 69 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
cy1ne	-394.967638	-395.005991	-394.290273	-394.409972	-394.483660	0.105841	0.067545
cy2ne	-394.955841	-395.000138	-394.277916	-394.397437	-394.471041	0.105735	0.066951
cy3ne_1	-394.968499	-394.992836	-394.295412	-394.414401	-394.487775	0.105217	0.066471
cy3ne_2	-394.967238	-394.992721	-394.294308	-394.413267	-394.486624	0.105399	0.066818
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
cy1ne	-394.938446	-394.261082	-394.380780	-394.454468	100	1.00	1.00
cy2ne	-394.933187	-394.255262	-394.374783	-394.448387	0.00	0.00	0.00
cy3ne_1	-394.926365	-394.253278	-394.372267	-394.445641	0.00	0.00	0.00
cy3ne_2	-394.925903	-394.252973	-394.371932	-394.445289	0.00	0.00	0.00
BoltzAvg:	-394.938427	-394.261066	-394.380768	-394.454457			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 66: QM properties for solution phase optimized tautomers/conformers of cytosine (**cyt**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
cy1ne1w_kicked48	-471.419663	-471.462819	-470.636877	-470.783773	-470.873930	0.133026	0.087763
cy1ne1w_kicked16	-471.419629	-471.462836	-470.636946	-470.783824	-470.873973	0.133102	0.088071
cy1ne1w_kicked08	-471.418693	-471.462679	-470.638310	-470.784791	-470.874739	0.133276	0.087968
cy1ne1w_kicked23	-471.415517	-471.462215	-470.634165	-470.781027	-470.871153	0.132998	0.087747
cy1ne1w_kicked94	-471.415255	-471.461979	-470.633599	-470.780681	-470.870893	0.133059	0.087624
cy1ne1w_kicked52	-471.415086	-471.461990	-470.633610	-470.780650	-470.870843	0.133013	0.087654
cy1ne1w_kicked79	-471.410055	-471.462854	-470.629198	-470.775917	-470.865956	0.133296	0.088608
cy1ne1w_kicked76	-471.411195	-471.462800	-470.631152	-470.777905	-470.867947	0.133407	0.088562
cy1ne1w_kicked18	-471.409349	-471.459865	-470.630216	-470.776208	-470.865905	0.132750	0.085855
cy1ne1w_kicked96	-471.416211	-471.460267	-470.636119	-470.782142	-470.871894	0.133090	0.087103
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
cy1ne1w_kicked48	-471.375056	-470.592270	-470.739166	-470.829323	0.22	0.04	0.04
cy1ne1w_kicked16	-471.374765	-470.592082	-470.738960	-470.829108	0.22	0.04	0.03
cy1ne1w_kicked08	-471.374711	-470.594328	-470.740809	-470.830757	0.51	0.16	0.18
cy1ne1w_kicked23	-471.374468	-470.593116	-470.739978	-470.830103	0.01	0.09	0.09
cy1ne1w_kicked94	-471.374355	-470.592699	-470.739781	-470.829993	0.01	0.06	0.08
cy1ne1w_kicked52	-471.374336	-470.592860	-470.739900	-470.830093	0.01	0.08	0.09
cy1ne1w_kicked79	-471.374246	-470.593389	-470.740108	-470.830148	0.00	0.16	0.09
cy1ne1w_kicked76	-471.374238	-470.594195	-470.740948	-470.830990	0.00	0.34	0.23
cy1ne1w_kicked18	-471.374010	-470.594876	-470.740869	-470.830566	0.00	0.02	0.15
cy1ne1w_kicked96	-471.373164	-470.593072	-470.739094	-470.828846	0.02	0.01	0.02
BoltzAvg:	-471.376594	-470.594097	-470.740437	-470.830394			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

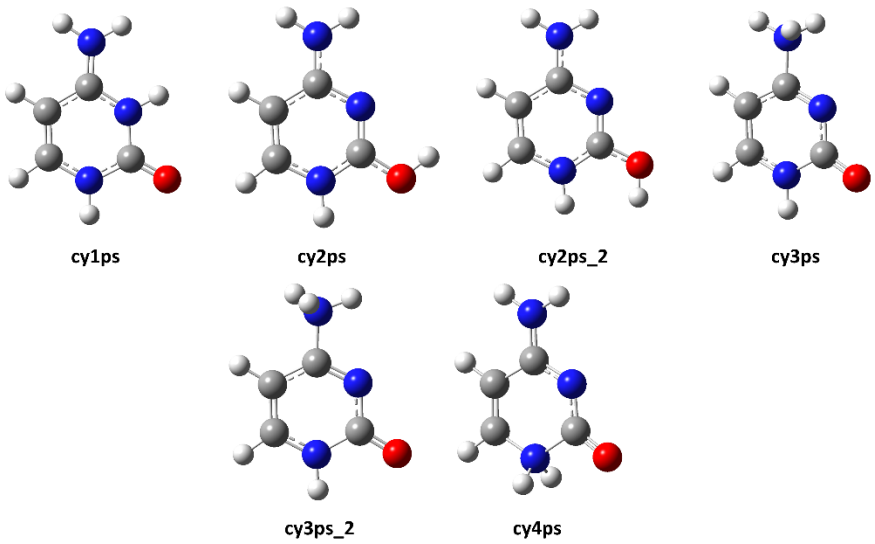


Figure 70: Tautomers for protonated cytosine (**cyt**) with relevant thermodynamic data provided in Table 67.

Table 67: QM properties for solution phase optimized tautomers/conformers of protonated cytosine (**cyt**) shown in Figure 70 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

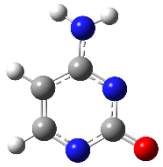
Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
cy1ps	-395.345621	-395.459764	-394.671772	-394.788049	-394.859911	0.119638	0.081513
cy2ps	-395.347024	-395.447367	-394.675977	-394.792187	-394.863980	0.119185	0.081078
cy2ps_2	-395.332891	-395.444293	-394.663047	-394.779479	-394.851313	0.119511	0.081571
cy3ps_2	-395.291072	-395.431653	-394.620031	-394.735863	-394.807410	0.120429	0.081757
cy3ps_1	-395.290299	-395.431365	-394.618975	-394.734842	-394.806404	0.119463	0.082180
cy4ps	-395.278055	-395.404661	-394.607925	-394.723911	-394.795428	0.119455	0.081006
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
cy1ps	-395.378251	-394.704402	-394.820679	-394.892541	0.01	1.00	1.00
cy2ps	-395.366289	-394.695242	-394.811452	-394.883244	0.99	0.00	0.00
cy2ps_2	-395.362722	-394.692878	-394.809311	-394.881144	0.00	0.00	0.00
cy3ps_2	-395.349896	-394.678855	-394.794687	-394.866233	0.00	0.00	0.00
cy3ps_1	-395.349185	-394.677861	-394.793729	-394.865290	0.00	0.00	0.00
cy4ps	-395.323655	-394.653526	-394.769512	-394.841028	0.00	0.00	0.00
BoltzAvg:	-395.378251	-394.704401	-394.820678	-394.892540			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 68: QM properties for solution phase optimized tautomers/conformers of protonated cytosine (**cyt**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
cy1ps1w_kicked12	-471.807371	-471.917233	-471.030728	-471.172995	-471.260796	0.146619	0.101350
cy1ps1w_kicked57	-471.804683	-471.917297	-471.028153	-471.170568	-471.258429	0.146721	0.101515
cy1ps1w_kicked96	-471.797548	-471.917432	-471.022515	-471.164550	-471.252205	0.146922	0.101653
cy1ps1w_kicked03	-471.808015	-471.915740	-471.031450	-471.173692	-471.261486	0.146630	0.100494
cy1ps1w_kicked45	-471.800931	-471.915713	-471.024771	-471.167285	-471.255172	0.146899	0.101198
cy1ps1w_kicked67	-471.799864	-471.916113	-471.024739	-471.166787	-471.254469	0.147007	0.101619
cy1ps1w_kicked07	-471.798165	-471.915353	-471.022904	-471.165042	-471.252740	0.146729	0.101001
cy1ps1w_kicked43	-471.798704	-471.915364	-471.023188	-471.165413	-471.253149	0.146763	0.101185
cy1ps1w_kicked13	-471.804890	-471.915032	-471.028603	-471.170879	-471.258662	0.146636	0.101043
cy1ps1w_kicked62	-471.800456	-471.915272	-471.024379	-471.166824	-471.254672	0.146724	0.101284
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
cy1ps1w_kicked12	-471.815883	-471.039240	-471.181507	-471.269309	0.31	0.12	0.11
cy1ps1w_kicked57	-471.815782	-471.039251	-471.181667	-471.269528	0.02	0.17	0.14
cy1ps1w_kicked96	-471.815779	-471.040747	-471.182781	-471.270436	0.00	0.41	0.36
cy1ps1w_kicked03	-471.815246	-471.038681	-471.180923	-471.268717	0.64	0.03	0.06
cy1ps1w_kicked45	-471.814515	-471.038354	-471.180868	-471.268756	0.00	0.04	0.06
cy1ps1w_kicked67	-471.814494	-471.039368	-471.181417	-471.269098	0.00	0.09	0.09
cy1ps1w_kicked07	-471.814352	-471.039091	-471.181229	-471.268927	0.00	0.05	0.07
cy1ps1w_kicked43	-471.814179	-471.038664	-471.180888	-471.268624	0.00	0.04	0.05
cy1ps1w_kicked13	-471.813989	-471.037702	-471.179979	-471.267762	0.03	0.02	0.02
cy1ps1w_kicked62	-471.813988	-471.037911	-471.180356	-471.268204	0.00	0.03	0.03
BoltzAvg:	--471.815967	-471.039780	-471.181837	-471.269526			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.



d_cytosine_1ng_b

Figure 71: Tautomers for deprotonated cytosine (**cyt**) with relevant thermodynamic data provided in Table 69.

Table 69: QM properties for solution phase optimized tautomers/conformers of deprotonated cytosine (**cyt**) shown in Figure 71 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
d_cytosine_1ng_b	-394.402506	-394.525707	-393.716422	-393.841200	-393.917357	0.092439	0.054975
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
d_cytosine_1ng_b	-394.470732	-393.784648	-393.909425	-393.985582	1.00	1.00	1.00
BoltzAvg:	-394.470732	-393.784648	-393.909425	-393.985582			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 70: QM properties for solution phase optimized tautomers/conformers of deprotonated cytosine (**cyt**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
d_cytosine_1ng_kicked88	-470.867536	-470.984752	-470.075744	-470.227323	-470.319811	0.119210	0.072711
d_cytosine_1ng_kicked24	-470.860012	-470.984302	-470.069600	-470.221214	-470.313678	0.119082	0.072566
d_cytosine_1ng_kicked98	-470.866468	-470.984064	-470.076666	-470.227902	-470.320231	0.119139	0.072535
d_cytosine_1ng_kicked53	-470.866577	-470.983972	-470.076574	-470.227852	-470.320195	0.119127	0.072681
d_cytosine_1ng_kicked41	-470.866343	-470.983960	-470.076353	-470.227651	-470.320005	0.119152	0.072882
d_cytosine_1ng_kicked73	-470.862588	-470.983665	-470.072505	-470.224209	-470.316745	0.119151	0.072684
d_cytosine_1ng_kicked12	-470.863419	-470.983808	-470.073170	-470.224891	-470.317442	0.119200	0.073157
d_cytosine_1ng_kicked77	-470.856932	-470.984742	-470.067428	-470.218834	-470.311219	0.119523	0.074535
d_cytosine_1ng_kicked55	-470.856701	-470.981146	-470.067861	-470.218870	-470.311035	0.119711	0.072865
d_cytosine_1ng_kicked07	-470.857034	-470.981185	-470.068065	-470.219130	-470.311323	0.119745	0.073390
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
d_cytosine_1ng_kicked88	-470.912041	-470.120249	-470.271828	-470.364317	0.18	0.05	0.06
d_cytosine_1ng_kicked24	-470.911736	-470.121324	-470.272938	-470.365402	0.00	0.14	0.20
d_cytosine_1ng_kicked98	-470.911529	-470.121727	-470.272963	-470.365293	0.29	0.12	0.18
d_cytosine_1ng_kicked53	-470.911291	-470.121287	-470.272566	-470.364909	0.28	0.09	0.12
d_cytosine_1ng_kicked41	-470.911078	-470.121088	-470.272386	-470.364740	0.23	0.09	0.10
d_cytosine_1ng_kicked73	-470.910981	-470.120898	-470.272602	-470.365138	0.01	0.11	0.15
d_cytosine_1ng_kicked12	-470.910651	-470.120402	-470.272123	-470.364675	0.02	0.11	0.09
d_cytosine_1ng_kicked77	-470.910207	-470.120703	-470.272109	-470.364494	0.00	0.28	0.08
d_cytosine_1ng_kicked55	-470.908281	-470.119441	-470.270450	-470.362615	0.00	0.01	0.01
d_cytosine_1ng_kicked07	-470.907795	-470.118826	-470.269891	-470.362084	0.00	0.01	0.01
BoltzAvg:	-470.911691	-470.121116	-470.272536	-470.364962			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

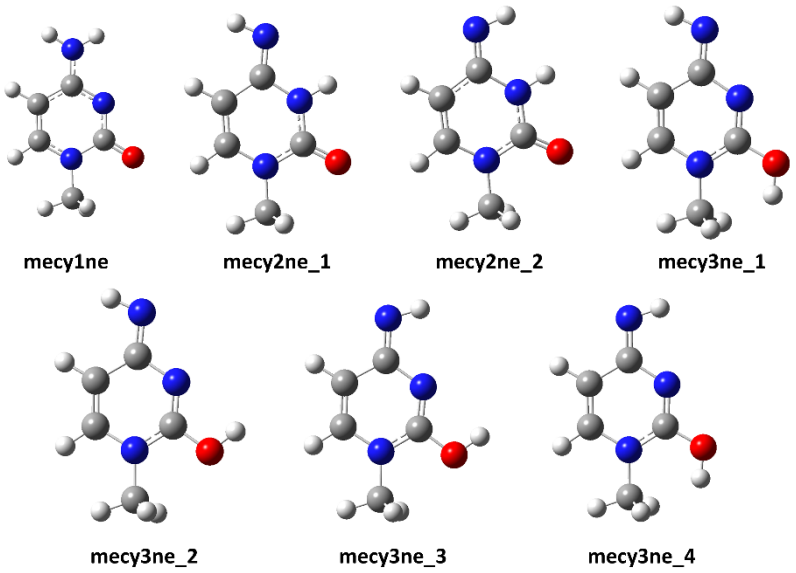


Figure 72: Tautomers for N1-metylcytosine (**1mC**) with relevant thermodynamic data provided in Table 71.

Table 71: QM properties for solution phase optimized tautomers/conformers of N1-metylcytosine (**1mC**) shown in Figure 72 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
mecy1ne	-434.283288	-434.317835	-433.525454	-433.656255	-433.736941	0.135385	0.093442
mecy2ne_1	-434.281389	-434.304877	-433.526358	-433.656709	-433.737182	0.135555	0.093850
mecy2ne_2	-434.278372	-434.303579	-433.523728	-433.653967	-433.734379	0.135857	0.094427
mecy3ne_3	-434.252853	-434.279020	-433.500660	-433.630751	-433.711005	0.135204	0.093585
mecy3ne_2	-434.247316	-434.278540	-433.495678	-433.626015	-433.706328	0.135169	0.093556
mecy3ne_4	-434.236408	-434.274366	-433.485909	-433.616199	-433.696479	0.133734	0.093010
mecy3ne_1	-434.229234	-434.273779	-433.479571	-433.610078	-433.690386	0.133899	0.093844
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
mecy1ne	-434.224393	-433.466559	-433.597360	-433.678045	1.00	1.00	1.00
mecy2ne_1	-434.211027	-433.455996	-433.586347	-433.666820	0.00	0.00	0.00
mecy2ne_2	-434.209152	-433.454507	-433.584747	-433.665159	0.00	0.00	0.00
mecy3ne_3	-434.185435	-433.433242	-433.563333	-433.643587	0.00	0.00	0.00
mecy3ne_2	-434.184984	-433.433346	-433.563682	-433.643996	0.00	0.00	0.00
mecy3ne_4	-434.181356	-433.430858	-433.561147	-433.641428	0.00	0.00	0.00
mecy3ne_1	-434.179935	-433.430272	-433.560779	-433.641087	0.00	0.00	0.00
BoltzAvg:	-434.224393	-433.466559	-433.597360	-433.678045			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91$kJ mol^{-1} not jet applied.							

Table 72: QM properties for solution phase optimized tautomers/conformers of N1-metylcytosine (**1mC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
mecy1ne1w_kicked83	-510.731818	-510.774172	-509.869727	-510.027737	-510.124887	0.162211	0.112755
mecy1ne1w_kicked41	-510.731945	-510.774197	-509.869858	-510.027817	-510.124944	0.162236	0.112797
mecy1ne1w_kicked89	-510.731881	-510.774448	-509.870503	-510.028052	-510.124989	0.162631	0.113604
mecy1ne1w_kicked67	-510.735481	-510.774688	-509.872050	-510.030088	-510.127264	0.162660	0.113955
mecy1ne1w_kicked22	-510.733316	-510.774570	-509.871381	-510.028961	-510.125913	0.162706	0.113887
mecy1ne1w_kicked12	-510.731044	-510.773819	-509.868986	-510.027153	-510.124351	0.162372	0.113189
mecy1ne1w_kicked54	-510.734897	-510.774630	-509.871836	-510.029769	-510.126894	0.162693	0.114210
mecy1ne1w_kicked85	-510.727309	-510.774118	-509.866773	-510.024542	-510.121556	0.162858	0.114398
mecy1ne1w_kicked57	-510.726480	-510.774765	-509.865091	-510.022866	-510.119874	0.163144	0.115335
mecy1ne1w_kicked19	-510.726948	-510.774126	-509.866386	-510.024208	-510.121244	0.162934	0.114716
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
mecy1ne1w_kicked83	-510.661417	-509.799326	-509.957336	-510.054486	0.04	0.11	0.21
mecy1ne1w_kicked41	-510.661400	-509.799313	-509.957272	-510.054399	0.04	0.10	0.19
mecy1ne1w_kicked89	-510.660844	-509.799465	-509.957015	-510.053951	0.04	0.10	0.12
mecy1ne1w_kicked67	-510.660733	-509.797302	-509.955340	-510.052517	0.45	0.03	0.03
mecy1ne1w_kicked22	-510.660683	-509.798748	-509.956328	-510.053280	0.11	0.06	0.06
mecy1ne1w_kicked12	-510.660630	-509.798573	-509.956740	-510.053938	0.02	0.08	0.12
mecy1ne1w_kicked54	-510.660420	-509.797359	-509.955292	-510.052417	0.30	0.03	0.02
mecy1ne1w_kicked85	-510.659720	-509.799184	-509.956953	-510.053968	0.00	0.18	0.12
mecy1ne1w_kicked57	-510.659430	-509.798041	-509.955816	-510.052824	0.00	0.11	0.04
mecy1ne1w_kicked19	-510.659410	-509.798848	-509.956670	-510.053705	0.00	0.18	0.09
BoltzAvg:	-510.663228	-509.799043	-509.956883	-510.053974			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

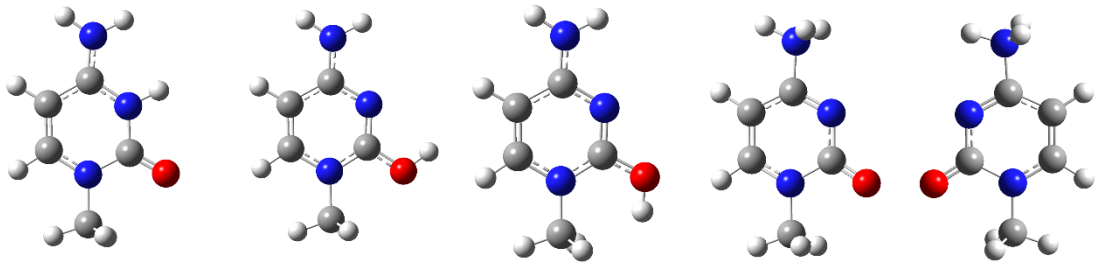


Figure 73: Tautomers for protonated N1-metylcytosine (**1mC**) with relevant thermodynamic data provided in Table 73.

Table 73: QM properties for solution phase optimized tautomers/conformers of protonated N1-metylcytosine (**1mC**) shown in Figure 73 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
mecy1ps	-434.666819	-434.773210	-433.911576	-434.039309	-434.118346	0.149276	0.107742
mecy2ps_1	-434.667302	-434.760273	-433.914962	-434.042548	-434.121487	0.148834	0.107088
mecy2ps_2	-434.652690	-434.755925	-433.901824	-434.029550	-434.108484	0.148602	0.106905
mecy3ps_1	-434.612308	-434.745474	-433.859943	-433.987352	-434.066139	0.148965	0.108231
mecy3ps_2	-434.612989	-434.745885	-433.860668	-433.987971	-434.066695	0.150195	0.109071
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
mecy1ps	-434.665468	-433.910225	-434.037958	-434.116995	1.00	1.00	1.00
mecy2ps_1	-434.653185	-433.900845	-434.028431	-434.107371	0.00	0.00	0.00
mecy2ps_2	-434.649020	-433.898154	-434.025880	-434.104814	0.00	0.00	0.00
mecy3ps_1	-434.637243	-433.884878	-434.012287	-434.091074	0.00	0.00	0.00
mecy3ps_2	-434.636814	-433.884493	-434.011796	-434.090520	0.00	0.00	0.00
BoltzAvg:	-434.665468	-433.910224	-434.037958	-434.116995			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 74: QM properties for solution phase optimized tautomers/conformers protonated N1-metylcytosine (**1mC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
mecy1ps1w_kicked22	-511.132571	-511.230508	-510.274406	-510.427826	-510.522679	0.176145	0.127049
mecy1ps1w_kicked30	-511.125506	-511.230016	-510.267638	-510.421479	-510.516505	0.176152	0.126982
mecy1ps1w_kicked34	-511.130443	-511.230176	-510.272338	-510.425871	-510.520775	0.176256	0.127460
mecy1ps1w_kicked78	-511.119667	-511.229180	-510.263247	-510.416720	-510.511560	0.175499	0.127863
mecy1ps1w_kicked87	-511.111613	-511.227802	-510.252102	-510.407316	-510.502905	0.176594	0.127442
mecy1ps1w_kicked96	-511.109073	-511.227916	-510.251300	-510.406026	-510.501385	0.176574	0.127699
mecy1ps1w_kicked93	-511.111759	-511.227834	-510.253014	-510.407978	-510.503455	0.176547	0.127687
mecy1ps1w_kicked23	-511.100092	-511.227758	-510.243513	-510.398161	-510.493480	0.176455	0.127761
mecy1ps1w_kicked02	-511.111498	-511.227815	-510.252019	-510.407229	-510.502815	0.176665	0.127910
mecy1ps1w_kicked32	-511.109922	-511.227852	-510.251889	-510.406708	-510.502112	0.176559	0.127953
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
mecy1ps1w_kicked22	-511.103459	-510.245294	-510.398714	-510.493567	0.88	0.11	0.18
mecy1ps1w_kicked30	-511.103034	-510.245165	-510.399007	-510.494033	0.00	0.17	0.29
mecy1ps1w_kicked34	-511.102716	-510.244611	-510.398145	-510.493048	0.12	0.09	0.10
mecy1ps1w_kicked78	-511.101317	-510.244898	-510.398371	-510.493211	0.00	0.36	0.12
mecy1ps1w_kicked87	-511.100360	-510.240848	-510.396063	-510.491652	0.00	0.01	0.02
mecy1ps1w_kicked96	-511.100217	-510.242444	-510.397170	-510.492529	0.00	0.05	0.06
mecy1ps1w_kicked93	-511.100147	-510.241402	-510.396366	-510.491844	0.00	0.02	0.03
mecy1ps1w_kicked23	-511.099997	-510.243418	-510.398066	-510.493385	0.00	0.14	0.15
mecy1ps1w_kicked02	-511.099905	-510.240426	-510.395636	-510.491222	0.00	0.01	0.01
mecy1ps1w_kicked32	-511.099899	-510.241866	-510.396685	-510.492089	0.00	0.04	0.04
BoltzAvg:	-511.103687	-510.244889	-510.398391	-510.493335			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

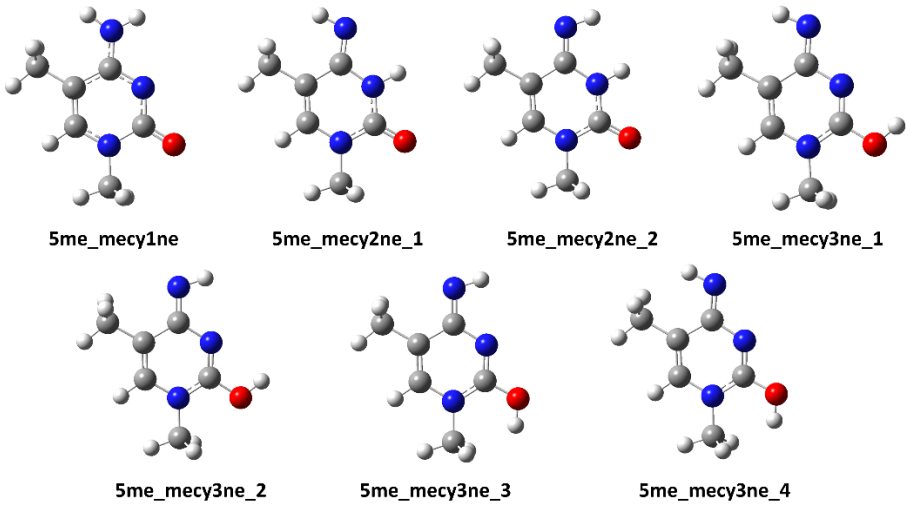


Figure 74: Tautomers for 1,3-dimethylcytosine (**1m5mC**) with relevant thermodynamic data provided in Table 75.

Table 75: QM properties for solution phase optimized tautomers/conformers of 1,3-dimethylcytosine (**1m5mC**) shown in Figure 74 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5me_mecy1ne	-473.606831	-473.641554	-472.768353	-472.910792	-472.998796	0.164878	0.120014
5me_mecy2ne_1	-473.606246	-473.629133	-472.770766	-472.912709	-473.000471	0.165048	0.120123
5me_mecy2ne_2	-473.603899	-473.627056	-472.768615	-472.910362	-472.998041	0.164877	0.119736
5me_mecy3ne_2	-473.579588	-473.603082	-472.746675	-472.888275	-472.975811	0.164411	0.119542
5me_mecy3ne_1	-473.572791	-473.603085	-472.740644	-472.882636	-472.970278	0.164521	0.119783
5me_mecy3ne_3	-473.563398	-473.598621	-472.732091	-472.873848	-472.961382	0.164548	0.120228
5me_mecy3ne_4	-473.554926	-473.598507	-472.724627	-472.866807	-472.954459	0.164664	0.120491
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5me_mecy1ne	-473.521540	-472.683062	-472.825501	-472.913506	1.00	1.00	1.00
5me_mecy2ne_1	-473.509010	-472.673530	-472.815473	-472.903235	0.00	0.00	0.00
5me_mecy2ne_2	-473.507320	-472.672036	-472.813783	-472.901462	0.00	0.00	0.00
5me_mecy3ne_2	-473.483540	-472.650627	-472.792227	-472.879763	0.00	0.00	0.00
5me_mecy3ne_1	-473.483302	-472.651156	-472.793147	-472.880790	0.00	0.00	0.00
5me_mecy3ne_3	-473.478393	-472.647086	-472.788843	-472.876377	0.00	0.00	0.00
5me_mecy3ne_4	-473.478016	-472.647716	-472.789897	-472.877548	0.00	0.00	0.00
BoltzAvg:	-473.521860	-472.683062	-472.825501	-472.913505			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 76: QM properties for solution phase optimized tautomers/conformers of 1,3-dimethylcytosine (**1m5mC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5Me_mecy1ne1w_kicked90	-550.055824	-550.098193	-549.113032	-549.282616	-549.387037	0.191907	0.139286
5Me_mecy1ne1w_kicked44	-550.059320	-550.099057	-549.115180	-549.284793	-549.389250	0.190945	0.140165
5Me_mecy1ne1w_kicked76	-550.055509	-550.097939	-549.112343	-549.282202	-549.386735	0.191898	0.139321
5Me_mecy1ne1w_kicked13	-550.050639	-550.098874	-549.108618	-549.277999	-549.382312	0.192171	0.140427
5Me_mecy1ne1w_kicked02	-550.051521	-550.098362	-549.110765	-549.280055	-549.384337	0.192092	0.140319
5Me_mecy1ne1w_kicked68	-550.056886	-550.097769	-549.114134	-549.283831	-549.388328	0.191972	0.139863
5Me_mecy1ne1w_kicked74	-550.051322	-550.098131	-549.110010	-549.279437	-549.383774	0.192102	0.140434
5Me_mecy1ne1w_kicked78	-550.047909	-550.095801	-549.106882	-549.275897	-549.380032	0.191898	0.138226
5Me_mecy1ne1w_kicked05	-550.044211	-550.095049	-549.103579	-549.272764	-549.376948	0.191864	0.137976
5Me_mecy1ne1w_kicked55	-550.047669	-550.095768	-549.106673	-549.275664	-549.379779	0.192020	0.139084
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5Me_mecy1ne1w_kicked90	-549.958907	-549.016114	-549.185699	-549.290120	0.06	0.06	0.12
5Me_mecy1ne1w_kicked44	-549.958892	-549.014751	-549.184365	-549.288822	0.64	0.11	0.03
5Me_mecy1ne1w_kicked76	-549.958618	-549.015452	-549.185311	-549.289843	0.04	0.05	0.09
5Me_mecy1ne1w_kicked13	-549.958447	-549.016426	-549.185806	-549.290120	0.00	0.16	0.12
5Me_mecy1ne1w_kicked02	-549.958043	-549.017287	-549.186577	-549.290859	0.00	0.34	0.26
5Me_mecy1ne1w_kicked68	-549.957906	-549.015154	-549.184850	-549.289347	0.24	0.05	0.05
5Me_mecy1ne1w_kicked74	-549.957697	-549.016385	-549.185812	-549.290149	0.00	0.18	0.12
5Me_mecy1ne1w_kicked78	-549.957575	-549.016548	-549.185563	-549.289698	0.00	0.01	0.08
5Me_mecy1ne1w_kicked05	-549.957073	-549.016442	-549.185626	-549.289810	0.00	0.01	0.09
5Me_mecy1ne1w_kicked55	-549.956684	-549.015689	-549.184679	-549.288794	0.00	0.01	0.03
BoltzAvg:	-549.960900	-549.016512	-549.185809	-549.290114			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

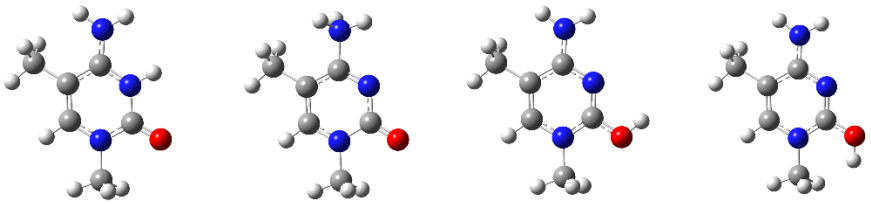


Figure 75: Tautomers for protonated 1,3-dimethylcytosine (**1m5mC**) with relevant thermodynamic data provided in Table 77.

Table 77: QM properties for solution phase optimized tautomers/conformers of protonated 1,3-dimethylcytosine (**1m5mC**) shown in Figure 75 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5me_mecy1ps	-473.994001	-474.097376	-473.157758	-473.297273	-473.383679	0.177139	0.133550
5me_mecy3ps_1	-473.995926	-474.085738	-473.162579	-473.301977	-473.388313	0.178179	0.133412
5me_mecy3ps_2	-473.981620	-474.081501	-473.149735	-473.289261	-473.375585	0.177402	0.134689
5me_mecy2ps	-473.939038	-474.070608	-473.105094	-473.244077	-473.330117	0.179233	0.134001
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5me_mecy1ps	-473.963826	-473.127584	-473.267099	-473.353504	1.00	1.00	1.00
5me_mecy3ps_1	-473.952326	-473.118979	-473.258377	-473.344713	0.00	0.00	0.00
5me_mecy3ps_2	-473.946812	-473.114928	-473.254453	-473.340778	0.00	0.00	0.00
5me_mecy2ps	-473.936607	-473.102663	-473.241646	-473.327686	0.00	0.00	0.00
BoltzAvg:	-473.963826	-473.127583	-473.267098	-473.353503			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 78: QM properties for solution phase optimized tautomers/conformers of protonated 1,3-dimethylcytosine (**1m5mC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5Me_mecy1ps1w_kicked01	-550.459168	-550.555452	-549.520111	-549.685253	-549.787435	0.205607	0.153048
5Me_mecy1ps1w_kicked12	-550.451602	-550.555070	-549.512878	-549.678455	-549.780811	0.205443	0.152911
5Me_mecy1ps1w_kicked62	-550.454442	-550.555090	-549.515571	-549.681013	-549.783320	0.205592	0.153255
5Me_mecy1ps1w_kicked45	-550.447297	-550.553989	-549.509801	-549.674992	-549.777160	0.205611	0.152686
5Me_mecy1ps1w_kicked08	-550.447861	-550.555116	-549.509423	-549.675064	-549.777430	0.205826	0.154048
5Me_mecy1ps1w_kicked47	-550.445176	-550.555214	-549.507753	-549.672989	-549.775164	0.206087	0.154188
5Me_mecy1ps1w_kicked55	-550.449613	-550.553556	-549.511065	-549.676690	-549.779056	0.205524	0.152730
5Me_mecy1ps1w_kicked77	-550.444487	-550.555191	-549.507293	-549.672457	-549.774588	0.206032	0.154371
5Me_mecy1ps1w_kicked25	-550.446803	-550.553086	-549.508587	-549.674040	-549.776313	0.205654	0.152888
5Me_mecy1ps1w_kicked59	-550.444675	-550.552974	-549.507240	-549.672359	-549.774472	0.205594	0.153224
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5Me_mecy1ps1w_kicked01	-550.402404	-549.463347	-549.628489	-549.730671	0.99	0.07	0.09
5Me_mecy1ps1w_kicked12	-550.402159	-549.463436	-549.629012	-549.731368	0.00	0.14	0.19
5Me_mecy1ps1w_kicked62	-550.401835	-549.462964	-549.628405	-549.730713	0.01	0.09	0.10
5Me_mecy1ps1w_kicked45	-550.401303	-549.463807	-549.628999	-549.731167	0.00	0.08	0.16
5Me_mecy1ps1w_kicked08	-550.401068	-549.462630	-549.628271	-549.730637	0.00	0.14	0.09
5Me_mecy1ps1w_kicked47	-550.401026	-549.463604	-549.628840	-549.731014	0.00	0.19	0.13
5Me_mecy1ps1w_kicked55	-550.400826	-549.462278	-549.627902	-549.730269	0.00	0.03	0.06
5Me_mecy1ps1w_kicked77	-550.400820	-549.463626	-549.628790	-549.730920	0.00	0.22	0.12
5Me_mecy1ps1w_kicked25	-550.400198	-549.461982	-549.627435	-549.729708	0.00	0.02	0.03
5Me_mecy1ps1w_kicked59	-550.399750	-549.462314	-549.627433	-549.729546	0.00	0.02	0.03
BoltzAvg:	-550.402845	-549.463342	-549.628632	-549.730872			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

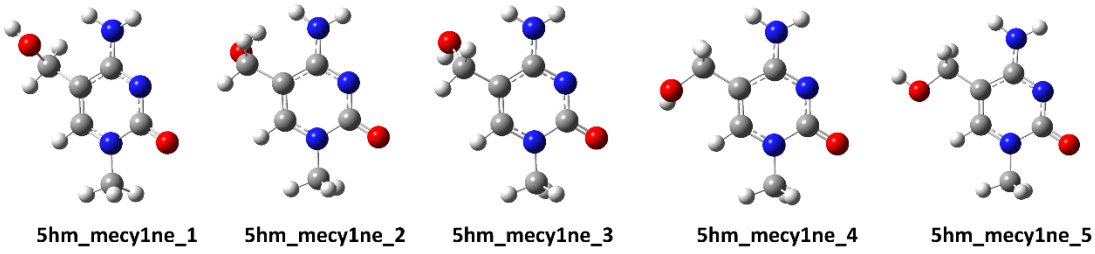


Figure 76: Conformers/tautomers for 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) with relevant thermodynamic data provided in Table 79.

Table 79: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) shown in Figure 76 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5hm_mecy1ne_3	-548.828056	-548.867992	-547.892392	-548.059190	-548.162107	0.171371	0.124353
5hm_mecy1ne_1	-548.825517	-548.866907	-547.890198	-548.057032	-548.159928	0.171161	0.124023
5hm_mecy1ne_2	-548.820836	-548.866418	-547.885302	-548.052391	-548.155382	0.171629	0.124470
5hm_mecy1ne_4	-548.820686	-548.865219	-547.886159	-548.052974	-548.155894	0.171254	0.123785
5hm_mecy1ne_5	-548.821298	-548.864131	-547.886092	-548.053329	-548.156466	0.170999	0.123072
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5hm_mecy1ne_3	-548.743639	-547.807975	-547.974773	-548.077689	0.91	0.47	0.40
5hm_mecy1ne_1	-548.742884	-547.807564	-547.974399	-548.077294	0.09	0.27	0.27
5hm_mecy1ne_2	-548.741948	-547.806414	-547.973503	-548.076494	0.00	0.12	0.11
5hm_mecy1ne_4	-548.741434	-547.806907	-547.973722	-548.076642	0.00	0.10	0.13
5hm_mecy1ne_5	-548.741059	-547.805853	-547.973090	-548.076227	0.00	0.04	0.09
BoltzAvg:	-548.743076	-547.807475	-547.974275	-548.077185			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 80: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
5MeOH_mecy1ne_1_kicked86	-625.274037	-625.323660	-624.235276	-624.428851	-624.548005	0.198181	0.143450
5MeOH_mecy1ne_1_kicked41	-625.274091	-625.323133	-624.234460	-624.428542	-624.547919	0.198154	0.143398
5MeOH_mecy1ne_1_kicked19	-625.268376	-625.323874	-624.229364	-624.423259	-624.542521	0.198468	0.144471
5MeOH_mecy1ne_1_kicked90	-625.269547	-625.323629	-624.230520	-624.423946	-624.543033	0.198458	0.144257
5MeOH_mecy1ne_1_kicked89	-625.269178	-625.323850	-624.230318	-624.424125	-624.543358	0.198396	0.144528
5MeOH_mecy1ne_1_kicked98	-625.269079	-625.322980	-624.231123	-624.424970	-624.544222	0.198400	0.144039
5MeOH_mecy1ne_1_kicked32	-625.275412	-625.322726	-624.235688	-624.429901	-624.549352	0.198247	0.143822
5MeOH_mecy1ne_1_kicked68	-625.268946	-625.323928	-624.230110	-624.423475	-624.542506	0.198566	0.145035
5MeOH_mecy1ne_1_kicked35	-625.269680	-625.323978	-624.230656	-624.423998	-624.543026	0.198521	0.145105
5MeOH_mecy1ne_1_kicked21	-625.270722	-625.323861	-624.231727	-624.424935	-624.543920	0.197390	0.145049
5MeOH_mecy1ne_2_kicked03	-625.272444	-625.323646	-624.231448	-624.425736	-624.545198	0.198731	0.144695
5MeOH_mecy1ne_2_kicked84	-625.268847	-625.322539	-624.228677	-624.423224	-624.542766	0.198503	0.143709
5MeOH_mecy1ne_2_kicked33	-625.271187	-625.323514	-624.232500	-624.425518	-624.544426	0.198604	0.144723
5MeOH_mecy1ne_2_kicked91	-625.268836	-625.322792	-624.229076	-624.423389	-624.542843	0.198569	0.144074
5MeOH_mecy1ne_2_kicked50	-625.264388	-625.322668	-624.226071	-624.420215	-624.539576	0.198533	0.144103
5MeOH_mecy1ne_2_kicked32	-625.269753	-625.322306	-624.230191	-624.424532	-624.544015	0.198395	0.143758
5MeOH_mecy1ne_2_kicked97	-625.263744	-625.323805	-624.225081	-624.418558	-624.537608	0.198910	0.145315
5MeOH_mecy1ne_2_kicked86	-625.264419	-625.323918	-624.225602	-624.419077	-624.538130	0.199019	0.145620
5MeOH_mecy1ne_2_kicked92	-625.262956	-625.323582	-624.223853	-624.417989	-624.537331	0.199092	0.145354
5MeOH_mecy1ne_2_kicked73	-625.264915	-625.323937	-624.226000	-624.419464	-624.538521	0.199104	0.145791
5MeOH_mecy1ne_3_kicked82	-625.279706	-625.324962	-624.238569	-624.432561	-624.551948	0.198254	0.143678
5MeOH_mecy1ne_3_kicked97	-625.276174	-625.323962	-624.235935	-624.430140	-624.549580	0.198216	0.143611
5MeOH_mecy1ne_3_kicked14	-625.271060	-625.324942	-624.231744	-624.425585	-624.544858	0.198564	0.144678
5MeOH_mecy1ne_3_kicked44	-625.277575	-625.323835	-624.237800	-624.431825	-624.551219	0.198390	0.144179
5MeOH_mecy1ne_3_kicked75	-625.275286	-625.324406	-624.237156	-624.430095	-624.549008	0.198671	0.144891
5MeOH_mecy1ne_3_kicked30	-625.271746	-625.324190	-624.233366	-624.427183	-624.546445	0.198593	0.144725
5MeOH_mecy1ne_3_kicked47	-625.273646	-625.324358	-624.236194	-624.428924	-624.547730	0.198690	0.145070
5MeOH_mecy1ne_3_kicked54	-625.276627	-625.324236	-624.238433	-624.431181	-624.549965	0.198636	0.144973
5MeOH_mecy1ne_3_kicked89	-625.278268	-625.323943	-624.238317	-624.432459	-624.551907	0.198599	0.144693
5MeOH_mecy1ne_3_kicked34	-625.272300	-625.324448	-624.234515	-624.428152	-624.547340	0.198726	0.145201
5MeOH_mecy1ne_4_kicked59	-625.273165	-625.322146	-624.233033	-624.426997	-624.546363	0.198263	0.143437
5MeOH_mecy1ne_4_kicked97	-625.265915	-625.321493	-624.229487	-624.422213	-624.541018	0.198347	0.143414
5MeOH_mecy1ne_4_kicked19	-625.267125	-625.321592	-624.230350	-624.423195	-624.542054	0.198357	0.143695
5MeOH_mecy1ne_4_kicked24	-625.266894	-625.321478	-624.230178	-624.422979	-624.541817	0.198319	0.143604
5MeOH_mecy1ne_4_kicked14	-625.265331	-625.321660	-624.228641	-624.422264	-624.541442	0.198425	0.144049
5MeOH_mecy1ne_4_kicked86	-625.270815	-625.321106	-624.231956	-624.426011	-624.545411	0.198389	0.143540
5MeOH_mecy1ne_4_kicked42	-625.263206	-625.321772	-624.225718	-624.419047	-624.538081	0.198676	0.144220
5MeOH_mecy1ne_4_kicked41	-625.265606	-625.321978	-624.227790	-624.421339	-624.540476	0.198622	0.144461
5MeOH_mecy1ne_4_kicked35	-625.264255	-625.322049	-624.226217	-624.419948	-624.539164	0.198738	0.144777
5MeOH_mecy1ne_4_kicked57	-625.270509	-625.321149	-624.231700	-624.425751	-624.545148	0.198526	0.143963
5MeOH_mecy1ne_5_kicked45	-625.273905	-625.321019	-624.233091	-624.427505	-624.547102	0.197784	0.142207
5MeOH_mecy1ne_5_kicked75	-625.270458	-625.320367	-624.230921	-624.425218	-624.544726	0.197824	0.142131
5MeOH_mecy1ne_5_kicked08	-625.271607	-625.319947	-624.232019	-624.426574	-624.546230	0.197993	0.142226
5MeOH_mecy1ne_5_kicked92	-625.269867	-625.320113	-624.230153	-624.424739	-624.544370	0.197961	0.142586
5MeOH_mecy1ne_5_kicked70	-625.271222	-625.319986	-624.231767	-624.426247	-624.545862	0.198058	0.142538
5MeOH_mecy1ne_5_kicked25	-625.271932	-625.320032	-624.232437	-624.426980	-624.546625	0.198069	0.142665
5MeOH_mecy1ne_5_kicked69	-625.266174	-625.320543	-624.228779	-624.422816	-624.542203	0.198097	0.143220
5MeOH_mecy1ne_5_kicked32	-625.265067	-625.320907	-624.226359	-624.420529	-624.539972	0.198340	0.143675
5MeOH_mecy1ne_5_kicked02	-625.265798	-625.320344	-624.227776	-624.421979	-624.541438	0.198195	0.143143

5MeOH_mecy1ne_5_kicked24	-625.264748	-625.320800	-624.226616	-624.420152	-624.539303	0.198264	0.143989
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5MeOH_mecy1ne_1_kicked86	-625.180210	-624.141449	-624.335025	-624.454178	0.01	0.03	0.06
5MeOH_mecy1ne_1_kicked41	-625.179735	-624.140104	-624.334186	-624.453563	0.00	0.02	0.03
5MeOH_mecy1ne_1_kicked19	-625.179403	-624.140391	-624.334286	-624.453549	0.00	0.04	0.03
5MeOH_mecy1ne_1_kicked90	-625.179372	-624.140346	-624.333771	-624.452859	0.00	0.01	0.02
5MeOH_mecy1ne_1_kicked89	-625.179322	-624.140462	-624.334268	-624.453501	0.00	0.04	0.03
5MeOH_mecy1ne_1_kicked98	-625.178941	-624.140985	-624.334832	-624.454084	0.00	0.04	0.06
5MeOH_mecy1ne_1_kicked32	-625.178904	-624.139180	-624.333393	-624.452844	0.02	0.01	0.02
5MeOH_mecy1ne_1_kicked68	-625.178893	-624.140057	-624.333422	-624.452453	0.00	0.02	0.01
5MeOH_mecy1ne_1_kicked35	-625.178873	-624.139849	-624.333191	-624.452219	0.00	0.02	0.01
5MeOH_mecy1ne_1_kicked21	-625.178812	-624.139818	-624.333026	-624.452010	0.00	0.04	0.01
5MeOH_mecy1ne_2_kicked03	-625.178951	-624.137955	-624.332243	-624.451705	0.00	0.01	0.00
5MeOH_mecy1ne_2_kicked84	-625.178830	-624.138661	-624.333207	-624.452749	0.00	0.01	0.01
5MeOH_mecy1ne_2_kicked33	-625.178791	-624.140103	-624.333122	-624.452029	0.00	0.01	0.01
5MeOH_mecy1ne_2_kicked91	-625.178718	-624.138958	-624.333271	-624.452725	0.00	0.01	0.01
5MeOH_mecy1ne_2_kicked50	-625.178565	-624.140248	-624.334391	-624.453753	0.00	0.03	0.04
5MeOH_mecy1ne_2_kicked32	-625.178548	-624.138985	-624.333327	-624.452809	0.00	0.01	0.01
5MeOH_mecy1ne_2_kicked97	-625.178490	-624.139827	-624.333304	-624.452353	0.00	0.02	0.01
5MeOH_mecy1ne_2_kicked86	-625.178298	-624.139481	-624.332956	-624.452009	0.00	0.01	0.01
5MeOH_mecy1ne_2_kicked92	-625.178228	-624.139125	-624.333261	-624.452603	0.00	0.02	0.01
5MeOH_mecy1ne_2_kicked73	-625.178146	-624.139231	-624.332695	-624.451752	0.00	0.01	0.00
5MeOH_mecy1ne_3_kicked82	-625.181284	-624.140148	-624.334139	-624.453526	0.36	0.02	0.03
5MeOH_mecy1ne_3_kicked97	-625.180351	-624.140111	-624.334317	-624.453756	0.03	0.02	0.04
5MeOH_mecy1ne_3_kicked14	-625.180264	-624.140948	-624.334788	-624.454061	0.00	0.07	0.06
5MeOH_mecy1ne_3_kicked44	-625.179656	-624.139880	-624.333906	-624.453299	0.16	0.02	0.02
5MeOH_mecy1ne_3_kicked75	-625.179515	-624.141384	-624.334324	-624.453237	0.02	0.03	0.02
5MeOH_mecy1ne_3_kicked30	-625.179465	-624.141085	-624.334901	-624.454164	0.00	0.08	0.06
5MeOH_mecy1ne_3_kicked47	-625.179288	-624.141836	-624.334566	-624.453372	0.00	0.05	0.03
5MeOH_mecy1ne_3_kicked54	-625.179263	-624.141069	-624.333817	-624.452601	0.04	0.02	0.01
5MeOH_mecy1ne_3_kicked89	-625.179250	-624.139299	-624.333441	-624.452889	0.34	0.02	0.02
5MeOH_mecy1ne_3_kicked34	-625.179247	-624.141461	-624.335098	-624.454287	0.00	0.13	0.07
5MeOH_mecy1ne_4_kicked59	-625.178709	-624.138577	-624.332541	-624.451907	0.00	0.00	0.01
5MeOH_mecy1ne_4_kicked97	-625.178079	-624.141651	-624.334377	-624.453182	0.00	0.01	0.02
5MeOH_mecy1ne_4_kicked19	-625.177897	-624.141122	-624.333967	-624.452826	0.00	0.01	0.02
5MeOH_mecy1ne_4_kicked24	-625.177874	-624.141158	-624.333959	-624.452797	0.00	0.01	0.01
5MeOH_mecy1ne_4_kicked14	-625.177611	-624.140921	-624.334545	-624.453723	0.00	0.03	0.04
5MeOH_mecy1ne_4_kicked86	-625.177566	-624.138706	-624.332761	-624.452162	0.00	0.00	0.01
5MeOH_mecy1ne_4_kicked42	-625.177552	-624.140063	-624.333393	-624.452426	0.00	0.01	0.01
5MeOH_mecy1ne_4_kicked41	-625.177517	-624.139701	-624.333250	-624.452387	0.00	0.01	0.01
5MeOH_mecy1ne_4_kicked35	-625.177272	-624.139234	-624.332965	-624.452181	0.00	0.01	0.01
5MeOH_mecy1ne_4_kicked57	-625.177186	-624.138377	-624.332429	-624.451825	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked45	-625.178812	-624.137998	-624.332412	-624.452008	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked75	-625.178236	-624.138699	-624.332995	-624.452503	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked08	-625.177721	-624.138134	-624.332689	-624.452344	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked92	-625.177527	-624.137813	-624.332399	-624.452030	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked70	-625.177448	-624.137993	-624.332473	-624.452088	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked25	-625.177367	-624.137872	-624.332415	-624.452061	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked69	-625.177323	-624.139929	-624.333966	-624.453352	0.00	0.01	0.03
5MeOH_mecy1ne_5_kicked32	-625.177232	-624.138525	-624.332694	-624.452137	0.00	0.00	0.01
5MeOH_mecy1ne_5_kicked02	-625.177201	-624.139179	-624.333382	-624.452841	0.00	0.01	0.02
5MeOH_mecy1ne_5_kicked24	-625.176811	-624.138679	-624.332215	-624.451366	0.00	0.00	0.00
BoltzAvg:	-625.183130	-624.140834	-624.334186	-624.453388			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

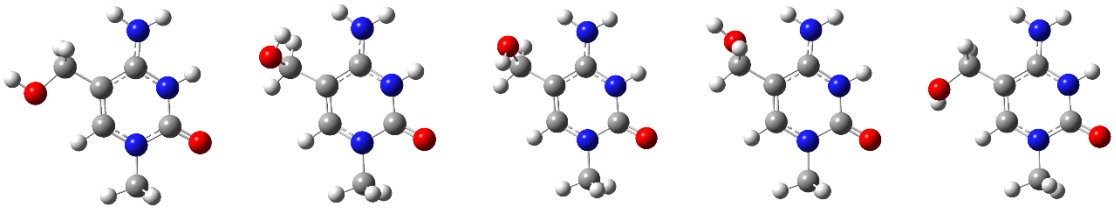


Figure 77: Conformers/tautomers for protonated 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) with relevant thermodynamic data provided in Table 81.

Table 81: QM properties for solution phase optimized tautomers/conformers of protonated 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) shown in Figure 77 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5hm_mecy1ps_9	-549.211061	-549.322499	-548.278432	-548.442203	-548.543465	0.185411	0.138218
5hm_mecy1ps_10	-549.212317	-549.321409	-548.279748	-548.443558	-548.544825	0.185191	0.138008
5hm_mecy1ps_8	-549.200878	-549.320882	-548.268707	-548.432776	-548.534126	0.185451	0.138116
5hm_mecy1ps_11	-549.203997	-549.319933	-548.272388	-548.436049	-548.537247	0.185181	0.137863
5hm_mecy1ps_7	-549.208188	-549.318761	-548.275613	-548.439687	-548.541097	0.184884	0.137112
5hm_mecy1ps_17	-549.211665	-549.310113	-548.281825	-548.445406	-548.546559	0.184699	0.137545
5hm_mecy1ps_21	-549.212459	-549.309006	-548.282623	-548.446241	-548.547384	0.184415	0.136947
5hm_mecy1ps_19	-549.210901	-549.307496	-548.281165	-548.445072	-548.546377	0.184347	0.136349
5hm_mecy1ps_20	-549.206522	-549.308582	-548.277734	-548.441246	-548.542346	0.184777	0.137487
5hm_mecy1ps_18	-549.202592	-549.308510	-548.273198	-548.437090	-548.538325	0.185450	0.138699
5hm_mecy1ps_12	-549.197426	-549.305756	-548.269089	-548.432726	-548.533827	0.184677	0.137673
5hm_mecy1ps_14	-549.198094	-549.304653	-548.269806	-548.433508	-548.534613	0.184518	0.137423
5hm_mecy1ps_16	-549.188583	-549.304113	-548.260526	-548.424555	-548.525784	0.184931	0.137701
5hm_mecy1ps_15	-549.192287	-549.304341	-548.265108	-548.428756	-548.529853	0.184838	0.138078
5hm_mecy1ps_13	-549.196895	-549.303283	-548.268702	-548.432750	-548.534056	0.184498	0.137090
5hm_mecy1ps_2	-549.166564	-549.297029	-548.236081	-548.399023	-548.499802	0.185336	0.138391
5hm_mecy1ps_3	-549.167026	-549.295960	-548.236591	-548.399642	-548.500451	0.185214	0.138326
5hm_mecy1ps_4	-549.148281	-549.292794	-548.219111	-548.382220	-548.483031	0.185408	0.137509
5hm_mecy1ps_6	-549.151936	-549.291666	-548.221972	-548.385474	-548.486486	0.185251	0.137130
5hm_mecy1ps_1	-549.144931	-549.294050	-548.215302	-548.378804	-548.479763	0.186418	0.139639
5hm_mecy1ps_5	-549.145044	-549.293912	-548.215415	-548.378907	-548.479867	0.186450	0.139995
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5hm_mecy1ps_9	-549.184281	-548.251652	-548.415423	-548.516685	0.01	0.47	0.44
5hm_mecy1ps_10	-549.183401	-548.250832	-548.414642	-548.515909	0.04	0.21	0.19
5hm_mecy1ps_8	-549.182766	-548.250595	-548.414664	-548.516014	0.00	0.20	0.22
5hm_mecy1ps_11	-549.182070	-548.250461	-548.414122	-548.515319	0.00	0.10	0.10
5hm_mecy1ps_7	-549.181649	-548.249075	-548.413148	-548.514559	0.00	0.03	0.05
5hm_mecy1ps_17	-549.172568	-548.242729	-548.406309	-548.507462	0.23	0.00	0.00
5hm_mecy1ps_21	-549.172059	-548.242223	-548.405841	-548.506985	0.54	0.00	0.00
5hm_mecy1ps_19	-549.171147	-548.241411	-548.405318	-548.506623	0.19	0.00	0.00
5hm_mecy1ps_20	-549.171095	-548.242307	-548.405819	-548.506919	0.00	0.00	0.00
5hm_mecy1ps_18	-549.169811	-548.240417	-548.404308	-548.505543	0.00	0.00	0.00

5hm_mecy1ps_12	-549.168083	-548.239746	-548.403382	-548.504484	0.00	0.00	0.00
5hm_mecy1ps_14	-549.167230	-548.238943	-548.402644	-548.503749	0.00	0.00	0.00
5hm_mecy1ps_16	-549.166412	-548.238355	-548.402383	-548.503613	0.00	0.00	0.00
5hm_mecy1ps_15	-549.166263	-548.239084	-548.402732	-548.503829	0.00	0.00	0.00
5hm_mecy1ps_13	-549.166193	-548.238000	-548.402048	-548.503354	0.00	0.00	0.00
5hm_mecy1ps_2	-549.158638	-548.228155	-548.391097	-548.491876	0.00	0.00	0.00
5hm_mecy1ps_3	-549.157634	-548.227198	-548.390249	-548.491058	0.00	0.00	0.00
5hm_mecy1ps_4	-549.155285	-548.226115	-548.389224	-548.490035	0.00	0.00	0.00
5hm_mecy1ps_6	-549.154536	-548.224572	-548.388074	-548.489086	0.00	0.00	0.00
5hm_mecy1ps_1	-549.154411	-548.224782	-548.388284	-548.489243	0.00	0.00	0.00
5hm_mecy1ps_5	-549.153917	-548.224287	-548.387779	-548.488740	0.00	0.00	0.00
BoltzAvg:	-549.183696	-548.251081	-548.414877	-548.516150			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 82: QM properties for solution phase optimized tautomers/conformers of protonated 1-methyl-5-hydroxymethylcytosine (**1m5hmC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5MeOH_mecy1ps_7_kicked15	-625.673564	-625.776580	-624.638050	-624.827816	-624.945046	0.211623	0.156317
5MeOH_mecy1ps_7_kicked40	-625.671239	-625.776214	-624.635805	-624.825687	-624.942960	0.212144	0.157340
5MeOH_mecy1ps_7_kicked50	-625.652654	-625.775842	-624.616649	-624.807786	-624.925536	0.212194	0.157102
5MeOH_mecy1ps_7_kicked54	-625.660948	-625.776078	-624.626495	-624.816571	-624.933884	0.212331	0.157404
5MeOH_mecy1ps_7_kicked10	-625.663977	-625.776134	-624.628966	-624.819206	-624.936619	0.212222	0.157567
5MeOH_mecy1ps_7_kicked55	-625.661589	-625.776090	-624.626907	-624.817076	-624.934432	0.212375	0.157648
5MeOH_mecy1ps_7_kicked36	-625.661757	-625.775067	-624.627672	-624.817531	-624.934755	0.211893	0.156774
5MeOH_mecy1ps_7_kicked30	-625.643875	-625.775306	-624.609193	-624.799620	-624.917061	0.211977	0.157328
5MeOH_mecy1ps_7_kicked99	-625.650581	-625.774640	-624.614376	-624.805321	-624.922999	0.211858	0.156757
5MeOH_mecy1ps_7_kicked85	-625.649242	-625.773282	-624.613231	-624.804650	-624.922562	0.211432	0.155526
5MeOH_mecy1ps_8_kicked72	-625.666379	-625.779874	-624.630671	-624.820283	-624.937380	0.212412	0.158373
5MeOH_mecy1ps_8_kicked31	-625.663205	-625.778308	-624.628178	-624.818107	-624.935338	0.212308	0.157415
5MeOH_mecy1ps_8_kicked80	-625.664076	-625.778346	-624.628999	-624.818873	-624.936079	0.212391	0.157800
5MeOH_mecy1ps_8_kicked54	-625.652884	-625.778416	-624.619437	-624.809234	-624.926360	0.212532	0.158159
5MeOH_mecy1ps_8_kicked45	-625.654017	-625.778270	-624.619893	-624.809964	-624.927213	0.212667	0.158244
5MeOH_mecy1ps_8_kicked42	-625.658485	-625.776909	-624.623859	-624.813974	-624.931257	0.212299	0.157320
5MeOH_mecy1ps_8_kicked10	-625.652078	-625.778470	-624.618764	-624.808543	-624.925655	0.212842	0.159325
5MeOH_mecy1ps_8_kicked46	-625.655019	-625.777014	-624.621356	-624.811213	-624.928363	0.212417	0.157993
5MeOH_mecy1ps_8_kicked94	-625.654875	-625.776982	-624.620492	-624.810699	-624.928011	0.212642	0.158035
5MeOH_mecy1ps_8_kicked71	-625.642513	-625.777849	-624.607715	-624.798037	-624.915390	0.212705	0.159007
5MeOH_mecy1ps_9_kicked72	-625.677085	-625.780140	-624.641548	-624.830926	-624.947971	0.211877	0.156826
5MeOH_mecy1ps_9_kicked70	-625.671318	-625.779739	-624.635949	-624.825745	-624.942969	0.211863	0.156597
5MeOH_mecy1ps_9_kicked77	-625.674230	-625.779796	-624.638795	-624.828291	-624.945374	0.211871	0.156737
5MeOH_mecy1ps_9_kicked94	-625.662562	-625.779910	-624.628797	-624.818251	-624.935268	0.212029	0.156926
5MeOH_mecy1ps_9_kicked12	-625.676897	-625.780124	-624.641407	-624.830778	-624.947818	0.211949	0.157196
5MeOH_mecy1ps_9_kicked95	-625.663311	-625.779830	-624.629163	-624.818788	-624.935890	0.212114	0.157328
5MeOH_mecy1ps_9_kicked21	-625.663143	-625.779862	-624.629170	-624.818703	-624.935759	0.212251	0.157736
5MeOH_mecy1ps_9_kicked22	-625.665798	-625.779879	-624.630810	-624.820765	-624.938030	0.212242	0.157804
5MeOH_mecy1ps_9_kicked52	-625.664906	-625.778388	-624.630111	-624.820006	-624.937235	0.212315	0.157540
5MeOH_mecy1ps_9_kicked54	-625.667948	-625.778154	-624.632418	-624.822159	-624.939342	0.212168	0.157397
5MeOH_mecy1ps_10_kicked30	-625.671876	-625.778608	-624.636644	-624.826451	-624.943651	0.211638	0.155959
5MeOH_mecy1ps_10_kicked45	-625.668756	-625.778638	-624.633691	-624.823643	-624.940898	0.211586	0.156130
5MeOH_mecy1ps_10_kicked68	-625.677261	-625.779080	-624.641809	-624.831274	-624.948334	0.211744	0.156581
5MeOH_mecy1ps_10_kicked93	-625.677175	-625.779056	-624.641722	-624.831216	-624.948296	0.211873	0.157100
5MeOH_mecy1ps_10_kicked74	-625.666475	-625.778755	-624.631582	-624.821568	-624.938827	0.211968	0.157177
5MeOH_mecy1ps_10_kicked41	-625.666380	-625.779874	-624.630672	-624.820282	-624.937379	0.212412	0.158375
5MeOH_mecy1ps_10_kicked37	-625.663359	-625.778860	-624.629624	-624.819147	-624.936173	0.212131	0.157555
5MeOH_mecy1ps_10_kicked31	-625.649649	-625.777979	-624.614151	-624.804460	-624.921844	0.212092	0.157946

5MeOH_mecy1ps_10_kicked59	-625.664649	-625.776930	-624.630041	-624.819811	-624.936963	0.211966	0.157266
5MeOH_mecy1ps_10_kicked69	-625.656112	-625.776243	-624.621123	-624.811815	-624.929358	0.212036	0.156600
5MeOH_mecy1ps_11_kicked68	-625.669885	-625.777750	-624.635282	-624.824593	-624.941589	0.212082	0.157452
5MeOH_mecy1ps_11_kicked83	-625.655018	-625.777354	-624.622376	-624.811717	-624.928663	0.212326	0.158042
5MeOH_mecy1ps_11_kicked67	-625.659029	-625.777250	-624.625028	-624.814855	-624.932045	0.212491	0.158066
5MeOH_mecy1ps_11_kicked06	-625.645578	-625.776607	-624.613056	-624.802420	-624.919386	0.212000	0.157491
5MeOH_mecy1ps_11_kicked58	-625.649753	-625.776745	-624.616407	-624.806011	-624.923102	0.212215	0.157707
5MeOH_mecy1ps_11_kicked04	-625.648762	-625.776647	-624.615559	-624.805159	-624.922248	0.212207	0.157832
5MeOH_mecy1ps_11_kicked23	-625.657662	-625.776396	-624.624700	-624.814077	-624.931068	0.212543	0.158283
5MeOH_mecy1ps_11_kicked60	-625.645228	-625.774485	-624.610189	-624.801138	-624.918810	0.211944	0.156707
5MeOH_mecy1ps_11_kicked27	-625.658638	-625.775751	-624.624813	-624.814393	-624.931470	0.212301	0.157982
5MeOH_mecy1ps_11_kicked78	-625.658680	-625.775683	-624.624810	-624.814433	-624.931525	0.212327	0.158105
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5MeOH_mecy1ps_7_kicked15	-625.620263	-624.584749	-624.774514	-624.891745	0.01	0.00	0.00
5MeOH_mecy1ps_7_kicked40	-625.618874	-624.583440	-624.773322	-624.890595	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked50	-625.618740	-624.582735	-624.773872	-624.891622	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked54	-625.618674	-624.584222	-624.774298	-624.891611	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked10	-625.618567	-624.583556	-624.773796	-624.891209	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked55	-625.618442	-624.583760	-624.773929	-624.891285	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked36	-625.618293	-624.584207	-624.774067	-624.891291	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked30	-625.617978	-624.583296	-624.773724	-624.891165	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked99	-625.617883	-624.581678	-624.772623	-624.890301	0.00	0.00	0.00
5MeOH_mecy1ps_7_kicked85	-625.617756	-624.581745	-624.773164	-624.891077	0.00	0.00	0.00
5MeOH_mecy1ps_8_kicked72	-625.621501	-624.585793	-624.775405	-624.892502	0.00	0.02	0.01
5MeOH_mecy1ps_8_kicked31	-625.620893	-624.585867	-624.775795	-624.893026	0.00	0.01	0.01
5MeOH_mecy1ps_8_kicked80	-625.620546	-624.585469	-624.775343	-624.892549	0.00	0.01	0.01
5MeOH_mecy1ps_8_kicked54	-625.620257	-624.586810	-624.776607	-624.893733	0.00	0.04	0.02
5MeOH_mecy1ps_8_kicked45	-625.620026	-624.585902	-624.775973	-624.893222	0.00	0.02	0.01
5MeOH_mecy1ps_8_kicked42	-625.619589	-624.584964	-624.775078	-624.892362	0.00	0.00	0.01
5MeOH_mecy1ps_8_kicked10	-625.619145	-624.585831	-624.775610	-624.892722	0.00	0.03	0.01
5MeOH_mecy1ps_8_kicked46	-625.619021	-624.585358	-624.775215	-624.892365	0.00	0.01	0.01
5MeOH_mecy1ps_8_kicked94	-625.618947	-624.584564	-624.774771	-624.892083	0.00	0.01	0.00
5MeOH_mecy1ps_8_kicked71	-625.618842	-624.584044	-624.774366	-624.891719	0.00	0.01	0.00
5MeOH_mecy1ps_9_kicked72	-625.623314	-624.587777	-624.777156	-624.894200	0.21	0.03	0.04
5MeOH_mecy1ps_9_kicked70	-625.623142	-624.587773	-624.777569	-624.894793	0.00	0.05	0.07
5MeOH_mecy1ps_9_kicked77	-625.623059	-624.587624	-624.777120	-624.894203	0.01	0.03	0.04
5MeOH_mecy1ps_9_kicked94	-625.622984	-624.589219	-624.778673	-624.895690	0.00	0.14	0.19
5MeOH_mecy1ps_9_kicked12	-625.622928	-624.587438	-624.776809	-624.893849	0.17	0.03	0.03
5MeOH_mecy1ps_9_kicked95	-625.622502	-624.588355	-624.777979	-624.895081	0.00	0.10	0.10
5MeOH_mecy1ps_9_kicked21	-625.622126	-624.588153	-624.777685	-624.894742	0.00	0.10	0.07
5MeOH_mecy1ps_9_kicked22	-625.622075	-624.587087	-624.777042	-624.894307	0.00	0.07	0.04
5MeOH_mecy1ps_9_kicked52	-625.620848	-624.586053	-624.775949	-624.893177	0.00	0.01	0.01
5MeOH_mecy1ps_9_kicked54	-625.620757	-624.585227	-624.774968	-624.892151	0.00	0.00	0.00
5MeOH_mecy1ps_10_kicked30	-625.622649	-624.587417	-624.777224	-624.894424	0.00	0.02	0.05
5MeOH_mecy1ps_10_kicked45	-625.622508	-624.587443	-624.777395	-624.894650	0.00	0.03	0.06
5MeOH_mecy1ps_10_kicked68	-625.622499	-624.587047	-624.776512	-624.893572	0.30	0.01	0.02
5MeOH_mecy1ps_10_kicked93	-625.621956	-624.586503	-624.775997	-624.893077	0.29	0.01	0.01
5MeOH_mecy1ps_10_kicked74	-625.621578	-624.586685	-624.776671	-624.893930	0.00	0.03	0.03
5MeOH_mecy1ps_10_kicked41	-625.621499	-624.585791	-624.775402	-624.892498	0.00	0.02	0.01
5MeOH_mecy1ps_10_kicked37	-625.621305	-624.587569	-624.777092	-624.894118	0.00	0.05	0.04
5MeOH_mecy1ps_10_kicked31	-625.620033	-624.584535	-624.774844	-624.892228	0.00	0.01	0.00
5MeOH_mecy1ps_10_kicked59	-625.619664	-624.585056	-624.774826	-624.891978	0.00	0.00	0.00
5MeOH_mecy1ps_10_kicked69	-625.619643	-624.584654	-624.775346	-624.892889	0.00	0.01	0.01
5MeOH_mecy1ps_11_kicked68	-625.620298	-624.585695	-624.775006	-624.892002	0.00	0.00	0.00
5MeOH_mecy1ps_11_kicked83	-625.619312	-624.586670	-624.776011	-624.892957	0.00	0.02	0.01
5MeOH_mecy1ps_11_kicked67	-625.619184	-624.585183	-624.775010	-624.892200	0.00	0.01	0.00
5MeOH_mecy1ps_11_kicked06	-625.619116	-624.586594	-624.775959	-624.892925	0.00	0.01	0.01
5MeOH_mecy1ps_11_kicked58	-625.619038	-624.585692	-624.775296	-624.892387	0.00	0.01	0.01
5MeOH_mecy1ps_11_kicked04	-625.618815	-624.585612	-624.775212	-624.892301	0.00	0.01	0.01
5MeOH_mecy1ps_11_kicked23	-625.618113	-624.585150	-624.774528	-624.891518	0.00	0.00	0.00
5MeOH_mecy1ps_11_kicked60	-625.617778	-624.582740	-624.773688	-624.891360	0.00	0.00	0.00
5MeOH_mecy1ps_11_kicked27	-625.617769	-624.583944	-624.773523	-624.890600	0.00	0.00	0.00

5MeOH_mecy1ps_11_kicked78	-625.617578	-624.583707	-624.773331	-624.890423	0.00	0.00	0.00
BoltzAvg:	-625.624515	-624.587802	-624.777285	-624.894347			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.							

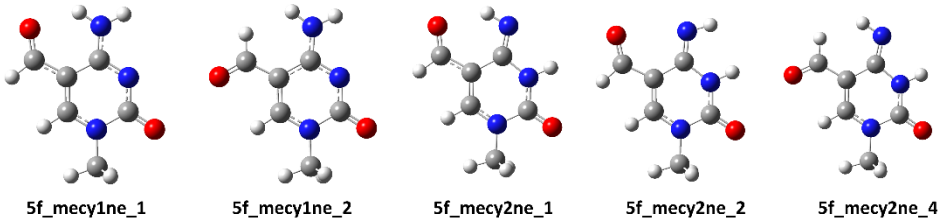


Figure 78: Conformers/tautomers for 1-methyl-5-formylcytosine (**1m5fC**) with relevant thermodynamic data provided in Table 83.

Table 83: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-formylcytosine (**1m5fC**) shown in Figure 78 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_mecy1ne_1	-547.627046	-547.660417	-546.693412	-546.857496	-546.958646	0.146896	0.101500
5f_mecy1ne_2	-547.617300	-547.654075	-546.683507	-546.847516	-546.948664	0.147024	0.100818
5f_mecy2ne_1	-547.617149	-547.646516	-546.686984	-546.850446	-546.951298	0.147565	0.102109
5f_mecy2ne_4	-547.613838	-547.641608	-546.683258	-546.846651	-546.947505	0.146359	0.101856
5f_mecy2ne_2	-547.601829	-547.640871	-546.673682	-546.836935	-546.937643	0.147549	0.102012
5f_mecy3ne_1	-547.580890	-547.620035	-546.654226	-546.817732	-546.918462	0.146936	0.101422
5f_mecy3ne_3	-547.586468	-547.616226	-546.659505	-546.822648	-546.923335	0.146637	0.100712
5f_mecy3ne_4	-547.578654	-547.614832	-546.651640	-546.815095	-546.915837	0.145886	0.101384
5f_mecy3ne_2	-547.578654	-547.614832	-546.651640	-546.815095	-546.915837	0.145886	0.101384
5f_mecy3ne_6	-547.562127	-547.614817	-546.637513	-546.801208	-546.901941	0.147020	0.101612
5f_mecy3ne_8	-547.570193	-547.611214	-546.645101	-546.808438	-546.909147	0.146754	0.101025
5f_mecy3ne_7	-547.558076	-547.611433	-546.634316	-546.797556	-546.898098	0.146838	0.101247
5f_mecy3ne_5	-547.560594	-547.609708	-546.635765	-546.799379	-546.900103	0.147073	0.101634
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_mecy1ne_1	-547.558917	-546.625283	-546.789367	-546.890517	1.00	1.00	1.00
5f_mecy1ne_2	-547.553257	-546.619463	-546.783473	-546.884621	0.00	0.00	0.00
5f_mecy2ne_1	-547.544407	-546.614243	-546.777704	-546.878556	0.00	0.00	0.00
5f_mecy2ne_4	-547.539752	-546.609172	-546.772565	-546.873418	0.00	0.00	0.00
5f_mecy2ne_2	-547.538859	-546.610712	-546.773964	-546.874672	0.00	0.00	0.00
5f_mecy3ne_1	-547.518613	-546.591948	-546.755455	-546.856184	0.00	0.00	0.00
5f_mecy3ne_3	-547.515514	-546.588551	-546.751694	-546.852381	0.00	0.00	0.00
5f_mecy3ne_4	-547.513448	-546.586434	-546.749889	-546.850631	0.00	0.00	0.00
5f_mecy3ne_2	-547.513448	-546.586434	-546.749889	-546.850631	0.00	0.00	0.00
5f_mecy3ne_6	-547.513205	-546.588591	-546.752286	-546.853019	0.00	0.00	0.00
5f_mecy3ne_8	-547.510189	-546.585097	-546.748434	-546.849143	0.00	0.00	0.00
5f_mecy3ne_7	-547.510186	-546.586426	-546.749666	-546.850209	0.00	0.00	0.00
5f_mecy3ne_5	-547.508074	-546.583245	-546.746859	-546.847583	0.00	0.00	0.00
BoltzAvg:	-547.558903	-546.625270	-546.789356	-546.890506			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 84: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-formylcytosine (**1m5fC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_mecy1ne1w_kicked88	-624.077470	-624.117039	-623.038846	-623.230137	-623.347749	0.174194	0.121226
5f_mecy1ne1w_kicked08	-624.076092	-624.116946	-623.038795	-623.229695	-623.347125	0.174332	0.121820
5f_mecy1ne1w_kicked72	-624.073888	-624.116296	-623.036429	-623.227720	-623.345307	0.174269	0.121405
5f_mecy1ne1w_kicked41	-624.069552	-624.117066	-623.032639	-623.223696	-623.341159	0.174477	0.122415
5f_mecy1ne1w_kicked10	-624.075671	-624.116144	-623.038059	-623.229501	-623.347181	0.174423	0.121787
5f_mecy1ne1w_kicked06	-624.070637	-624.115821	-623.033601	-623.224850	-623.342388	0.174408	0.121503
5f_mecy1ne1w_kicked26	-624.075510	-624.115892	-623.037902	-623.229279	-623.346921	0.174416	0.121973
5f_mecy1ne1w_kicked87	-624.069474	-624.116575	-623.034093	-623.225027	-623.342440	0.174567	0.122674
5f_mecy1ne1w_kicked21	-624.065197	-624.113976	-623.029274	-623.220081	-623.337407	0.174097	0.120437
5f_mecy1ne1w_kicked38	-624.066273	-624.114454	-623.030534	-623.221210	-623.338486	0.174187	0.120993
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_mecy1ne1w_kicked88	-623.995813	-622.957189	-623.148480	-623.266092	0.39	0.06	0.10
5f_mecy1ne1w_kicked08	-623.995126	-622.957828	-623.148729	-623.266159	0.20	0.10	0.10
5f_mecy1ne1w_kicked72	-623.994891	-622.957432	-623.148724	-623.266310	0.03	0.08	0.12
5f_mecy1ne1w_kicked41	-623.994651	-622.957737	-623.148795	-623.266258	0.00	0.17	0.12
5f_mecy1ne1w_kicked10	-623.994357	-622.956746	-623.148188	-623.265867	0.21	0.06	0.08
5f_mecy1ne1w_kicked06	-623.994318	-622.957283	-623.148531	-623.266069	0.00	0.06	0.09
5f_mecy1ne1w_kicked26	-623.993919	-622.956310	-623.147688	-623.265330	0.16	0.04	0.04
5f_mecy1ne1w_kicked87	-623.993901	-622.958520	-623.149454	-623.266867	0.00	0.39	0.22
5f_mecy1ne1w_kicked21	-623.993539	-622.957616	-623.148424	-623.265749	0.00	0.02	0.07
5f_mecy1ne1w_kicked38	-623.993461	-622.957723	-623.148398	-623.265674	0.00	0.03	0.06
BoltzAvg:	-623.997704	-622.957772	-623.148748	-623.266214			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.							

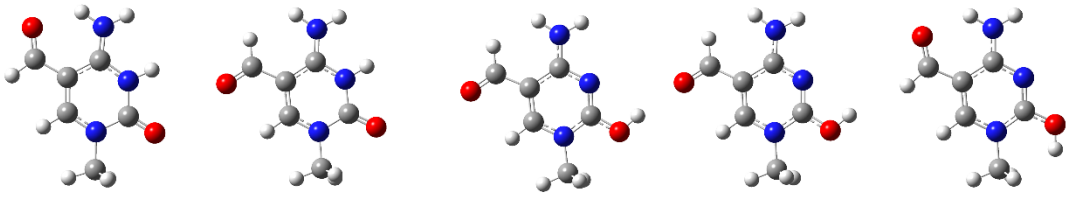


Figure 79: Conformers/tautomers for protonated 1-methyl-5-formylcytosine (**1m5fC**) with relevant thermodynamic data provided in Table 85.

Table 85: QM properties for solution phase optimized tautomers/conformers of protonated 1-methyl-5-formylcytosine (**1m5fC**) shown in Figure 79 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_mecy1ps_1	-547.998862	-548.109147	-547.070234	-547.231146	-547.330554	0.160519	0.114739
5f_mecy1ps_2	-547.985362	-548.102814	-547.056977	-547.217681	-547.317002	0.160569	0.114056
5f_mecy3ps_2	-547.984796	-548.089229	-547.059029	-547.219626	-547.318868	0.158722	0.113093
5f_mecy3ps_3	-547.984796	-548.089229	-547.059029	-547.219626	-547.318868	0.158722	0.113093
5f_mecy3ps_4	-547.980421	-548.090732	-547.056122	-547.217036	-547.316338	0.160182	0.114802
5f_mecy3ps_1	-547.970311	-548.084536	-547.046158	-547.206898	-547.306135	0.160074	0.113596

5f_mecy4ps_1	-547.961943	-548.084931	-547.032653	-547.194181	-547.293726	0.160172	0.114386
5f_mecy2ps_3	-547.955968	-548.081274	-547.029658	-547.190110	-547.289228	0.160833	0.114877
5f_mecy4ps_4	-547.956171	-548.080260	-547.026967	-547.188349	-547.287847	0.160524	0.114579
5f_mecy2ps_1	-547.952314	-548.080399	-547.025674	-547.186234	-547.285416	0.161413	0.114761
5f_mecy4ps_3	-547.947440	-548.076735	-547.018807	-547.180086	-547.279494	0.160902	0.115128
5f_mecy2ps_2	-547.934234	-548.075726	-547.008415	-547.168749	-547.267775	0.161319	0.114589
5f_mecy4ps_2	-547.951991	-548.076179	-547.024015	-547.185148	-547.284515	0.161032	0.115728
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_mecy1ps_1	-547.994408	-547.065780	-547.226692	-547.326100	1.00	1.00	1.00
5f_mecy1ps_2	-547.988758	-547.060373	-547.221078	-547.320398	0.00	0.00	0.00
5f_mecy3ps_2	-547.976136	-547.050369	-547.210966	-547.310208	0.00	0.00	0.00
5f_mecy3ps_3	-547.976136	-547.050369	-547.210966	-547.310208	0.00	0.00	0.00
5f_mecy3ps_4	-547.975930	-547.051632	-547.212545	-547.311848	0.00	0.00	0.00
5f_mecy3ps_1	-547.970940	-547.046787	-547.207527	-547.306764	0.00	0.00	0.00
5f_mecy4ps_1	-547.970545	-547.041254	-547.202783	-547.302328	0.00	0.00	0.00
5f_mecy2ps_3	-547.966397	-547.040087	-547.200539	-547.299657	0.00	0.00	0.00
5f_mecy4ps_4	-547.965681	-547.036477	-547.197859	-547.297357	0.00	0.00	0.00
5f_mecy2ps_1	-547.965638	-547.038999	-547.199559	-547.298740	0.00	0.00	0.00
5f_mecy4ps_3	-547.961607	-547.032974	-547.194253	-547.293661	0.00	0.00	0.00
5f_mecy2ps_2	-547.961137	-547.035318	-547.195652	-547.294678	0.00	0.00	0.00
5f_mecy4ps_2	-547.960451	-547.032475	-547.193608	-547.292975	0.00	0.00	0.00
BoltzAvg:	-547.994394	-547.065763	-547.226677	-547.326087			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 86: QM properties for solution phase optimized tautomers/conformers protonated 1-methyl-5-formylcytosine (**1m5fC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_mecy1ps1w_kicked57	-624.460504	-624.567471	-623.428982	-623.615859	-623.731208	0.187570	0.133957
5f_mecy1ps1w_kicked22	-624.464971	-624.567752	-623.433186	-623.619815	-623.735067	0.187585	0.134324
5f_mecy1ps1w_kicked14	-624.457706	-624.567494	-623.426262	-623.613294	-623.728701	0.187740	0.134371
5f_mecy1ps1w_kicked40	-624.454855	-624.565793	-623.423797	-623.610845	-623.726237	0.187848	0.134678
5f_mecy1ps1w_kicked34	-624.461525	-624.565952	-623.430018	-623.616816	-623.732126	0.187819	0.134879
5f_mecy1ps1w_kicked86	-624.451991	-624.565898	-623.421202	-623.608284	-623.723682	0.188100	0.135160
5f_mecy1ps1w_kicked87	-624.453044	-624.565827	-623.422122	-623.609227	-623.724644	0.188063	0.135437
5f_mecy1ps1w_kicked67	-624.441272	-624.563987	-623.410748	-623.598462	-623.714117	0.188205	0.134717
5f_mecy1ps1w_kicked41	-624.434235	-624.564113	-623.402844	-623.591042	-623.706885	0.188131	0.134966
5f_mecy1ps1w_kicked10	-624.448466	-624.562380	-623.417361	-623.604749	-623.720302	0.187759	0.133389
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_mecy1ps1w_kicked57	-624.433514	-623.401992	-623.588869	-623.704218	0.02	0.24	0.33
5f_mecy1ps1w_kicked22	-624.433428	-623.401643	-623.588271	-623.703524	0.94	0.17	0.16
5f_mecy1ps1w_kicked14	-624.433123	-623.401679	-623.588711	-623.704118	0.00	0.28	0.29
5f_mecy1ps1w_kicked40	-624.431115	-623.400057	-623.587104	-623.702497	0.00	0.06	0.05
5f_mecy1ps1w_kicked34	-624.431073	-623.399566	-623.586363	-623.701674	0.04	0.03	0.02
5f_mecy1ps1w_kicked86	-624.430738	-623.399949	-623.587030	-623.702428	0.00	0.07	0.05
5f_mecy1ps1w_kicked87	-624.430390	-623.399468	-623.586573	-623.701990	0.00	0.07	0.03
5f_mecy1ps1w_kicked67	-624.429270	-623.398746	-623.586460	-623.702116	0.00	0.03	0.04
5f_mecy1ps1w_kicked41	-624.429147	-623.397755	-623.585953	-623.701797	0.00	0.03	0.03
5f_mecy1ps1w_kicked10	-624.428991	-623.397885	-623.585273	-623.700827	0.00	0.00	0.01
BoltzAvg:	-624.434822	-623.401462	-623.588277	-623.703612			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

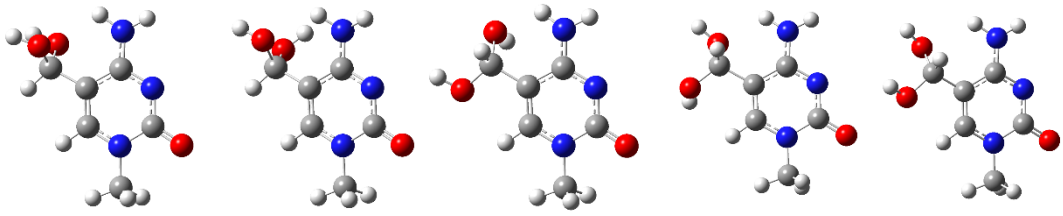


Figure 80: Conformers/tautomers for 1-methy-5-dihydroxymethylcytosine (**1m5dhmC**) with relevant thermodynamic data provided in Table 87.

Table 87: QM properties for solution phase optimized tautomers/conformers of 1-methy-5-dihydroxymethylcytosine (**1m5dhmC**) shown in Figure 80 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
1m5dhmCa4	-624.054468	-624.105212	-623.023711	-623.214849	-623.332678	0.176630	0.126861
1m5dhmCha	-624.063276	-624.105920	-623.031998	-623.222792	-623.340527	0.177011	0.128071
1m5dhmCgb	-624.060775	-624.105573	-623.029147	-623.220062	-623.337787	0.176823	0.127782
1m5dhmCna	-624.055726	-624.104959	-623.025368	-623.216339	-623.334119	0.176680	0.127229
1m5dhmCjb	-624.059987	-624.105867	-623.029450	-623.220101	-623.337734	0.177111	0.128364
1m5dhmChb	-624.057194	-624.104163	-623.025826	-623.216964	-623.334815	0.176912	0.127012
1m5dhmC10	-624.054853	-624.103711	-623.024605	-623.215555	-623.333327	0.176801	0.127000
1m5dhmC7	-624.057296	-624.105493	-623.027128	-623.217702	-623.335299	0.177320	0.128805
1m5dhmCa5	-624.056422	-624.103515	-623.025333	-623.216151	-623.333823	0.177011	0.127683
1m5dhmC6	-624.056092	-624.103561	-623.025855	-623.216664	-623.334382	0.176970	0.127840
1m5dhmCa6	-624.053336	-624.102167	-623.022857	-623.213821	-623.331569	0.176494	0.126671
1m5dhmCa9	-624.046745	-624.100057	-623.016090	-623.207230	-623.324993	0.177688	0.127898
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
1m5dhmCa4	-623.978351	-622.947594	-623.138732	-623.256561	0.00	0.22	0.34
1m5dhmCha	-623.977849	-622.946570	-623.137365	-623.255099	0.90	0.11	0.07
1m5dhmCgb	-623.977791	-622.946163	-623.137078	-623.254803	0.05	0.07	0.05
1m5dhmCna	-623.977730	-622.947373	-623.138343	-623.256123	0.00	0.19	0.22
1m5dhmCjb	-623.977503	-622.946966	-623.137617	-623.255249	0.05	0.16	0.09
1m5dhmChb	-623.977151	-622.945783	-623.136922	-623.254772	0.00	0.03	0.05
1m5dhmC10	-623.976711	-622.946463	-623.137413	-623.255185	0.00	0.05	0.08
1m5dhmC7	-623.976688	-622.946520	-623.137095	-623.254691	0.00	0.11	0.05
1m5dhmCa5	-623.975832	-622.944743	-623.135561	-623.253232	0.00	0.01	0.01
1m5dhmC6	-623.975721	-622.945484	-623.136293	-623.254011	0.00	0.03	0.02
1m5dhmCa6	-623.975496	-622.945017	-623.135981	-623.253729	0.00	0.01	0.02
1m5dhmCa9	-623.972159	-622.941504	-623.132644	-623.250407	0.00	0.00	0.00
BoltzAvg:	-623.977663-	-622.946906	-623.137916	-623.255720			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 88: QM properties for solution phase optimized tautomers/conformers of 1-methy-5-dihydroxymethylcytosine (**1m5dhmC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5dhmCa_4_aas	-700.505012	-700.562010	-699.370092	-699.588027	-699.722141	0.203430	0.146079
5dhmCgb_ade	-700.512460	-700.562574	-699.375118	-699.593308	-699.727547	0.203778	0.147473
5dhmCha_acg	-700.511163	-700.563073	-699.377384	-699.594278	-699.728011	0.204076	0.148124
5dhmCha_aay	-700.514957	-700.562839	-699.378018	-699.596030	-699.730247	0.204007	0.147998
5dhmCha_adm	-700.511080	-700.563127	-699.377550	-699.594247	-699.727891	0.204134	0.148327
5dhmCha_abz	-700.511475	-700.562145	-699.375821	-699.593851	-699.728057	0.203956	0.147602
5dhmCha_ack	-700.511173	-700.561893	-699.375135	-699.593428	-699.727736	0.203923	0.147527
5dhmCha_acu	-700.507086	-700.562689	-699.373622	-699.590325	-699.723935	0.204217	0.148349
5dhmCgb_abw	-700.508634	-700.561843	-699.372625	-699.590779	-699.724969	0.203895	0.147617
5dhmC7_aaj	-700.506731	-700.562997	-699.373817	-699.590402	-699.723949	0.204442	0.148773
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5dhmCa_4_aas	-700.415931	-699.281012	-699.498946	-699.633060	0.00	0.16	0.46
5dhmCgb_ade	-700.415101	-699.277759	-699.495949	-699.630188	0.04	0.02	0.02
5dhmCha_acg	-700.414949	-699.281170	-699.498064	-699.631797	0.07	0.19	0.12
5dhmCha_aay	-700.414841	-699.277902	-699.495913	-699.630130	0.71	0.03	0.02
5dhmCha_adm	-700.414800	-699.281270	-699.497967	-699.631611	0.06	0.18	0.10
5dhmCha_abz	-700.414543	-699.278889	-699.496918	-699.631125	0.07	0.06	0.06
5dhmCha_ack	-700.414366	-699.278328	-699.496621	-699.630930	0.05	0.05	0.05
5dhmCha_acu	-700.414340	-699.280876	-699.497579	-699.631189	0.00	0.11	0.06
5dhmCgb_abw	-700.414226	-699.278217	-699.496372	-699.630561	0.00	0.04	0.03
5dhmC7_aaj	-700.414224	-699.281310	-699.497895	-699.631443	0.00	0.17	0.08
BoltzAvg:	-700.418609	-699.281024	-699.498116	-699.632098			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

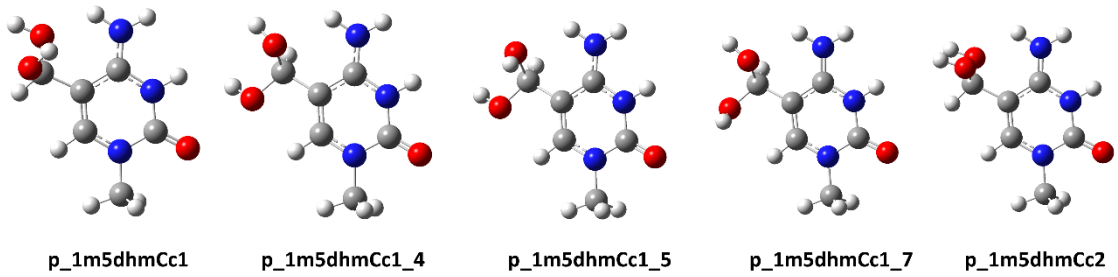


Figure 81: Conformers/tautomers for protonated 1-methy-5-dihydroxymethylcytosine (**1m5dhmC**) with relevant thermodynamic data provided in Table 89.

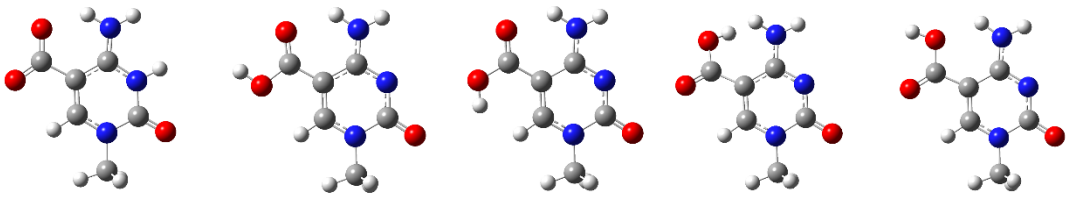
Table 89: QM properties for solution phase optimized tautomers/conformers of protonated 1-methy-5-dihydroxymethylcytosine (**1m5dhmC**) shown in Figure 81 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
p_1m5dhmCc_2	-624.445213	-624.559008	-623.417353	-623.605213	-623.721269	0.190576	0.140670
p_1m5dhmCc1_5	-624.448115	-624.559674	-623.419967	-623.607509	-623.723476	0.190766	0.141529
p_1m5dhmCc_1	-624.443583	-624.559162	-623.415068	-623.602892	-623.718942	0.190639	0.141335
p_1m5dhmCc1_7	-624.445119	-624.559719	-623.417634	-623.605036	-623.720913	0.191134	0.142201
p_1m5dhmCc1_4	-624.444513	-624.558865	-623.417044	-623.604725	-623.720719	0.190639	0.141377
p_1m5dhmC2a	-624.444620	-624.558006	-623.416505	-623.604394	-623.720484	0.190739	0.140908
p_1m5dhmCc1_10	-624.437643	-624.559141	-623.410853	-623.598201	-623.714039	0.191339	0.142611
p_1m5dhmCc_9	-624.435198	-624.557023	-623.407429	-623.595162	-623.711155	0.190661	0.141205
p_1m5dhmCc_10	-624.436475	-624.555982	-623.409255	-623.597147	-623.713241	0.190847	0.140805
p_1m5dhmC2oa	-624.444055	-624.546674	-623.418893	-623.606588	-623.722535	0.189971	0.140113
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
p_1m5dhmCc_2	-624.418338	-623.390477	-623.578338	-623.694393	0.06	0.19	0.33
p_1m5dhmCc1_5	-624.418145	-623.389997	-623.577539	-623.693506	0.61	0.15	0.13
p_1m5dhmCc_1	-624.417827	-623.389312	-623.577136	-623.693186	0.01	0.10	0.09
p_1m5dhmCc1_7	-624.417518	-623.390033	-623.577435	-623.693312	0.04	0.17	0.10
p_1m5dhmCc1_4	-624.417488	-623.390019	-623.577700	-623.693694	0.03	0.18	0.16
p_1m5dhmC2a	-624.417098	-623.388983	-623.576873	-623.692962	0.03	0.04	0.07
p_1m5dhmCc1_10	-624.416530	-623.389740	-623.577088	-623.692926	0.00	0.14	0.07
p_1m5dhmCc_9	-624.415818	-623.388049	-623.575781	-623.691774	0.00	0.02	0.02
p_1m5dhmCc_10	-624.415177	-623.387957	-623.575849	-623.691943	0.00	0.01	0.02
p_1m5dhmC2oa	-624.406561	-623.381399	-623.569095	-623.685041	0.23	0.00	0.00
BoltzAvg:	-624.417796	-623.389921	-623.577610	-623.693624			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 90: QM properties for solution phase optimized tautomers/conformers protonated 1-methy-5-dihydroxymethylcytosine (**1m5dhmC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
p_5dhmC2_aae	-700.906872	-701.016434	-699.776231	-699.989933	-700.121862	0.217096	0.159450
p_5dhmCc_264	-700.902725	-701.016367	-699.772278	-699.986245	-700.118286	0.217054	0.159417
p_5dhmCc_223	-700.909706	-701.016743	-699.778968	-699.992556	-700.124451	0.217131	0.160035
p_5dhmC2_abp	-700.896118	-701.016551	-699.767074	-699.980650	-700.112474	0.217317	0.159977
p_5dhmCc1_573	-700.913047	-701.017554	-699.781851	-699.995108	-700.126905	0.217578	0.161129
p_5dhmCc1_539	-700.912850	-701.017512	-699.781684	-699.994948	-700.126749	0.217561	0.161113
p_5dhmCc_287	-700.898210	-701.016419	-699.768189	-699.982220	-700.114273	0.217348	0.160290
p_5dhmCc_202	-700.897168	-701.016463	-699.767640	-699.981422	-700.113350	0.217454	0.160431
p_5dhmC_2a65	-700.909282	-701.015748	-699.778287	-699.991845	-700.123740	0.217353	0.160140
p_5dhmCc1_4_ads	-700.908817	-701.016789	-699.778384	-699.991800	-700.123635	0.217602	0.161199
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
p_5dhmC2_aae	-700.856984	-699.726343	-699.940045	-700.071974	0.00	0.07	0.11
p_5dhmCc_264	-700.856950	-699.726502	-699.940470	-700.072511	0.00	0.12	0.19
p_5dhmCc_223	-700.856708	-699.725970	-699.939558	-700.071453	0.04	0.07	0.06
p_5dhmC2_abp	-700.856574	-699.727530	-699.941106	-700.072929	0.00	0.26	0.30
p_5dhmCc1_573	-700.856425	-699.725229	-699.938486	-700.070283	0.50	0.04	0.02
p_5dhmCc1_539	-700.856399	-699.725233	-699.938497	-700.070299	0.43	0.04	0.02
p_5dhmCc_287	-700.856129	-699.726108	-699.940139	-700.072192	0.00	0.16	0.14
p_5dhmCc_202	-700.856032	-699.726504	-699.940286	-700.072215	0.00	0.17	0.14
p_5dhmC_2a65	-700.855608	-699.724612	-699.938171	-700.070066	0.02	0.01	0.01
p_5dhmCc1_4_ads	-700.855590	-699.725157	-699.938573	-700.070407	0.02	0.05	0.02
BoltzAvg:	-700.859083	-699.726643	-699.940354	-700.072269			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							



5ca_mecy1ne_1 5ca_mecy2ne_1 5ca_mecy2ne_2 5ca_mecy2ne_3 5ca_mecy2ne_4

Figure 82: Conformers/tautomers for 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 91.

Table 91: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-carboxycytosine (**1m5caC**) shown in Figure 82 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
5ca_mecy1ne_1	-622.855955	-622.920564	-621.822092	-622.012436	-622.129566	0.153275	0.105871
5ca_mecy2ne_1	-622.883851	-622.918576	-621.854974	-622.043450	-622.159677	0.152708	0.105118
5ca_mecy2ne_4	-622.879905	-622.916252	-621.850452	-622.038950	-622.155208	0.152857	0.105081
5ca_mecy2ne_2	-622.867682	-622.912515	-621.840542	-622.029281	-622.145508	0.153146	0.105573
5ca_mecy2ne_3	-622.864931	-622.908091	-621.837866	-622.026063	-622.142067	0.153128	0.105613
5ca_mecy3ne_2	-622.845273	-622.906355	-621.814085	-622.004182	-622.121137	0.152933	0.105706
5ca_mecy3ne_1	-622.831195	-622.901883	-621.801565	-621.991812	-622.108781	0.152889	0.105573
5ca_mecy4ne_1	-622.832242	-622.898133	-621.801042	-621.990307	-622.106860	0.152794	0.105412

Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	Boltz_pop (E_{CBS})	Boltz_pop (H_{CBS})	Boltz_pop (G_{CBS})
5ca_mecy1ne_1	-622.814693	-621.780830	-621.971174	-622.088304	0.00	0.29	0.25
5ca_mecy2ne_1	-622.813458	-621.784581	-621.973058	-622.089285	0.99	0.68	0.71
5ca_mecy2ne_4	-622.811171	-621.781717	-621.970215	-622.086474	0.01	0.03	0.04
5ca_mecy2ne_2	-622.806942	-621.779801	-621.968541	-622.084767	0.00	0.01	0.01
5ca_mecy2ne_3	-622.802478	-621.775413	-621.963610	-622.079614	0.00	0.00	0.00
5ca_mecy3ne_2	-622.800649	-621.769461	-621.959558	-622.076513	0.00	0.00	0.00
5ca_mecy3ne_1	-622.796310	-621.766680	-621.956927	-622.073896	0.00	0.00	0.00
5ca_mecy4ne_1	-622.792721	-621.761521	-621.950786	-622.067339	0.00	0.00	0.00
BoltzAvg:	-622.814368	-621.784359	-621.972694	-622.088911			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91$kJ mol^{-1} not jet applied.							

Table 92: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-carboxycytosine (**1m5caC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
5ca_mecy2ne_1_kicked02	-699.334061	-699.376653	-698.202334	-698.416950	-698.549173	0.179765	0.125375
5ca_mecy1ne_1_kicked81	-699.304574	-699.377793	-698.168130	-698.384754	-698.517935	0.180234	0.125480
5ca_mecy2ne_1_kicked52	-699.326851	-699.374623	-698.196129	-698.411443	-698.543930	0.180176	0.126031
5ca_mecy2ne_1_kicked98	-699.326975	-699.375166	-698.194877	-698.410263	-698.542775	0.179981	0.125369
5ca_mecy1ne_1_kicked83	-699.301812	-699.377978	-698.166394	-698.383274	-698.516561	0.181067	0.127396
5ca_mecy2ne_1_kicked100	-699.326369	-699.375187	-698.194114	-698.409570	-698.542105	0.180051	0.125747
5ca_mecy1ne_1_kicked90	-699.310535	-699.378038	-698.173732	-698.390010	-698.523077	0.180203	0.125405
5ca_mecy1ne_1_kicked76	-699.303990	-699.377133	-698.166638	-698.383864	-698.517288	0.180169	0.125262
5ca_mecy1ne_1_kicked28	-699.308086	-699.377575	-698.170577	-698.387475	-698.520798	0.180096	0.125243
5ca_mecy2ne_1_kicked91	-699.333145	-699.374104	-698.200244	-698.416061	-698.548810	0.179929	0.124945
5ca_mecy2ne_1_kicked92	-699.322673	-699.372582	-698.191623	-698.406582	-698.538872	0.179719	0.124187
5ca_mecy2ne_4_kicked07	-699.330248	-699.374888	-698.197906	-698.412647	-698.544954	0.179918	0.125102
5ca_mecy1ne_1_kicked45	-699.300631	-699.376445	-698.164837	-698.381755	-698.515060	0.180744	0.126589
5ca_mecy1ne_1_kicked02	-699.300686	-699.376498	-698.165007	-698.381800	-698.515069	0.180762	0.126601
5ca_mecy2ne_1_kicked90	-699.331272	-699.374375	-698.198537	-698.414141	-698.546763	0.178918	0.125642
5ca_mecy1ne_1_kicked74	-699.310855	-699.377453	-698.173129	-698.390474	-698.523996	0.180937	0.126512
5ca_mecy1ne_1_kicked55	-699.308303	-699.377676	-698.170604	-698.387490	-698.520802	0.180474	0.126160
5ca_mecy2ne_1_kicked49	-699.332369	-699.373961	-698.199526	-698.415279	-698.547982	0.180177	0.125574
5ca_mecy2ne_1_kicked34	-699.324313	-699.372393	-698.193236	-698.408272	-698.540616	0.179997	0.124717
5ca_mecy1ne_1_kicked94	-699.300541	-699.375101	-698.163097	-698.380736	-698.514335	0.180331	0.125203
5ca_mecy2ne_1_kicked13	-699.325104	-699.372350	-698.192898	-698.408576	-698.541216	0.180087	0.124833
5ca_mecy2ne_4_kicked20	-699.330302	-699.374846	-698.198098	-698.412687	-698.544933	0.180045	0.125924
5ca_mecy2ne_4_kicked91	-699.322288	-699.372956	-698.189476	-698.404980	-698.537569	0.180138	0.125302
5ca_mecy2ne_4_kicked57	-699.323005	-699.372415	-698.191715	-698.407022	-698.539521	0.180234	0.126160
5ca_mecy2ne_4_kicked17	-699.323697	-699.372925	-698.191194	-698.406528	-698.539052	0.180179	0.125794
5ca_mecy2ne_4_kicked100	-699.327287	-699.372155	-698.194017	-698.409665	-698.542332	0.179910	0.124918
5ca_mecy2ne_4_kicked62	-699.331059	-699.372992	-698.196548	-698.412180	-698.544863	0.179895	0.124950
5ca_mecy2ne_4_kicked47	-699.326642	-699.371851	-698.193267	-698.409136	-698.541890	0.180108	0.125321
5ca_mecy2ne_4_kicked61	-699.327371	-699.371543	-698.194859	-698.410277	-698.542838	0.180164	0.125350
5ca_mecy2ne_4_kicked92	-699.326999	-699.371887	-698.193402	-698.409308	-698.542080	0.178954	0.125562
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	Boltz_pop (E_{CBS})	Boltz_pop (H_{CBS})	Boltz_pop (G_{CBS})
5ca_mecy2ne_1_kicked02	-699.251278	-698.119551	-698.334167	-698.466390	0.48	0.20	0.19
5ca_mecy1ne_1_kicked81	-699.252313	-698.115868	-698.332492	-698.465674	0.00	0.06	0.09
5ca_mecy2ne_1_kicked52	-699.248592	-698.117869	-698.333184	-698.465670	0.00	0.12	0.09
5ca_mecy2ne_1_kicked98	-699.249797	-698.117699	-698.333085	-698.465597	0.00	0.07	0.08
5ca_mecy1ne_1_kicked83	-699.250582	-698.115164	-698.332045	-698.465332	0.00	0.14	0.06

5ca_mecy2ne_1_kicked100	-699.249440	-698.117185	-698.332641	-698.465176	0.00	0.06	0.05
5ca_mecy1ne_1_kicked90	-699.252633	-698.115830	-698.332108	-698.465175	0.00	0.04	0.05
5ca_mecy1ne_1_kicked76	-699.251871	-698.114519	-698.331745	-698.465169	0.00	0.03	0.05
5ca_mecy1ne_1_kicked28	-699.252332	-698.114823	-698.331721	-698.465044	0.00	0.03	0.05
5ca_mecy2ne_1_kicked91	-699.249159	-698.116258	-698.332075	-698.464824	0.33	0.02	0.04
5ca_mecy2ne_1_kicked92	-699.248395	-698.117345	-698.332303	-698.464594	0.00	0.01	0.03
5ca_mecy2ne_4_kicked07	-699.249786	-698.117444	-698.332185	-698.464492	0.01	0.02	0.03
5ca_mecy1ne_1_kicked45	-699.249856	-698.114062	-698.330981	-698.464285	0.00	0.03	0.02
5ca_mecy1ne_1_kicked02	-699.249897	-698.114218	-698.331011	-698.464280	0.00	0.03	0.02
5ca_mecy2ne_1_kicked90	-699.248733	-698.115998	-698.331603	-698.464225	0.04	0.06	0.02
5ca_mecy1ne_1_kicked74	-699.250941	-698.113215	-698.330560	-698.464082	0.00	0.02	0.02
5ca_mecy1ne_1_kicked55	-699.251516	-698.113817	-698.330703	-698.464015	0.00	0.02	0.02
5ca_mecy2ne_1_kicked49	-699.248387	-698.115544	-698.331297	-698.464000	0.14	0.01	0.02
5ca_mecy2ne_1_kicked34	-699.247676	-698.116599	-698.331635	-698.463979	0.00	0.01	0.02
5ca_mecy1ne_1_kicked94	-699.249898	-698.112454	-698.330092	-698.463691	0.00	0.01	0.01
5ca_mecy2ne_1_kicked13	-699.247517	-698.115310	-698.330989	-698.463628	0.00	0.00	0.01
5ca_mecy2ne_4_kicked20	-699.248922	-698.116718	-698.331307	-698.463553	0.01	0.01	0.01
5ca_mecy2ne_4_kicked91	-699.247654	-698.114842	-698.330346	-698.462935	0.00	0.00	0.01
5ca_mecy2ne_4_kicked57	-699.246254	-698.114964	-698.330271	-698.462770	0.00	0.01	0.00
5ca_mecy2ne_4_kicked17	-699.247131	-698.114629	-698.329962	-698.462486	0.00	0.00	0.00
5ca_mecy2ne_4_kicked100	-699.247237	-698.113968	-698.329615	-698.462283	0.00	0.00	0.00
5ca_mecy2ne_4_kicked62	-699.248042	-698.113531	-698.329163	-698.461846	0.00	0.00	0.00
5ca_mecy2ne_4_kicked47	-699.246530	-698.113155	-698.329024	-698.461778	0.00	0.00	0.00
5ca_mecy2ne_4_kicked61	-699.246193	-698.113681	-698.329099	-698.461660	0.00	0.00	0.00
5ca_mecy2ne_4_kicked92	-699.246325	-698.112728	-698.328633	-698.461405	0.00	0.00	0.00
BoltzAvg:	-699.254945	-698.118382	-698.332840	-698.465261			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

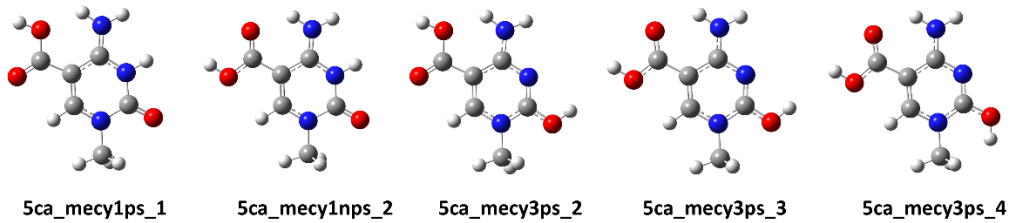


Figure 83: Conformers/tautomers for protonated 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 93.

Table 93: QM properties for solution phase optimized tautomers/conformers of protonated 1-methyl-5-carboxycytosine (**1m5caC**) shown in Figure 83 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5ca_mecy1ps_2	-623.260215	-623.368213	-622.235866	-622.421073	-622.535484	0.166383	0.118618
5ca_mecy1ps_1	-623.254651	-623.366021	-622.229894	-622.415013	-622.529386	0.166665	0.118781
5ca_mecy3ps_3	-623.257373	-623.354926	-622.235590	-622.420630	-622.534922	0.165989	0.118371
5ca_mecy3ps_2	-623.252970	-623.352765	-622.230722	-622.415744	-622.530045	0.166325	0.118720
5ca_mecy3ps_4	-623.242327	-623.350213	-622.222207	-622.407404	-622.521703	0.166028	0.118449
5ca_mecy3ps_1	-623.238354	-623.348122	-622.217689	-622.402861	-622.517161	0.166281	0.118701
5ca_mecy2ps_1	-623.214106	-623.340361	-622.192158	-622.376977	-622.491141	0.167303	0.118852
5ca_mecy2ps_2	-623.206816	-623.338795	-622.184717	-622.369383	-622.483474	0.167508	0.119455
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5ca_mecy1ps_2	-623.249595	-622.225247	-622.410454	-622.524864	0.64	0.96	0.96
5ca_mecy1ps_1	-623.247240	-622.222483	-622.407602	-622.521974	0.00	0.04	0.04
5ca_mecy3ps_3	-623.236555	-622.214772	-622.399812	-622.514105	0.35	0.00	0.00
5ca_mecy3ps_2	-623.234045	-622.211797	-622.396819	-622.511120	0.00	0.00	0.00
5ca_mecy3ps_4	-623.231764	-622.211644	-622.396842	-622.511140	0.00	0.00	0.00
5ca_mecy3ps_1	-623.229421	-622.208757	-622.393928	-622.508228	0.00	0.00	0.00
5ca_mecy2ps_1	-623.221509	-622.199561	-622.384380	-622.498545	0.00	0.00	0.00
5ca_mecy2ps_2	-623.219340	-622.197241	-622.381907	-622.495999	0.00	0.00	0.00
BoltzAvg:	-623.249416	-622.225106	-622.410321	-622.524735			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 94: QM properties for solution phase optimized tautomers/conformers of protonated 1-methyl-5-carboxycytosine (**1m5caC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5ca_mecy1ps_1_kicked14	-699.715012	-699.825866	-698.587229	-698.798280	-698.928558	0.193608	0.138982
5ca_mecy1ps_1_kicked30	-699.711900	-699.825386	-698.584329	-698.795503	-698.925820	0.193457	0.138622
5ca_mecy1ps_1_kicked93	-699.714529	-699.825815	-698.586640	-698.797810	-698.928139	0.193638	0.139200
5ca_mecy1ps_1_kicked85	-699.713755	-699.825693	-698.585891	-698.797090	-698.927426	0.193611	0.139136
5ca_mecy1ps_1_kicked36	-699.720857	-699.824559	-698.592880	-698.803756	-698.933998	0.193318	0.138145
5ca_mecy1ps_1_kicked75	-699.707789	-699.824159	-698.581449	-698.792368	-698.922577	0.193351	0.138291
5ca_mecy1ps_1_kicked77	-699.708168	-699.822747	-698.581994	-698.792861	-698.923050	0.193816	0.138633
5ca_mecy1ps_1_kicked01	-699.708483	-699.822929	-698.582280	-698.793143	-698.923329	0.193759	0.138948
5ca_mecy1ps_1_kicked38	-699.702007	-699.818713	-698.575671	-698.786772	-698.917083	0.192976	0.136503
5ca_mecy1ps_1_kicked26	-699.699257	-699.820679	-698.571502	-698.783654	-698.914368	0.193933	0.138655
5ca_mecy1ps_2_kicked13	-699.718909	-699.827421	-698.591621	-698.802742	-698.933050	0.193134	0.138272
5ca_mecy1ps_2_kicked90	-699.719038	-699.827456	-698.591640	-698.802824	-698.933160	0.193155	0.138496
5ca_mecy1ps_2_kicked98	-699.718566	-699.827384	-698.591361	-698.802447	-698.932736	0.193248	0.138674
5ca_mecy1ps_2_kicked91	-699.718155	-699.827240	-698.590800	-698.802070	-698.932429	0.193337	0.138810
5ca_mecy1ps_2_kicked68	-699.717988	-699.826305	-698.590954	-698.802278	-698.932675	0.193351	0.138163
5ca_mecy1ps_2_kicked40	-699.725960	-699.826680	-698.598555	-698.809448	-698.939679	0.193456	0.138549
5ca_mecy1ps_2_kicked83	-699.716321	-699.826336	-698.589395	-698.800767	-698.931182	0.193497	0.138588
5ca_mecy1ps_2_kicked05	-699.713098	-699.826461	-698.587291	-698.798235	-698.928439	0.193568	0.138779
5ca_mecy1ps_2_kicked93	-699.723974	-699.826426	-698.596645	-698.807639	-698.937910	0.193494	0.138815
5ca_mecy1ps_2_kicked30	-699.712541	-699.824927	-698.586967	-698.797900	-698.928108	0.194188	0.139270
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5ca_mecy1ps_1_kicked14	-699.686884	-698.559102	-698.770152	-698.900430	0.00	0.01	0.01
5ca_mecy1ps_1_kicked30	-699.686764	-698.559192	-698.770367	-698.900684	0.00	0.01	0.01
5ca_mecy1ps_1_kicked93	-699.686615	-698.558726	-698.769896	-698.900225	0.00	0.01	0.01
5ca_mecy1ps_1_kicked85	-699.686557	-698.558693	-698.769891	-698.900228	0.00	0.01	0.01
5ca_mecy1ps_1_kicked36	-699.686414	-698.558437	-698.769313	-698.899555	0.00	0.00	0.00
5ca_mecy1ps_1_kicked75	-699.685868	-698.559529	-698.770447	-698.900656	0.00	0.01	0.01
5ca_mecy1ps_1_kicked77	-699.684114	-698.557940	-698.768807	-698.898995	0.00	0.00	0.00
5ca_mecy1ps_1_kicked01	-699.683981	-698.557779	-698.768642	-698.898827	0.00	0.00	0.00
5ca_mecy1ps_1_kicked38	-699.682210	-698.555874	-698.766974	-698.897286	0.00	0.00	0.00
5ca_mecy1ps_1_kicked26	-699.682024	-698.554269	-698.766420	-698.897134	0.00	0.00	0.00
5ca_mecy1ps_2_kicked13	-699.689149	-698.561862	-698.772982	-698.903290	0.00	0.16	0.18
5ca_mecy1ps_2_kicked90	-699.688960	-698.561562	-698.772746	-698.903082	0.00	0.16	0.14
5ca_mecy1ps_2_kicked98	-699.688710	-698.561506	-698.772592	-698.902880	0.00	0.14	0.12
5ca_mecy1ps_2_kicked91	-699.688430	-698.561076	-698.772345	-698.902705	0.00	0.12	0.10
5ca_mecy1ps_2_kicked68	-699.688142	-698.561109	-698.772432	-698.902829	0.00	0.07	0.11
5ca_mecy1ps_2_kicked40	-699.688131	-698.560726	-698.771618	-698.901849	0.86	0.03	0.04
5ca_mecy1ps_2_kicked83	-699.687748	-698.560822	-698.772193	-698.902609	0.00	0.07	0.09
5ca_mecy1ps_2_kicked05	-699.687682	-698.561875	-698.772819	-698.903023	0.00	0.13	0.13
5ca_mecy1ps_2_kicked93	-699.687611	-698.560283	-698.771276	-698.901548	0.13	0.03	0.03
5ca_mecy1ps_2_kicked30	-699.685657	-698.560083	-698.771016	-698.901224	0.00	0.02	0.02
BoltzAvg:	-699.690201	-698.561268	-698.772394	-698.902710			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

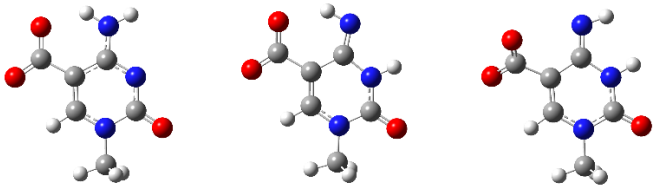


Figure 84: Conformers/tautomers for deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) with relevant thermodynamic data provided in Table 95.

Table 95: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) shown in Figure 84 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
d_5caC_1ng	-622.350453	-622.463556	-621.309520	-621.504391	-621.623919	0.139721	0.092580
d_5caC_2ng	-622.336608	-622.448326	-621.298552	-621.492723	-621.611904	0.140430	0.093477
d_5caC_3ng	-622.307523	-622.440025	-621.269427	-621.464506	-621.583991	0.140165	0.092645
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
d_5caC_1ng	-622.370976	-621.330043	-621.524914	-621.644441	1.00	1.00	1.00
d_5caC_2ng	-622.354849	-621.316793	-621.510964	-621.630145	0.00	0.00	0.00
d_5caC_3ng	-622.347380	-621.309284	-621.504363	-621.623848	0.00	0.00	0.00
BoltzAvg:	-622.370976	-621.330043	-621.524914	-621.644441			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 96: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-carboxycytosine (**1m5caC**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
d_5caC_1ng_kicked83	-698.812454	-698.921137	-697.667790	-697.888881	-698.024461	0.166524	0.112160
d_5caC_1ng_kicked63	-698.808527	-698.920816	-697.662309	-697.884169	-698.020079	0.166676	0.112145
d_5caC_1ng_kicked75	-698.807192	-698.920693	-697.661944	-697.883548	-698.019342	0.166744	0.112296
d_5caC_1ng_kicked88	-698.812228	-698.921220	-697.667462	-697.888571	-698.024165	0.166831	0.112887
d_5caC_1ng_kicked26	-698.801392	-698.920834	-697.656688	-697.878357	-698.014142	0.166909	0.112533
d_5caC_1ng_kicked31	-698.807855	-698.920723	-697.662141	-697.883890	-698.019747	0.166689	0.112451
d_5caC_1ng_kicked60	-698.806375	-698.921557	-697.663491	-697.884616	-698.020188	0.167196	0.113532
d_5caC_1ng_kicked55	-698.808851	-698.920927	-697.664255	-697.885973	-698.021818	0.167147	0.112943
d_5caC_1ng_kicked27	-698.806913	-698.919948	-697.661928	-697.883712	-698.019567	0.166795	0.112327
d_5caC_1ng_kicked86	-698.805083	-698.920154	-697.662201	-697.883103	-698.018623	0.166958	0.112911
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
d_5caC_1ng_kicked83	-698.808977	-697.664312	-697.885403	-698.020984	0.55	0.08	0.11
d_5caC_1ng_kicked63	-698.808671	-697.662453	-697.884313	-698.020223	0.01	0.03	0.05
d_5caC_1ng_kicked75	-698.808397	-697.663149	-697.884753	-698.020547	0.00	0.05	0.07
d_5caC_1ng_kicked88	-698.808333	-697.663567	-697.884677	-698.020270	0.40	0.06	0.05
d_5caC_1ng_kicked26	-698.808301	-697.663596	-697.885266	-698.021051	0.00	0.09	0.12
d_5caC_1ng_kicked31	-698.808272	-697.662558	-697.884307	-698.020164	0.00	0.04	0.05
d_5caC_1ng_kicked60	-698.808025	-697.665141	-697.886267	-698.021838	0.01	0.43	0.28

d_5caC_1ng_kicked55	-698.807984	-697.663389	-697.885106	-698.020951	0.03	0.09	0.11
d_5caC_1ng_kicked27	-698.807621	-697.662636	-697.884420	-698.020275	0.00	0.03	0.05
d_5caC_1ng_kicked86	-698.807243	-697.664362	-697.885263	-698.020784	0.00	0.09	0.09
BoltzAvg:	-698.812568	-697.664292	-697.885404	-698.021025			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

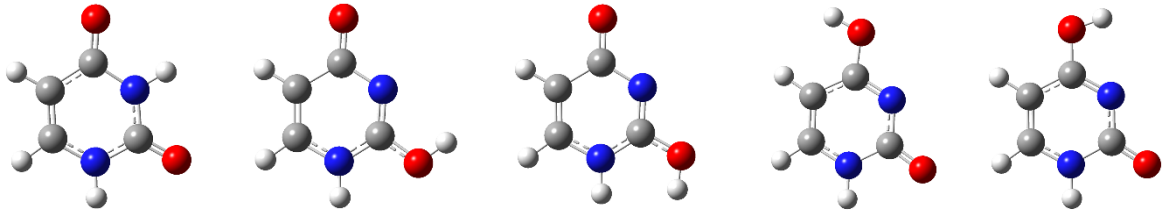


Figure 85: Conformers/tautomers for uracil (**ura**) with relevant thermodynamic data provided in Table 97.

Table 97: QM properties for solution phase optimized tautomers/conformers of uracil (**ura**) shown in Figure 85 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
ur1ne	-414.853149	-414.879906	-414.168330	-414.293678	-414.370915	0.093576	0.056083
ur3ne_2	-414.832786	-414.864735	-414.149621	-414.275108	-414.352338	0.093177	0.055847
ur3ne_1	-414.821121	-414.863190	-414.138963	-414.264668	-414.341926	0.093232	0.055988
ur2ne_1	-414.820361	-414.857319	-414.138383	-414.263979	-414.341208	0.092927	0.055248
ur2ne_2	-414.802545	-414.853714	-414.122189	-414.248096	-414.325389	0.093158	0.055622
ur4ne_1	-414.832495	-414.852931	-414.153152	-414.277995	-414.354912	0.092611	0.055215
ur4ne_2	-414.830661	-414.852595	-414.151672	-414.276444	-414.353308	0.092722	0.055445
ur4ne_3	-414.823788	-414.852142	-414.145609	-414.270492	-414.347362	0.092872	0.055743
ur4ne_4	-414.823743	-414.851954	-414.145628	-414.270498	-414.347370	0.093147	0.056065
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
ur1ne	-414.823823	-414.139004	-414.264352	-414.341589	1.00	1.00	1.00
ur3ne_2	-414.808888	-414.125722	-414.251210	-414.328440	0.00	0.00	0.00
ur3ne_1	-414.807202	-414.125043	-414.250749	-414.328006	0.00	0.00	0.00
ur2ne_1	-414.802071	-414.120092	-414.245689	-414.322918	0.00	0.00	0.00
ur2ne_2	-414.798092	-414.117736	-414.243643	-414.320936	0.00	0.00	0.00
ur4ne_1	-414.797716	-414.118372	-414.243215	-414.320133	0.00	0.00	0.00
ur4ne_2	-414.797150	-414.118161	-414.242933	-414.319797	0.00	0.00	0.00
ur4ne_3	-414.796399	-414.118220	-414.243103	-414.319973	0.00	0.00	0.00
ur4ne_4	-414.795889	-414.117774	-414.242644	-414.319516	0.00	0.00	0.00
BoltzAvg:	-414.823823	-414.139004	-414.264352	-414.341589			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 98: QM properties for solution phase optimized tautomers/conformers of uracil (**ura**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
ur1ne1w_kicked96	-491.300052	-491.335433	-490.511209	-490.663808	-490.757513	0.120783	0.075709
ur1ne1w_kicked53	-491.297593	-491.335063	-490.510455	-490.662069	-490.755310	0.120538	0.075365
ur1ne1w_kicked68	-491.298857	-491.335674	-490.511439	-490.663573	-490.757049	0.121038	0.076050
ur1ne1w_kicked54	-491.298860	-491.335112	-490.511386	-490.663138	-490.756455	0.120645	0.075560
ur1ne1w_kicked66	-491.300044	-491.335142	-490.512431	-490.664176	-490.757487	0.120701	0.075650
ur1ne1w_kicked18	-491.302189	-491.335362	-490.514454	-490.666076	-490.759352	0.120817	0.075960
ur1ne1w_kicked78	-491.300304	-491.335459	-490.511098	-490.663825	-490.757590	0.120860	0.076133
ur1ne1w_kicked99	-491.297951	-491.335085	-490.510712	-490.662384	-490.755657	0.120684	0.075826
ur1ne1w_kicked04	-491.295567	-491.335822	-490.508003	-490.660304	-490.753854	0.120965	0.076584
ur1ne1w_kicked93	-491.299052	-491.335096	-490.510404	-490.663116	-490.756850	0.120918	0.075923
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{Solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
ur1ne1w_kicked96	-491.259724	-490.470882	-490.623480	-490.717185	0.08	0.08	0.10
ur1ne1w_kicked53	-491.259698	-490.472560	-490.624175	-490.717415	0.01	0.09	0.13
ur1ne1w_kicked68	-491.259624	-490.472206	-490.624340	-490.717815	0.05	0.17	0.19
ur1ne1w_kicked54	-491.259552	-490.472078	-490.623830	-490.717147	0.03	0.08	0.10
ur1ne1w_kicked66	-491.259492	-490.471879	-490.623624	-490.716935	0.08	0.06	0.08
ur1ne1w_kicked18	-491.259402	-490.471668	-490.623289	-490.716565	0.60	0.05	0.05
ur1ne1w_kicked78	-491.259326	-490.470120	-490.622846	-490.716612	0.09	0.06	0.05
ur1ne1w_kicked99	-491.259259	-490.472020	-490.623692	-490.716965	0.01	0.08	0.08
ur1ne1w_kicked04	-491.259238	-490.471673	-490.623975	-490.717525	0.00	0.24	0.14
ur1ne1w_kicked93	-491.259173	-490.470525	-490.623236	-490.716971	0.04	0.07	0.08
BoltzAvg:	-491.261192	-490.471987	-490.623835	-490.717265			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

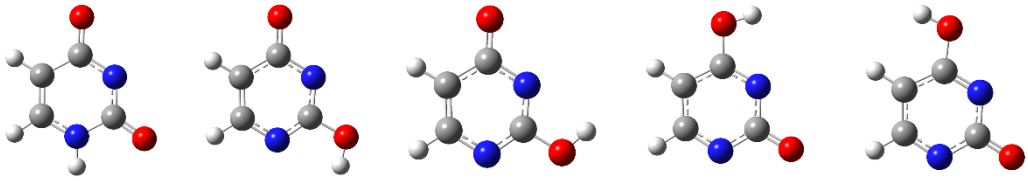


Figure 86: Conformers/tautomers for deprotonated uracil (**ura**) with relevant thermodynamic data provided in Table 99.

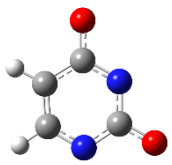
Table 99: QM properties for solution phase optimized tautomers/conformers of deprotonated uracil (**ura**) shown in Figure 86 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
ur1ng	-414.286895	-414.407002	-413.591646	-413.723647	-413.804075	0.080144	0.043049
ur2ng_2	-414.286678	-414.392929	-413.594724	-413.725479	-413.805334	0.079781	0.042799
ur2ng_1	-414.284223	-414.392766	-413.592473	-413.723315	-413.803192	0.079703	0.042686
ur3ng_1	-414.277987	-414.390603	-413.585154	-413.715735	-413.795495	0.079616	0.042641
ur3ng_2	-414.267016	-414.389187	-413.575268	-413.706057	-413.785865	0.079654	0.042736
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
ur1ng	-414.363953	-413.668704	-413.800705	-413.881133	0.19	1.00	1.00
ur2ng_2	-414.350130	-413.658176	-413.788932	-413.868786	0.73	0.00	0.00
ur2ng_1	-414.350080	-413.658331	-413.789172	-413.869049	0.08	0.00	0.00
ur3ng_1	-414.347962	-413.655129	-413.785709	-413.865469	0.00	0.00	0.00
ur3ng_2	-414.346451	-413.654703	-413.785492	-413.865300	0.00	0.00	0.00
BoltzAvg:	-414.363953	-413.668704	-413.800705	-413.881133			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91$kJ mol^{-1} not jet applied.							

Table 100: QM properties for solution phase optimized tautomers/conformers of deprotonated uracil (**ura**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	**corr. ΔH**[a] (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG**[a,b] (UB3LYP-D3/ 6-31+G(d,p))
ur1ng1w_kicked25	-490.747399	-490.864933	-489.947815	-490.106202	-490.202756	0.106836	0.062768
ur1ng1w_kicked19	-490.747491	-490.864785	-489.948483	-490.107218	-490.203939	0.107213	0.062842
ur1ng1w_kicked79	-490.747474	-490.865053	-489.947989	-490.106352	-490.202906	0.106976	0.063261
ur1ng1w_kicked08	-490.747712	-490.864971	-489.948084	-490.106457	-490.203003	0.106986	0.063285
ur1ng1w_kicked97	-490.747083	-490.864806	-489.948916	-490.107324	-490.203882	0.107206	0.063156
ur1ng1w_kicked64	-490.743108	-490.864046	-489.944198	-490.102956	-490.199638	0.106947	0.062426
ur1ng1w_kicked59	-490.749271	-490.864444	-489.950126	-490.108514	-490.205067	0.106951	0.062858
ur1ng1w_kicked29	-490.747909	-490.864406	-489.948886	-490.107370	-490.203971	0.106981	0.062873
ur1ng1w_kicked53	-490.748014	-490.863917	-489.948870	-490.107494	-490.204145	0.106954	0.062585
ur1ng1w_kicked74	-490.741607	-490.863442	-489.942982	-490.101865	-490.198608	0.106923	0.062152
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
ur1ng1w_kicked25	-490.802165	-490.002580	-490.160968	-490.257522	0.03	0.08	0.07
ur1ng1w_kicked19	-490.801943	-490.002935	-490.161670	-490.258391	0.12	0.15	0.17
ur1ng1w_kicked79	-490.801792	-490.002307	-490.160671	-490.257224	0.04	0.09	0.05
ur1ng1w_kicked08	-490.801686	-490.002058	-490.160431	-490.256978	0.04	0.07	0.04
ur1ng1w_kicked97	-490.801650	-490.003483	-490.161891	-490.258450	0.11	0.22	0.18
ur1ng1w_kicked64	-490.801620	-490.002710	-490.161468	-490.258151	0.00	0.10	0.13
ur1ng1w_kicked59	-490.801586	-490.002440	-490.160829	-490.257382	0.39	0.07	0.06
ur1ng1w_kicked29	-490.801533	-490.002510	-490.160994	-490.257595	0.12	0.08	0.07
ur1ng1w_kicked53	-490.801332	-490.002188	-490.160812	-490.257463	0.15	0.06	0.06
ur1ng1w_kicked74	-490.801290	-490.002664	-490.161547	-490.258291	0.00	0.09	0.16
BoltzAvg:	-490.803851	-490.002775	-490.161351	-490.258003			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.



dd_uracil_2ng

Figure 87: Conformers/tautomers for twice deprotonated uracil (**ura**) with relevant thermodynamic data provided in Table 101.

Table 101: QM properties for solution phase optimized tautomers/conformers of twice deprotonated uracil (**ura**) shown in Figure 87 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	**corr. ΔH**[a] (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG**[a,b] (UB3LYP-D3/ 6-31+G(d,p))
dd_uracil_2ng	-413.570489	-413.917813	-412.854968	-412.996116	-413.080835	0.066640	0.030088
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
dd_uracil_2ng	-413.887725	-413.172204	-413.313352	-413.398071	1.00	1.00	1.00
BoltzAvg:	-413.887725	-413.172204	-413.313352	-413.398071			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 102: QM properties for solution phase optimized tautomers/conformers of twice deprotonated uracil (**ura**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
dd_uracil_2ng_kicked57	-490.046060	-490.378231	-489.227756	-489.394175	-489.494580	0.093093	0.049907
dd_uracil_2ng_kicked43	-490.047125	-490.378274	-489.228659	-489.395167	-489.495631	0.093173	0.050026
dd_uracil_2ng_kicked19	-490.048923	-490.378004	-489.229435	-489.396610	-489.497353	0.093222	0.049997
dd_uracil_2ng_kicked03	-490.040501	-490.377484	-489.222594	-489.389678	-489.490342	0.093200	0.049866
dd_uracil_2ng_kicked50	-490.042728	-490.377063	-489.224843	-489.391997	-489.492712	0.093451	0.049738
dd_uracil_2ng_kicked93	-490.049146	-490.376617	-489.231612	-489.397974	-489.498354	0.093288	0.049411
dd_uracil_2ng_kicked44	-490.045495	-490.377588	-489.228595	-489.395078	-489.495487	0.093670	0.050522
dd_uracil_2ng_kicked32	-490.049128	-490.376835	-489.232167	-489.398737	-489.499215	0.093318	0.049877
dd_uracil_2ng_kicked45	-490.040150	-490.376345	-489.223490	-489.390471	-489.491101	0.093271	0.049461
dd_uracil_2ng_kicked26	-490.042474	-490.376481	-489.225508	-489.392514	-489.493166	0.093331	0.049734
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
dd_uracil_2ng_kicked57	-490.328324	-489.510021	-489.676440	-489.776844	0.00	0.10	0.07
dd_uracil_2ng_kicked43	-490.328248	-489.509783	-489.676291	-489.776754	0.01	0.09	0.07
dd_uracil_2ng_kicked19	-490.328007	-489.508519	-489.675694	-489.776437	0.09	0.06	0.05
dd_uracil_2ng_kicked03	-490.327618	-489.509711	-489.676795	-489.777459	0.00	0.16	0.14
dd_uracil_2ng_kicked50	-490.327325	-489.509441	-489.676594	-489.777309	0.00	0.09	0.12
dd_uracil_2ng_kicked93	-490.327206	-489.509672	-489.676034	-489.776414	0.25	0.03	0.05
dd_uracil_2ng_kicked44	-490.327066	-489.510166	-489.676649	-489.777058	0.01	0.13	0.09
dd_uracil_2ng_kicked32	-490.326958	-489.509997	-489.676567	-489.777045	0.63	0.09	0.09
dd_uracil_2ng_kicked45	-490.326884	-489.510224	-489.677204	-489.777834	0.00	0.14	0.20
dd_uracil_2ng_kicked26	-490.326747	-489.509781	-489.676787	-489.777439	0.00	0.12	0.13
BoltzAvg:	-490.329622	-489.509890	-489.676662	-489.777258			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

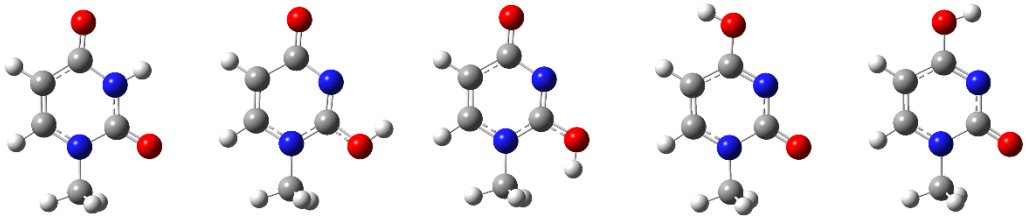


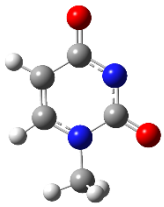
Figure 88: Conformers/tautomers for 1-methyluracil (**1mU**) with relevant thermodynamic data provided in Table 103.

Table 103: QM properties for solution phase optimized tautomers/conformers of 1-methyluracil (**1mU**) shown in Figure 88 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
meur1ne	-454.168160	-454.192094	-453.402361	-453.538929	-453.623220	0.123326	0.082390
meur3ne_2	-454.148889	-454.177323	-453.385046	-453.521711	-453.605974	0.122565	0.081680
meur3ne_1	-454.137566	-454.175927	-453.374743	-453.511580	-453.595843	0.122791	0.081920
meur2ne_1	-454.134419	-454.169327	-453.371225	-453.507966	-453.592229	0.122727	0.081580
meur2ne_2	-454.116246	-454.164648	-453.354872	-453.491801	-453.576072	0.122487	0.081304
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
meur1ne	-454.109704	-453.343905	-453.480473	-453.564764	1.00	1.00	1.00
meur3ne_2	-454.095643	-453.331800	-453.468465	-453.552728	0.00	0.00	0.00
meur3ne_1	-454.094007	-453.331184	-453.468021	-453.552284	0.00	0.00	0.00
meur2ne_1	-454.087747	-453.324554	-453.461294	-453.545557	0.00	0.00	0.00
meur2ne_2	-454.083344	-453.321969	-453.458899	-453.543170	0.00	0.00	0.00
BoltzAvg:	-454.109704	-453.343905	-453.480473	-453.564764			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 104: QM properties for solution phase optimized tautomers/conformers of 1-methyluracil (**1mU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
meur1ne_kicked53	-530.615250	-530.647969	-529.746240	-529.909743	-530.010337	0.150619	0.101969
meur1ne_kicked92	-530.610792	-530.648227	-529.741832	-529.905476	-530.006137	0.150621	0.103034
meur1ne_kicked14	-530.613928	-530.646421	-529.745897	-529.908747	-530.009071	0.150283	0.101348
meur1ne_kicked91	-530.609777	-530.648272	-529.741113	-529.904735	-530.005369	0.150985	0.103414
meur1ne_kicked16	-530.614417	-530.647399	-529.745337	-529.909044	-530.009727	0.150692	0.102542
meur1ne_kicked29	-530.615442	-530.647902	-529.745378	-529.909270	-530.010065	0.150841	0.103067
meur1ne_kicked97	-530.614501	-530.647878	-529.745439	-529.909026	-530.009680	0.150803	0.103071
meur1ne_kicked54	-530.613561	-530.647550	-529.745149	-529.908615	-530.009184	0.150784	0.102785
meur1ne_kicked47	-530.613867	-530.647450	-529.745334	-529.908881	-530.009491	0.150676	0.102696
meur1ne_kicked90	-530.609594	-530.647512	-529.741893	-529.905288	-530.005820	0.150607	0.102768
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
meur1ne_kicked53	-530.546000	-529.676989	-529.840492	-529.941086	0.27	0.09	0.19
meur1ne_kicked92	-530.545193	-529.676234	-529.839877	-529.940538	0.00	0.16	0.11
meur1ne_kicked14	-530.545073	-529.677041	-529.839892	-529.940216	0.07	0.03	0.08
meur1ne_kicked91	-530.544858	-529.676194	-529.839817	-529.940450	0.00	0.15	0.10
meur1ne_kicked16	-530.544857	-529.675777	-529.839484	-529.940167	0.14	0.06	0.07
meur1ne_kicked29	-530.544835	-529.674771	-529.838663	-529.939457	0.20	0.04	0.03
meur1ne_kicked97	-530.544807	-529.675745	-529.839332	-529.939986	0.13	0.08	0.06
meur1ne_kicked54	-530.544765	-529.676353	-529.839818	-529.940388	0.08	0.09	0.09
meur1ne_kicked47	-530.544754	-529.676221	-529.839768	-529.940377	0.11	0.09	0.09
meur1ne_kicked90	-530.544744	-529.677043	-529.840437	-529.940969	0.00	0.20	0.17
BoltzAvg:	-530.547076	-529.676611	-529.839987	-529.940558			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.							



meur1ng

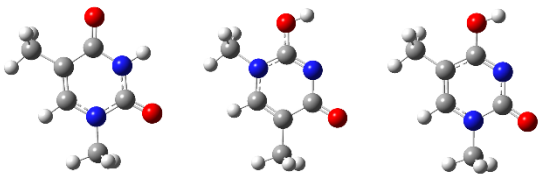
Figure 89: Conformers/tautomers for deprotonated 1-methyluracil (**1mU**) with relevant thermodynamic data provided in Table 105.

Table 105: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyluracil (**1mU**) shown in Figure 89 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
meur1ng	-453.599498	-453.718026	-452.824723	-452.967373	-453.054589	0.109538	0.068632
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
meur1ng	-453.649394	-452.874619	-453.017269	-453.104485	1.00	1.00	1.00
BoltzAvg:	-453.649394	-452.874619	-453.017269	-453.104485			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 106: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyluracil (**1mU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
meur1ng1w_kicked47	-530.062246	-530.175626	-529.183356	-529.352488	-529.455878	0.136505	0.088331
meur1ng1w_kicked64	-530.055839	-530.175185	-529.177281	-529.346743	-529.450251	0.136502	0.088492
meur1ng1w_kicked41	-530.059590	-530.176121	-529.180319	-529.349422	-529.452804	0.136765	0.089554
meur1ng1w_kicked05	-530.059776	-530.176142	-529.180436	-529.349552	-529.452938	0.136754	0.089673
meur1ng1w_kicked04	-530.060625	-530.175146	-529.181595	-529.351020	-529.454538	0.136652	0.088717
meur1ng1w_kicked58	-530.061690	-530.175599	-529.182849	-529.352015	-529.455423	0.136608	0.089185
meur1ng1w_kicked81	-530.055839	-530.175224	-529.177256	-529.346711	-529.450211	0.136577	0.088886
meur1ng1w_kicked82	-530.060530	-530.176159	-529.180989	-529.350175	-529.453603	0.136806	0.089993
meur1ng1w_kicked89	-530.060497	-530.175978	-529.181625	-529.351128	-529.454689	0.137073	0.090035
meur1ng1w_kicked48	-530.056764	-530.175650	-529.178130	-529.347260	-529.450654	0.136731	0.089754
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
meur1ng1w_kicked47	-530.087295	-529.208406	-529.377538	-529.480927	0.43	0.10	0.18
meur1ng1w_kicked64	-530.086693	-529.208136	-529.377597	-529.481105	0.00	0.14	0.22
meur1ng1w_kicked41	-530.086567	-529.207296	-529.376400	-529.479781	0.02	0.08	0.05
meur1ng1w_kicked05	-530.086469	-529.207129	-529.376246	-529.479631	0.02	0.08	0.05
meur1ng1w_kicked04	-530.086429	-529.207398	-529.376823	-529.480342	0.10	0.07	0.10
meur1ng1w_kicked58	-530.086414	-529.207572	-529.376738	-529.480147	0.27	0.09	0.08
meur1ng1w_kicked81	-530.086338	-529.207755	-529.377210	-529.480710	0.00	0.13	0.15
meur1ng1w_kicked82	-530.086166	-529.206625	-529.375811	-529.479239	0.04	0.07	0.03
meur1ng1w_kicked89	-530.085943	-529.207070	-529.376573	-529.480135	0.12	0.14	0.08
meur1ng1w_kicked48	-530.085896	-529.207262	-529.376392	-529.479787	0.00	0.10	0.06
BoltzAvg:	-530.088384	-529.207743	-529.377046	-529.480513			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							



5me_meur1ne 5me_meur2ne 5me_meur3ne

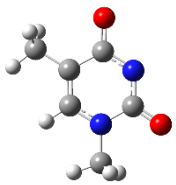
Figure 90: Conformers/tautomers for 1,5-dimethyluracil (**1m5mU**) with relevant thermodynamic data provided in Table 107.

Table 107: QM properties for solution phase optimized tautomers/conformers of 1,5-dimethyluracil (**1m5mU**) shown in Figure 90 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	**corr. ΔH**[a] (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG**[a,b] (UB3LYP-D3/ 6-31+G(d,p))
5me_meur1ne	-493.493841	-493.516434	-492.647418	-492.795479	-492.887034	0.152396	0.107717
5me_meur3ne	-493.473086	-493.501216	-492.628308	-492.776473	-492.867984	0.152014	0.107885
5me_meur2ne	-493.461394	-493.494131	-492.617539	-492.765714	-492.857210	0.151844	0.107296
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5me_meur1ne	-493.408717	-492.562295	-492.710355	-492.801910	1.00	1.00	1.00
5me_meur3ne	-493.393331	-492.548553	-492.696718	-492.788228	0.00	0.00	0.00
5me_meur2ne	-493.386835	-492.542980	-492.691154	-492.782650	0.00	0.00	0.00
BoltzAvg:	-493.408717	-492.562295	-492.710355	-492.801910			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 108: QM properties for solution phase optimized tautomers/conformers of 1,5-dimethyluracil (**1m5mU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	**corr. ΔH**[a] (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG**[a,b] (UB3LYP-D3/ 6-31+G(d,p))
5Me_meur1ne1w_kicked71	-569.939726	-569.972325	-568.990903	-569.165766	-569.273563	0.179906	0.127522
5Me_meur1ne1w_kicked47	-569.935735	-569.972934	-568.986155	-569.161353	-569.269292	0.180061	0.128386
5Me_meur1ne1w_kicked10	-569.941284	-569.972372	-568.990214	-569.165744	-569.273864	0.179888	0.127987
5Me_meur1ne1w_kicked99	-569.935401	-569.972566	-568.986209	-569.161355	-569.269261	0.180043	0.128276
5Me_meur1ne1w_kicked88	-569.937513	-569.973058	-568.988298	-569.163204	-569.271015	0.180037	0.128946
5Me_meur1ne1w_kicked81	-569.939576	-569.971484	-568.990238	-569.165225	-569.273094	0.179835	0.127572
5Me_meur1ne1w_kicked39	-569.940996	-569.971925	-568.990626	-569.165995	-569.274035	0.179989	0.128154
5Me_meur1ne1w_kicked60	-569.939715	-569.971249	-568.991054	-569.165449	-569.273056	0.179635	0.127508
5Me_meur1ne1w_kicked63	-569.935833	-569.972423	-568.987480	-569.162342	-569.270130	0.180126	0.129140
5Me_meur1ne1w_kicked86	-569.932831	-569.970552	-568.984325	-569.158840	-569.266464	0.179905	0.127282
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5Me_meur1ne1w_kicked71	-569.844803	-568.895981	-569.070843	-569.178640	0.43	0.10	0.18
5Me_meur1ne1w_kicked47	-569.844548	-568.894968	-569.070165	-569.178104	0.00	0.14	0.22
5Me_meur1ne1w_kicked10	-569.844385	-568.893315	-569.068845	-569.176966	0.02	0.08	0.05
5Me_meur1ne1w_kicked99	-569.844290	-568.895098	-569.070244	-569.178150	0.02	0.08	0.05
5Me_meur1ne1w_kicked88	-569.844112	-568.894897	-569.069803	-569.177614	0.10	0.07	0.10
5Me_meur1ne1w_kicked81	-569.843912	-568.894574	-569.069561	-569.177431	0.27	0.09	0.08
5Me_meur1ne1w_kicked39	-569.843771	-568.893401	-569.068770	-569.176810	0.00	0.13	0.15
5Me_meur1ne1w_kicked60	-569.843741	-568.895080	-569.069475	-569.177082	0.04	0.07	0.03
5Me_meur1ne1w_kicked63	-569.843283	-568.894930	-569.069792	-569.177580	0.12	0.14	0.08
5Me_meur1ne1w_kicked86	-569.843270	-568.894764	-569.069279	-569.176902	0.00	0.10	0.06
BoltzAvg:	-569.846450	-568.895183	-569.070075	-569.177912			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							



5me_meur1ng

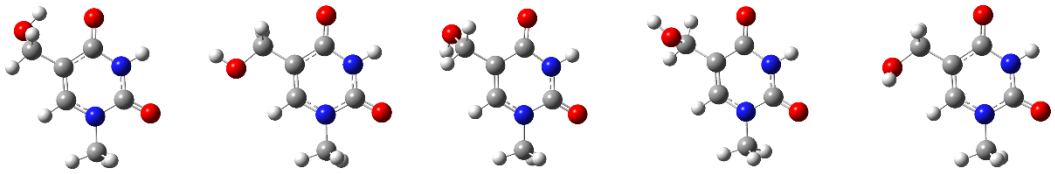
Figure 91: Conformers/tautomers for deprotonated 1,5-dimethyluracil (**1m5mU**) with relevant thermodynamic data provided in Table 109.

Table 109: QM properties for solution phase optimized tautomers/conformers of deprotonated 1,5-dimethyluracil (**1m5mU**) shown in Figure 91 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5Me_meur1ng	-492.924003	-493.041205	-492.069749	-492.223517	-492.317852	0.138895	0.094897
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5Me_meur1ng	-492.946308	-492.092055	-492.245822	-492.340157	1.00	1.00	1.00
BoltzAvg:	-492.946308	-492.092055	-492.245822	-492.340157			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 110: QM properties for solution phase optimized tautomers/conformers of deprotonated 1,5-dimethyluracil (**1m5mU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5Me_meur1ng1w_kicked82	-569.385269	-569.499887	-568.426109	-568.606426	-568.716968	0.165841	0.114881
5Me_meur1ng1w_kicked24	-569.384713	-569.499778	-568.425686	-568.605979	-568.716500	0.165911	0.115077
5Me_meur1ng1w_kicked88	-569.386731	-569.499051	-568.428019	-568.608357	-568.718896	0.165861	0.114770
5Me_meur1ng1w_kicked18	-569.380280	-569.498595	-568.421886	-568.602618	-568.713315	0.165841	0.114719
5Me_meur1ng1w_kicked100	-569.385858	-569.498563	-568.427218	-568.607743	-568.718376	0.166024	0.114818
5Me_meur1ng1w_kicked01	-569.384461	-569.498779	-568.425275	-568.606260	-568.717091	0.166224	0.115461
5Me_meur1ng1w_kicked46	-569.383131	-569.498117	-568.424907	-568.605615	-568.716337	0.166085	0.115184
5Me_meur1ng1w_kicked94	-569.381537	-569.499467	-568.424295	-568.604636	-568.715148	0.166459	0.116609
5Me_meur1ng1w_kicked72	-569.378381	-569.498419	-568.421516	-568.601959	-568.712530	0.166230	0.115639
5Me_meur1ng1w_kicked53	-569.378691	-569.498408	-568.421919	-568.602285	-568.712816	0.166348	0.116062
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5Me_meur1ng1w_kicked82	-569.385006	-568.425846	-568.606163	-568.716705	0.06	0.09	0.12
5Me_meur1ng1w_kicked24	-569.384701	-568.425674	-568.605967	-568.716488	0.04	0.08	0.10
5Me_meur1ng1w_kicked88	-569.384281	-568.425569	-568.605907	-568.716446	0.49	0.06	0.09
5Me_meur1ng1w_kicked18	-569.383875	-568.425481	-568.606213	-568.716910	0.00	0.09	0.15
5Me_meur1ng1w_kicked100	-569.383745	-568.425105	-568.605631	-568.716264	0.28	0.04	0.08
5Me_meur1ng1w_kicked01	-569.383318	-568.424132	-568.605117	-568.715949	0.07	0.05	0.05
5Me_meur1ng1w_kicked46	-569.382933	-568.424709	-568.605418	-568.716139	0.03	0.05	0.07
5Me_meur1ng1w_kicked94	-569.382858	-568.425615	-568.605956	-568.716468	0.01	0.22	0.09
5Me_meur1ng1w_kicked72	-569.382780	-568.425915	-568.606358	-568.716930	0.00	0.17	0.15
5Me_meur1ng1w_kicked53	-569.382346	-568.425573	-568.605939	-568.716470	0.00	0.14	0.09
BoltzAvg:	-569.386411	-568.425571	-568.605985	-568.716569			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							



5hm_meur1ne_2 **5hm_meur1ne_3** **5hm_meur1ne_4** **5hm_meur1ne_5** **5hm_meur1ne_6**

Figure 92: Conformers/tautomers for 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with relevant thermodynamic data provided in Table 111.

Table 111: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-hydroxymethyluracil (**1m5hmU**) shown in Figure 92 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5hm_meur1ne_2	-568.713703	-568.742857	-567.769852	-567.942671	-568.049302	0.159324	0.112891
5hm_meur1ne_4	-568.706543	-568.742101	-567.763698	-567.936463	-568.043066	0.159045	0.112464
5hm_meur1ne_5	-568.707037	-568.741083	-567.763954	-567.936854	-568.043488	0.158941	0.111839
5hm_meur1ne_6	-568.709540	-568.740924	-567.766748	-567.939272	-568.045813	0.159095	0.112255
5hm_meur1ne_3	-568.711544	-568.739919	-567.767873	-567.940757	-568.047487	0.158778	0.111641
5hm_meur1ne_tau3_13	-568.690817	-568.727107	-567.748661	-567.921582	-568.028150	0.158843	0.112509
5hm_meur1ne_tau3_5	-568.687471	-568.726707	-567.746090	-567.918989	-568.025542	0.158749	0.112356
5hm_meur1ne_tau4_14	-568.684519	-568.727126	-567.743746	-567.916173	-568.022488	0.158625	0.112817
5hm_meur1ne_tau3_3	-568.687299	-568.725702	-567.745800	-567.918818	-568.025402	0.158606	0.112014
5hm_meur1ne_tau4_12	-568.688082	-568.725377	-567.747019	-567.919633	-568.026106	0.158540	0.112010
5hm_meur1ne_tau4_4	-568.682920	-568.726023	-567.742253	-567.914820	-568.021170	0.158553	0.112779
5hm_meur1ne_tau3_10	-568.689736	-568.724335	-567.747890	-567.920852	-568.027507	0.158349	0.111599
5hm_meur1ne_tau4_5	-568.674532	-568.723940	-567.733960	-567.907064	-568.013633	0.159296	0.112880
5hm_meur1ne_tau4_11	-568.673845	-568.722083	-567.734221	-567.907043	-568.013531	0.158536	0.112114
5hm_meur1ne_tau2_12	-568.682844	-568.720779	-567.741517	-567.914441	-568.021014	0.158585	0.112091
5hm_meur1ne_tau2_9	-568.680092	-568.718178	-567.738921	-567.911938	-568.018616	0.158265	0.110895
5hm_meur1ne_tau2_7	-568.677960	-568.719118	-567.737603	-567.910280	-568.016784	0.158668	0.111933
5hm_meur1ne_tau2_11	-568.672668	-568.719342	-567.732510	-567.905374	-568.011916	0.158764	0.112293
5hm_meur1ne_tau2_3	-568.673326	-568.718320	-567.732842	-567.905881	-568.012474	0.158673	0.111792
5hm_meur1ne_tau1_10	-568.665209	-568.716101	-567.725638	-567.898772	-568.005373	0.158502	0.112087
5hm_meur1ne_tau1_9	-568.654471	-568.714620	-567.716122	-567.889154	-567.995690	0.158594	0.112061
5hm_meur1ne_tau1_12	-568.655310	-568.713599	-567.716598	-567.889829	-567.996433	0.158495	0.111546
5hm_meur1ne_tau1_8	-568.662633	-568.713669	-567.723277	-567.896500	-568.003203	0.158457	0.111696
5hm_meur1ne_tau1_7	-568.660246	-568.714586	-567.721750	-567.894596	-568.001095	0.158840	0.112679
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5hm_meur1ne_2	-568.629966	-567.686115	-567.858934	-567.965565	0.85	0.28	0.21
5hm_meur1ne_4	-568.629637	-567.686792	-567.859557	-567.966160	0.00	0.44	0.39
5hm_meur1ne_5	-568.629244	-567.686161	-567.859062	-567.965695	0.00	0.16	0.24
5hm_meur1ne_6	-568.628669	-567.685877	-567.858402	-567.964942	0.02	0.09	0.11
5hm_meur1ne_3	-568.628278	-567.684607	-567.857491	-567.964221	0.12	0.03	0.05
5hm_meur1ne_tau3_13	-568.614598	-567.672441	-567.845363	-567.951930	0.00	0.00	0.00
5hm_meur1ne_tau3_5	-568.614351	-567.672971	-567.845869	-567.952422	0.00	0.00	0.00
5hm_meur1ne_tau4_14	-568.614309	-567.673536	-567.845963	-567.952278	0.00	0.00	0.00
5hm_meur1ne_tau3_3	-568.613688	-567.672189	-567.845207	-567.951792	0.00	0.00	0.00
5hm_meur1ne_tau4_12	-568.613367	-567.672304	-567.844918	-567.951391	0.00	0.00	0.00
5hm_meur1ne_tau4_4	-568.613244	-567.672577	-567.845145	-567.951494	0.00	0.00	0.00
5hm_meur1ne_tau3_10	-568.612736	-567.670890	-567.843852	-567.950507	0.00	0.00	0.00
5hm_meur1ne_tau4_5	-568.611060	-567.670488	-567.843592	-567.950161	0.00	0.00	0.00
5hm_meur1ne_tau4_11	-568.609969	-567.670346	-567.843167	-567.949656	0.00	0.00	0.00

5hm_meur1ne_tau2_12	-568.608688	-567.667361	-567.840285	-567.946858	0.00	0.00	0.00
5hm_meur1ne_tau2_9	-568.607283	-567.666112	-567.839129	-567.945807	0.00	0.00	0.00
5hm_meur1ne_tau2_7	-568.607185	-567.666828	-567.839506	-567.946009	0.00	0.00	0.00
5hm_meur1ne_tau2_11	-568.607049	-567.666891	-567.839756	-567.946297	0.00	0.00	0.00
5hm_meur1ne_tau2_3	-568.606528	-567.666044	-567.839083	-567.945676	0.00	0.00	0.00
5hm_meur1ne_tau1_10	-568.604014	-567.664443	-567.837576	-567.944178	0.00	0.00	0.00
5hm_meur1ne_tau1_9	-568.602559	-567.664210	-567.837242	-567.943778	0.00	0.00	0.00
5hm_meur1ne_tau1_12	-568.602053	-567.663341	-567.836571	-567.943176	0.00	0.00	0.00
5hm_meur1ne_tau1_8	-568.601973	-567.662617	-567.835840	-567.942542	0.00	0.00	0.00
5hm_meur1ne_tau1_7	-568.601907	-567.663410	-567.836257	-567.942755	0.00	0.00	0.00
BoltzAvg:	-568.629511	-567.686302	-567.859085	-567.965695			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 112: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5OHMe_meur1ne_2_kicked92	-645.165497	-645.201608	-644.118540	-644.317513	-644.440153	0.186624	0.134367
5OHMe_meur1ne_2_kicked52	-645.165473	-645.201594	-644.118171	-644.317316	-644.440035	0.186609	0.134366
5OHMe_meur1ne_2_kicked80	-645.158218	-645.200242	-644.111441	-644.310406	-644.433020	0.186649	0.133809
5OHMe_meur1ne_2_kicked99	-645.158373	-645.200274	-644.111341	-644.310476	-644.433168	0.186741	0.134182
5OHMe_meur1ne_2_kicked79	-645.160135	-645.198601	-644.113239	-644.312929	-644.435835	0.186636	0.132696
5OHMe_meur1ne_2_kicked43	-645.162266	-645.199113	-644.113879	-644.313510	-644.436410	0.186553	0.133324
5OHMe_meur1ne_2_kicked01	-645.159146	-645.198097	-644.111089	-644.310994	-644.433981	0.186435	0.132392
5OHMe_meur1ne_2_kicked51	-645.158825	-645.198051	-644.110749	-644.310703	-644.433711	0.186411	0.132351
5OHMe_meur1ne_2_kicked63	-645.160326	-645.197420	-644.114087	-644.313312	-644.436044	0.186060	0.131737
5OHMe_meur1ne_2_kicked77	-645.154444	-645.198465	-644.107977	-644.307860	-644.430826	0.186706	0.132986
5OHMe_meur1ne_3_kicked96	-645.153803	-645.197319	-644.108930	-644.307435	-644.429886	0.186121	0.131798
5OHMe_meur1ne_3_kicked84	-645.158674	-645.195477	-644.110590	-644.310824	-644.434035	0.186038	0.130879
5OHMe_meur1ne_3_kicked23	-645.156016	-645.196015	-644.108563	-644.308298	-644.431284	0.185904	0.131497
5OHMe_meur1ne_3_kicked66	-645.159790	-645.195408	-644.111437	-644.311728	-644.434996	0.185907	0.131257
5OHMe_meur1ne_3_kicked97	-645.154848	-645.196864	-644.108382	-644.307528	-644.430254	0.186186	0.132775
5OHMe_meur1ne_3_kicked58	-645.158807	-645.194969	-644.111250	-644.311402	-644.434596	0.185815	0.130883
5OHMe_meur1ne_3_kicked75	-645.152987	-645.195651	-644.106630	-644.306529	-644.429567	0.186178	0.131635
5OHMe_meur1ne_3_kicked25	-645.154595	-645.196079	-644.108456	-644.308049	-644.430954	0.186202	0.132208
5OHMe_meur1ne_3_kicked57	-645.156508	-645.195227	-644.110534	-644.309603	-644.432306	0.185987	0.131395
5OHMe_meur1ne_3_kicked54	-645.157803	-645.194350	-644.111776	-644.311042	-644.433853	0.185796	0.130574
5OHMe_meur1ne_4_kicked83	-645.154850	-645.200216	-644.107643	-644.306811	-644.429510	0.186368	0.133811
5OHMe_meur1ne_4_kicked31	-645.148280	-645.198361	-644.103645	-644.302415	-644.424934	0.186337	0.132542
5OHMe_meur1ne_4_kicked77	-645.151648	-645.197628	-644.105023	-644.305020	-644.428067	0.186288	0.131936
5OHMe_meur1ne_4_kicked75	-645.148114	-645.198457	-644.103574	-644.302347	-644.424860	0.186479	0.132796
5OHMe_meur1ne_4_kicked29	-645.151849	-645.197645	-644.104987	-644.305056	-644.428138	0.186337	0.132147
5OHMe_meur1ne_4_kicked32	-645.147361	-645.198525	-644.103278	-644.301784	-644.424187	0.186507	0.133113
5OHMe_meur1ne_4_kicked78	-645.152895	-645.196693	-644.107574	-644.306746	-644.429434	0.186059	0.131316
5OHMe_meur1ne_4_kicked65	-645.148054	-645.198241	-644.102343	-644.302132	-644.425043	0.186577	0.132911
5OHMe_meur1ne_4_kicked86	-645.153249	-645.197115	-644.106637	-644.306681	-644.429753	0.186137	0.131800
5OHMe_meur1ne_4_kicked52	-645.148073	-645.198413	-644.103195	-644.302093	-644.424652	0.186505	0.133103
5OHMe_meur1ne_5_kicked65	-645.165497	-645.201608	-644.118540	-644.317513	-644.440153	0.186624	0.134367
5OHMe_meur1ne_5_kicked08	-645.165473	-645.201594	-644.118171	-644.317317	-644.440035	0.186609	0.134367
5OHMe_meur1ne_5_kicked02	-645.158218	-645.200242	-644.111441	-644.310407	-644.433020	0.186649	0.133809
5OHMe_meur1ne_5_kicked90	-645.152415	-645.198402	-644.105920	-644.305184	-644.427897	0.186225	0.132723
5OHMe_meur1ne_5_kicked24	-645.155865	-645.199179	-644.108283	-644.307591	-644.430359	0.186243	0.133643
5OHMe_meur1ne_5_kicked41	-645.156741	-645.198118	-644.109809	-644.309398	-644.432237	0.186422	0.132597
5OHMe_meur1ne_5_kicked46	-645.151739	-645.198045	-644.105189	-644.304486	-644.427227	0.186332	0.132777
5OHMe_meur1ne_5_kicked14	-645.154541	-645.196759	-644.106912	-644.307201	-644.430341	0.186172	0.131728

5OHMe_meur1ne_5_kicked11	-645.149170	-645.197166	-644.103028	-644.303001	-644.425980	0.186470	0.132243
5OHMe_meur1ne_5_kicked37	-645.156177	-645.197468	-644.108355	-644.308096	-644.431029	0.186226	0.132567
5OHMe_meur1ne_6_kicked97	-645.155070	-645.197485	-644.109969	-644.308515	-644.430990	0.186252	0.132209
5OHMe_meur1ne_6_kicked50	-645.155781	-645.197449	-644.110450	-644.309069	-644.431583	0.186219	0.132254
5OHMe_meur1ne_6_kicked84	-645.153789	-645.197294	-644.108436	-644.307116	-644.429644	0.186205	0.132511
5OHMe_meur1ne_6_kicked47	-645.156136	-645.197424	-644.110833	-644.309426	-644.431932	0.186294	0.132755
5OHMe_meur1ne_6_kicked10	-645.152344	-645.197257	-644.108069	-644.306349	-644.428698	0.186281	0.132679
5OHMe_meur1ne_6_kicked31	-645.151717	-645.197687	-644.106076	-644.305072	-644.427693	0.186474	0.133166
5OHMe_meur1ne_6_kicked35	-645.156734	-645.196428	-644.109534	-644.309390	-644.432433	0.186248	0.132089
5OHMe_meur1ne_6_kicked22	-645.156737	-645.195958	-644.110090	-644.309836	-644.432810	0.186091	0.131650
5OHMe_meur1ne_6_kicked43	-645.150896	-645.196633	-644.105455	-644.304970	-644.427807	0.186444	0.132511
5OHMe_meur1ne_6_kicked11	-645.151089	-645.196977	-644.105296	-644.304854	-644.427718	0.186677	0.132896
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5OHMe_meur1ne_2_kicked92	-645.067241	-644.020283	-644.219257	-644.341897	0.26	0.11	0.04
5OHMe_meur1ne_2_kicked52	-645.067228	-644.019926	-644.219071	-644.341790	0.23	0.10	0.03
5OHMe_meur1ne_2_kicked80	-645.066433	-644.019656	-644.218621	-644.341235	0.00	0.03	0.02
5OHMe_meur1ne_2_kicked99	-645.066092	-644.019060	-644.218195	-644.340887	0.00	0.03	0.01
5OHMe_meur1ne_2_kicked79	-645.065905	-644.019009	-644.218699	-644.341605	0.00	0.01	0.03
5OHMe_meur1ne_2_kicked43	-645.065789	-644.017402	-644.217033	-644.339933	0.00	0.00	0.00
5OHMe_meur1ne_2_kicked01	-645.065705	-644.017648	-644.217553	-644.340539	0.00	0.00	0.01
5OHMe_meur1ne_2_kicked51	-645.065700	-644.017624	-644.217577	-644.340586	0.00	0.00	0.01
5OHMe_meur1ne_2_kicked63	-645.065683	-644.019445	-644.218669	-644.341401	0.00	0.01	0.02
5OHMe_meur1ne_2_kicked77	-645.065479	-644.019012	-644.218895	-644.341861	0.00	0.02	0.04
5OHMe_meur1ne_3_kicked96	-645.065521	-644.020648	-644.219152	-644.341603	0.00	0.01	0.03
5OHMe_meur1ne_3_kicked84	-645.064598	-644.016514	-644.216748	-644.339959	0.00	0.00	0.00
5OHMe_meur1ne_3_kicked23	-645.064518	-644.017065	-644.216800	-644.339786	0.00	0.00	0.00
5OHMe_meur1ne_3_kicked66	-645.064151	-644.015798	-644.216089	-644.339357	0.00	0.00	0.00
5OHMe_meur1ne_3_kicked97	-645.064089	-644.017623	-644.216769	-644.339495	0.00	0.00	0.00
5OHMe_meur1ne_3_kicked58	-645.064086	-644.016529	-644.216681	-644.339874	0.00	0.00	0.00
5OHMe_meur1ne_3_kicked75	-645.064016	-644.017659	-644.217558	-644.340596	0.00	0.00	0.01
5OHMe_meur1ne_3_kicked25	-645.063871	-644.017732	-644.217325	-644.340230	0.00	0.00	0.01
5OHMe_meur1ne_3_kicked57	-645.063832	-644.017857	-644.216927	-644.339630	0.00	0.00	0.00
5OHMe_meur1ne_3_kicked54	-645.063776	-644.017749	-644.217015	-644.339826	0.00	0.00	0.00
5OHMe_meur1ne_4_kicked83	-645.066405	-644.019199	-644.218367	-644.341066	0.00	0.03	0.02
5OHMe_meur1ne_4_kicked31	-645.065819	-644.021184	-644.219954	-644.342474	0.00	0.04	0.07
5OHMe_meur1ne_4_kicked77	-645.065692	-644.019068	-644.219064	-644.342112	0.00	0.02	0.05
5OHMe_meur1ne_4_kicked75	-645.065661	-644.021121	-644.219894	-644.342407	0.00	0.04	0.06
5OHMe_meur1ne_4_kicked29	-645.065498	-644.018636	-644.218705	-644.341788	0.00	0.01	0.03
5OHMe_meur1ne_4_kicked32	-645.065412	-644.021328	-644.219835	-644.342237	0.00	0.05	0.05
5OHMe_meur1ne_4_kicked78	-645.065377	-644.020056	-644.219227	-644.341915	0.00	0.01	0.04
5OHMe_meur1ne_4_kicked65	-645.065330	-644.019618	-644.219408	-644.342319	0.00	0.04	0.06
5OHMe_meur1ne_4_kicked86	-645.065315	-644.018704	-644.218747	-644.341819	0.00	0.01	0.03
5OHMe_meur1ne_4_kicked52	-645.065310	-644.020432	-644.219330	-644.341888	0.00	0.03	0.04
5OHMe_meur1ne_5_kicked65	-645.067241	-644.020284	-644.219257	-644.341897	0.26	0.11	0.04
5OHMe_meur1ne_5_kicked08	-645.067227	-644.019925	-644.219071	-644.341789	0.23	0.10	0.03
5OHMe_meur1ne_5_kicked02	-645.066433	-644.019656	-644.218621	-644.341235	0.00	0.03	0.02
5OHMe_meur1ne_5_kicked90	-645.065679	-644.019184	-644.218448	-644.341161	0.00	0.01	0.02
5OHMe_meur1ne_5_kicked24	-645.065536	-644.017954	-644.217262	-644.340029	0.00	0.01	0.01
5OHMe_meur1ne_5_kicked41	-645.065521	-644.018590	-644.218178	-644.341017	0.00	0.01	0.01
5OHMe_meur1ne_5_kicked46	-645.065268	-644.018718	-644.218015	-644.340756	0.00	0.01	0.01
5OHMe_meur1ne_5_kicked14	-645.065031	-644.017402	-644.217691	-644.340830	0.00	0.00	0.01
5OHMe_meur1ne_5_kicked11	-645.064923	-644.018781	-644.218754	-644.341733	0.00	0.01	0.03
5OHMe_meur1ne_5_kicked37	-645.064901	-644.017079	-644.216820	-644.339753	0.00	0.00	0.00
5OHMe_meur1ne_6_kicked97	-645.065276	-644.020175	-644.218721	-644.341197	0.00	0.01	0.02
5OHMe_meur1ne_6_kicked50	-645.065195	-644.019864	-644.218483	-644.340996	0.00	0.01	0.01
5OHMe_meur1ne_6_kicked84	-645.064783	-644.019430	-644.218111	-644.340638	0.00	0.01	0.01
5OHMe_meur1ne_6_kicked47	-645.064669	-644.019365	-644.217959	-644.340465	0.00	0.01	0.01

5OHMe_meur1ne_6_kicked10	-645.064578	-644.020304	-644.218583	-644.340932	0.00	0.01	0.01
5OHMe_meur1ne_6_kicked31	-645.064521	-644.018880	-644.217876	-644.340497	0.00	0.01	0.01
5OHMe_meur1ne_6_kicked35	-645.064339	-644.017138	-644.216995	-644.340038	0.00	0.00	0.01
5OHMe_meur1ne_6_kicked22	-645.064308	-644.017660	-644.217407	-644.340381	0.00	0.00	0.01
5OHMe_meur1ne_6_kicked43	-645.064122	-644.018681	-644.218196	-644.341033	0.00	0.01	0.01
5OHMe_meur1ne_6_kicked11	-645.064081	-644.018288	-644.217846	-644.340710	0.00	0.01	0.01
BoltzAvg:	-645.067892	-644.020224	-644.218980	-644.341629			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

5hm_meur1ng_6 **5hm_meur1ng_7** **5hm_meur1ng_10** **5hm_meur1ng_11** **5hm_meur1ng_12**

Figure 93: Conformers/tautomers for deprotonated 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with relevant thermodynamic data provided in Table 113.

Table 113: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-hydroxymethyluracil (**1m5hmU**) shown in Figure 93 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5hm_meur1ng_12	-568.156108	-568.270072	-567.204038	-567.382461	-567.491855	0.145568	0.099823
5hm_meur1ng_11	-568.142348	-568.268319	-567.190627	-567.369256	-567.478721	0.145447	0.099238
5hm_meur1ng_10	-568.138993	-568.267285	-567.187217	-567.366042	-567.475559	0.145387	0.099096
5hm_meur1ng_7	-568.145088	-568.267023	-567.193679	-567.372174	-567.481632	0.145467	0.098984
5hm_meur1ng_6	-568.143318	-568.266034	-567.191523	-567.370367	-567.480012	0.145190	0.098392
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5hm_meur1ng_12	-568.170249	-567.218179	-567.396601	-567.505996	1.00	0.60	0.45
5hm_meur1ng_11	-568.169081	-567.217360	-567.395989	-567.505455	0.00	0.21	0.25
5hm_meur1ng_10	-568.168189	-567.216413	-567.395238	-567.504755	0.00	0.09	0.12
5hm_meur1ng_7	-568.168039	-567.216630	-567.395125	-567.504583	0.00	0.06	0.10
5hm_meur1ng_6	-568.167642	-567.215847	-567.394691	-567.504336	0.00	0.03	0.08
BoltzAvg:	-568.169642	-567.217578	-567.396029	-567.505439			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 114: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-hydroxymethyluracil (**1m5hmU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5OHMe_meur1ng_10_kicked03	-644.613653	-644.729712	-643.558115	-643.762575	-643.887962	0.172774	0.121437
5OHMe_meur1ng_10_kicked28	-644.615227	-644.729853	-643.559238	-643.763869	-643.889360	0.172851	0.121713
5OHMe_meur1ng_10_kicked89	-644.616352	-644.726997	-643.559962	-643.765162	-643.890839	0.172384	0.119230
5OHMe_meur1ng_10_kicked37	-644.610744	-644.727744	-643.555310	-643.759661	-643.884951	0.172851	0.120971
5OHMe_meur1ng_10_kicked02	-644.599642	-644.725645	-643.543154	-643.748515	-643.874232	0.172144	0.118931
5OHMe_meur1ng_10_kicked64	-644.601563	-644.724834	-643.545361	-643.750802	-643.876548	0.172299	0.118621
5OHMe_meur1ng_10_kicked72	-644.601373	-644.726952	-643.545485	-643.750127	-643.875562	0.172499	0.120875
5OHMe_meur1ng_10_kicked20	-644.595809	-644.724399	-643.539978	-643.745704	-643.871552	0.172289	0.118753
5OHMe_meur1ng_10_kicked84	-644.592713	-644.724714	-643.537066	-643.742082	-643.867605	0.172430	0.119274
5OHMe_meur1ng_10_kicked43	-644.597655	-644.723883	-643.542007	-643.747804	-643.873736	0.172375	0.118462
5OHMe_meur1ng_11_kicked01	-644.602202	-644.726689	-643.545917	-643.751070	-643.876722	0.172277	0.119280
5OHMe_meur1ng_11_kicked48	-644.604054	-644.725858	-643.548016	-643.753266	-643.878953	0.172336	0.118754
5OHMe_meur1ng_11_kicked11	-644.604707	-644.728093	-643.548843	-643.753373	-643.878760	0.172687	0.121148
5OHMe_meur1ng_11_kicked60	-644.599559	-644.725532	-643.543645	-643.749354	-643.875239	0.172510	0.119269
5OHMe_meur1ng_11_kicked23	-644.603211	-644.725396	-643.547272	-643.752619	-643.878350	0.172440	0.119338
5OHMe_meur1ng_11_kicked64	-644.600792	-644.725411	-643.544831	-643.750497	-643.876377	0.172693	0.119422

5OHMe_meur1ng_11_kicked34	-644.600604	-644.724945	-643.545106	-643.750668	-643.876512	0.172564	0.119057
5OHMe_meur1ng_11_kicked24	-644.595467	-644.725231	-643.541371	-643.746622	-643.872287	0.172698	0.119523
5OHMe_meur1ng_11_kicked67	-644.597859	-644.725822	-643.543075	-643.748335	-643.874006	0.172818	0.120227
5OHMe_meur1ng_11_kicked85	-644.600453	-644.724920	-643.544761	-643.750390	-643.876265	0.172587	0.119388
5OHMe_meur1ng_12_kicked89	-644.615418	-644.728103	-643.558543	-643.763556	-643.889159	0.172198	0.118937
5OHMe_meur1ng_12_kicked34	-644.615407	-644.728040	-643.558564	-643.763560	-643.889146	0.172208	0.119277
5OHMe_meur1ng_12_kicked22	-644.610726	-644.728187	-643.556350	-643.760849	-643.886221	0.172765	0.119832
5OHMe_meur1ng_12_kicked46	-644.616867	-644.727455	-643.560255	-643.765354	-643.890999	0.172454	0.119396
5OHMe_meur1ng_12_kicked78	-644.611827	-644.727343	-643.555546	-643.760915	-643.886697	0.172644	0.119542
5OHMe_meur1ng_12_kicked31	-644.616352	-644.726997	-643.559963	-643.765163	-643.890839	0.172384	0.119229
5OHMe_meur1ng_12_kicked77	-644.613651	-644.726524	-643.557840	-643.763189	-643.888954	0.172438	0.119231
5OHMe_meur1ng_12_kicked43	-644.614367	-644.726558	-643.557937	-643.763542	-643.889418	0.172658	0.119570
5OHMe_meur1ng_12_kicked50	-644.614130	-644.726575	-643.558171	-643.763538	-643.889306	0.172513	0.119592
5OHMe_meur1ng_12_kicked72	-644.609348	-644.726879	-643.554747	-643.759853	-643.885474	0.172658	0.119942
5OHMe_meur1ng_6_kicked40	-644.604066	-644.724210	-643.547499	-643.752827	-643.878636	0.171938	0.118078
5OHMe_meur1ng_6_kicked95	-644.603863	-644.724117	-643.547211	-643.752571	-643.878388	0.171870	0.118177
5OHMe_meur1ng_6_kicked82	-644.605827	-644.723492	-643.549594	-643.754991	-643.880823	0.172113	0.118256
5OHMe_meur1ng_6_kicked93	-644.604176	-644.723210	-643.547416	-643.753441	-643.879566	0.172234	0.118548
5OHMe_meur1ng_6_kicked19	-644.602151	-644.722649	-643.546508	-643.752239	-643.878246	0.172192	0.118178
5OHMe_meur1ng_6_kicked02	-644.599984	-644.723830	-643.545480	-643.750826	-643.876611	0.172554	0.119369
5OHMe_meur1ng_6_kicked72	-644.602078	-644.722662	-643.546255	-643.752064	-643.878107	0.172253	0.118668
5OHMe_meur1ng_6_kicked96	-644.598037	-644.722974	-643.543733	-643.749131	-643.874950	0.172311	0.119027
5OHMe_meur1ng_6_kicked87	-644.598595	-644.723136	-643.544440	-643.749834	-643.875647	0.172436	0.119401
5OHMe_meur1ng_6_kicked59	-644.597617	-644.723002	-643.543216	-643.748698	-643.874564	0.172397	0.119289
5OHMe_meur1ng_7_kicked56	-644.605510	-644.725209	-643.549390	-643.754391	-643.880026	0.172247	0.118831
5OHMe_meur1ng_7_kicked62	-644.605361	-644.725097	-643.549156	-643.754193	-643.879840	0.172228	0.118858
5OHMe_meur1ng_7_kicked82	-644.607395	-644.724477	-643.551617	-643.756661	-643.882309	0.172337	0.118934
5OHMe_meur1ng_7_kicked86	-644.606736	-644.723986	-643.551017	-643.756186	-643.881900	0.172302	0.118749
5OHMe_meur1ng_7_kicked49	-644.601197	-644.724055	-643.545755	-643.751122	-643.876881	0.172326	0.118928
5OHMe_meur1ng_7_kicked38	-644.604656	-644.724212	-643.548741	-643.754296	-643.880185	0.172493	0.119090
5OHMe_meur1ng_7_kicked61	-644.599078	-644.723849	-643.545313	-643.750309	-643.875925	0.172414	0.118876
5OHMe_meur1ng_7_kicked03	-644.600378	-644.724342	-643.546110	-643.751226	-643.876889	0.172645	0.119752
5OHMe_meur1ng_7_kicked41	-644.604082	-644.723929	-643.548337	-643.753893	-643.879791	0.172530	0.119348
5OHMe_meur1ng_7_kicked81	-644.601966	-644.724862	-643.547912	-643.752826	-643.878378	0.172807	0.120338
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5OHMe_meur1ng_10_kicked03	-644.608275	-643.552737	-643.757196	-643.882584	0.01	0.18	0.04
5OHMe_meur1ng_10_kicked28	-644.608140	-643.552152	-643.756783	-643.882273	0.05	0.16	0.03
5OHMe_meur1ng_10_kicked89	-644.607767	-643.551377	-643.756577	-643.882253	0.23	0.02	0.03
5OHMe_meur1ng_10_kicked37	-644.606773	-643.551339	-643.755690	-643.880980	0.00	0.02	0.01
5OHMe_meur1ng_10_kicked02	-644.606714	-643.550226	-643.755587	-643.881304	0.00	0.01	0.01
5OHMe_meur1ng_10_kicked64	-644.606213	-643.550011	-643.755452	-643.881198	0.00	0.00	0.01
5OHMe_meur1ng_10_kicked72	-644.606077	-643.550189	-643.754831	-643.880266	0.00	0.01	0.00
5OHMe_meur1ng_10_kicked20	-644.605646	-643.549815	-643.755541	-643.881389	0.00	0.00	0.01
5OHMe_meur1ng_10_kicked84	-644.605440	-643.549792	-643.754809	-643.880332	0.00	0.00	0.00
5OHMe_meur1ng_10_kicked43	-644.605421	-643.549773	-643.755570	-643.881502	0.00	0.00	0.01
5OHMe_meur1ng_11_kicked01	-644.607409	-643.551124	-643.756277	-643.881929	0.00	0.02	0.02
5OHMe_meur1ng_11_kicked48	-644.607104	-643.551067	-643.756316	-643.882003	0.00	0.01	0.02
5OHMe_meur1ng_11_kicked11	-644.606945	-643.551081	-643.755610	-643.880998	0.00	0.03	0.01
5OHMe_meur1ng_11_kicked60	-644.606263	-643.550349	-643.756058	-643.881943	0.00	0.01	0.02
5OHMe_meur1ng_11_kicked23	-644.606058	-643.550120	-643.755466	-643.881198	0.00	0.01	0.01
5OHMe_meur1ng_11_kicked64	-644.605989	-643.550028	-643.755693	-643.881573	0.00	0.01	0.01
5OHMe_meur1ng_11_kicked34	-644.605888	-643.550390	-643.755952	-643.881796	0.00	0.01	0.02
5OHMe_meur1ng_11_kicked24	-644.605708	-643.551613	-643.756864	-643.882529	0.00	0.02	0.04
5OHMe_meur1ng_11_kicked67	-644.605595	-643.550810	-643.756071	-643.881741	0.00	0.02	0.02
5OHMe_meur1ng_11_kicked85	-644.605532	-643.549841	-643.755469	-643.881345	0.00	0.01	0.01
5OHMe_meur1ng_12_kicked89	-644.609166	-643.552291	-643.757305	-643.882908	0.04	0.03	0.06

5OHMe_meur1ng_12_kicked34	-644.608763	-643.551920	-643.756915	-643.882502	0.04	0.03	0.04
5OHMe_meur1ng_12_kicked22	-644.608355	-643.553979	-643.758479	-643.883850	0.00	0.13	0.16
5OHMe_meur1ng_12_kicked46	-644.608059	-643.551447	-643.756546	-643.882190	0.27	0.02	0.03
5OHMe_meur1ng_12_kicked78	-644.607801	-643.551520	-643.756889	-643.882670	0.00	0.03	0.05
5OHMe_meur1ng_12_kicked31	-644.607768	-643.551378	-643.756578	-643.882255	0.23	0.02	0.03
5OHMe_meur1ng_12_kicked77	-644.607293	-643.551482	-643.756830	-643.882595	0.03	0.03	0.04
5OHMe_meur1ng_12_kicked43	-644.606988	-643.550557	-643.756162	-643.882039	0.05	0.02	0.02
5OHMe_meur1ng_12_kicked50	-644.606983	-643.551024	-643.756391	-643.882159	0.05	0.02	0.03
5OHMe_meur1ng_12_kicked72	-644.606937	-643.552336	-643.757442	-643.883062	0.00	0.07	0.07
5OHMe_meur1ng_6_kicked40	-644.606132	-643.549565	-643.754893	-643.880702	0.00	0.00	0.01
5OHMe_meur1ng_6_kicked95	-644.605940	-643.549287	-643.754647	-643.880464	0.00	0.00	0.00
5OHMe_meur1ng_6_kicked82	-644.605236	-643.549003	-643.754400	-643.880232	0.00	0.00	0.00
5OHMe_meur1ng_6_kicked93	-644.604662	-643.547902	-643.753928	-643.880052	0.00	0.00	0.00
5OHMe_meur1ng_6_kicked19	-644.604471	-643.548828	-643.754559	-643.880566	0.00	0.00	0.00
5OHMe_meur1ng_6_kicked02	-644.604461	-643.549958	-643.755303	-643.881088	0.00	0.01	0.01
5OHMe_meur1ng_6_kicked72	-644.603994	-643.548172	-643.753981	-643.880023	0.00	0.00	0.00
5OHMe_meur1ng_6_kicked96	-644.603947	-643.549643	-643.755041	-643.880860	0.00	0.00	0.01
5OHMe_meur1ng_6_kicked87	-644.603735	-643.549579	-643.754974	-643.880787	0.00	0.00	0.01
5OHMe_meur1ng_6_kicked59	-644.603713	-643.549313	-643.754794	-643.880660	0.00	0.00	0.01
5OHMe_meur1ng_7_kicked56	-644.606378	-643.550258	-643.755259	-643.880894	0.00	0.00	0.01
5OHMe_meur1ng_7_kicked62	-644.606239	-643.550033	-643.755071	-643.880718	0.00	0.00	0.01
5OHMe_meur1ng_7_kicked82	-644.605543	-643.549765	-643.754809	-643.880456	0.00	0.00	0.00
5OHMe_meur1ng_7_kicked86	-644.605237	-643.549518	-643.754687	-643.880401	0.00	0.00	0.00
5OHMe_meur1ng_7_kicked49	-644.605127	-643.549684	-643.755052	-643.880811	0.00	0.00	0.01
5OHMe_meur1ng_7_kicked38	-644.605122	-643.549208	-643.754762	-643.880651	0.00	0.00	0.01
5OHMe_meur1ng_7_kicked61	-644.604973	-643.551208	-643.756204	-643.881820	0.00	0.01	0.02
5OHMe_meur1ng_7_kicked03	-644.604590	-643.550321	-643.755438	-643.881100	0.00	0.01	0.01
5OHMe_meur1ng_7_kicked41	-644.604581	-643.548836	-643.754392	-643.880290	0.00	0.00	0.00
5OHMe_meur1ng_7_kicked81	-644.604524	-643.550470	-643.755384	-643.880936	0.00	0.01	0.01
BoltzAvg:	-644.610443	-643.552329	-643.756865	-643.882355			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.

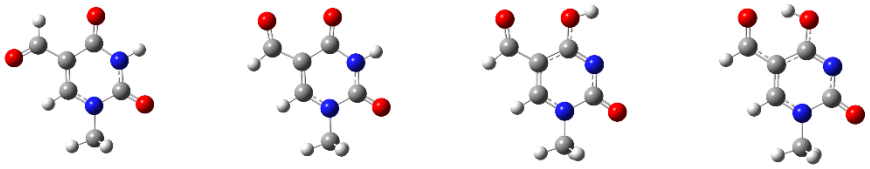


Figure 94: Conformers/tautomers for 1-methyl-5-formyluracil (**1m5fU**) with relevant thermodynamic data provided in Table 115.

Table 115: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-formyluracil (**1m5fU**) shown in Figure 94 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

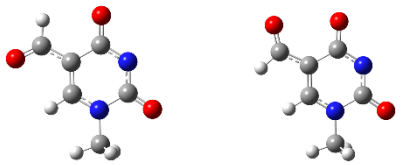
Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_meur1ne_1	-567.502135	-567.528826	-566.561471	-566.731081	-566.835784	0.133608	0.089224
5f_meur1ne_2	-567.492074	-567.528864	-566.553042	-566.722595	-566.827165	0.134916	0.089646
5f_meur3ne_3	-567.483414	-567.518514	-566.545274	-566.715185	-566.819811	0.134180	0.089554
5f_meur3ne_2	-567.476222	-567.513223	-566.538431	-566.708129	-566.812666	0.134253	0.089213
5f_meur3ne_1	-567.482791	-567.513147	-566.543677	-566.713462	-566.818145	0.133338	0.089214
5f_meur3ne_4	-567.469700	-567.509228	-566.531267	-566.701278	-566.805974	0.133448	0.089503
5f_meur2ne_2	-567.468222	-567.505070	-566.530608	-566.700394	-566.805081	0.133024	0.088446
5f_meur2ne_1	-567.456611	-567.505021	-566.520457	-566.690083	-566.794578	0.134263	0.088869
5f_meur2ne_3	-567.450250	-567.500034	-566.514702	-566.684678	-566.789372	0.134296	0.088877
5f_meur2ne_4	-567.438014	-567.499940	-566.503841	-566.673651	-566.778148	0.134443	0.089272
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_meur1ne_1	-567.439602	-566.498939	-566.668548	-566.773251	1.00	0.45	0.25
5f_meur1ne_2	-567.439218	-566.500185	-566.669738	-566.774309	0.00	0.55	0.75
5f_meur3ne_3	-567.428960	-566.490820	-566.660731	-566.765357	0.00	0.00	0.00
5f_meur3ne_2	-567.424010	-566.486219	-566.655917	-566.760454	0.00	0.00	0.00
5f_meur3ne_1	-567.423933	-566.484819	-566.654604	-566.759286	0.00	0.00	0.00
5f_meur3ne_4	-567.419725	-566.481293	-566.651303	-566.756000	0.00	0.00	0.00
5f_meur2ne_2	-567.416624	-566.479010	-566.648796	-566.753483	0.00	0.00	0.00
5f_meur2ne_1	-567.416152	-566.479999	-566.649625	-566.754119	0.00	0.00	0.00
5f_meur2ne_3	-567.411157	-566.475609	-566.645585	-566.750279	0.00	0.00	0.00
5f_meur2ne_4	-567.410668	-566.476495	-566.646305	-566.750802	0.00	0.00	0.00
BoltzAvg:	-567.439449	-566.499922	-566.669475	-566.774048			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 116: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-formyluracil (**1m5fU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_meur1ne1w_1_kicked71	-643.949190	-643.983888	-642.905980	-643.101970	-643.222757	0.161682	0.107966
5f_meur1ne1w_1_kicked34	-643.947039	-643.983673	-642.902704	-643.099669	-643.220848	0.161754	0.108057
5f_meur1ne1w_1_kicked08	-643.949574	-643.983962	-642.906274	-643.102259	-643.223048	0.161861	0.108626
5f_meur1ne1w_1_kicked80	-643.948996	-643.984108	-642.904065	-643.101064	-643.222273	0.162091	0.108812
5f_meur1ne1w_1_kicked91	-643.944733	-643.983899	-642.900674	-643.097428	-643.218508	0.161996	0.108768
5f_meur1ne1w_1_kicked14	-643.945575	-643.983827	-642.901419	-643.098234	-643.219347	0.161887	0.108726
5f_meur1ne1w_1_kicked31	-643.942297	-643.984076	-642.899443	-643.096010	-643.216994	0.162195	0.109146
5f_meur1ne1w_1_kicked96	-643.944029	-643.984461	-642.900984	-643.097480	-643.218447	0.162475	0.109674
5f_meur1ne1w_1_kicked27	-643.939965	-643.982595	-642.897283	-643.093366	-643.214138	0.161941	0.107838
5f_meur1ne1w_1_kicked23	-643.947894	-643.983472	-642.903706	-643.100568	-643.221722	0.162016	0.108871
5f_meur1ne1w_2_kicked80	-643.937310	-643.984046	-642.894210	-643.091155	-643.212204	0.162223	0.109336
5f_meur1ne1w_2_kicked68	-643.937751	-643.984030	-642.894111	-643.091186	-643.212295	0.162215	0.109336
5f_meur1ne1w_2_kicked20	-643.940079	-643.983697	-642.899170	-643.095109	-643.215767	0.162204	0.109015
5f_meur1ne1w_2_kicked07	-643.937805	-643.984328	-642.895993	-643.092480	-643.213347	0.162386	0.109763
5f_meur1ne1w_2_kicked34	-643.939388	-643.984483	-642.895739	-643.092515	-643.213585	0.162185	0.109980
5f_meur1ne1w_2_kicked27	-643.933195	-643.984520	-642.891737	-643.088302	-643.209186	0.162516	0.110034
5f_meur1ne1w_2_kicked52	-643.933877	-643.983981	-642.891828	-643.088568	-643.209505	0.162171	0.109529
5f_meur1ne1w_2_kicked84	-643.936041	-643.983500	-642.893100	-643.090139	-643.211291	0.162078	0.109160
5f_meur1ne1w_2_kicked37	-643.939180	-643.984275	-642.895269	-643.092104	-643.213208	0.162252	0.109962
5f_meur1ne1w_2_kicked66	-643.936857	-643.983377	-642.893443	-643.090516	-643.211691	0.162079	0.109117
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_meur1ne1w_1_kicked71	-643.875922	-642.832712	-643.028702	-643.149489	0.29	0.02	0.05
5f_meur1ne1w_1_kicked34	-643.875616	-642.831281	-643.028246	-643.149426	0.04	0.02	0.05
5f_meur1ne1w_1_kicked08	-643.875336	-642.832036	-643.028021	-643.148811	0.39	0.02	0.02
5f_meur1ne1w_1_kicked80	-643.875296	-642.830365	-643.027365	-643.148573	0.17	0.01	0.02
5f_meur1ne1w_1_kicked91	-643.875131	-642.831072	-643.027826	-643.148907	0.00	0.02	0.03
5f_meur1ne1w_1_kicked14	-643.875101	-642.830945	-643.027759	-643.148872	0.01	0.02	0.03
5f_meur1ne1w_1_kicked31	-643.874930	-642.832076	-643.028643	-643.149627	0.00	0.05	0.06
5f_meur1ne1w_1_kicked96	-643.874787	-642.831743	-643.028239	-643.149205	0.00	0.04	0.04
5f_meur1ne1w_1_kicked27	-643.874757	-642.832075	-643.028158	-643.148930	0.00	0.01	0.03
5f_meur1ne1w_1_kicked23	-643.874601	-642.830413	-643.027275	-643.148430	0.10	0.01	0.02
5f_meur1ne1w_2_kicked80	-643.874710	-642.831610	-643.028555	-643.149604	0.00	0.05	0.06
5f_meur1ne1w_2_kicked68	-643.874694	-642.831054	-643.028130	-643.149238	0.00	0.04	0.04
5f_meur1ne1w_2_kicked20	-643.874682	-642.833773	-643.029712	-643.150370	0.00	0.09	0.13
5f_meur1ne1w_2_kicked07	-643.874565	-642.832753	-643.029240	-643.150108	0.00	0.12	0.10
5f_meur1ne1w_2_kicked34	-643.874503	-642.830855	-643.027630	-643.148700	0.00	0.04	0.02
5f_meur1ne1w_2_kicked27	-643.874486	-642.833027	-643.029593	-643.150477	0.00	0.21	0.14
5f_meur1ne1w_2_kicked52	-643.874452	-642.832403	-643.029142	-643.150080	0.00	0.12	0.09
5f_meur1ne1w_2_kicked84	-643.874340	-642.831399	-643.028438	-643.149590	0.00	0.05	0.06
5f_meur1ne1w_2_kicked37	-643.874313	-642.830402	-643.027237	-643.148341	0.00	0.03	0.01
5f_meur1ne1w_2_kicked66	-643.874260	-642.830846	-643.027919	-643.149094	0.00	0.03	0.03
BoltzAvg:	-643.878152	-642.832644	-643.028855	-643.149706			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91$kJ mol^{-1} not jet applied.



5f_meur1ng_1 5f_meur1ng_2

Figure 95: Conformers/tautomers for deprotonated 1-methyl-5-formyluracil (**1m5fU**) with relevant thermodynamic data provided in Table 117.

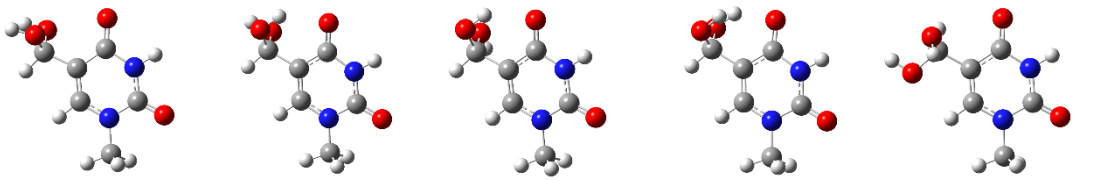
Table 117: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-formyluracil (**1m5fU**) shown in Figure 95 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_meur1ng_1	-566.952746	-567.060780	-566.001770	-566.177247	-566.284856	0.120033	0.076170
5f_meur1ng_2	-566.939071	-567.060583	-565.989754	-566.165210	-566.272656	0.121254	0.076521
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_meur1ng_1	-566.984610	-566.033633	-566.209110	-566.316719	1.00	0.49	0.27
5f_meur1ng_2	-566.984062	-566.034745	-566.210201	-566.317646	0.00	0.51	0.73
BoltzAvg:	-566.984413	-566.034483	-566.209940	-566.317394			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 118: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-formyluracil (**1m5fU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5f_meur1ng1w_1_kicked90	-643.411084	-643.518261	-642.355757	-642.557825	-642.681643	0.147975	0.095629
5f_meur1ng1w_1_kicked67	-643.412029	-643.518328	-642.356478	-642.558517	-642.682317	0.148210	0.096106
5f_meur1ng1w_1_kicked100	-643.412779	-643.517763	-642.357750	-642.559862	-642.683671	0.148097	0.095816
5f_meur1ng1w_1_kicked99	-643.408093	-643.517498	-642.353553	-642.556130	-642.680180	0.148431	0.096436
5f_meur1ng1w_1_kicked01	-643.407272	-643.518189	-642.353741	-642.555828	-642.679656	0.148648	0.097192
5f_meur1ng1w_1_kicked86	-643.409064	-643.516551	-642.354264	-642.556736	-642.680723	0.148319	0.095857
5f_meur1ng1w_1_kicked59	-643.407114	-643.516301	-642.353138	-642.555352	-642.679214	0.148312	0.095816
5f_meur1ng1w_1_kicked43	-643.409636	-643.516715	-642.354619	-642.557259	-642.681333	0.148502	0.096350
5f_meur1ng1w_1_kicked89	-643.404038	-643.516122	-642.348437	-642.551213	-642.675289	0.148360	0.095796
5f_meur1ng1w_1_kicked05	-643.404195	-643.516848	-642.351003	-642.553149	-642.676964	0.148425	0.096525
5f_meur1ng1w_2_kicked70	-643.397638	-643.518209	-642.344061	-642.546068	-642.669716	0.148150	0.095991
5f_meur1ng1w_2_kicked19	-643.398642	-643.518333	-642.344713	-642.546750	-642.670401	0.148275	0.096142
5f_meur1ng1w_2_kicked45	-643.398940	-643.517468	-642.345703	-642.547821	-642.671491	0.148341	0.096299
5f_meur1ng1w_2_kicked26	-643.399047	-643.518115	-642.345488	-642.547675	-642.671454	0.148381	0.097055
5f_meur1ng1w_2_kicked13	-643.399483	-643.517168	-642.345225	-642.547751	-642.671623	0.148553	0.097262
5f_meur1ng1w_2_kicked35	-643.390005	-643.516220	-642.337297	-642.539713	-642.663434	0.148480	0.096392
5f_meur1ng1w_2_kicked52	-643.396056	-643.516512	-642.342979	-642.545383	-642.669201	0.148473	0.096707
5f_meur1ng1w_2_kicked87	-643.395931	-643.516494	-642.342662	-642.545147	-642.668999	0.148488	0.096699
5f_meur1ng1w_2_kicked39	-643.392210	-643.516265	-642.338644	-642.541307	-642.665159	0.148557	0.096475
5f_meur1ng1w_2_kicked53	-643.391507	-643.517073	-642.340240	-642.542271	-642.665877	0.148683	0.097401
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5f_meur1ng1w_1_kicked90	-643.422632	-642.367305	-642.569373	-642.693191	0.08	0.03	0.05
5f_meur1ng1w_1_kicked67	-643.422222	-642.366671	-642.568710	-642.692510	0.16	0.02	0.02
5f_meur1ng1w_1_kicked100	-643.421947	-642.366918	-642.569030	-642.692839	0.65	0.02	0.03

5f_meur1ng1w_1_kicked99	-643.421062	-642.366523	-642.569099	-642.693149	0.02	0.04	0.05
5f_meur1ng1w_1_kicked01	-643.420997	-642.367466	-642.569553	-642.693381	0.01	0.09	0.06
5f_meur1ng1w_1_kicked86	-643.420694	-642.365895	-642.568367	-642.692354	0.03	0.01	0.02
5f_meur1ng1w_1_kicked59	-643.420485	-642.366509	-642.568723	-642.692585	0.01	0.01	0.02
5f_meur1ng1w_1_kicked43	-643.420365	-642.365347	-642.567988	-642.692061	0.05	0.01	0.01
5f_meur1ng1w_1_kicked89	-643.420326	-642.364725	-642.567500	-642.691576	0.00	0.00	0.01
5f_meur1ng1w_1_kicked05	-643.420323	-642.367131	-642.569278	-642.693092	0.00	0.04	0.04
5f_meur1ng1w_2_kicked70	-643.422218	-642.368641	-642.570648	-642.694296	0.00	0.11	0.15
5f_meur1ng1w_2_kicked19	-643.422191	-642.368262	-642.570299	-642.693950	0.00	0.08	0.11
5f_meur1ng1w_2_kicked45	-643.421169	-642.367932	-642.570049	-642.693719	0.00	0.07	0.08
5f_meur1ng1w_2_kicked26	-643.421060	-642.367501	-642.569688	-642.693467	0.00	0.11	0.06
5f_meur1ng1w_2_kicked13	-643.419906	-642.365648	-642.568174	-642.692046	0.00	0.03	0.01
5f_meur1ng1w_2_kicked35	-643.419828	-642.367120	-642.569537	-642.693258	0.00	0.04	0.05
5f_meur1ng1w_2_kicked52	-643.419805	-642.366728	-642.569133	-642.692950	0.00	0.04	0.04
5f_meur1ng1w_2_kicked87	-643.419795	-642.366527	-642.569011	-642.692863	0.00	0.04	0.03
5f_meur1ng1w_2_kicked39	-643.419790	-642.366224	-642.568887	-642.692738	0.00	0.02	0.03
5f_meur1ng1w_2_kicked53	-643.419672	-642.368405	-642.570436	-642.694042	0.00	0.21	0.12
BoltzAvg:	-643.424844	-642.367769	-642.569799	-642.693471			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M} = +7.91 kJ\ mol^{-1}$ not jet applied.							



5dhm_meur1ne_a_1 5dhm_meur1ne_a_5 5dhm_meur1ne_a_9 5dhm_meur1ne_a_10 5dhm_meur1ne_a1_2

Figure 96: Conformers/tautomers for 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) with relevant thermodynamic data provided in Table 119.

Table 119: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) shown in Figure 96 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5dhm_meur1ne_a_10	-643.946962	-643.981863	-642.907279	-643.104207	-643.225671	0.164656	0.115592
5dhm_meur1ne_a_5	-643.945950	-643.981416	-642.906846	-643.103780	-643.225245	0.164734	0.116092
5dhm_meur1ne_a1_2	-643.943042	-643.980046	-642.903985	-643.100871	-643.222367	0.164636	0.115598
5dhm_meur1ne_a_9_b	-643.939663	-643.979387	-642.901363	-643.098248	-643.219664	0.164718	0.115514
5dhm_meur1ne_a1_4	-643.939442	-643.979246	-642.900857	-643.097959	-643.219522	0.164729	0.115925
5dhm_meur1ne_a1_3	-643.946026	-643.978589	-642.906578	-643.103698	-643.225304	0.164369	0.115286
5dhm_meur1ne_a1_1	-643.940310	-643.979632	-642.901885	-643.098659	-643.220049	0.164902	0.116419
5dhm_meur1ne_a1_11	-643.934555	-643.978947	-642.897031	-643.093698	-643.215043	0.164888	0.116410
5dhm_meur1ne_a1_12	-643.938244	-643.978089	-642.899943	-643.096859	-643.218335	0.164717	0.115710
5dhm_meur1ne_a_1	-643.928276	-643.977895	-642.891413	-643.088475	-643.209931	0.164544	0.115564
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5dhm_meur1ne_a_10	-643.866271	-642.826588	-643.023516	-643.144981	0.43	0.26	0.31
5dhm_meur1ne_a_5	-643.865324	-642.826219	-643.023154	-643.144619	0.27	0.28	0.21
5dhm_meur1ne_a1_2	-643.864448	-642.825391	-643.022277	-643.143773	0.01	0.07	0.09
5dhm_meur1ne_a_9_b	-643.863873	-642.825573	-643.022458	-643.143874	0.00	0.07	0.10
5dhm_meur1ne_a1_4	-643.863321	-642.824736	-643.021838	-643.143401	0.00	0.06	0.06
5dhm_meur1ne_a1_3	-643.863303	-642.823855	-643.020975	-643.142581	0.29	0.02	0.02
5dhm_meur1ne_a1_1	-643.863213	-642.824788	-643.021562	-643.142952	0.00	0.06	0.04
5dhm_meur1ne_a1_11	-643.862537	-642.825013	-643.021680	-643.143025	0.00	0.06	0.04
5dhm_meur1ne_a1_12	-643.862379	-642.824078	-643.020994	-643.142471	0.00	0.02	0.02
5dhm_meur1ne_a_1	-643.862331	-642.825468	-643.022530	-643.143986	0.00	0.10	0.11
BoltzAvg:	-643.865495	-642.825832	-643.022764	-643.144225			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 120: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
1m5dhmUa_10_acx	-720.390350	-720.438382	-719.247609	-719.470985	-719.608532	0.192579	0.137796
1m5dhmUa_5_abm	-720.391857	-720.437021	-719.249766	-719.473501	-719.611184	0.192071	0.136547
1m5dhmUa_10_aba	-720.390404	-720.438306	-719.247861	-719.470955	-719.608385	0.192787	0.137938
1m5dhmUa1_2_acm	-720.386230	-720.436547	-719.243817	-719.467185	-719.604760	0.191957	0.136356
1m5dhmUa1_2_abm	-720.387157	-720.436834	-719.245337	-719.468448	-719.605964	0.191094	0.136653
1m5dhmUa_5_abc	-720.392173	-720.436981	-719.249898	-719.473381	-719.610963	0.192220	0.137051
1m5dhmUa_10_aaw	-720.387881	-720.437509	-719.245730	-719.469504	-719.607208	0.192538	0.137709
1m5dhmUa_5_abn	-720.387030	-720.436805	-719.246187	-719.469838	-719.607480	0.192394	0.137569
1m5dhmUa1_1_abn	-720.384056	-720.436073	-719.242508	-719.465686	-719.603164	0.192141	0.136959
1m5dhmUa1_2_adi	-720.384386	-720.435733	-719.242606	-719.466523	-719.604335	0.192288	0.136660
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
1m5dhmUa_10_acx	-720.300586	-719.157845	-719.381222	-719.518769	0.03	0.09	0.07
1m5dhmUa_5_abm	-720.300474	-719.158384	-719.382118	-719.519802	0.51	0.13	0.22
1m5dhmUa_10_aba	-720.300368	-719.157825	-719.380918	-719.518348	0.03	0.06	0.05
1m5dhmUa1_2_acm	-720.300191	-719.157777	-719.381146	-719.518720	0.00	0.04	0.07
1m5dhmUa1_2_abm	-720.300181	-719.158361	-719.381472	-719.518987	0.00	0.17	0.09
1m5dhmUa_5_abc	-720.299930	-719.157655	-719.381137	-719.518719	0.41	0.06	0.07
1m5dhmUa_10_aaw	-720.299800	-719.157649	-719.381423	-719.519127	0.01	0.13	0.11
1m5dhmUa_5_abn	-720.299236	-719.158392	-719.382044	-719.519686	0.01	0.24	0.19
1m5dhmUa1_1_abn	-720.299114	-719.157566	-719.380743	-719.518222	0.00	0.03	0.04
1m5dhmUa1_2_adi	-720.299073	-719.157292	-719.381209	-719.519021	0.00	0.05	0.09
BoltzAvg:	-720.302496	-719.158017	-719.381552	-719.519203			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

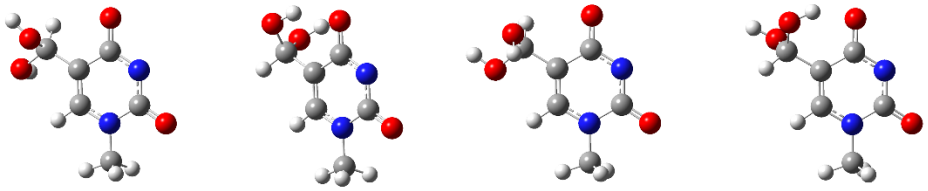


Figure 97: Conformers/tautomers for deprotonated 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) with relevant thermodynamic data provided in Table 121.

Table 121: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) shown in Figure 97 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

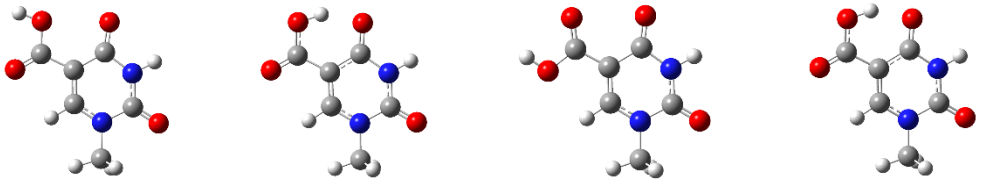
Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5dhm_meur1ng_a1_8	-643.397596	-643.511689	-642.349768	-642.552240	-642.676467	0.151333	0.103773
5dhm_meur1ng_f2_12	-643.391659	-643.511041	-642.344231	-642.546662	-642.670845	0.151010	0.103253
5dhm_meur1ng_f1_7	-643.379373	-643.507303	-642.331352	-642.534352	-642.658864	0.150793	0.102320
5dhm_meur1ng_7a	-643.373278	-643.505188	-642.325915	-642.529007	-642.653513	0.151043	0.102715
5dhm_meur1ng_d1_6	-643.368842	-643.496949	-642.311667	-642.516399	-642.641837	0.150274	0.102094
5dhm_meur1ng_d1_1	-643.356330	-643.494435	-642.298183	-642.503390	-642.628842	0.150381	0.101963
5dhm_meur1ng_e_2	-643.349718	-643.480985	-642.293894	-642.498949	-642.624424	0.149562	0.101268
5dhm_meur1ng_e2_10	-643.348058	-643.481393	-642.290678	-642.496044	-642.621487	0.150061	0.102002
5dhm_meur1ng_e2_7	-643.340114	-643.478649	-642.284447	-642.489776	-642.615192	0.149735	0.101741
5dhm_meur1ng_c2_2	-643.327112	-643.473625	-642.271925	-642.477017	-642.602513	0.149883	0.101741
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5dhm_meur1ng_a1_8	-643.407916	-642.360088	-642.562560	-642.686787	1.00	0.49	0.43
5dhm_meur1ng_f2_12	-643.407788	-642.360360	-642.562791	-642.686974	0.00	0.49	0.53
5dhm_meur1ng_f1_7	-643.404983	-642.356962	-642.559962	-642.684473	0.00	0.02	0.04
5dhm_meur1ng_7a	-643.402473	-642.355109	-642.558202	-642.682708	0.00	0.00	0.01
5dhm_meur1ng_d1_6	-643.394855	-642.337680	-642.542411	-642.667850	0.00	0.00	0.00
5dhm_meur1ng_d1_1	-643.392472	-642.334325	-642.539532	-642.664984	0.00	0.00	0.00
5dhm_meur1ng_e_2	-643.379717	-642.323893	-642.528948	-642.654423	0.00	0.00	0.00
5dhm_meur1ng_e2_10	-643.379391	-642.322010	-642.527377	-642.652820	0.00	0.00	0.00
5dhm_meur1ng_e2_7	-643.376908	-642.321241	-642.526569	-642.651985	0.00	0.00	0.00
5dhm_meur1ng_c2_2	-643.371884	-642.316697	-642.521789	-642.647285	0.00	0.00	0.00
BoltzAvg:	-643.407781	-642.360182	-642.562597	-642.686776			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 122: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-dihydroxymethyluracil (**1m5dhmU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
d_1m5dhmU_a1_8_abz	-719.856350	-719.969427	-718.703838	-718.932915	-719.073347	0.177897	0.122800
d_1m5dhmU_a1_8_acw	-719.856687	-719.969371	-718.704132	-718.933244	-719.073701	0.177745	0.122776
d_1m5dhmU_a1_8_aah	-719.856232	-719.969240	-718.703618	-718.932738	-719.073190	0.177771	0.122680
d_1m5dhmU_a1_8_aaf	-719.856726	-719.972181	-718.705953	-718.934173	-719.074296	0.178469	0.125728
d_1m5dhmU_a1_8_aca	-719.857756	-719.972296	-718.707116	-718.935277	-719.075373	0.178523	0.125896
d_1m5dhmU_a1_8_ado	-719.856323	-719.969258	-718.703778	-718.932907	-719.073373	0.177874	0.122870
d_1m5dhmU_a1_8_acb	-719.856806	-719.972209	-718.706072	-718.934275	-719.074391	0.178520	0.125893
d_1m5dhmU_a1_8_aao	-719.856056	-719.969462	-718.703541	-718.932644	-719.073084	0.177923	0.123163
d_1m5dhmU_f2_12_aax	-719.850414	-719.968855	-718.698369	-718.927397	-719.067784	0.177863	0.122887
d_1m5dhmU_a1_8_aad	-719.857552	-719.968893	-718.705314	-718.934434	-719.074885	0.178219	0.123639
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
d_1m5dhmU_a1_8_abz	-719.846627	-718.694114	-718.923191	-719.063623	0.04	0.02	0.10
d_1m5dhmU_a1_8_acw	-719.846595	-718.694040	-718.923152	-719.063609	0.06	0.02	0.10
d_1m5dhmU_a1_8_aah	-719.846560	-718.693946	-718.923066	-719.063518	0.03	0.01	0.09
d_1m5dhmU_a1_8_aaf	-719.846453	-718.695680	-718.923900	-719.064023	0.11	0.29	0.15
d_1m5dhmU_a1_8_aca	-719.846400	-718.695760	-718.923921	-719.064017	0.35	0.32	0.15
d_1m5dhmU_a1_8_ado	-719.846388	-718.693842	-718.922972	-719.063438	0.04	0.01	0.08
d_1m5dhmU_a1_8_acb	-719.846316	-718.695582	-718.923785	-719.063900	0.12	0.29	0.14
d_1m5dhmU_a1_8_aao	-719.846299	-718.693784	-718.922887	-719.063328	0.03	0.02	0.07
d_1m5dhmU_f2_12_aax	-719.845968	-718.693923	-718.922950	-719.063338	0.00	0.01	0.07
d_1m5dhmU_a1_8_aad	-719.845254	-718.693015	-718.922136	-719.062587	0.21	0.01	0.03
BoltzAvg:	-719.849755	-718.695228	-718.923460	-719.063679			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.



5ca_meur1ne_1 5ca_meur1ne_2 5ca_meur1ne_4 5ca_meur2ne_2

Figure 98: Conformers for 1-methyl-5-carboxyuracil (**1m5caU**) with relevant thermodynamic data provided in Table 123.

Table 123: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-carboxyuracil (**1m5caU**) shown in Figure 98 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5ca_meur2ne_2	-642.762212	-642.793314	-641.727428	-641.921487	-642.041171	0.140541	0.093395
5ca_meur1ne_2	-642.762214	-642.793313	-641.727423	-641.921487	-642.041175	0.139551	0.093944
5ca_meur1ne_4	-642.750498	-642.787330	-641.715525	-641.909567	-642.029257	0.140593	0.093227
5ca_meur1ne_1	-642.753966	-642.787143	-641.718599	-641.912583	-642.032257	0.140816	0.093640
5ca_meur1ne_3	-642.730917	-642.780358	-641.697964	-641.892148	-642.011770	0.138801	0.094279
5ca_meur2ne_1	-642.673343	-642.760158	-641.632688	-641.829364	-641.950095	0.140900	0.093941
5ca_meur3ne_2	-642.642275	-642.751750	-641.603339	-641.800062	-641.920783	0.140505	0.093205
5ca_meur3ne_1	-642.642138	-642.750429	-641.602196	-641.799050	-641.919807	0.140263	0.092671
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5ca_meur2ne_2	-642.699919	-641.665135	-641.859194	-641.978878	0.50	0.26	0.64
5ca_meur1ne_2	-642.699369	-641.664578	-641.858642	-641.978330	0.50	0.74	0.36
5ca_meur1ne_4	-642.694103	-641.659130	-641.853172	-641.972862	0.00	0.00	0.00
5ca_meur1ne_1	-642.693503	-641.658136	-641.852120	-641.971794	0.00	0.00	0.00
5ca_meur1ne_3	-642.686079	-641.653126	-641.847310	-641.966932	0.00	0.00	0.00
5ca_meur2ne_1	-642.666217	-641.625563	-641.822239	-641.942969	0.00	0.00	0.00
5ca_meur3ne_2	-642.658545	-641.619609	-641.816332	-641.937053	0.00	0.00	0.00
5ca_meur3ne_1	-642.657758	-641.617816	-641.814670	-641.935427	0.00	0.00	0.00
BoltzAvg:	-642.699710	-641.664927	-641.858988	-641.978673			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 124: QM properties for solution phase optimized tautomers/conformers of 1-methyl-5-carboxyuracil (**1m5caU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5ca_meur1ne_2_kicked37	-719.208166	-719.248342	-718.070325	-718.291289	-718.427263	0.167513	0.113050
5ca_meur1ne_2_kicked32	-719.201328	-719.248590	-718.063799	-718.284955	-718.420983	0.167785	0.113830
5ca_meur1ne_2_kicked68	-719.207110	-719.248045	-718.070065	-718.290919	-718.426834	0.167936	0.113459
5ca_meur1ne_2_kicked07	-719.202703	-719.248360	-718.065246	-718.286344	-718.422323	0.168076	0.114120
5ca_meur1ne_2_kicked90	-719.203254	-719.248627	-718.065940	-718.286877	-718.422820	0.168167	0.114552
5ca_meur1ne_2_kicked01	-719.202926	-719.248651	-718.066174	-718.287080	-718.423016	0.168347	0.115032
5ca_meur1ne_2_kicked76	-719.203449	-719.247099	-718.065153	-718.286386	-718.422459	0.167838	0.113551
5ca_meur1ne_2_kicked03	-719.201177	-719.247480	-718.064794	-718.285634	-718.421530	0.167936	0.113937
5ca_meur1ne_2_kicked14	-719.205612	-719.247721	-718.067215	-718.288516	-718.424621	0.167793	0.114184
5ca_meur1ne_2_kicked10	-719.201387	-719.247577	-718.064637	-718.285667	-718.421647	0.167986	0.114129
5ca_meur2ne_2_kicked09	-719.210267	-719.249072	-718.072957	-718.293299	-718.429030	0.167350	0.112710
5ca_meur2ne_2_kicked05	-719.206527	-719.248441	-718.069023	-718.289987	-718.425951	0.167473	0.112543
5ca_meur2ne_2_kicked10	-719.205977	-719.247933	-718.067160	-718.288588	-718.424757	0.167549	0.112804
5ca_meur2ne_2_kicked65	-719.213366	-719.248592	-718.075956	-718.296263	-718.431991	0.167629	0.113583
5ca_meur2ne_2_kicked91	-719.206029	-719.247951	-718.067133	-718.288580	-718.424757	0.167604	0.113044

5ca_meur2ne_2_kicked60	-719.207272	-719.247872	-718.068955	-718.290299	-718.426453	0.167741	0.113165
5ca_meur2ne_2_kicked01	-719.207755	-719.248356	-718.069032	-718.290071	-718.426085	0.167831	0.113685
5ca_meur2ne_2_kicked31	-719.207457	-719.247891	-718.069085	-718.290438	-718.426598	0.167786	0.113296
5ca_meur2ne_2_kicked08	-719.210855	-719.248302	-718.071801	-718.293227	-718.429427	0.168046	0.113779
5ca_meur2ne_2_kicked84	-719.202788	-719.248560	-718.065308	-718.286403	-718.422380	0.168043	0.114103
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5ca_meur1ne_2_kicked37	-719.135292	-717.997451	-718.218414	-718.354389	0.01	0.06	0.07
5ca_meur1ne_2_kicked32	-719.134760	-717.997231	-718.218387	-718.354415	0.00	0.10	0.07
5ca_meur1ne_2_kicked68	-719.134586	-717.997542	-718.218395	-718.354311	0.00	0.05	0.06
5ca_meur1ne_2_kicked07	-719.134240	-717.996782	-718.217881	-718.353860	0.00	0.06	0.04
5ca_meur1ne_2_kicked90	-719.134075	-717.996761	-718.217698	-718.353641	0.00	0.06	0.03
5ca_meur1ne_2_kicked01	-719.133619	-717.996866	-718.217772	-718.353709	0.00	0.09	0.03
5ca_meur1ne_2_kicked76	-719.133548	-717.995252	-718.216486	-718.352558	0.00	0.01	0.01
5ca_meur1ne_2_kicked03	-719.133543	-717.997161	-718.218000	-718.353897	0.00	0.05	0.04
5ca_meur1ne_2_kicked14	-719.133537	-717.995141	-718.216441	-718.352546	0.00	0.02	0.01
5ca_meur1ne_2_kicked10	-719.133448	-717.996698	-718.217728	-718.353708	0.00	0.05	0.03
5ca_meur2ne_2_kicked09	-719.136362	-717.999052	-718.219394	-718.355125	0.04	0.10	0.15
5ca_meur2ne_2_kicked05	-719.135898	-717.998394	-718.219358	-718.355322	0.00	0.09	0.19
5ca_meur2ne_2_kicked10	-719.135129	-717.996312	-718.217740	-718.353908	0.00	0.03	0.04
5ca_meur2ne_2_kicked65	-719.135009	-717.997599	-718.217906	-718.353634	0.88	0.04	0.03
5ca_meur2ne_2_kicked91	-719.134907	-717.996011	-718.217458	-718.353636	0.00	0.02	0.03
5ca_meur2ne_2_kicked60	-719.134707	-717.996389	-718.217734	-718.353888	0.00	0.03	0.04
5ca_meur2ne_2_kicked01	-719.134671	-717.995948	-718.216987	-718.353001	0.00	0.02	0.02
5ca_meur2ne_2_kicked31	-719.134595	-717.996223	-718.217576	-718.353735	0.00	0.03	0.03
5ca_meur2ne_2_kicked08	-719.134523	-717.995470	-718.216896	-718.353095	0.06	0.02	0.02
5ca_meur2ne_2_kicked84	-719.134457	-717.996976	-718.218071	-718.354048	0.00	0.07	0.05
BoltzAvg:	-719.137765	-717.997875	-718.218463	-718.354353			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

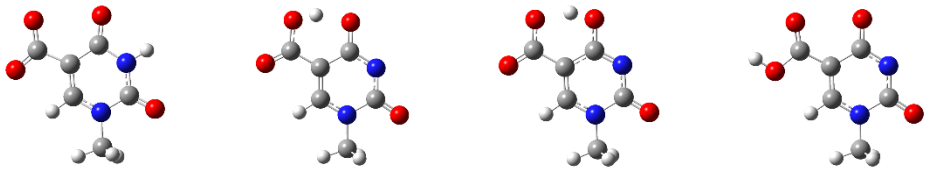


Figure 99: Conformers/tautomers for deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) with relevant thermodynamic data provided in Table 125.

Table 125: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) shown in Figure 99 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

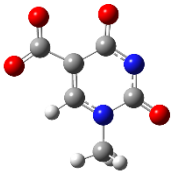
Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5ca_meur1ng_3	-642.228332	-642.330347	-641.184394	-641.383756	-641.506150	-0.079443	0.079443
5ca_meur1ng_1	-642.204100	-642.331566	-641.155575	-641.356451	-641.479603	-0.081074	0.081074
5ca_meur2ng_1	-642.193704	-642.318169	-641.148925	-641.348988	-641.471650	-0.080499	0.080499
5ca_meur1ng_4	-642.198373	-642.317701	-641.153198	-641.353198	-641.475867	-0.080468	0.080468
5ca_meur1ng_2	-642.184206	-642.313366	-641.141526	-641.341464	-641.463981	-0.080704	0.080704
5ca_meur1ng_5	-642.224859	-642.329844	-641.179860	-641.379689	-641.502302	-0.079579	0.079579
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5ca_meur1ng_3	-642.250904	-641.206967	-641.406328	-641.528722	0.98	0.31	0.72
5ca_meur1ng_1	-642.250492	-641.201967	-641.402843	-641.525995	0.00	0.55	0.04
5ca_meur2ng_1	-642.237670	-641.192890	-641.392954	-641.515616	0.00	0.00	0.00
5ca_meur1ng_4	-642.237233	-641.192058	-641.392058	-641.514727	0.00	0.00	0.00
5ca_meur1ng_2	-642.232662	-641.189982	-641.389920	-641.512437	0.00	0.00	0.00
5ca_meur1ng_5	-642.250265	-641.205266	-641.405095	-641.527708	0.02	0.14	0.24
BoltzAvg:	-642.251731	-641.206705	-641.406003	-641.528366			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 126: QM properties for solution phase optimized tautomers/conformers of deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
5ca_meur1ng_1_kicked54	-718.667555	-718.790508	-717.515164	-717.742441	-717.881769	0.154889	0.101585
5ca_meur1ng_1_kicked62	-718.663395	-718.790305	-717.512518	-717.739652	-717.878861	0.155001	0.101389
5ca_meur1ng_1_kicked99	-718.666528	-718.789606	-717.514007	-717.741773	-717.881282	0.154893	0.100693
5ca_meur1ng_1_kicked74	-718.668077	-718.789317	-717.515950	-717.743034	-717.882251	0.154561	0.100444
5ca_meur1ng_1_kicked49	-718.663316	-718.789128	-717.511209	-717.738934	-717.878421	0.154706	0.100718
5ca_meur1ng_1_kicked05	-718.667102	-718.789264	-717.515205	-717.742311	-717.881530	0.154697	0.100984
5ca_meur1ng_1_kicked06	-718.662689	-718.788844	-717.510715	-717.738260	-717.877653	0.154721	0.100774
5ca_meur1ng_1_kicked42	-718.666347	-718.788444	-717.513603	-717.741029	-717.880441	0.154790	0.100955
5ca_meur1ng_1_kicked68	-718.665122	-718.788542	-717.511866	-717.739426	-717.878895	0.154879	0.101056
5ca_meur1ng_1_kicked78	-718.657518	-718.787394	-717.505142	-717.733092	-717.872613	0.154870	0.100409
5ca_meur1ng_3_kicked82	-718.685055	-718.787504	-717.536665	-717.762733	-717.901362	0.153358	0.099993
5ca_meur1ng_3_kicked26	-718.685962	-718.786637	-717.537936	-717.764181	-717.902892	0.153354	0.099864
5ca_meur1ng_3_kicked92	-718.680229	-718.786218	-717.532561	-717.759080	-717.897882	0.153311	0.099752
5ca_meur1ng_3_kicked28	-718.684373	-718.786447	-717.536204	-717.762496	-717.901213	0.153465	0.100041
5ca_meur1ng_3_kicked94	-718.682556	-718.785643	-717.534089	-717.760682	-717.899523	0.153337	0.099674

5ca_meur1ng_3_kicked31	-718.684169	-718.786303	-717.536158	-717.762710	-717.901564	0.153493	0.100365
5ca_meur1ng_3_kicked68	-718.679276	-718.785430	-717.531365	-717.757808	-717.896558	0.153124	0.099554
5ca_meur1ng_3_kicked11	-718.679602	-718.786050	-717.532399	-717.758602	-717.897279	0.153617	0.100176
5ca_meur1ng_3_kicked79	-718.681632	-718.785939	-717.534705	-717.760895	-717.899584	0.153428	0.100113
5ca_meur1ng_3_kicked64	-718.679887	-718.786101	-717.532982	-717.759016	-717.897637	0.153596	0.100321
5ca_meur1ng_5_kicked88	-718.678476	-718.787149	-717.531486	-717.757713	-717.896429	0.153821	0.100788
5ca_meur1ng_5_kicked90	-718.683175	-718.786564	-717.534168	-717.760608	-717.899441	0.153513	0.100640
5ca_meur1ng_5_kicked06	-718.679463	-718.785563	-717.529864	-717.756547	-717.895452	0.153206	0.099758
5ca_meur1ng_5_kicked89	-718.683984	-718.786759	-717.535244	-717.761546	-717.900312	0.153659	0.100983
5ca_meur1ng_5_kicked17	-718.678641	-718.785448	-717.529142	-717.755871	-717.894793	0.153156	0.099914
5ca_meur1ng_5_kicked99	-718.677572	-718.785407	-717.529483	-717.756069	-717.894937	0.153563	0.100161
5ca_meur1ng_5_kicked29	-718.677355	-718.785064	-717.529967	-717.756301	-717.895064	0.153335	0.099937
5ca_meur1ng_5_kicked02	-718.679220	-718.785473	-717.529980	-717.757057	-717.896147	0.153558	0.100369
5ca_meur1ng_5_kicked07	-718.681242	-718.786451	-717.532475	-717.759308	-717.898314	0.153879	0.101414
5ca_meur1ng_5_kicked81	-718.676885	-718.784956	-717.529731	-717.755858	-717.894540	0.153394	0.100135
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
5ca_meur1ng_1_kicked54	-718.688923	-717.536532	-717.763809	-717.903137	0.00	0.03	0.03
5ca_meur1ng_1_kicked62	-718.688916	-717.538038	-717.765172	-717.904381	0.00	0.09	0.11
5ca_meur1ng_1_kicked99	-718.688913	-717.536392	-717.764158	-717.903667	0.00	0.02	0.05
5ca_meur1ng_1_kicked74	-718.688873	-717.536746	-717.763831	-717.903047	0.00	0.01	0.03
5ca_meur1ng_1_kicked49	-718.688410	-717.536303	-717.764028	-717.903515	0.00	0.02	0.04
5ca_meur1ng_1_kicked05	-718.688280	-717.536383	-717.763489	-717.902708	0.00	0.01	0.02
5ca_meur1ng_1_kicked06	-718.688070	-717.536096	-717.763641	-717.903034	0.00	0.02	0.03
5ca_meur1ng_1_kicked42	-718.687489	-717.534745	-717.762171	-717.901583	0.00	0.00	0.01
5ca_meur1ng_1_kicked68	-718.687486	-717.534230	-717.761790	-717.901258	0.00	0.00	0.00
5ca_meur1ng_1_kicked78	-718.686985	-717.534609	-717.762558	-717.902079	0.00	0.00	0.01
5ca_meur1ng_3_kicked82	-718.687511	-717.539121	-717.765188	-717.903817	0.11	0.06	0.06
5ca_meur1ng_3_kicked26	-718.686773	-717.538748	-717.764992	-717.903703	0.56	0.05	0.05
5ca_meur1ng_3_kicked92	-718.686466	-717.538798	-717.765317	-717.904118	0.00	0.07	0.08
5ca_meur1ng_3_kicked28	-718.686406	-717.538237	-717.764528	-717.903245	0.09	0.03	0.03
5ca_meur1ng_3_kicked94	-718.685969	-717.537501	-717.764095	-717.902936	0.02	0.02	0.02
5ca_meur1ng_3_kicked31	-718.685938	-717.537927	-717.764480	-717.903333	0.14	0.05	0.04
5ca_meur1ng_3_kicked68	-718.685876	-717.537965	-717.764408	-717.903158	0.00	0.03	0.03
5ca_meur1ng_3_kicked11	-718.685874	-717.538671	-717.764874	-717.903551	0.00	0.04	0.05
5ca_meur1ng_3_kicked79	-718.685826	-717.538899	-717.765090	-717.903778	0.02	0.06	0.06
5ca_meur1ng_3_kicked64	-718.685780	-717.538875	-717.764909	-717.903530	0.00	0.05	0.04
5ca_meur1ng_5_kicked88	-718.686361	-717.539371	-717.765598	-717.904314	0.00	0.15	0.10
5ca_meur1ng_5_kicked90	-718.685924	-717.536917	-717.763357	-717.902189	0.01	0.02	0.01
5ca_meur1ng_5_kicked06	-718.685805	-717.536206	-717.762889	-717.901794	0.00	0.01	0.01
5ca_meur1ng_5_kicked89	-718.685776	-717.537036	-717.763338	-717.902104	0.04	0.02	0.01
5ca_meur1ng_5_kicked17	-718.685534	-717.536035	-717.762764	-717.901686	0.00	0.01	0.01
5ca_meur1ng_5_kicked99	-718.685246	-717.537158	-717.763743	-717.902611	0.00	0.02	0.02
5ca_meur1ng_5_kicked29	-718.685127	-717.537739	-717.764073	-717.902836	0.00	0.02	0.02
5ca_meur1ng_5_kicked02	-718.685104	-717.535864	-717.762940	-717.902030	0.00	0.01	0.01
5ca_meur1ng_5_kicked07	-718.685037	-717.536270	-717.763103	-717.902109	0.00	0.03	0.01
5ca_meur1ng_5_kicked81	-718.684821	-717.537667	-717.763794	-717.902476	0.00	0.02	0.01
BoltzAvg:	-718.691107	-717.538508	-717.764721	-717.903546			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.



dd_5ca_meur

Figure 100: Conformers/tautomers for twice deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) with relevant thermodynamic data provided in Table 127.

Table 127: QM properties for solution phase optimized tautomers/conformers of twice deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) shown in Figure 100 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
dd_5ca_meur	-641.521623	-641.855743	-640.459062	-640.667919	-640.794959	0.114247	0.068202
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
dd_5ca_meur	-641.787541	-640.724980	-640.933837	-641.060877	1.00	1.00	1.00
BoltzAvg:	-641.787541	-640.724980	-640.933837	-641.060877			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

Table 128: QM properties for solution phase optimized tautomers/conformers of twice deprotonated 1-methyl-5-carboxyuracil (**1m5caU**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
dd_5caU_2ng_kicked82	-717.998463	-718.315625	-716.832314	-717.067152	-717.210223	0.141206	0.088553
dd_5caU_2ng_kicked69	-717.992696	-718.315217	-716.827830	-717.062612	-717.205576	0.141296	0.088259
dd_5caU_2ng_kicked04	-717.992156	-718.314427	-716.825044	-717.060185	-717.203299	0.140929	0.087755
dd_5caU_2ng_kicked86	-717.988564	-718.314310	-716.822583	-717.058147	-717.201462	0.141202	0.087731
dd_5caU_2ng_kicked39	-717.996709	-718.314034	-716.831263	-717.065774	-717.208626	0.140975	0.087613
dd_5caU_2ng_kicked73	-717.983360	-718.313777	-716.818279	-717.053748	-717.197018	0.141109	0.087432
dd_5caU_2ng_kicked65	-717.992398	-718.313560	-716.825914	-717.061188	-717.204325	0.140992	0.087268
dd_5caU_2ng_kicked67	-717.995484	-718.313976	-716.830166	-717.064765	-717.207647	0.141190	0.088209
dd_5caU_2ng_kicked81	-717.986823	-718.313241	-716.820434	-717.056001	-717.199243	0.141097	0.087518
dd_5caU_2ng_kicked58	-717.988372	-718.313410	-716.823611	-717.058508	-717.201503	0.141033	0.087815
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
dd_5caU_2ng_kicked82	-718.227072	-717.060922	-717.295761	-717.438832	0.79	0.15	0.08
dd_5caU_2ng_kicked69	-718.226958	-717.062092	-717.296874	-717.439837	0.01	0.30	0.23
dd_5caU_2ng_kicked04	-718.226672	-717.059560	-717.294702	-717.437815	0.00	0.03	0.03
dd_5caU_2ng_kicked86	-718.226579	-717.060598	-717.296162	-717.439476	0.00	0.13	0.16
dd_5caU_2ng_kicked39	-718.226421	-717.060975	-717.295486	-717.438338	0.15	0.04	0.05
dd_5caU_2ng_kicked73	-718.226345	-717.061265	-717.296733	-717.440003	0.00	0.18	0.28
dd_5caU_2ng_kicked65	-718.226292	-717.059808	-717.295082	-717.438219	0.00	0.03	0.04
dd_5caU_2ng_kicked67	-718.225767	-717.060449	-717.295047	-717.437930	0.05	0.04	0.03
dd_5caU_2ng_kicked81	-718.225723	-717.059335	-717.294901	-717.438143	0.00	0.03	0.04
dd_5caU_2ng_kicked58	-718.225595	-717.060834	-717.295731	-717.438726	0.00	0.08	0.07
BoltzAvg:	-718.229303	-717.061229	-717.296209	-717.439351			
[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.							

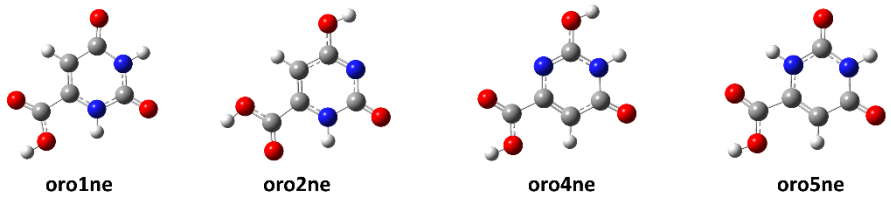


Figure 101: Conformers/tautomers for orotic acid (**oro**) with relevant thermodynamic data provided in Table 129.

Table 129: QM properties for solution phase optimized tautomers/conformers of orotic acid (**oro**) shown in Figure 101 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
oro5ne	-603.440099	-603.469412	-602.486325	-602.669218	-602.781872	0.110472	0.065889
oro1ne	-603.437560	-603.468331	-602.483571	-602.666410	-602.779059	0.110587	0.066116
oro2ne	-603.418799	-603.454178	-602.466773	-602.649724	-602.762329	0.110159	0.065951
oro6ne	-603.379451	-603.451758	-602.420657	-602.605715	-602.719322	0.110012	0.065468
oro3ne	-603.410114	-603.448372	-602.458387	-602.641616	-602.754312	0.110070	0.065543
oro4ne	-603.402656	-603.447459	-602.451857	-602.634786	-602.747310	0.109992	0.065408
oro8ne	-603.365512	-603.445702	-602.407793	-602.593067	-602.706690	0.110287	0.065849
oro7ne	-603.391217	-603.443230	-602.439038	-602.621977	-602.734498	0.109697	0.065120
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
oro5ne	-603.403523	-602.449749	-602.632643	-602.745296	0.95	0.83	0.84
oro1ne	-603.402215	-602.448226	-602.631065	-602.743713	0.05	0.17	0.16
oro2ne	-603.388227	-602.436201	-602.619152	-602.731757	0.00	0.00	0.00
oro6ne	-603.386290	-602.427497	-602.612555	-602.726161	0.00	0.00	0.00
oro3ne	-603.382829	-602.431102	-602.614332	-602.727027	0.00	0.00	0.00
oro4ne	-603.382051	-602.431252	-602.614181	-602.726705	0.00	0.00	0.00
oro8ne	-603.379853	-602.422134	-602.607408	-602.721031	0.00	0.00	0.00
oro7ne	-603.378110	-602.425931	-602.608870	-602.721391	0.00	0.00	0.00
BoltzAvg:	-603.403460	-602.449496	-602.632393	-602.745047			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm \rightarrow 1M}_{0K \rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 130: QM properties for solution phase optimized tautomers/conformers of orotic acid (**oro**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
oro1ne_kicked57	-679.890230	-679.928729	-678.832684	-679.041720	-679.170418	0.137421	0.086753
oro1ne_kicked39	-679.886384	-679.928259	-678.829265	-679.038297	-679.166968	0.137522	0.086747
oro1ne_kicked02	-679.887259	-679.928372	-678.829984	-679.039038	-679.167722	0.137496	0.086912
oro1ne_kicked50	-679.886709	-679.928316	-678.829516	-679.038561	-679.167239	0.137547	0.087016
oro1ne_kicked91	-679.884438	-679.923209	-678.827855	-679.037171	-679.165965	0.137376	0.085297
oro1ne_kicked81	-679.883941	-679.923908	-678.827030	-679.036216	-679.164971	0.137628	0.086405
oro1ne_kicked75	-679.885152	-679.923642	-678.827702	-679.037698	-679.166769	0.138218	0.086652
oro1ne_kicked04	-679.883443	-679.923539	-678.825205	-679.035387	-679.164528	0.137969	0.086611
oro1ne_kicked51	-679.881593	-679.923590	-678.825093	-679.034711	-679.163590	0.138077	0.086712
oro1ne_kicked45	-679.884803	-679.923619	-678.828445	-679.038006	-679.166883	0.138126	0.086755
oro5ne_kicked59	-679.892328	-679.929493	-678.835034	-679.044124	-679.172831	0.137364	0.086095
oro5ne_kicked03	-679.891194	-679.929365	-678.833926	-679.043088	-679.171821	0.137470	0.086688
oro5ne_kicked37	-679.889248	-679.929146	-678.832312	-679.041420	-679.170110	0.137474	0.086473
oro5ne_kicked45	-679.891729	-679.929386	-678.834619	-679.043613	-679.172274	0.137447	0.086736
oro5ne_kicked44	-679.889683	-679.929207	-678.832663	-679.041797	-679.170504	0.137531	0.086819
oro5ne_kicked74	-679.890298	-679.929281	-678.833133	-679.042284	-679.171000	0.137561	0.087031

oro5ne_kicked48	-679.888564	-679.924809	-678.831811	-679.041111	-679.169902	0.137512	0.085818
oro5ne_kicked06	-679.887161	-679.924340	-678.830756	-679.040042	-679.168783	0.137553	0.085760
oro5ne_kicked81	-679.888721	-679.924765	-678.832164	-679.041414	-679.170186	0.137621	0.086229
oro5ne_kicked25	-679.888383	-679.924948	-678.831560	-679.040864	-679.169658	0.137759	0.086553
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
oro1ne_kicked57	-679.841976	-678.784430	-678.993466	-679.122164	0.03	0.05	0.04
oro1ne_kicked39	-679.841512	-678.784392	-678.993425	-679.122096	0.00	0.04	0.04
oro1ne_kicked02	-679.841460	-678.784185	-678.993239	-679.121923	0.00	0.04	0.03
oro1ne_kicked50	-679.841300	-678.784107	-678.993152	-679.121829	0.00	0.04	0.03
oro1ne_kicked91	-679.837912	-678.781329	-678.990645	-679.119439	0.00	0.00	0.00
oro1ne_kicked81	-679.837503	-678.780592	-678.989778	-679.118533	0.00	0.00	0.00
oro1ne_kicked75	-679.836990	-678.779540	-678.989536	-679.118607	0.00	0.00	0.00
oro1ne_kicked04	-679.836928	-678.778690	-678.988871	-679.118013	0.00	0.00	0.00
oro1ne_kicked51	-679.836878	-678.780378	-678.989997	-679.118876	0.00	0.00	0.00
oro1ne_kicked45	-679.836864	-678.780505	-678.990066	-679.118943	0.00	0.00	0.00
oro5ne_kicked59	-679.843398	-678.786104	-678.995194	-679.123901	0.41	0.16	0.24
oro5ne_kicked03	-679.842677	-678.785409	-678.994571	-679.123304	0.14	0.14	0.13
oro5ne_kicked37	-679.842673	-678.785737	-678.994845	-679.123535	0.02	0.14	0.16
oro5ne_kicked45	-679.842650	-678.785540	-678.994534	-679.123194	0.23	0.13	0.11
oro5ne_kicked44	-679.842388	-678.785369	-678.994502	-679.123209	0.04	0.13	0.12
oro5ne_kicked74	-679.842250	-678.785085	-678.994236	-679.122953	0.06	0.12	0.09
oro5ne_kicked48	-679.838991	-678.782238	-678.991538	-679.120329	0.02	0.00	0.01
oro5ne_kicked06	-679.838580	-678.782175	-678.991462	-679.120202	0.01	0.00	0.00
oro5ne_kicked81	-679.838536	-678.781979	-678.991229	-679.120001	0.02	0.00	0.00
oro5ne_kicked25	-679.838395	-678.781572	-678.990876	-679.119670	0.01	0.00	0.00
BoltzAvg:	-679.843953	-678.785400	-678.994485	-679.123181			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91$kJ mol^{-1} not jet applied.

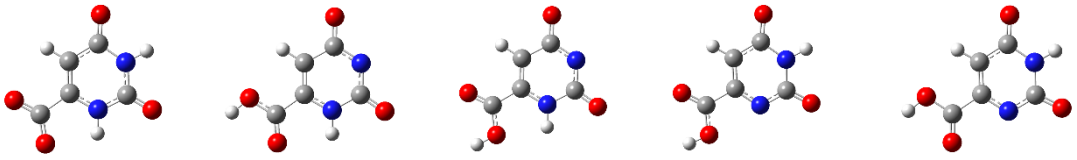


Figure 102: Conformers/tautomers for deprotonated orotic acid (**oro**) with relevant thermodynamic data provided in Table 131.

Table 131: QM properties for solution phase optimized tautomers/conformers of deprotonated orotic acid (**oro**) shown in Figure 102 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

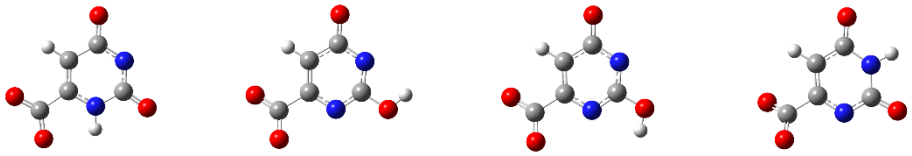
Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
oro1ng	-602.926245	-603.024665	-601.957704	-602.147277	-602.263448	0.097759	0.053953
oro4ng	-602.901622	-603.004437	-601.936496	-602.124876	-602.240374	0.097205	0.053569
oro5ng	-602.898965	-603.004244	-601.933943	-602.122450	-602.237980	0.097136	0.053389
oro2ng	-602.883882	-603.000853	-601.918753	-602.108418	-602.224391	0.097093	0.052927
oro3ng	-602.881894	-602.999795	-601.916527	-602.106106	-602.222068	0.097110	0.052961
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
oro1ng	-602.970712	-602.002171	-602.191744	-602.307915	1.00	1.00	1.00
oro4ng	-602.950868	-601.985741	-602.174122	-602.289619	0.00	0.00	0.00
oro5ng	-602.950855	-601.985833	-602.174340	-602.289871	0.00	0.00	0.00
oro2ng	-602.947926	-601.982797	-602.172462	-602.288434	0.00	0.00	0.00
oro3ng	-602.946834	-601.981467	-602.171047	-602.287008	0.00	0.00	0.00
BoltzAvg:	-602.970712	-602.002171	-602.1917440	-602.307915			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 132: QM properties for solution phase optimized tautomers/conformers of deprotonated orotic acid (**oro**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
oro1ng_kicked57	-679.385272	-679.481294	-678.313517	-678.529396	-678.661624	0.124701	0.074113
oro1ng_kicked94	-679.385477	-679.481333	-678.313697	-678.529577	-678.661810	0.124749	0.074317
oro1ng_kicked20	-679.385750	-679.481499	-678.313932	-678.529717	-678.661922	0.124832	0.074566
oro1ng_kicked65	-679.380581	-679.480799	-678.309048	-678.525538	-678.658040	0.124912	0.074228
oro1ng_kicked28	-679.380948	-679.481540	-678.310557	-678.526482	-678.658739	0.125122	0.075073
oro1ng_kicked78	-679.375024	-679.480021	-678.302462	-678.519239	-678.651846	0.124774	0.073562
oro1ng_kicked45	-679.377340	-679.480617	-678.305558	-678.521784	-678.654139	0.124998	0.074369
oro1ng_kicked49	-679.375158	-679.480032	-678.302799	-678.519476	-678.652037	0.124791	0.073800
oro1ng_kicked31	-679.368148	-679.478968	-678.297100	-678.513212	-678.645546	0.124492	0.072760
oro1ng_kicked19	-679.373149	-679.479826	-678.301115	-678.517867	-678.650442	0.124879	0.073701
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
oro1ng_kicked57	-679.407181	-678.335426	-678.551305	-678.683533	0.27	0.10	0.10
oro1ng_kicked94	-679.407016	-678.335236	-678.551116	-678.683349	0.33	0.09	0.08
oro1ng_kicked20	-679.406933	-678.335115	-678.550900	-678.683105	0.37	0.09	0.06
oro1ng_kicked65	-679.406571	-678.335038	-678.551528	-678.684029	0.01	0.15	0.16
oro1ng_kicked28	-679.406467	-678.336076	-678.552001	-678.684258	0.01	0.37	0.21
oro1ng_kicked78	-679.406459	-678.333896	-678.550673	-678.683280	0.00	0.04	0.07
oro1ng_kicked45	-679.406248	-678.334466	-678.550693	-678.683047	0.00	0.06	0.06
oro1ng_kicked49	-679.406232	-678.333873	-678.550550	-678.683111	0.00	0.04	0.06
oro1ng_kicked31	-679.406208	-678.335160	-678.551272	-678.683605	0.00	0.03	0.11
oro1ng_kicked19	-679.406125	-678.334090	-678.550843	-678.683417	0.00	0.05	0.09
BoltzAvg:	-679.409124	-678.335308	-678.551308	-678.683650			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.



dd_oro2ng_1 dd_oro2ng_2 dd_oro2ng_3 dd_oro2ng_4

Figure 103: Conformers/tautomers for twice deprotonated orotic acid (**oro**) with relevant thermodynamic data provided in Table 133.

Table 133: QM properties for solution phase optimized tautomers/conformers of twice deprotonated orotic acid (**oro**) shown in Figure 103 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
dd_oro2ng_1	-602.248690	-602.550834	-601.264823	-601.463071	-601.583430	0.084335	0.040844
dd_oro2ng_4	-602.238092	-602.548898	-601.252294	-601.450347	-601.570580	0.084312	0.040929
dd_oro2ng_3	-602.214839	-602.532358	-601.233395	-601.431342	-601.551428	0.083934	0.040662
dd_oro2ng_2	-602.213041	-602.532304	-601.231711	-601.429636	-601.549721	0.083942	0.040701
dd_oro2ng_5	-602.203582	-602.529480	-601.221965	-601.419590	-601.539535	0.083797	0.040597
dd_oro2ng_6	-602.196263	-602.528289	-601.215380	-601.413228	-601.533258	0.083909	0.040634
dd_oro2ng_9	-602.208655	-602.522379	-601.227506	-601.425087	-601.544901	0.083472	0.040220
dd_oro2ng_10	-602.177042	-602.516388	-601.193894	-601.392232	-601.512339	0.083665	0.040324
dd_oro2ng_8	-602.174538	-602.516363	-601.191419	-601.389939	-601.510095	0.083728	0.040301
dd_oro2ng_7	-602.181064	-602.514172	-601.199232	-601.397477	-601.517485	0.083582	0.039941
Molecule	$G_{298\_qh\_B3LYP}$	$G_{solv,\ 298,\ DPLNO(TZ)}$	$G_{solv,\ 298,\ DPLNO(QZ)}$	$G_{solv,\ 298,\ DPLNO(CBS)}$	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
dd_oro2ng_1	-602.509990	-601.526124	-601.724371	-601.844730	1.00	0.99	0.99
dd_oro2ng_4	-602.507969	-601.522171	-601.720224	-601.840457	0.00	0.01	0.01
dd_oro2ng_3	-602.491696	-601.510252	-601.708199	-601.828285	0.00	0.00	0.00
dd_oro2ng_2	-602.491603	-601.510273	-601.708197	-601.828283	0.00	0.00	0.00
dd_oro2ng_5	-602.488883	-601.507266	-601.704891	-601.824836	0.00	0.00	0.00
dd_oro2ng_6	-602.487655	-601.506772	-601.704620	-601.824650	0.00	0.00	0.00
dd_oro2ng_9	-602.482159	-601.501010	-601.698591	-601.818404	0.00	0.00	0.00
dd_oro2ng_10	-602.476064	-601.492916	-601.691253	-601.811361	0.00	0.00	0.00
dd_oro2ng_8	-602.476062	-601.492943	-601.691463	-601.811619	0.00	0.00	0.00
dd_oro2ng_7	-602.474231	-601.492399	-601.690644	-601.810652	0.00	0.00	0.00
BoltzAvg:	-602.509990	-601.526065	-601.72432010	-601.844685			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

Table 134: QM properties for solution phase optimized tautomers/conformers of twice deprotonated orotic acid (**oro**) with one explicit water molecule calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (ne = neutral, ps = cation, ng = anion).

Molecule	E_{tot}[a] (UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVTZ)	E_{tot}[a] (DLPNO-CCSD(T)/ cc-pVQZ)	E_{tot}[a] (DLPNO-CCSD(T)/ CBS)	corr. ΔH[a] (UB3LYP-D3/ 6-31+G(d,p))	corr. ΔG[a,b] (UB3LYP-D3/ 6-31+G(d,p))
oro2ng1w_1_kicked75	-678.716020	-679.008995	-677.629691	-677.853861	-677.990135	0.111292	0.060942
oro2ng1w_1_kicked39	-678.711273	-679.008455	-677.624120	-677.848937	-677.985495	0.111311	0.060764
oro2ng1w_1_kicked64	-678.717804	-679.008833	-677.630219	-677.854284	-677.990509	0.111186	0.061262
oro2ng1w_1_kicked15	-678.716140	-679.008952	-677.629792	-677.853940	-677.990205	0.111426	0.061396
oro2ng1w_1_kicked85	-678.716110	-679.008972	-677.629575	-677.853770	-677.990051	0.111455	0.061465
oro2ng1w_1_kicked09	-678.713191	-679.008427	-677.626560	-677.850562	-677.986751	0.111303	0.061543
oro2ng1w_1_kicked38	-678.717429	-679.008739	-677.629894	-677.854015	-677.990278	0.110143	0.061969
oro2ng1w_1_kicked21	-678.717711	-679.007806	-677.630555	-677.854963	-677.991328	0.111337	0.061037
oro2ng1w_1_kicked96	-678.717605	-679.007956	-677.631070	-677.855043	-677.991205	0.111290	0.061232
oro2ng1w_1_kicked30	-678.711621	-679.007923	-677.624901	-677.849411	-677.985820	0.111416	0.061316

oro2ng1w_4_kicked69	-678.710496	-679.008039	-677.620475	-677.844495	-677.980719	0.111228	0.061465
oro2ng1w_4_kicked11	-678.709350	-679.007981	-677.619538	-677.843573	-677.979830	0.111240	0.061638
oro2ng1w_4_kicked68	-678.702554	-679.007430	-677.613667	-677.837735	-677.973968	0.111134	0.061266
oro2ng1w_4_kicked48	-678.709381	-679.007673	-677.620755	-677.844584	-677.980644	0.111466	0.061718
oro2ng1w_4_kicked83	-678.702642	-679.007393	-677.614006	-677.837955	-677.974162	0.111239	0.061613
oro2ng1w_4_kicked85	-678.710594	-679.007265	-677.621821	-677.845699	-677.981843	0.111390	0.061650
oro2ng1w_4_kicked88	-678.709002	-679.007725	-677.620571	-677.844363	-677.980428	0.110400	0.062248
oro2ng1w_4_kicked75	-678.703198	-679.005921	-677.615652	-677.839494	-677.975568	0.111107	0.060455
oro2ng1w_4_kicked59	-678.710099	-679.006343	-677.622307	-677.845742	-677.981651	0.111265	0.060975
oro2ng1w_4_kicked51	-678.708673	-679.006990	-677.620015	-677.844040	-677.980217	0.111433	0.061687
Molecule	**$G_{298\_qh\_B3LYP}$**	**$G_{solv,\ 298,\ DPLNO(TZ)}$**	**$G_{solv,\ 298,\ DPLNO(QZ)}$**	**$G_{solv,\ 298,\ DPLNO(CBS)}$**	**Boltz_pop (E_{CBS})**	**Boltz_pop (H_{CBS})**	**Boltz_pop (G_{CBS})**
oro2ng1w_1_kicked75	-678.948053	-677.861723	-678.085894	-678.222168	0.08	0.20	0.27
oro2ng1w_1_kicked39	-678.947691	-677.860537	-678.085355	-678.221913	0.00	0.12	0.20
oro2ng1w_1_kicked64	-678.947571	-677.859986	-678.084051	-678.220276	0.12	0.04	0.04
oro2ng1w_1_kicked15	-678.947556	-677.861208	-678.085355	-678.221621	0.09	0.16	0.15
oro2ng1w_1_kicked85	-678.947507	-677.860972	-678.085167	-678.221448	0.07	0.14	0.12
oro2ng1w_1_kicked09	-678.946884	-677.860253	-678.084255	-678.220444	0.00	0.06	0.04
oro2ng1w_1_kicked38	-678.946770	-677.859236	-678.083357	-678.219619	0.09	0.13	0.02
oro2ng1w_1_kicked21	-678.946769	-677.859613	-678.084020	-678.220386	0.29	0.03	0.04
oro2ng1w_1_kicked96	-678.946724	-677.860190	-678.084163	-678.220324	0.25	0.04	0.04
oro2ng1w_1_kicked30	-678.946607	-677.859887	-678.084397	-678.220806	0.00	0.06	0.06
oro2ng1w_4_kicked69	-678.946574	-677.856553	-678.080574	-678.216797	0.00	0.00	0.00
oro2ng1w_4_kicked11	-678.946343	-677.856532	-678.080566	-678.216823	0.00	0.00	0.00
oro2ng1w_4_kicked68	-678.946164	-677.857277	-678.081345	-678.217578	0.00	0.00	0.00
oro2ng1w_4_kicked48	-678.945955	-677.857329	-678.081158	-678.217218	0.00	0.00	0.00
oro2ng1w_4_kicked83	-678.945780	-677.857143	-678.081092	-678.217300	0.00	0.00	0.00
oro2ng1w_4_kicked85	-678.945615	-677.856842	-678.080720	-678.216864	0.00	0.00	0.00
oro2ng1w_4_kicked88	-678.945477	-677.857046	-678.080838	-678.216903	0.00	0.01	0.00
oro2ng1w_4_kicked75	-678.945466	-677.857920	-678.081761	-678.217836	0.00	0.00	0.00
oro2ng1w_4_kicked59	-678.945368	-677.857576	-678.081011	-678.216920	0.00	0.00	0.00
oro2ng1w_4_kicked51	-678.945303	-677.856645	-678.080670	-678.216847	0.00	0.00	0.00
BoltzAvg:	-678.949815	-677.860843	-678.085106	-678.221460			

[a]: Using solution phase optimized SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) geometries; [b]: Standard state correction of $\Delta G^{1atm\rightarrow 1M}_{0K\rightarrow 298K} = +7.91 kJ\ mol^{-1}$ not jet applied.

4.6 REFERENCES

[1] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, 333, 1300-1303.

[2] 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.

[3] G. Schürmann, M. Cossi, V. Barone, J. Tomasi, *J. Phys. Chem. A* **1998**, 102, 6706-6712.

[4] J. Ho, M. L. Coote, *Theor Chem Acc* **2009**, 125, 3.

[5] J. Ho, *Aust. J. Chem.* **2014**, 67, 1441-1460.

[6] P. G. Seybold, G. C. Shields, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, 5, 290-297.

[7] G. C. Shields, P. G. Seybold, *Computational Approaches for the Prediction of PKa Values*, Taylor & Francis Group, **2017**.

[8] Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2008**, 4, 1849-1868.

[9] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, J. Gauss, *J. Chem. Phys.* **2004**, 120, 4129-4141.

[10] E. K. Pokon, M. D. Liptak, S. Feldgus, G. C. Shields, *J. Phys. Chem. A* **2001**, 105, 10483-10487.

[11] P. C. Redfern, P. Zapol, L. A. Curtiss, K. Raghavachari, *J. Phys. Chem. A* **2000**, 104, 5850-5854.

[12] J. A. M. Jr., M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **1999**, 110, 2822-2827.

[13] R. Casasnovas, J. Ortega-Castro, J. Frau, J. Donoso, F. Muñoz, *Int J Quantum Chem* **2014**, 114, 1350-1363.

[14] S. Mirzaei, M. V. Ivanov, Q. K. Timerghazin, *J. Phys. Chem. A* **2019**, *123*, 9498-9504.

[15] I. Tunon, D. Rinaldi, M. F. Ruiz-Lopez, J. L. Rivail, *J. Phys. Chem.* **1995**, 99, 3798-3805.

[16] B. Thapa, H. B. Schlegel, *J. Phys. Chem. A* **2015**, 119, 5134-5144.

[17] E. L. Jones, A. J. Mlotkowski, S. P. Hebert, H. B. Schlegel, C. S. Chow, *J. Phys. Chem. A* **2022**, 126, 1518-1529.

[18] V. Korotenko, unpublished results **2022**.

[19] N. S. W. Jonasson, R. Janßen, A. Menke, F. L. Zott, H. Zipse, L. J. Daumann, *ChemBioChem* **2021**, 22, 3333-3340.

[20] F. L. Zott, V. Korotenko, H. Zipse, *ChemBioChem* **2022**, 23, e202100651.

[21] N. L. Haworth, Q. Wang, M. L. Coote, *J. Phys. Chem. A* **2017**, 121, 5217-5225.

[22] D. Šakić, M. Hanževački, D. M. Smith, V. Vrček, *Org. Biomol. Chem.* **2015**, 13, 11740-11752.

[23] M. Saunders, *J. Comput. Chem.* **2004**, 25, 621-626.

[24] D. Šakić, H. Zipse, V. Vrček, *Org. Biomol. Chem.* **2011**, 9, 4336-4346.

[25] Y. Nakagawa, K. Uehara, N. Mizuno, *Inorg. Chem.* **2005**, 44, 9068-9075.

[26] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *J. Am. Chem. Soc.* **2013**, 135, 15813-15822.

[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] R. P. Johnson, A. M. Fleming, R. T. Perera, C. J. Burrows, H. S. White, *J. Am. Chem. Soc.* **2017**, 139, 2750-2756.

[29] S. Sappa, D. Dey, B. Sudhamalla, K. Islam, *J. Am. Chem. Soc.* **2021**, 143, 11891-11896.

[30] R. C. A. Dubini, A. Schön, M. Müller, T. Carell, P. Rovó, *Nuc. Ac. Res.* **2020**, 48, 8796-8807.

[31] 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.

[32] M. Sumino, A. Ohkubo, H. Taguchi, K. Seio, M. Sekine, *Bioorganic Med. Chem. Lett.* **2008**, 18, 274-277.

[33] Q. Dai, P. J. Sanstead, C. S. Peng, D. Han, C. He, A. Tokmakoff, *ACS Chem. Biol.* **2016**, 11, 470-477.

[34] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer, *Bioinformatics* **2006**, 22, 3067-3074.
[35] B. Ashwood, P. J. Sanstead, Q. Dai, C. He, A. Tokmakoff, *J. Phys. Chem. B* **2020**, *124*, 627-640.
[36] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, 8, 328-330.
[37] A. D. Becke, *J. Chem. Phys.* **1993**, 98, 5648-5652.
[38] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, 132, 154104.
[39] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, 54, 724.
[40] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, 72, 650.
[41] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, 113, 6378-6396.
[42] R. F. Ribeiro, A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2011**, 115, 14556-14562.
[43] J. P. Merrick, D. Moran, L. Radom, *J. Phys. Chem. A* **2007**, 111, 11683-11700.
[44] S. Grimme, *Chem. Eur. J.* **2012**, 18, 9955-9964.
[45] G. Luchini, J. Alegre-Requena, I. Funes-Ardoiz, R. Paton, *F1000Research* **2020**, 9.
[46] 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, Wallingford, CT, **2009**.
[47] A. Altun, F. Neese, G. Bistoni, *Beilstein J. Org. Chem.* **2018**, 14, 919-929.
[48] M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2017**, 146, 164105.
[49] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, 8, e1327.
[50] T. H. Dunning Jr., *J. Chem. Phys.* **1989**, 90, 1007-1023.
[51] F. Neese, E. F. Valeev, *J. Chem. Theory Comput.* **2011**, 7, 33-43.
[52] S. Ganguly, K. K. Kundu, *Can. J. Chem.* **1994**, 72, 1120-1126.
[53] P. A. Levene, L. W. Bass, H. S. Simms, *J. Biol. Chem.* **1926**, *70*, 229-241.
[54] D. Shugar, J. J. Fox, *Biochim Biophys Acta* **1952**, 9, 199-218.
[55] R. Stewart, M. G. Harris, *J. Org. Chem.* **1978**, 43, 3123-3126.
[56] C. J. La Francois, Y. H. Jang, T. Cagin, W. A. Goddard, L. C. Sowers, *Chem. Res. Tox.* **2000**, 13, 462-470.
[57] T. Ito, R. Kurihara, N. Utsumi, Y. Hamaguchi, K. Tanabe, S.-i. Nishimoto, *Chem. Com.* **2013**, 49, 10281-10283.
[58] J. J. Fox, D. Van Praag, I. Wempen, I. L. Doerr, L. Cheong, J. E. Knoll, M. L. Eidinoff, A. Bendich, G. B. Brown, *J. Am. Chem. Soc.* **1959**, 81, 178-187.
[59] K. N. Ganesh, V. A. Kumar, D. A. Barawkar, K. G. Rajeev, T. P. Prakash, P. S. Pallan, V. S. Rana, *Ind. J. Chem.* **1997**, 36, 519-524.
[60] M. R. Loeb, S. S. Cohen, *J. Biol. Chem.* **1959**, 234, 364-369.
[61] H. Lönnberg, *Tetrahedron* **1982**, *38*, 1517-1521.
[62] N. Koji, S. Noboru, Y. Futaba, *Bul. Chem. Soc. Jap.* **1961**, 34, 53-57.
[63] H. J. Nestler, E. R. Garrett, *J. Pharm. Sci.* **1968**, *57*, 1117-1125.
[64] E. R. Tucci, B. E. Doody, N. C. Li, *J. Phys. Chem.* **1961**, 65, 1570-1574.
[65] R. Ramautar, J. S. Toraño, G. W. Somsen, G. J. de Jong, *Electrophoresis* **2010**, 31, 2319-2327.
[66] S. P. Hebert, H. B. Schlegel, *Chem. Res. Toxi.* **2019**, 32, 2295-2304.

5 TET-LIKE OXIDATION IN 5-METHYLCYTOSINE AND DERIVATIVES

Niko S. W. Jonasson, Rachel Janßen, Annika Meneke, Fabian L. Zott, Hendrik Zipse and Lena J Daumann

ChemBioChem, **2021**, 22, 3333–3340. - Published by Wiley-VCH GmbH.

DOI and link to article: https://doi.org/10.1002/cbic.202100420

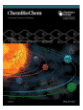
Cover feature: https://doi.org/10.1002/cbic.202100509

Author contributions: N.J. R.J. and A.M. contributed equally to this work. The project was conceived by N.J., R.J., A.M., F.Z., H.Z. and L.D. The experimental study was performed by N.J., R.J. and A.M. Synthesis was assisted by F.Z. The computational study was performed by F.Z. The manuscript was jointly written by N.J., R.J., A.M., F.Z., H.Z. and L.D. The experimental part of the SI was prepared by N.J., R.J. and A.M, the computational part was prepared by F.Z.

Additional Information: The supporting information of this article is presented in an altered version, to fit into the structure of this thesis. The original file can be accessed under the flowing link: https://chemistry-europe.onlinelibrary.wiley.com/202100420sup.pdf

Note: This section contains individual numbering of figures, tables and references.

Special Collection



TET-Like Oxidation in 5-Methylcytosine and Derivatives: A Computational and Experimental Study

Niko S. W. Jonasson[+],[a] Rachel Janßen[+],[a] Annika Menke[+],[a] Fabian L. Zott,[a] Hendrik Zipse,[a] and Lena J. Daumann*[a]

The epigenetic marker 5-methylcytosine (5mC) is an important factor in DNA modification and epigenetics. It can be modified through a three-step oxidation performed by ten-eleven-translocation (TET) enzymes and we have previously reported that the iron(IV)-oxo complex $[Fe(O)(Py_5Me_2H)]^{2+}$ (**1**) can oxidize 5mC. Here, we report the reactivity of this iron(IV)-oxo complex towards a wider scope of methylated cytosine and uracil derivatives relevant for synthetic DNA applications, such as 1-methylcytosine (1mC), 5-methyl-*iso*-cytosine (5m*i*C) and thymine (T/5mU). The observed kinetic parameters are corroborated by calculation of the C–H bond energies at the reactive sites which was found to be an efficient tool for reaction rate prediction of **1** towards methylated DNA bases. We identified oxidation products of methylated cytosine derivatives using HPLC-MS and GC-MS. Thereby, we shed light on the impact of the methyl group position and resulting C–H bond dissociation energies on reactivity towards TET-like oxidation.

Introduction

Using DNA as information storage for non-biological data has experienced a considerable development in recent years.[1] DNA not only provides an immensely high density of information, but its durability allows to store information over decades and centuries.[1c] Besides the canonical nucleobases C, G, T and A, epigenetics extend the 'DNA alphabet' with the epigenetic markers 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) as fifth and sixth letter.[2] In nature, these additional nucleobases are formed by direct methylation of cytosine which causes the gene to be silenced. Oxidation of the methyl group can alter or remove the epigenetic marker and therefore introduces a second layer of information.[3] This has also been put to use in DNA information storage systems.[4] Aside from nature, using unnatural orthogonal nucleobases pairs in synthetic DNA systems has expanded the 'DNA alphabet' up to eight letters within one system, called hachimoji DNA, increasing the density of information storable in DNA even further (Figure 1A).[5] Furthermore, the synthetic nucleoside *N1*-methylpseudouridine (1mΨ, Figure 1B), which consists of a 1-methyluracil (1mU) nucleobase fragment bound at its 5 position to ribose, was used in the Covid-19 mRNA vaccines by Pfizer/BioNTech (Comirnaty, BNT162b2) and Moderna (Spikevax, mRNA-1273).[6] The use of 1mΨ in mRNA has been reported to increase protein expression compared to *pseudo*-uridine Ψ and therefore likely contributes to the high efficacy of the mentioned vaccines.[7]

Using epigenetic markers or synthetic DNA bases has increased the potential for DNA information storage, however, the idea of epigenetic manipulation of synthetic DNA bases has not been employed yet. In natural epigenetics, ten-eleven translocation (TET) enzymes are involved in the oxidation of the methyl group in 5mC. TET enzymes belong to the superfamily of iron(II)/α-KG dependent non-heme enzymes and use an iron(IV)-oxo moiety as the catalytically active species for the stepwise transformation of 5mC to 5hmC, then to 5-formylcytosine (5fC) and finally to 5-carboxycytosine (5caC). We have recently shown that a synthetic iron(IV)-oxo complex (**1**, Figure 2) is capable of performing the same reaction on nucleobase,[9] nucleoside, and even nucleotide substrates.[10] It has been shown that hydroxyl radicals are capable of oxidizing 5mC, however, in addition to hydroxylation of the methyl group yielding 5hmC and/or 5fC, oxidation of the 5,6-double bond in 5mC was observed.[11] In epigenetic sequencing applications, 5hmC is oxidized to 5fC with potassium perruthenate ($KRuO_4$). Under these conditions, 5mC is unaffected. In this work, we explored the chemistry of the biomimetic system (**1**) towards synthetic DNA bases 1mC and *iso*-cytosine (as methylated 5m*i*C) occurring in hachimoji DNA and RNA, respectively, as well as methylated uracil derivatives. We present data on the diverse reactivity of different methylated nucleobases with the biomimetic compound **1** and present calculations of C–H bond dissociation energies (BDEs) as a viable method to predict the corresponding reactivity. The obtained results open up the possibility to install methyl groups with different reactivity towards oxidation and thus tunability for reaction, as well as to

[a] *N. S. W. Jonasson,[+] R. Janßen,[+] A. Menke,[+] F. L. Zott, H. Zipse, L. J. Daumann*
Department of Chemistry,
Ludwig-Maximilians-University Munich
Butenandtstr. 5–13
81377 München (Germany)
E-mail: lena.daumann@lmu.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.202100420

This article is part of a joint ChemBioChem-ChemMedChem Special Collection on Chemical Epigenetics available at bit.ly/chemepi2021.

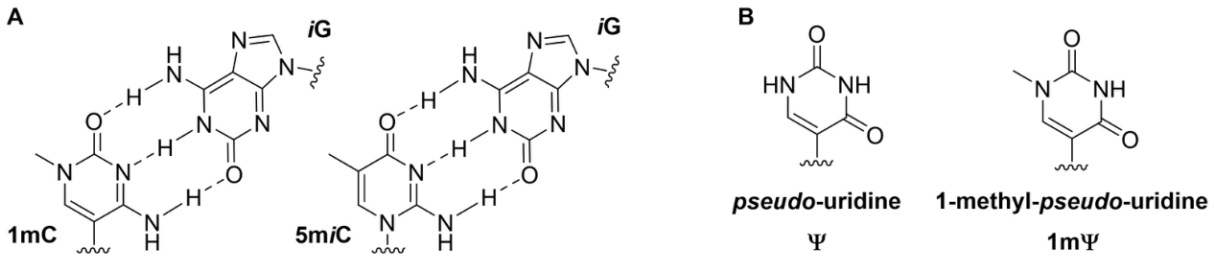


Figure 1. A) Base pairs 1-methylcytidine (1mC) – *iso*-guanosine (*i*G) and 5-methyl-*iso*-cytidine (5m*i*C) – *iso*-guanosine used in hachimoji DNA and other synthetic DNA applications.[5a,8] B) Nucleobase fragments in *pseudo*-uridine (Ψ) and 1-methyl-*pseudo*-uridine (1mΨ).

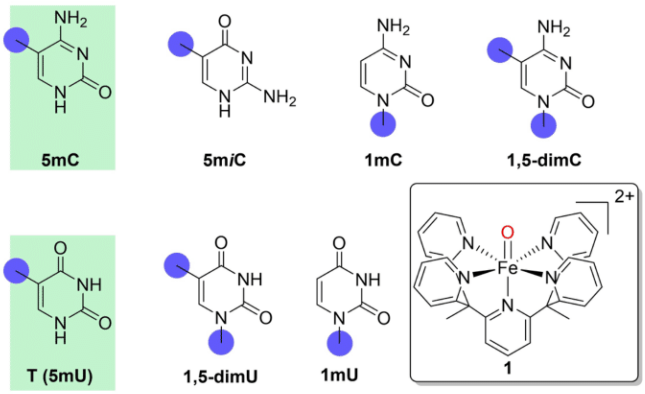


Figure 2. Methylated cytosine and uracil derivatives used in this work: 5-methylcytosine (5mC), 5-methyl-*iso*-cytosine (5miC), 1-methylcytosine (1mC), 1,5-dimethylcytosine (1,5dimC), thymine (T), 1-methyluracil (1mU), 1,5-dimethyluracil (1,5dimU). The studied methyl groups are marked with blue circles, the naturally occurring nucleobases have a green background. Used iron(IV)-oxo complex $[Fe(O)(Py_5Me_2H)]^{2+}$ (**1**) shown in rounded rectangle.

draw conclusions on the suitability of certain methylated DNA species in nature.

Results and Discussion

As substrates we used the naturally occurring 5-methylcytosine (5mC) and thymine/5-methyluracil (T/5mU) in addition to the synthetic nucleobases 5-methyl-*iso*-cytosine (5m*i*C), 1-methylcytosine (1mC), 1,5-dimethylcytosine (1,5dimC), 1-methyluracil (1mU) and 1,5-dimethyluracil (1,5dimU, Figure 2). The cytosine derivatives were also analyzed for the products formed when reacted with **1** using HPLC-MS and GC-MS. We evaluate which nucleobases could be useful for synthetic biology and epigenetics with respect to their ability to be further modified by TET enzymes and their biomimetic complexes.

UV Vis kinetics

We have previously reported that the absorbance of **1** at $\lambda = 718$ nm can be used to measure the initial reaction rates (data points used for rate calculation: minute 1–2) and then to determine the rate constants k_S of the individual substrates with **1**.[9] To confirm this is valid for the substrates used in this work, we monitored the observed relative absorbance of each substrate: in all reactions the initial absorbance was not significantly decreased after 2 min reaction time (>85–90%), indicating that only small amounts of **1** had been consumed. The only exception with 80% of the initial absorbance of **1** was with the rapidly reacting 5m*i*C (Figures S1 and S2). As the absorption decrease was still reliably linear within the monitored timeframe (Figure 3B) we deemed this data analysis to be a suitable approximation. Additionally, we analyzed the reaction mixture of 5m*i*C and **1** after 2 min reaction time using GC-MS and found mostly unreacted 5m*i*C (see Figure S15). For all other substrates the linear consumption of **1** was observed for a much longer timeframe. We were therefore confident to monitor almost exclusively the reaction of **1** with the respective starting material and not that of any further oxidized products.

For all substrates we found a decrease of the absorbance at $\lambda = 718$ nm over the monitored time frame (Figure 3A, compare also Supporting Information Figures S1 and S2 for full UV-Vis spectra and control reactions). Using the method of initial rates, reaction rates were calculated from the observed decrease in absorbance by linear regression. We then calculated the corresponding rate constants k_S (Figure 4) using the second order rate law found by Jonasson and Daumann for the reaction of 5mC with **1**. This rate law was applied for all substrates used in this work (Eq. 1):

$$v = k_s[S][1] \quad (1)$$

where v is the observed reaction rate, $[S]$ the concentration of the substrates and $[1]$ the concentration of **1**. When the amount of **1** was varied from 1–9 equivalents, a linear dependence of the reaction rates on the amount of **1** (Figure 5) was observed for all substrates and none showed any saturation behavior (as had been previously observed for 5hmC).[9] This confirms that the chosen concentration ranges are suitable for our purposes. The linear increase in reaction rate upon increase of the added amount of **1** indicates a rate law of first order for **1** for all substrates. This agrees with our previous findings concerning the reaction of **1** with 5mC and justifies use of Eq. (1) for the calculation of k_S.[9]

The *N*-methylated substrates 1mC and 1mU reacted significantly slower than all other compounds (Figures 4 and 5). The uracil derivatives 1mU, T, and 1,5dimU reacted faster than their

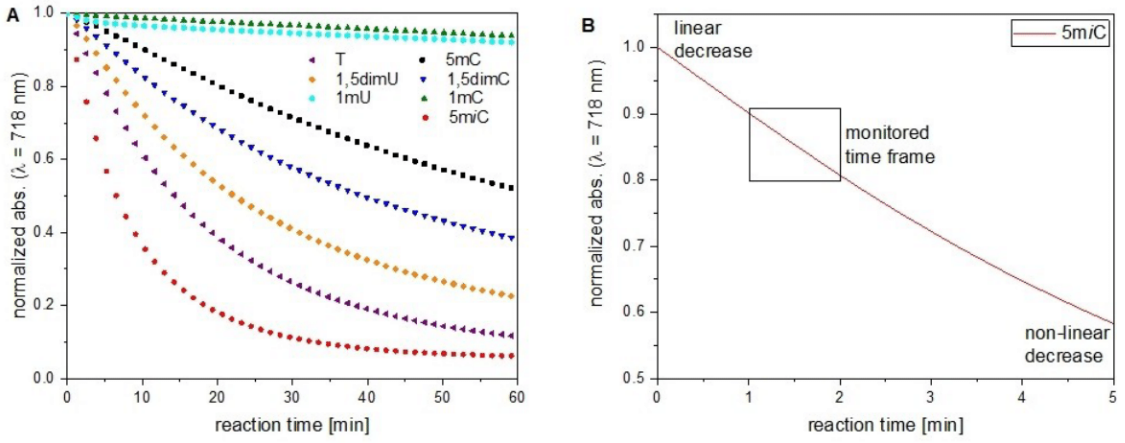


Figure 3. UV/Vis spectroscopy kinetics: A) Plot of the development of the absorbance of a series of reactions. B) Linear and non-linear decrease of the absorbance at λ = 718 nm. Conditions: [S] = 1 mM, [**1**] = 5 mM, H_2O, 30 °C.

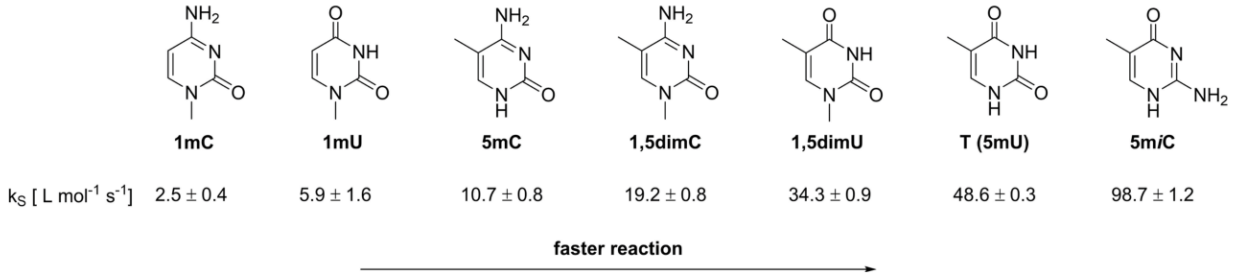


Figure 4. Observed rate constants k_s of the substrates at the following conditions: [S] = 1 mM, [**1**] = 5 mM, H_2O, 30 °C. Rate constants k_s were calculated using a second order rate equation (Eq. 1).

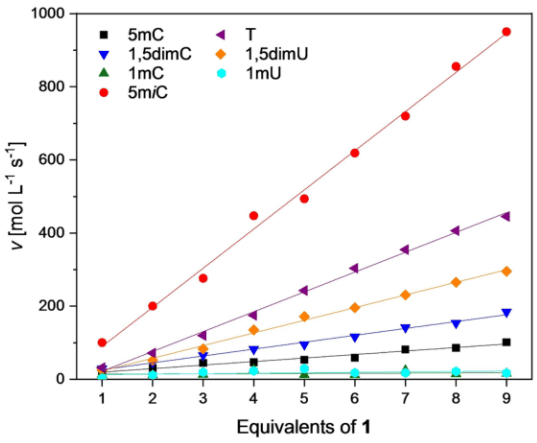


Figure 5. Plot of the measured reaction rates *v* of the reaction of **1** with nucleobase substrates. For a zoomed-in version of 5mC, 1mC, and 1,5dimC results see Supporting Information Figure S3[B]. 5mC (black squares), 5m*i*C (red dots), 1mC (green triangles), 1,5dimC (blue inverted triangles), T (purple twisted triangles), 1,5dimU (orange diamonds), and 1mU (cyan circles). Conditions: [S] = 1 mM, [**1**] = 1–9 mM, H_2O, 30 °C. The Supporting Information contains a second set of measurements for 5mC, 5m*i*C, 1mC, and 1,5dimC (Figure S3[A]).

cytosine counterparts 1mC, 5mC and 1,5dimC, respectively. When comparing mono- vs. dimethylation, a significant difference between uracil- and cytosine derived substrates was noted: The dimethylated compounds 1,5dimC and 1,5dimU showed divergent reactivity: whereas 1,5dimC reacts faster than 5mC, 1,5dimU reacts slower than its monomethylated counterpart T. We found that in the case of dimethylated substrates, reactivity can be attributed almost completely to the methyl group bound to the carbon atom at position 5 (*vide infra* for details). 5m*i*C then shows the fastest reaction rates *v* by a large margin.

Clearly, the nature of the substituents on the 1, 2, and 4 position influences the reactivity of the methyl groups present. It can be summarized that an amine group at position 4 (as in 5mC) slows the reactivity, whereas a carbonyl moiety (as in T and 1,5mU) increases it. Also, a guanidine moiety (amine substituent at position 2, as in 5m*i*C) increases reactivity compared to a urea moiety (as in 5mC or T). Methylation of the 1 position can influence the reactivity both ways: in the case of an exocyclic amine on position 4 (as in 1,5dimC), the reaction rate is increased. If, however, two carbonyl functions are present (as in 1,5dimU), the rate is decreased when the 1 position is methylated. These observations imply that the heterocyclic, conjugated ring system is capable of relaying electronic information from the substituents to the methyl groups. Effects

stemming from steric interactions and coordination of substrates and products to **1** might also influence the reactivity.

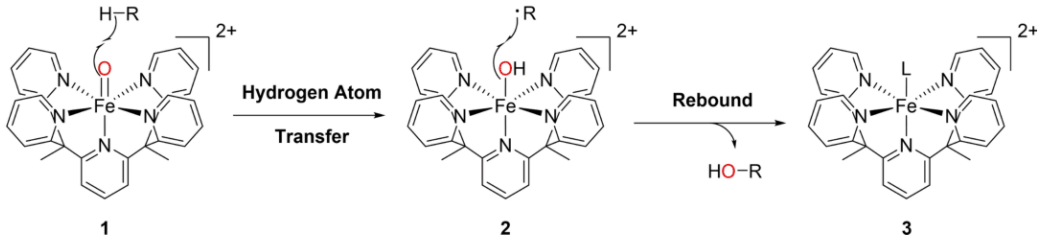
BDE calculation

In a next step, we wanted to rationalize the above described trends. As iron(IV)-oxo compounds both in synthetic[12] and enzymatic context,[13] as well as **1** in particular,[9,14] are reported to react *via* a hydrogen atom transfer from an aliphatic C–H bond (Scheme 1), BDEs are commonly believed to play an integral part concerning reaction rates.

We therefore calculated the relevant BDEs of the substrates (Figure S17). All quantum mechanics (QM) results are reported at the SMD(H_2O)/DLPNO-CCSD(T)/CBS//(U)B3LYP-D3/6-31+G-(d,p) level of theory as stated in the supporting information (for thermodynamic data see Tables S5–10).[15] A graphical representation of all aqueous phase BDE values at the relevant C–H bonds in the above-mentioned substrates is shown in Figure 6 (also compare gas phase BDE values, Figures S18–19). A significant difference in BDE values can be observed between *N1*- and *C5*-methylated compounds: methyl groups connected to another carbon atom possess BDE values of 379–387 kJ mol^{-1}, whereas the methyl groups situated on the *N1*-nitrogen atom show much higher BDEs of 413–416 kJ mol^{-1}. The BDE of 5m*i*C is somewhat lower than all other compounds, including its closest structural relatives T (ΔBDE = 3.7 kJ mol^{-1}) and 5mC (ΔBDE = 7.9 kJ mol^{-1}). The site-specific BDE values of doubly methylated compounds generally compare to their mono-methylated parent species. The aqueous phase BDE value of 1hmC is the highest of all calculated compounds at 421.0 kJ mol^{-1}, although we note that there is a substantial solvation effect on this value (see Table S2).

Comparing BDEs to observed rate constants

The calculated BDE values were found to predict reactivity of the substrates perfectly: low BDEs correspond to high rate constants k_s. The only exception from the observed flawless correlation is 1mU, which reacts slightly faster than its BDE reactivity would predict. We hypothesize that this observation is



Scheme 1. Two-step reaction mechanism for **1** with aliphatic C–H bonds (indicated as R–H) as postulated by Daumann and Jonasson and Chantarojsiri et al.[9,14] The transferred oxygen atom is marked in red. L = solvent, substrate, or product molecule that completes the coordination sphere of **3**.

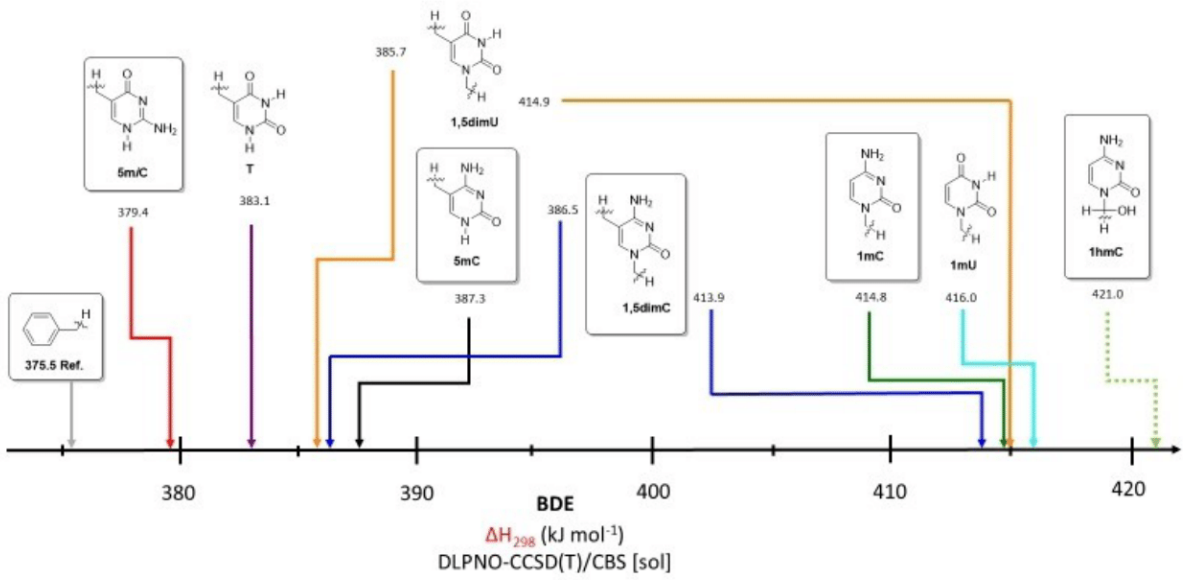


Figure 6. Aqueous phase ($\Delta H_{sol} = \Delta H_{298} + \Delta G_{solv}$) RCH_2–H bond dissociation energies (BDEs) of the relevant substrates calculated at the DLPNO-CCSD(T)/CBS level of theory. Dotted lines indicate the BDEs of molecules that were calculated but not included in experiments. For comparison of calculated BDEs of gas and aqueous phase see Figures S18–19.

due to the method of data analysis. We used the absorbance values between 1–2 min reaction time for the calculation of reaction time, however, as the reaction of 1mC and 1mU is very slow, small variations (possibly due to uncompleted mixing upon starting the experiment) influence the observed reactivity strongly. When the decrease in absorbance for 1mC and 1mU is compared for the entire length of the experiment, a very similar behavior is observed. This would fit very well with the BDEs of 1mC and 1mU being very similar at 415 and 416 kJ mol^{-1}, respectively. Nonetheless, the behavior of 1mU does fit very well within the broader trend described above, even if the data analysis is not perfectly suited to its behavior.

The correlation of BDE and k_S is in particular remarkable for the situation in the twice-methylated substrates 1,5dimC and 1,5dimU (Table 1). As described above we found that *N*-methylation increased the reaction rate in the cytosine derivative whereas the opposite was observed for the uracil derivative. This behavior is mirrored in the corresponding BDE values: the C–H bond on the carbon-bound methyl group in 1,5dimC is found to possess a lower BDE than 5mC, implying the experimentally confirmed increased reactivity. In the case of 1,5dimU a higher BDE was calculated, matching its lower reactivity compared to T.

The perfect correlation of calculated BDEs to observed reaction rates provides further evidence to corroborate the previously postulated two-step reaction pathway of **1** with aliphatic C–H bonds (Scheme 1).[9] In the first step in this mechanism, a hydrogen atom is transferred from the substrates R–Me(H) to the iron compound, generating an iron(III)-hydroxido species (**2**) and a carbon-centered radical. These species then recombine in a rebound step to form the product R–OH and an iron(II)-species (**3**). It was found that C–H abstraction/hydrogen transfer is the rate limiting step.[9]

As the hydrogen atom transfer (HAT) step involves the breaking of the C–H bond in the substrate, the BDE should determine the reaction rate. We had previously demonstrated this for both 5mC and 5hmC; in this work we expanded the substrate scope significantly to include 1mC, 1,5dimC, 5m*i*C, T, 1mU, and 1,5dimU.

When plotting the aqueous phase BDE values versus ln(k_s), a linear Bell-Evans-Polanyi correlation can be observed for 5mC, 1,5dimC, 1,5dimU, T and 5m*i*C (Figure 7). 1mC and 1mU are outliers, probably due to them being *N*-methylated instead of C-methylated as all other compounds. As mentioned previously, for the doubly methylated compounds the reactivity of the methyl group at the 5-position prevails (*vide infra* for details on how we come to this conclusion). Repeating this type of analysis with gas phase BDE(C–H) values we obtain similar results (see Figure S23), thus, we conclude that the observed correlation reflects the intrinsic properties of the studied nucleobase substrates.

Table 1. Comparison of calculated BDEs and observed reaction rates.

Substrate	BDE [kJ mol^{-1}]	k_S [L mol^{-1} s^{-1}]
1mC	414.8	2.5 ± 0.4
1mU	416.0	5.9 ± 1.6
5mC	387.3	10.7 ± 0.8
1,5dimC	386.5	19.2 ± 0.8
1,5dimU	385.7	34.4 ± 0.9
T	383.1	48.6 ± 0.3
5m*i*C	379.4	98.7 ± 1.2

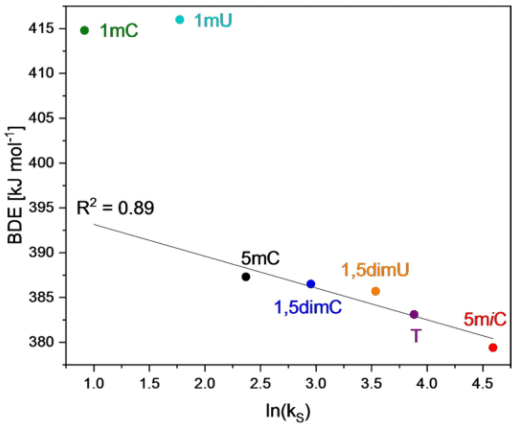


Figure 7. Plot of the calculated BDE values in aqueous phase against the observed rate constants k_S on a logarithmic scale ($R^2 = 0.891$) of the substrates 1mC, 1mU, 5mC, 1,5dimC, 1,5dimU, T, and 5m*i*C. In the case of the demethylated substrates 1,5dimC and 1,5dimU only the BDE of the carbon bound methyl groups are plotted. A similar plot of calculated BDE values in the gas phase against the observed rate constants k_S on a logarithmic scale can be found in Figure S23.

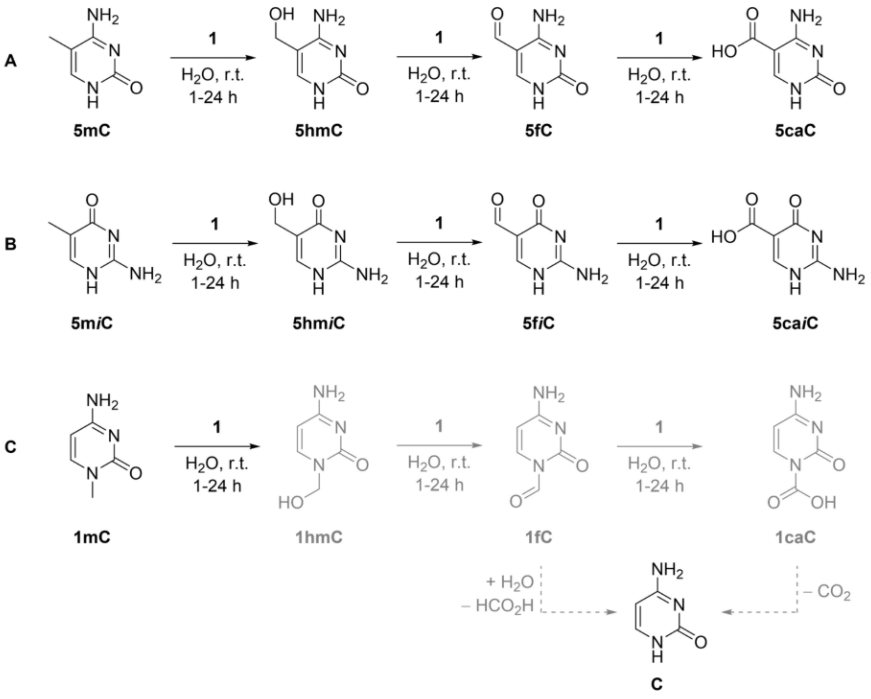
The presented data is a remarkable result, as it both provides evidence to confirm the previously postulated mechanism of **1** and shows that calculated BDEs can be reliably used to predict reaction rates for these types of substrates (in the absence of an enzyme's second coordination sphere). It is noteworthy that the reactivity of the natural nucleobase 5mC is in the middle of the observed spectrum: the *N*-methylated compounds 1mC or 1mU react significantly slower whereas 5m*i*C reacts significantly faster. This could be considered an indication on why 5mC can be considered an ideal epigenetic marker in nature: the reactivity of 5mC towards an iron(IV)-oxo moiety seems to be in a range that is both fast enough for efficient catalytic conversion by an enzyme and still slow enough to be controlled within a biological system. For example, 1mC is oxidized so slowly that even if its oxidized derivatives were stable (*vide infra*) it would not be a suitable substrate for enzymatic conversion. On the other hand, the reactivity of 5m*i*C seems to be so fast that it would be hard to control - an important factor in the delicate methylation equilibrium that is maintained by DNA methyltransferases (DNMT) and TET enzymes.[16] Whereas we do not claim that this behavior of 5mC and the other substrates towards iron(IV)-oxo species has been an evolutionary pressure resulting in the formation of 5mC epigenetics as we know it, we provide evidence that 5mC is indeed a perfect substrate for the task it performs. Regarding applications of natural and artificial methylated nucleobases in synthetic biology and DNA-based storage systems, their different reactivities could allow for additional layers of information and tunability. By incorporating two nucleobases of vastly different reactivity towards **1**, such as

1mC and 5m*i*C, into an artificial DNA strand, differentiation during oxidative sequencing could be used to drastically increase density of information.

Product Analysis (HPLC-MS/GC-MS)

We used both UHPLC-MS and GC-MS to identify the products formed in the reactions of **1** with the cytosine substrates, as these are most relevant to our research question (see Supporting Information, Table S1 and Figures S4–16). We also analyzed the product distribution of T oxidation products using GC-MS (see Figure S16).

In our recently published work on the reactivity of **1** towards 5mC we identified the oxidized derivatives 5hmC, 5fC, and 5caC, that are also formed by TET enzymes in DNA substrates, using GC-MS.[9] In this work we corroborated these results with HPLC-MS measurements (Scheme 2A). In the measurements of the reaction samples we found signals corresponding to all expected products (5hmC, 5fC, 5caC), but did also find an additional signal with a mass-to-charge ratio of 126.0622, which would correspond to 5mC. Due to the longer retention time (~13.2 min) than for the reference signal of 5mC (~7 min) but equal m/z ratio, we propose that this is a dimer of 5mC, probably formed during lyophilization (see Figures S4 and S5). In fact, a hemi-protonated dimer of 5mC (5mC-5mCH$^+$) has previously been isolated by us and was structurally characterized.[17] For reactions with 5m*i*C we found signals at m/z values corresponding to 5-hydroxymethyl *iso*-cytosine (5hm*i*C), 5-formyl *iso*-cytosine (5f*i*C), 5-carboxy *iso*-cytosine (5ca*i*C, Scheme 2B). When 5 equivalents of **1** were used, no significant difference in product distribution can be observed between the sample taken after 1 h and that taken after 24 h. This indicates that the reaction is complete after 1 h, which agrees with the observed high reaction rates and calculated low BDEs. When changing the substrate to 1mC only small amounts of cytosine (C) could be identified as a product using the standard procedure: conducting the reaction in water, filtration through silica, lyophilization, UHPLC measurement. However, traces of all oxidation products can be found using HPLC-MS when injecting samples before standard workup procedures were applied. We propose that the methyl group on 1mC is indeed oxidized to the expected products 1hmC, 1fC and 1cC, however, several pathways can lead to a quick decomposition towards cytosine. The first oxidation of 1mC leads to 1hmC which, as an hemiaminal, tends to equilibrate towards cytosine, when removing formaldehyde at reduced pressure (Scheme 2). This overall process, however, is endergonic (see Figure S21[A] and [B] and Table S3). Oxidation of 1hmC towards 1fC is proposed to be faster compared to the oxidation of 1mC, since the gas phase BDE of 1hmC is 4.3 kJ mol^{-1} lower than for 1mC. In this particular case, we observed that the gas phase BDE value represents a better estimation of the reaction trend than in solution phase. This is possibly due to formation of a cyclic hydrogen bond between the hydroxymethyl group and urea moiety (for details, please refer to Figure S20 and subsequent text). The oxidation products that follow, 1fC and 1caC, were both found to be exergonic and therefore unstable (see Scheme 2C). The calculated solution phase free energies of reaction ΔG for the



Scheme 2. Identified products in the reactions of 5mC, 5m*i*C, and 1mC with **1**. All structures in black were detected using HPLC-MS after 1 h and 24 h using both 1 and 5 equiv. of **1**, although in different ratios. 5m*i*C, 5hm*i*C and 5ca*i*C were also detected using GC-MS with 1 equiv. of **1** after 1 h. All structures in gray were detected on HPLC-MS by injection of the untreated reaction solution with 1 equiv. of **1** after 24 h and 44 h.

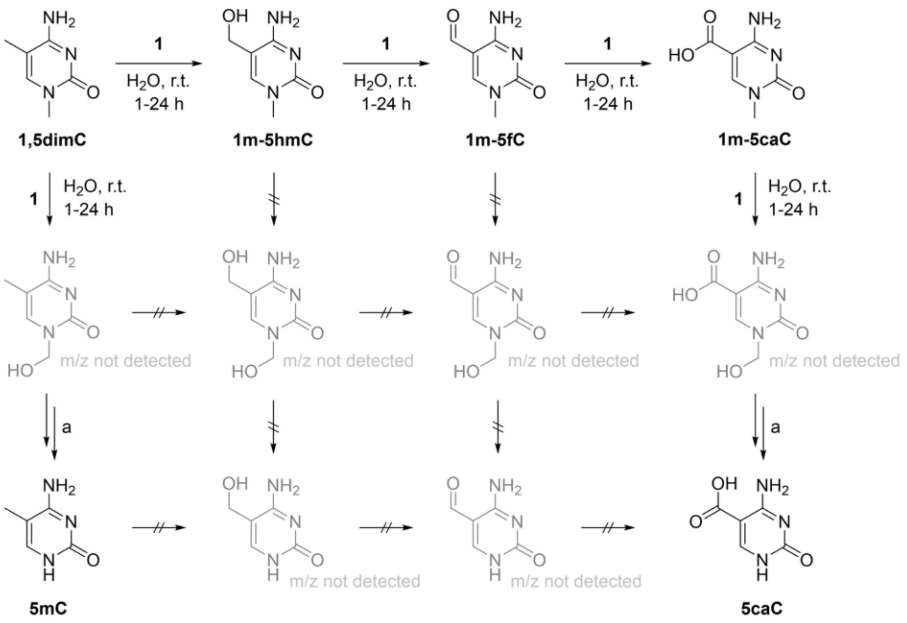
deformylation of 1fC is $-17.0\ kJ\,mol^{-1}$ (Figure S21[C]), for the decarboxylation of 5caC it is $-44.9\ kJ\,mol^{-1}$ (Figure S21[D]).

As a proof-of-concept for this hypothesis, free energies of reaction ΔG were calculated for the literature known deformylation of 6-hydroxymethyladenine (6hmA) and 6-formyladenine (6fA), which are oxidation products of naturally occurring 6-methyladenine (6mA) in mammalian DNA.[18] The deformylation of 6hmA towards adenine and formaldehyde is endergonic ($\Delta G_{298,H2O} = +16.5\ kJ\,mol^{-1}$, see Figure S22[B] and Table S4), while the deformylation of the following oxidative product, 6fA, is thermoneutral ($\Delta G_{298,H2O} = +1.1\ kJ\,mol^{-1}$, see Figure S22[C]). It has been reported, that 6hmA and 6fA are transient intermediates of the oxidation of 6 mA with hydrogen peroxide.[19] The aqueous phase BDE(C–H) value of 6 mA ($+396.4\ kJ\,mol^{-1}$, see Table S2) is $13.3\ kJ\,mol^{-1}$ lower than that of 6hmA ($+408.7\ kJ\,mol^{-1}$, see Table S2), which is in support of a consecutive oxidation cascade and therefore also of the proposed pathway for the oxidative demethylation of 1mC to C.

In the case of 1,5-dimC, the product composition is more convoluted, however, it matches with the expectations based on observed reaction rates and calculated BDEs (Scheme 3): in addition to the starting material we detected m/z values corresponding to 1-methyl-5-hydroxymethylcytosine (1m-5hmC), 1-methyl-5-formylcytosine (1m-5fC), and 1-methyl-5-carboxycytosine (1m-5caC). These products are expected as the carbon-bound methyl group at position 5 has a lower calculated BDE and should therefore be oxidized more readily. The observed products are also corresponding to those observed for reactions of 5mC and 5m*i*C with **1**. However, besides the oxidation products of position 5, we additionally detected both 5mC and 5caC (confirmed by m/z and retention time). We therefore propose, that for 1,5dimC hydroxylation on the 1-methyl group also occurs to some small extent, and this oxidation product then reacts to 5mC (Scheme 3), corresponding to our proposed mechanism for 1mC (Scheme 2). Similarly, 1m-5caC only offers the 1-methyl group as a substrate position for **1** to react with, so 5caC is formed *via* hydroxylation and subsequent reaction to 1m-5caC. The amounts of 1m-5hmC and 1m-5fC are too low to represent a significant target for 1-methyl-hydroxylation, therefore no 5hmC or 5fC are detected. Similarly, only small amounts of 5mC are detected and it is, as the calculated BDE values show, less readily oxidized to 5hmC and 5fC as its *N*-methylated counterpart 1,5dimC.

Conclusion

We have presented a comprehensive study of the reactivity of the biomimetic iron(IV)-oxo complex **1** towards a number of methylated cytosine and uracil substrates and compared the results to calculated BDE values. Using HPLC-MS and GC-MS we also identified the products of the reaction of **1** with the cytosine derivatives 5m*i*C, 1mC, and 1,5dimC. We also provided a reasonable explanation for the observed decomposition of oxidized 1mC derivatives 1hmC, 1fC, and 1caC by calculating deformylation and decarboxylation energy profiles and comparing these to the literature known pathway in 6hmA/6fA. In the case of 1,5dimC, combining the observed reaction rates, the calculated BDE values and the observations regarding 1mC allowed for a clear interpretation of the observed product distribution. We found that the reported reaction rates are in very good agreement with the calculated BDEs which therefore prove to be a very good predictor for the reactivity of **1** towards a broad range of methylated substrates. In summary the observed reaction rates towards **1** seemed to be dominated by the reactivity of the C-methylated part of the substrates. 5mC, T



Scheme 3. Identified products in the reaction of 1,5dimC with **1**. All structures in black were detected using HPLC-MS after 1 h and 24 h using both 1 and 5 equiv. of **1**, although in different ratios. a: For this step we propose a similar reaction sequence as for 1mC (see Scheme 2).

and 5m*i*C possess a distinctly different reactivity than the solely *N*-methylated compounds 1mC and 1mU. For the compounds 1,5dimC and 1,5dimU that are both *C*-methylated and *N*-methylated, their C–H reactivity towards oxidation by **1** is mostly determined by the methyl group bound to the carbon atom. Interestingly, diverging effects were observed for *N*-methylation in 1,5-dimC and 1,5-dimU: in the case of 1,5dimC the reactivity is higher than for its mono-methylated counterpart 5mC, whereas 1,5dimU was observed to react faster than T. The observed rates also suggest that the reactivity of the 5-methyl group on the epigenetic marker 5mC is ideal to fit its purpose in a delicate equilibrium maintained by a series of enzymes. While the low reactivity of 1mU towards **1** is certainly not the reason for the stability of the corresponding nucleoside *pseudo*-methyl-uridine (1mΨ) that is used in some Covid-19 mRNA vaccines, our analysis shows that 1mUs methyl group is rather inert towards oxidation reactions. These observations can be a useful tool in predicting the possibilities of using and manipulating methylated nucleobases in synthetic biology, e.g. data storage, and further understanding the mechanisms of epigenetics.

Acknowledgements

We thank Prof. Dr. Bernhard Lippert for a sample of 1,5dimC. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1309-325871075. N.S.W.J. thanks the Studienstiftung des Deutschen Volkes. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: 5-methylcytosine · computational chemistry · DNA methylation · epigenetics · synthetic biology

[1] a) L. Organick, S. D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, *Nat. Biotechnol.* **2018**, *36*, 242; b) R. Lopez, Y.-J. Chen, S. D. Ang, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Seelig, K. Strauss, L. Ceze, *Nat. Commun.* **2019**, *10*, 1–9; c) L. Ceze, J. Nivala, K. Strauss, *Nat. Rev. Genet.* **2019**, *20*, 456–466.
[2] A. Hofer, Z. J. Liu, S. Balasubramanian, *J. Am. Chem. Soc.* **2019**, *141*, 6420–6429.
[3] F. R. Traube, T. Carell, *RNA Biol.* **2017**, *14*, 1099–1107.
[4] C. Mayer, G. R. McInroy, P. Murat, P. Van Delft, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2016**, *55*, 11144–11148; *Angew. Chem.* **2016**, *128*, 11310–11314.
[5] a) S. Hoshika, N. A. Leal, M.-J. Kim, M.-S. Kim, N. B. Karalkar, H.-J. Kim, A. M. Bates, N. E. Watkins, H. A. SantaLucia, A. J. Meyer, *Science* **2019**, *363*, 884–887; b) S. A. Benner, A. M. Sismour, *Nat. Rev. Genet.* **2005**, *6*, 533–543.
[6] J. W. Park, P. N. Lagniton, Y. Liu, R.-H. Xu, *Int. J. Biol. Sci.* **2021**, *17*, 1446.
[7] O. Andries, S. Mc Cafferty, S. C. De Smedt, R. Weiss, N. N. Sanders, T. Kitada, *J. Controlled Release* **2015**, *217*, 337–344.
[8] A. M. Sismour, S. A. Benner, *Nucleic Acids Res.* **2005**, *33*, 5640–5646.
[9] N. S. Jonasson, L. J. Daumann, *Chem. Eur. J.* **2019**, *25*, 12091–12097.
[10] a) D. Schmidl, N. Jonasson, E. Korytiakova, T. Carell, L. Daumann, *Angew. Chem. Int. Ed.* **2021**, https://doi.org/10.1002/anie.202107277.
[11] a) M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934–937; b) G. S. Madugundu, J. Cadet, J. R. Wagner, *Nucleic Acids Res.* **2014**, *42*, 7450–7460; c) A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Res.* **2002**, *30*, 4068–4074.
[12] a) J. Kaizer, E. J. Klinker, N. Y. Oh, J.-U. Rohde, W. J. Song, A. Stubna, J. Kim, E. Münck, W. Nam, L. Que, *J. Am. Chem. Soc.* **2004**, *126*, 472–473; b) Y. Morimoto, J. Park, T. Suenobu, Y.-M. Lee, W. Nam, S. Fukuzumi, *Inorg. Chem.* **2012**, *51*, 10025–10036; c) A. Barbieri, O. Lanzalunga, A. Lapi, S. Di Stefano, *J. Org. Chem.* **2019**, *84*, 13549–13556; d) J. J. D. Sacramento, D. P. Goldberg, *Acc. Chem. Res.* **2018**, *51*, 2641–2652.
[13] a) J. C. Price, E. W. Barr, T. E. Glass, C. Krebs, J. M. Bollinger, *J. Am. Chem. Soc.* **2003**, *125*, 13008–13009; b) D. A. Proshlyakov, T. F. Henshaw, G. R. Monterosso, M. J. Ryle, R. P. Hausinger, *J. Am. Chem. Soc.* **2004**, *126*, 1022–1023; c) S. Kal, L. Que, *J. Biol. Inorg. Chem.* **2017**, *22*, 339–365.
[14] T. Chantarojsiri, Y. Sun, J. R. Long, C. J. Chang, *Inorg. Chem.* **2015**, *54*, 5879–5887.
[15] a) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652; b) S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104; c) R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724; d) R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650; e) A. Altun, F. Neese, G. Bistoni, *Beilstein J. Org. Chem.* **2018**, *14*, 919–929; f) M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2017**, *146*, 164105; g) F. Neese, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327; h) T. H. D. Jr., *J. Chem. Phys.* **1989**, *90*, 1007–1023; i) A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
[16] a) M. V. Greenberg, D. Bourc'his, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 590–607; b) Y. He, J. R. Ecker, *Annu. Rev. Genomics Hum. Genet.* **2015**, *16*, 55–77; c) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303; d) Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, *Science* **2011**, *333*, 1303–1307.
[17] A. Menke, R. C. A. Dubini, P. Mayer, P. Rovó, L. J. Daumann, *Eur. J. Inorg. Chem.* **2021**, 30–36.
[18] a) J. Xiong, T.-T. Ye, C.-J. Ma, Q.-Y. Cheng, B.-F. Yuan, Y.-Q. Feng, *Nucleic Acids Res.* **2019**, *47*, 1268–1277; b) Y. Fu, G. Jia, X. Pang, R. Wang, X. Wang, C. Li, *Nat. Commun.* **2013**, *4*, 1798.
[19] J. Wu, H. Xiao, T. Wang, T. Hong, B. Fu, D. Bai, Z. He, S. Peng, X. Xing, J. Hu, *Chem. Sci.* **2015**, *6*, 3013–3017.

Manuscript received: August 15, 2021
Revised manuscript received: September 8, 2021
Accepted manuscript online: September 9, 2021
Version of record online: September 23, 2021

5.1 Materials and Methods

Solvents and Chemicals

Chemicals were purchased from commercial sources (Sigma Aldrich, ABCR, Acros Organics, Alfa Aesar, TCI Chemicals, Oakwood Chemicals) or the LMU Munich chemical supply and used without further purification except for dichloromethane, diethyl ether, and hexanes. These were purchased from the LMU Munich chemical supply and were distilled once under reduced pressure prior to use.

Methods and Manipulations

All manipulations were carried out under ambient conditions if not stated otherwise. Air- and moisture-sensitive chemicals and absolute solvents were transferred via stainless-steel cannula or syringe. Organic solutions were concentrated by rotary evaporation at 40 °C. Analytical thin layer chromatography (TLC) was performed on pre-coated (silica gel, 0.25 mm, 60 Å pore-size, 230–400 mesh, Merck KGA) aluminum plates or which were impregnated with a fluorescent indicator (254 nm). TLC plates were visualized by exposure to ultraviolet light.

Unless otherwise indicated, a new batch of **1** was prepared freshly for each set of experiments according to the procedure described by Daumann and Jonasson[1], stored at -20 °C and it was not used longer than for three days. The anion exchange step was performed *in situ* only a few minutes before reaction start. OriginPro 2018G was used for data evaluation.

NMR spectroscopy

^{1}H NMR, ^{2}H NMR and ^{13}C NMR spectra were recorded at room temperature on Bruker Avance III (400 MHz) operating at 400 MHz for proton nuclei and 100 MHz for carbon nuclei. ^{1}H-chemical shifts are reported in ppm units relative to $CDCl_3$ (δH = 7.26), CD_3CN (δH = 1.94), D_2O (δH = 4.79) or DMSO-d_6 (δH = 2.50). ^{13}C chemical shifts are given in ppm units relative to $CDCl_3$ (δC = 77.16), CD_3CN (δC = 1.32) or DMSO-d_6 (δC = 39.52).[1] The following abbreviations were used: s = singlet, d = doublet, t = triplet, dd = doublet of doublets, dt = doublet of triplets, m = multiplet, br = broad. Coupling constants (*J*) are given in Hertz.

Elemental Analysis

Elemental analyses were determined with an Elementar vario micro.

UV Vis Kinetics

Instruments

UV/Vis full spectra including full spectra kinetics were recorded on an Agilent 8453 Diode Array Spectrophotometer with a thermostatted cuvette holder at room temperature. 10 mm quartz Suprasil cuvettes from Hellma® were used in these experiments. Initial rate kinetics of the reaction of the substrates with **1** were recorded on an Epoch2 Plate reader from Biotek using 96-well plates at 30 °C.

General Procedure for UV/Vis kinetics

For initial rate kinetics, the absorption values at 718 nm were collected every 7 s on an Epoch2 Plate reader and all experiments were performed twice. For data evaluation the absorption values from 63 s to 119 s were used for slope determination and linear regression was performed within a plot of the

averaged slope against the respective equivalents of **1**. Full UV-Vis spectra were collected every 60 s over a period of 1 h on an Agilent 8453 Diode Array Spectrophotometer.

Results and Additional Spectra

In addition to the initial rate kinetics described in the manuscript, also full UV-Vis spectra were recorded to gain information about the iron species that is formed within the reaction.

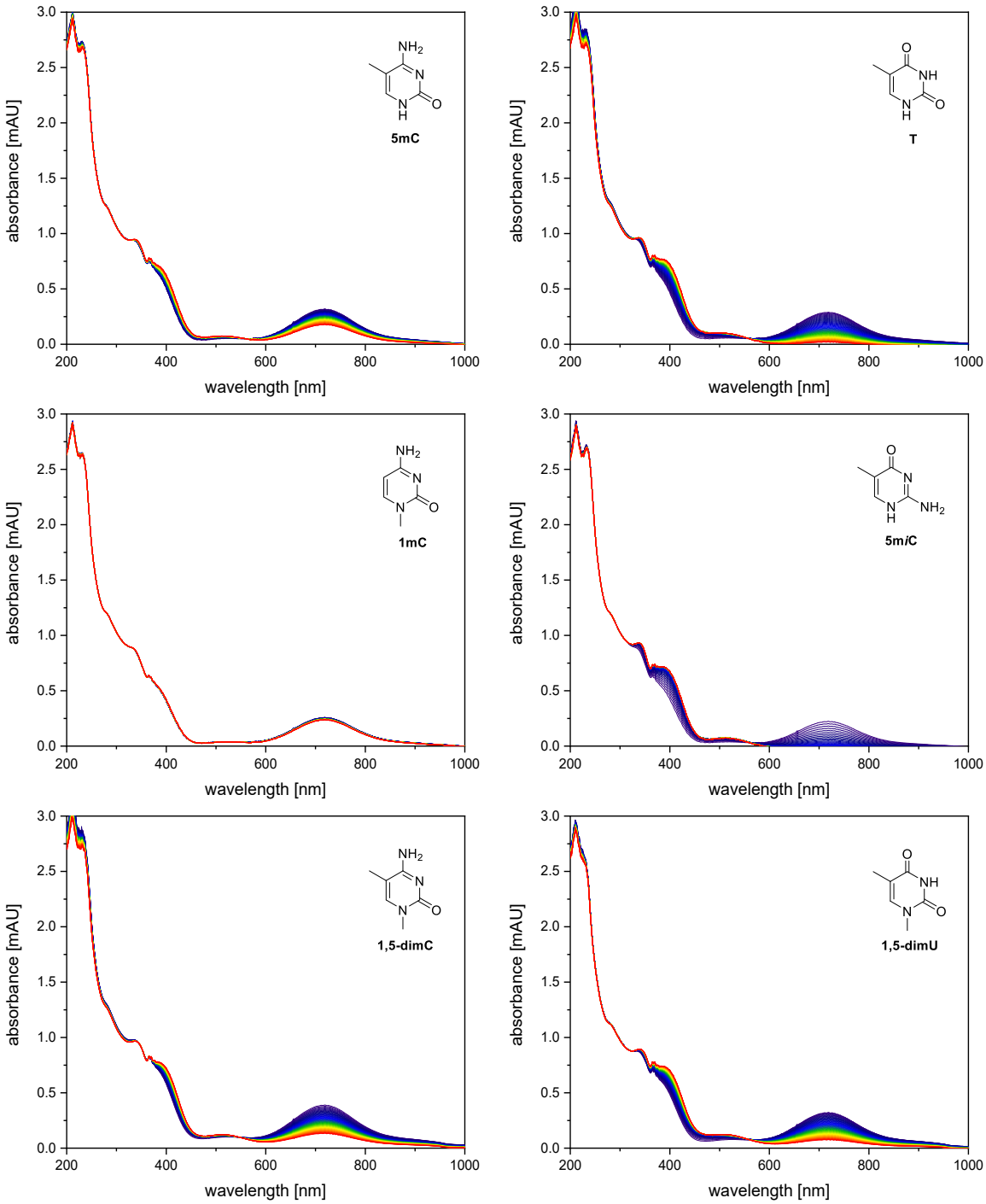


Figure S1: Kinetics of different substrates with **1** followed over 1 h recording full UV-Vis spectra. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 23 °C.

The characteristic absorption band of the iron(IV)-oxo moiety at 718 nm disappears within reaction whereas the appearance of a typical feature of the corresponding iron(III) species that is formed upon reaction can be observed.

Additionally, only the iron(IV)-oxo complex **1** without substrate was followed under the same conditions with UV-Vis spectroscopy as control reaction.

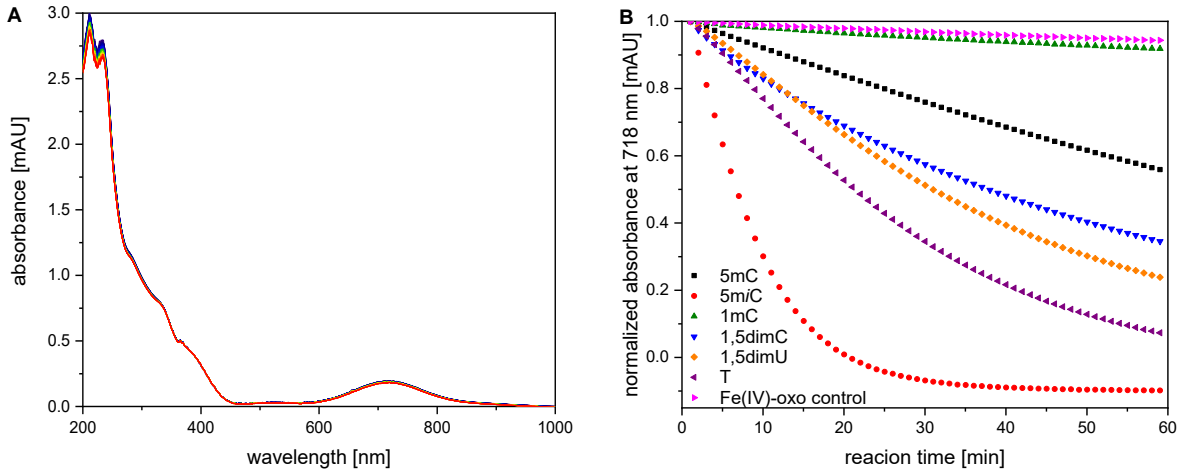


Figure S2: **A**) control reaction of **1** without substrate over 1 h. Conditions: [**1**] = 1 mM; H_2O, T = 23 °C. **B**) Plot of the development of the absorbance of a series of reactions including the Fe(IV)-oxo control reaction. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 23 °C.

The observed decrease of the iron(IV)-oxo absorption band at 718 nm is not caused by decomposition of the iron(IV)-oxo complex by itself but within a reaction with different substrates. With 1mC the reaction is marginally faster, showing some, albeit very slow, oxidation of this substrate.

In addition to the data shown in the manuscript, we performed an additional set of UV-Vis measurements at the same conditions to show the robustness of the system and look for signs of decomposition of **1**. The obtained reaction rates (Figure S3) are lower than those shown in the manuscript, indicating that **1** suffers some small amount of decomposition, however, the same trends can be observed. The batch of iron(IV)-oxo complex **1** used for these experiments had been stored for 5 days at -20 °C with several thaw-freeze cycles which is probably the cause for the observed decrease in reaction rates. This shows that **1** should always be made fresh for each measurement.

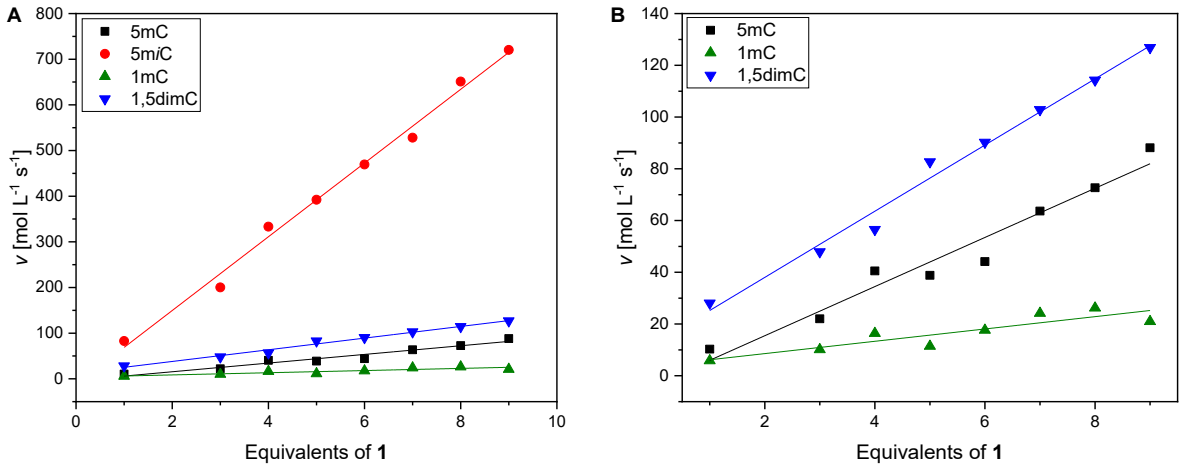


Figure S3: A) Second set of UV-Vis kinetic measurements of reaction rates at 1-9 equiv. of 1 towards 5mC, 5miC, 1mC, and 1,5dimC including linear fits. B) Zoomed in excerpt of the data presented in A. Conditions: [1] = 1-9 mM; [S] = 1 mM, H_2O, T = 30 °C.

Product Analysis by HPLC-MS and GC-MS

General Procedure

To an aqueous solution of the corresponding substrate (2.0 mM, 50.0 µL, 1.0 equiv.) a freshly prepared aqueous solution of **1** (10 mM, 100 µL, 1.0 equiv. or 500 µL, 5.0 equiv.) was added to give a total volume of 1.0 mL. The reaction mixture was shaked at 30 °C and after 1 h as well as 24 h of reaction time, a sample (200 µL) of the reaction solution was taken, filtered through silica and washed with water (6.0 mL).

High Performance Liquid Chromatography – Mass Spectrometry

Reaction samples were dried by lyophilisation. The entire amount of the obtained sample was suspended in 2.0 mL of water. Of the solution, 100 µL were diluted with 300 µL water, filtered and injected (10 µL) onto an Agilent® UHPLC 1260 InfinityII (G7115A 1260 DAD WR, G7116A 1260 MCT, G7167A 1260 Multisampler, G7104C1260 Flexible Pump) equipped with an ACE 5 C18-PFP (150 x 4.6 mm, ACE-1210-1546) column and coupled to an Agilent® 6530 C QTOF LC/MS (G1958-65171) system. Measurements were performed at 30 °C eluent temperature. As eluent system a water/methanol mixture with 0.1% formic acid was used with a gradient of 0-15% methanol with 0.1% formic acid over 15 min which was then kept at 15% methanol with 0.1% formic acid for 5 min. All used solvents were LC-MS grade. ESI-Mass spectra were recorded in scan mode between 50-250 m/z.

Liquid Chromatography – Mass Spectrometry

Oxidation products of 1mC were detected by injecting untreated reaction solution onto an Agilent 1100 SL system (G1313A ALS, G1316A COLCOM, G1316A VWD, G1312A Bin Pump) equipped with an ACE 5 C18-PFP (150 x 4.6 mm, ACE-1210-1546) column and coupled to a Bruker Daltonik HCTultra PTM Discovery system (ESI mode). As eluent system a water/methanol mixture with 0.1% formic acid was used with a gradient of 0-10% methanol with 0.1% formic acid over 10 min which was then kept at 15% methanol with 0.1% formic acid for 5 min. All used solvents were LC-MS grade. ESI-Mass spectra were recorded in scan mode between 100-1000 m/z.

Gas Chromatography – Mass Spectrometry

Reaction samples were dried by lyophilisation prior to derivatization. The entire amount of the obtained sample was suspended in 400 µl acetonitrile, 100 µl BSTFA were added and the mixture heated to 70 °C for 30 min. The samples were filtered and injected (1 µl, split depending on substrate concentration) onto an Agilent® 7920 GC equipped with a 30 m HP5-MS column (Agilent® 19091S-433UI) coupled to an Agilent® 5970 EI mass spectrometer. The injector temperature was set to 280 °C and the temperature of the ion source 230 °C. The initial oven temperature was 80 °C, held there for 2 min, ramped to 240 °C at 5 K/min and then held there for 20 min. Mass spectra were recorded in scan mode between 70-400 m/z.

Lyophilisation

Samples were lyophilised on a Christ Alpha 1-2 LDplus lyophilisator attached to a *vacuubrand* VACUU PURE® 10C screw pump.

HPLC-HRMS Traces

Signals at 3.5 min minutes occurring in all measurements are considered an impurity resulting from the removal of **1** and reacted product thereof using silica.

Table S1: Detected substrates and their respective products using UHPLC-MS including exact masses and retention times. Samples were prepared and measurements were performed as described above. Conditions: [**1**] = 1 mM or 5 mM; [S] = 1 mM, H_2O, T = 30 °C.

Substrate	**Products**	**Exact Mass [m/z]**	**Retention Time [min]**
5mC		126.0662	7.0
	5hmC	142.0611	5.0
	5fC	140.0455	8.2
	5caC	156.0404	6.0
5m*i*C		126.0662	9.8
	5hm*i*C	142.0611	5.7
	5f*i*C	140.0455	11.5
	5c*i*C	156.0404	13.8
1mC		126.0662	7.7
	C	112.0505	5.1

Substrate	**Products**	**Exact Mass [m/z]**	**Retention Time [min]**
1,5-dimC		140.0818	10.8
	1m-5hmC	156.0768	7.3
	1m-5fC	154.0611	11.7
	1m-5caC	170.0560	9.2
	5mC	126.0662	7.0
	5caC	156.0404	6.0

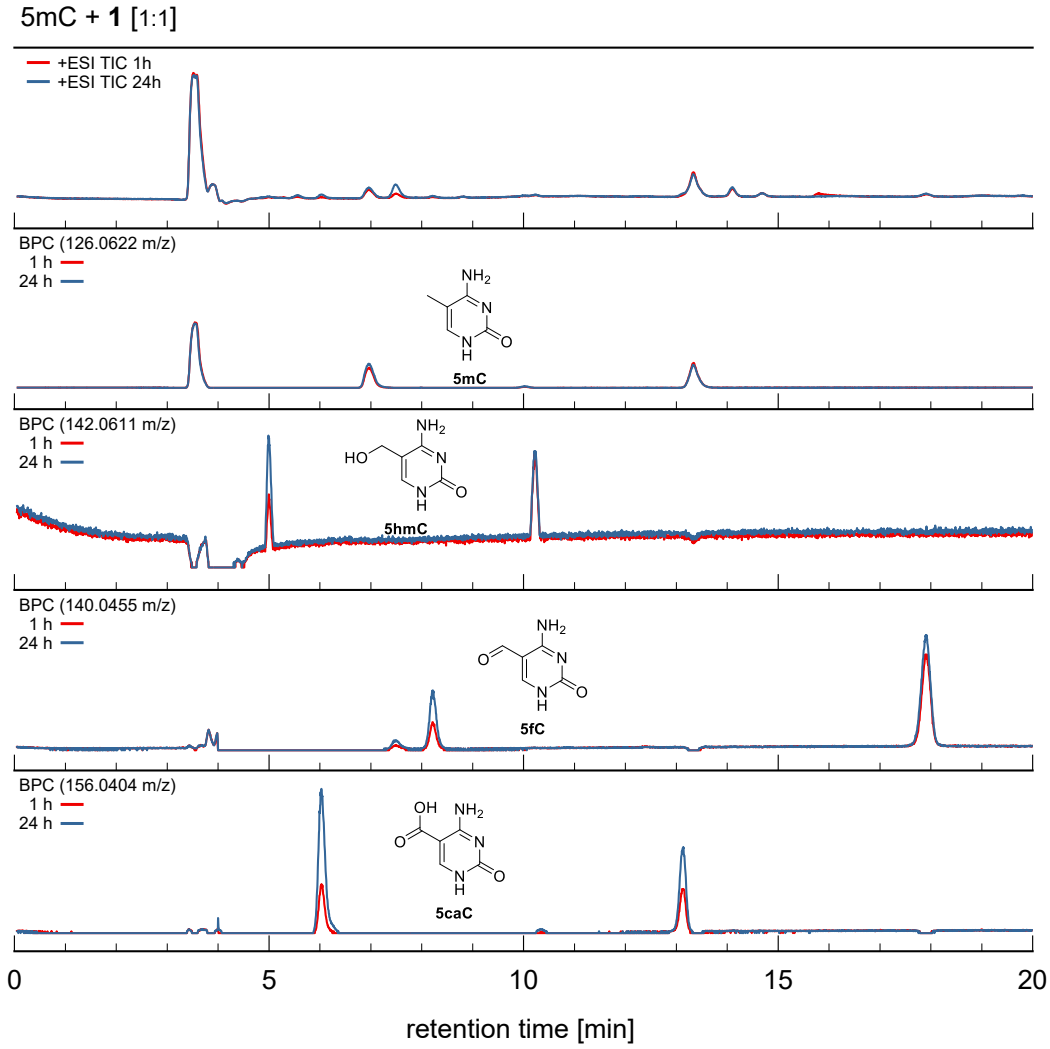


Figure S4: Excerpt of the HPLC-MS traces of the reaction of **1** with 5mC. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 5mC (126.0662 m/z), 5hmC (142.0611 m/z), 5fC (140.0455 m/z) and 5caC (156.0404 m/z) after 1 h and 24 h.

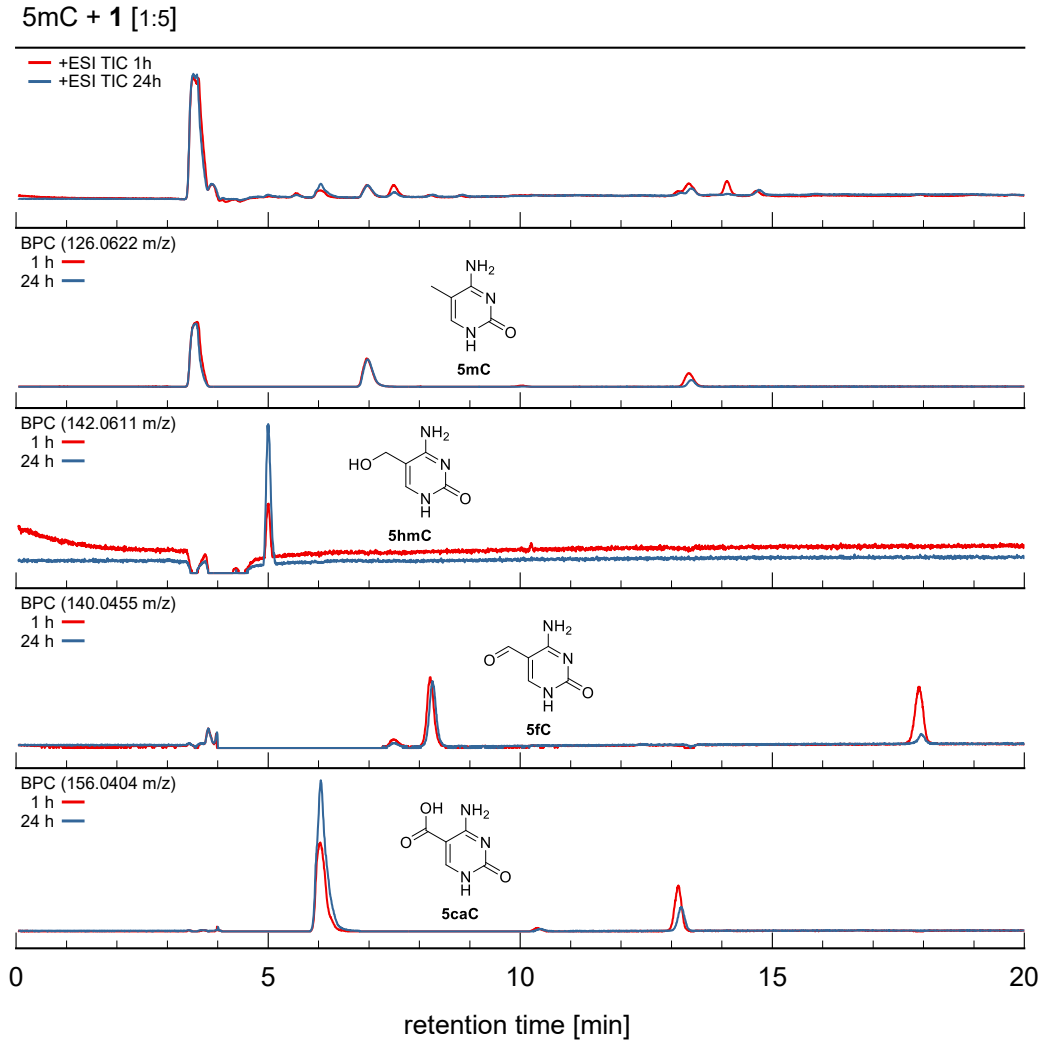


Figure S5: Excerpt of the HPLC-MS traces of the reaction of **1** with 5mC. Conditions: [**1**] = 5 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 5mC (126.0662 m/z), 5hmC (142.0611 m/z), 5fC (140.0455 m/z) and 5caC (156.0404 m/z) after 1 h and 24 h.

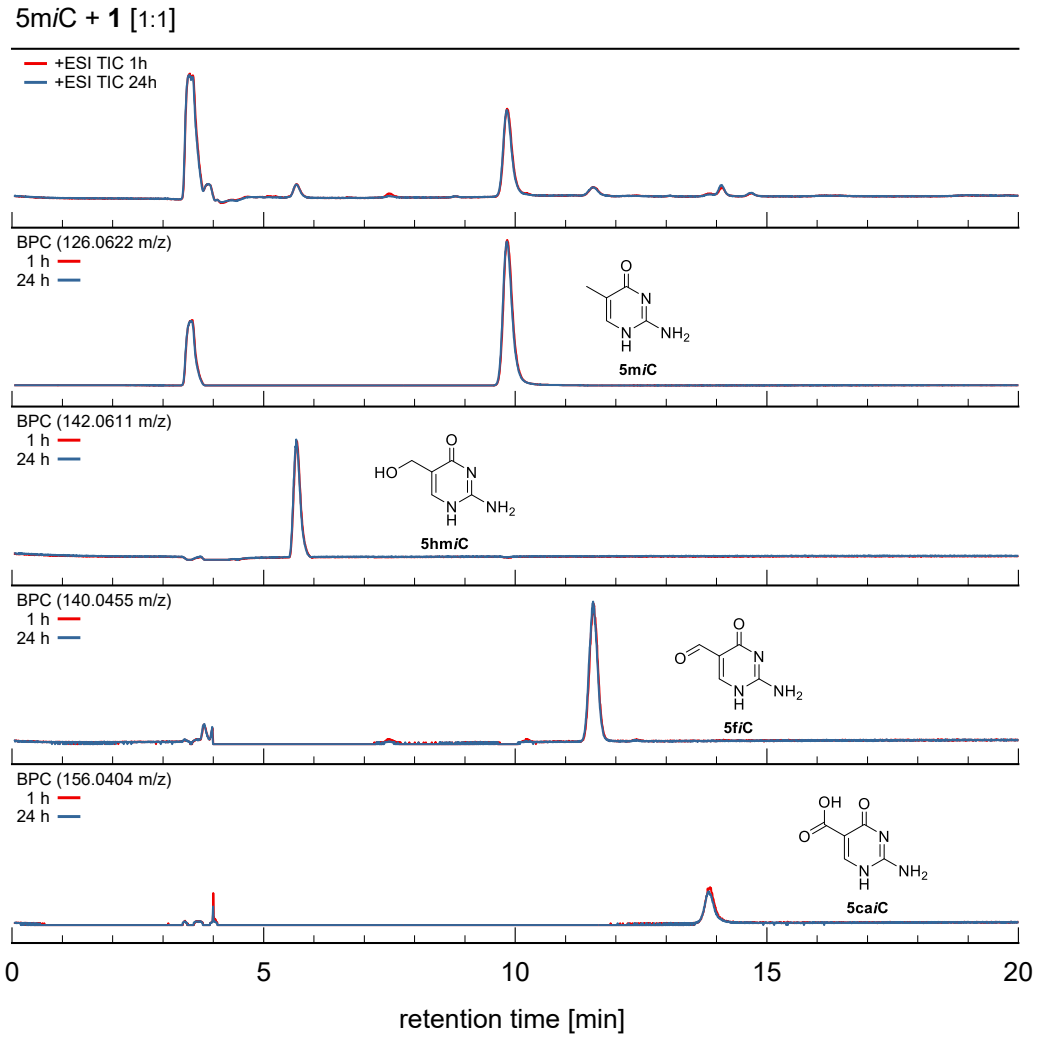


Figure S6: Excerpt of the HPLC-MS traces of the reaction of **1** with 5miC. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 5miC (126.0662 m/z), 5hmiC (142.0611 m/z), 5fiC (140.0455 m/z) and 5caiC (156.0404 m/z) after 1 h and 24 h.

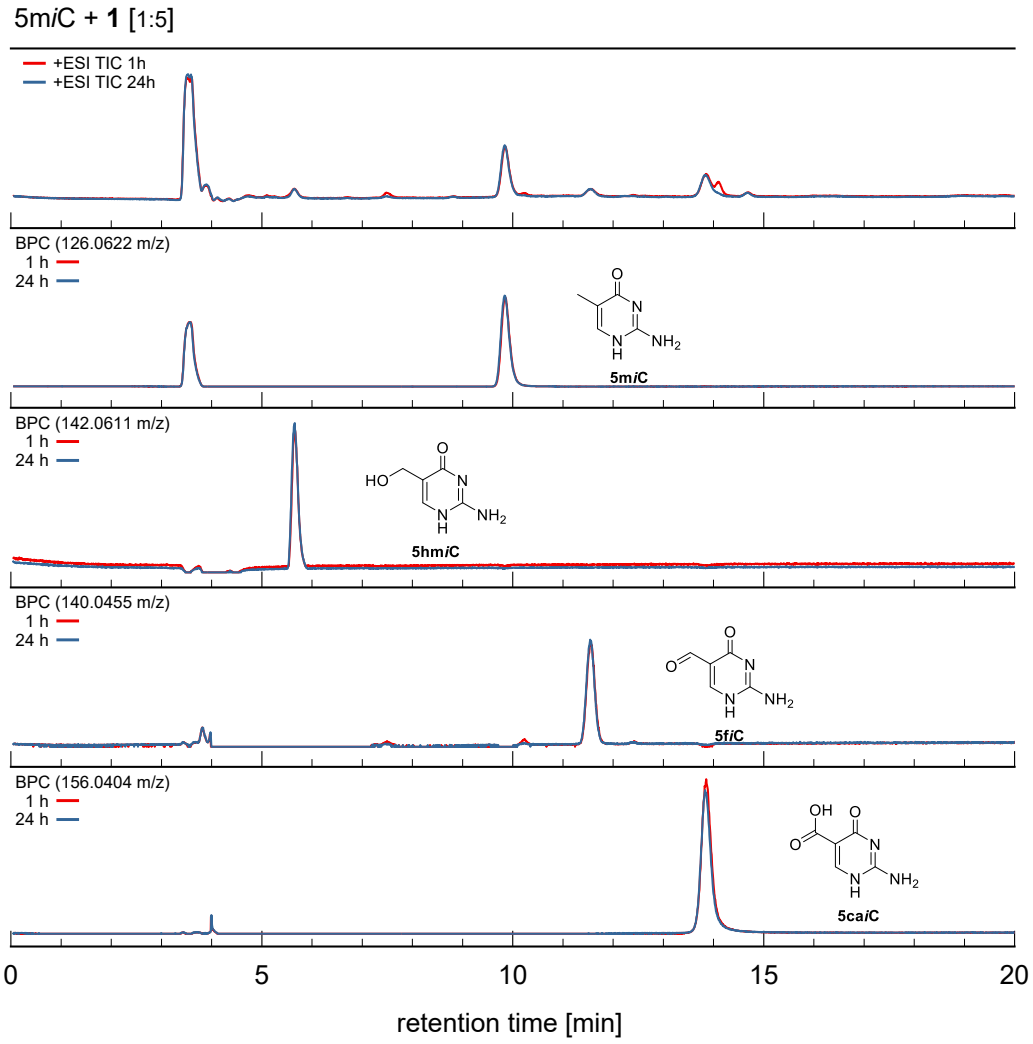


Figure S7: Excerpt of the HPLC-MS traces of the reaction of **1** with 5miC. Conditions: [**1**] = 5 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 5miC (126.0662 m/z), 5hmiC (142.0611 m/z), 5fiC (140.0455 m/z) and 5caiC (156.0404 m/z) after 1 h and 24 h.

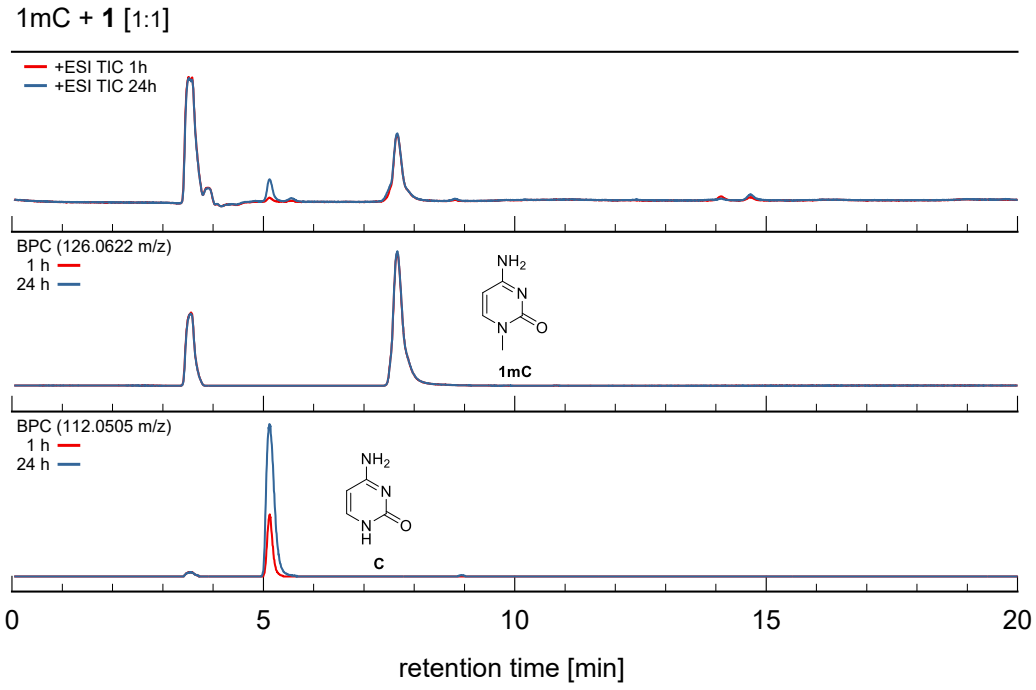


Figure S8: Excerpt of the HPLC-MS traces of the reaction of **1** with 1mC. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 1mC (126.0662 m/z) and C (112.0505 m/z) after 1 h and 24 h.

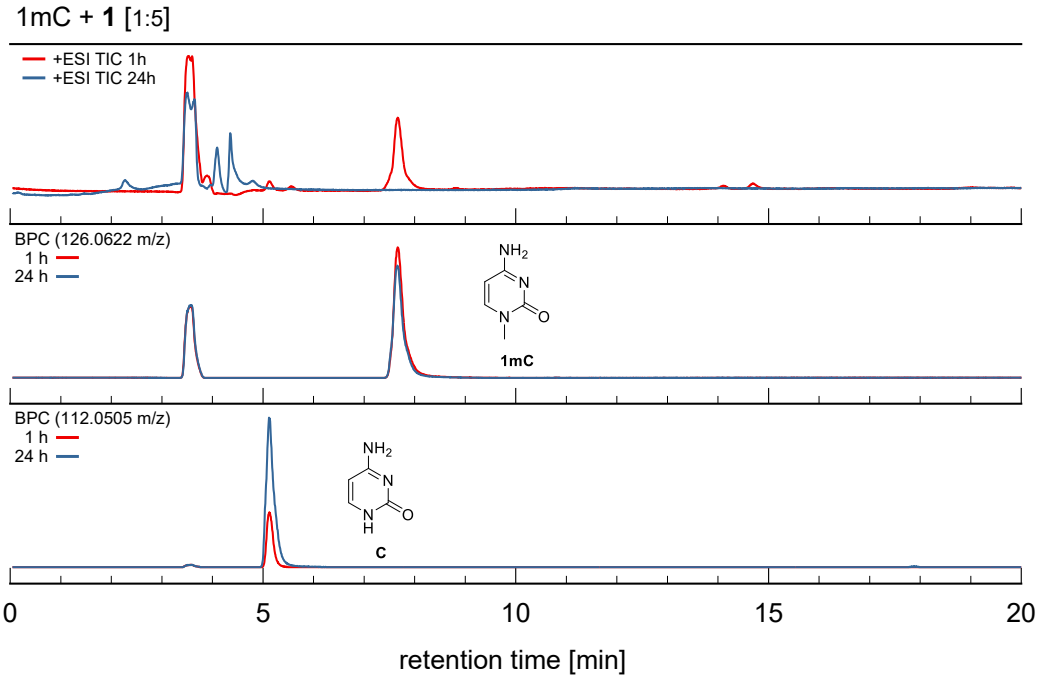


Figure S9: Excerpt of the HPLC-MS traces of the reaction of **1** with 1mC. Conditions: [**1**] = 5 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 1mC (126.0662 m/z) and C (112.0505 m/z) after 1 h and 24 h.

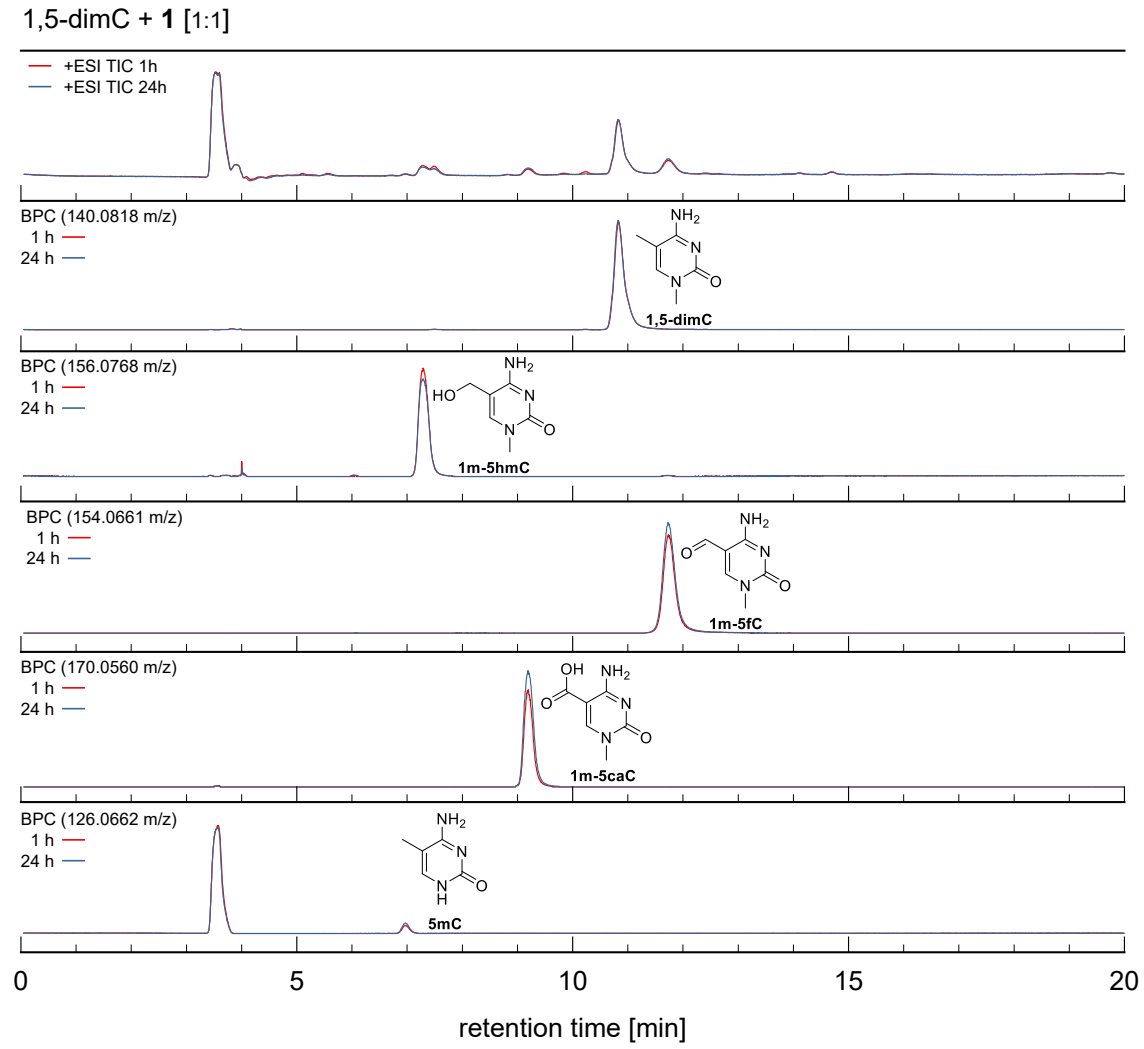


Figure S10: Excerpt of the HPLC-MS traces of the reaction of **1** with 1,5-dimC. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 1,5-dimC (140.0818 m/z), 1m-5hmC (156.0768 m/z), 1m-5fC (154.0661), 1m-5caC (170.0560 m/z), 5mC (126.0662 m/z) after 1 h and 24 h.

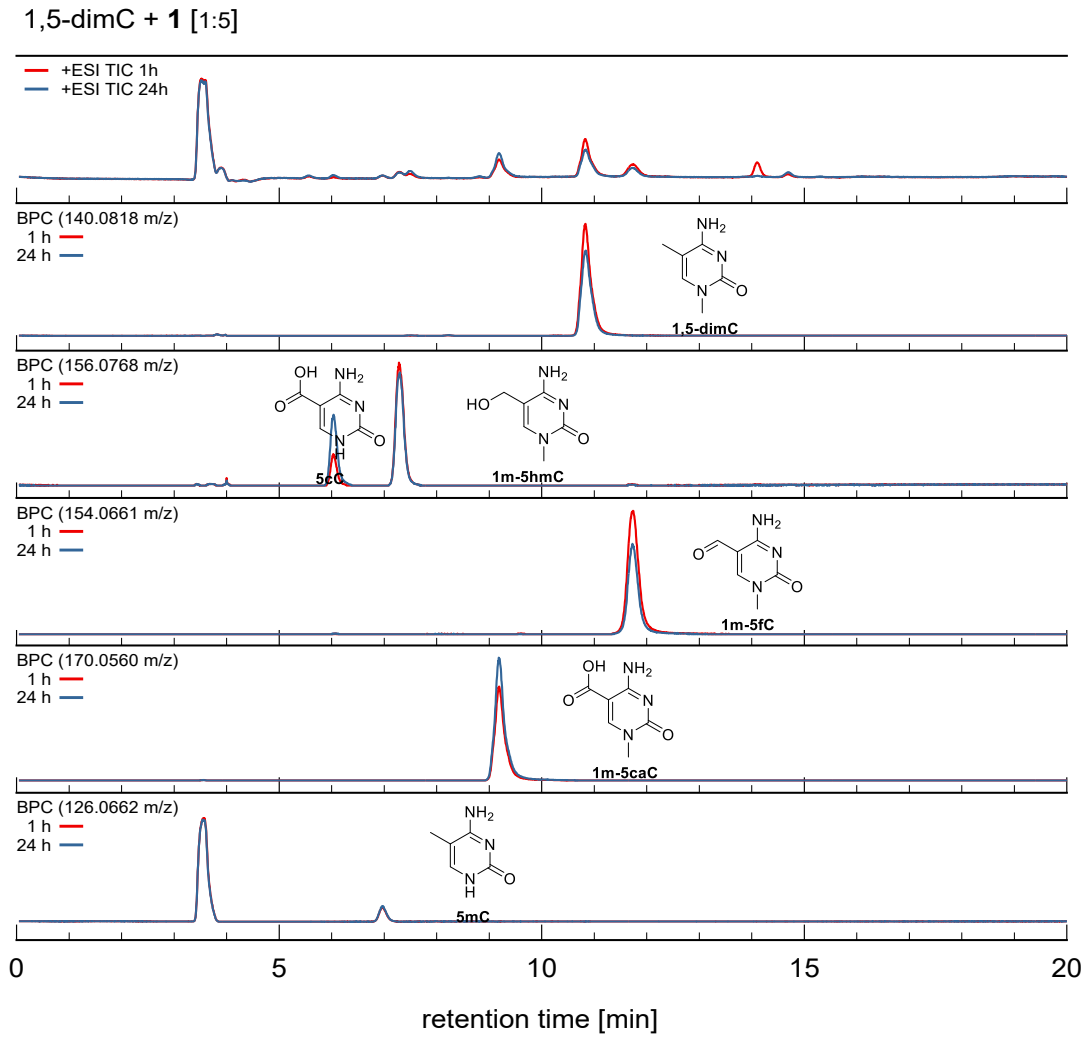


Figure S11: Excerpt of the HPLC-MS traces of the reaction of **1** with 1,5-dimC. Conditions: [**1**] = 5 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted base peak chromatograms (BPC) for 1,5-dimC (140.0818 m/z), 1m-5hmC (156.0768 m/z) + 5caC (156.0404 m/z), 1m-5fC (154.0661), 1m-5caC (170.0560 m/z), 5mC (126.0662 m/z) after 1 h and 24 h.

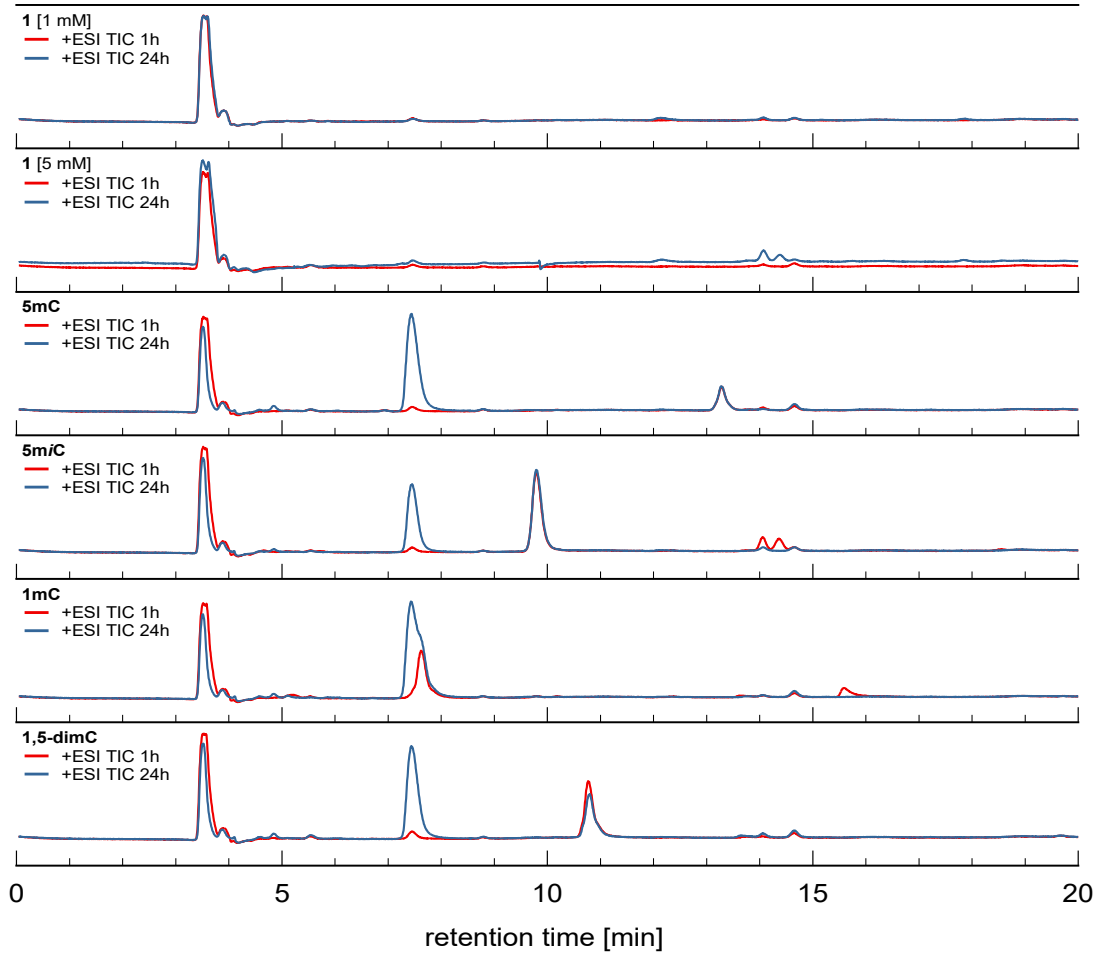


Figure S12: Displayed are the total ion chromatograms (TIC) of control reactions of 5mC, 5miC, 1mC, 1,5-dimC and **1** after 1 h and 24 h. Conditions: [S] = 1 mM, H_2O, T = 30 °C. No significant changes were observed. The signal at 7.4 min occurs independent of the substrate and is therefore considered an impurity originating from the column.

LC-LRMS Traces

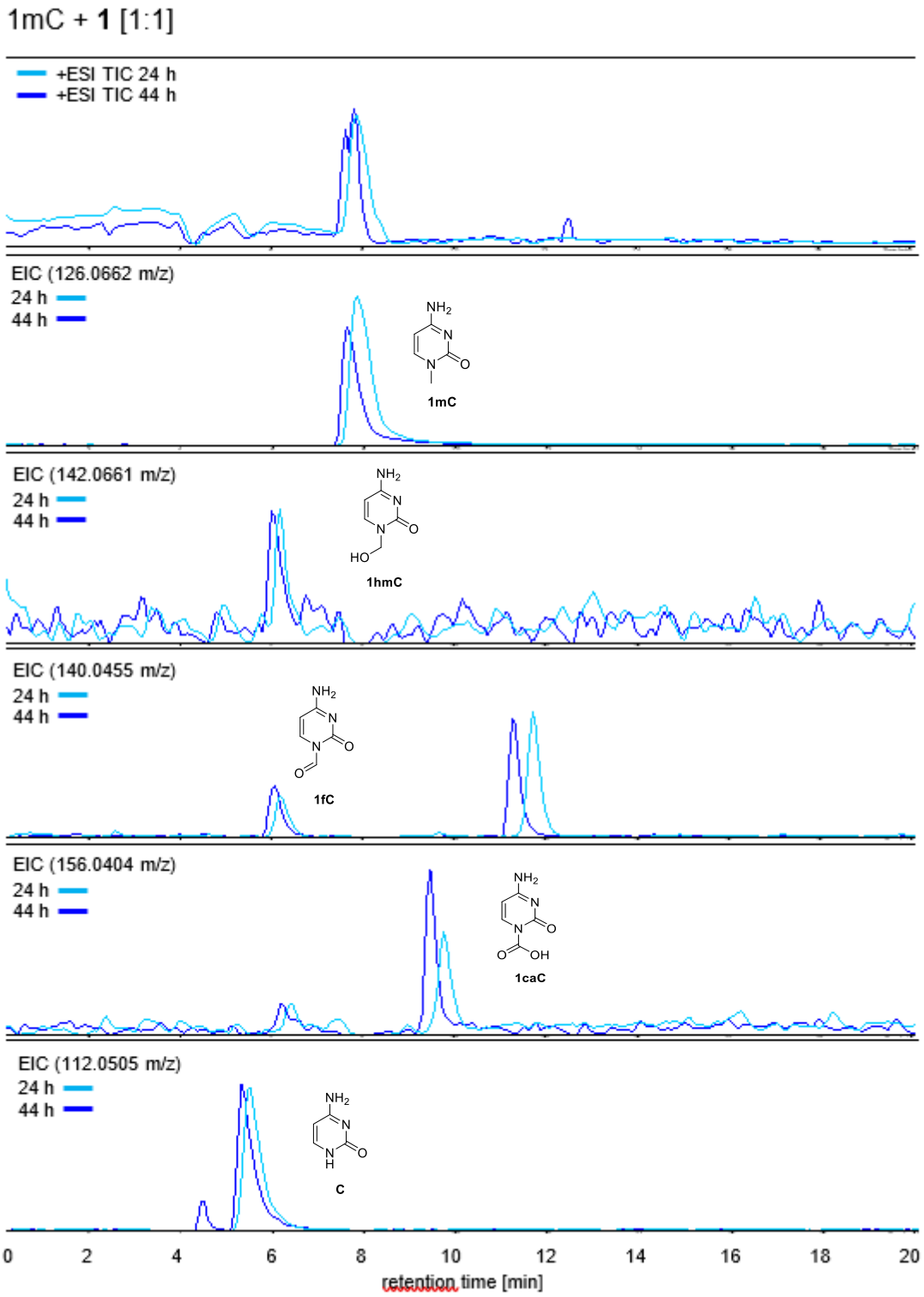


Figure S13: Excerpt of the LC-MS traces of the reaction of **1** with 1mC. Conditions: [**1**] = 1 mM; [S] = 1 mM, H_2O, T = 30 °C. Displayed are the total ion chromatogram (TIC) and extracted ion chromatograms (EIC) for 1mC (126.0662 m/z), 1hmC (142.0611 m/z), 1fC (140.0455 m/z), 1cC (156.0404 m/z) and C (112.0505 m/z) after 24 h and 44 h. The additional signal in the EIC of 140.0455 m/z was also found in control reactions (Figure S14) and is therefore not originated from reaction with **1.**

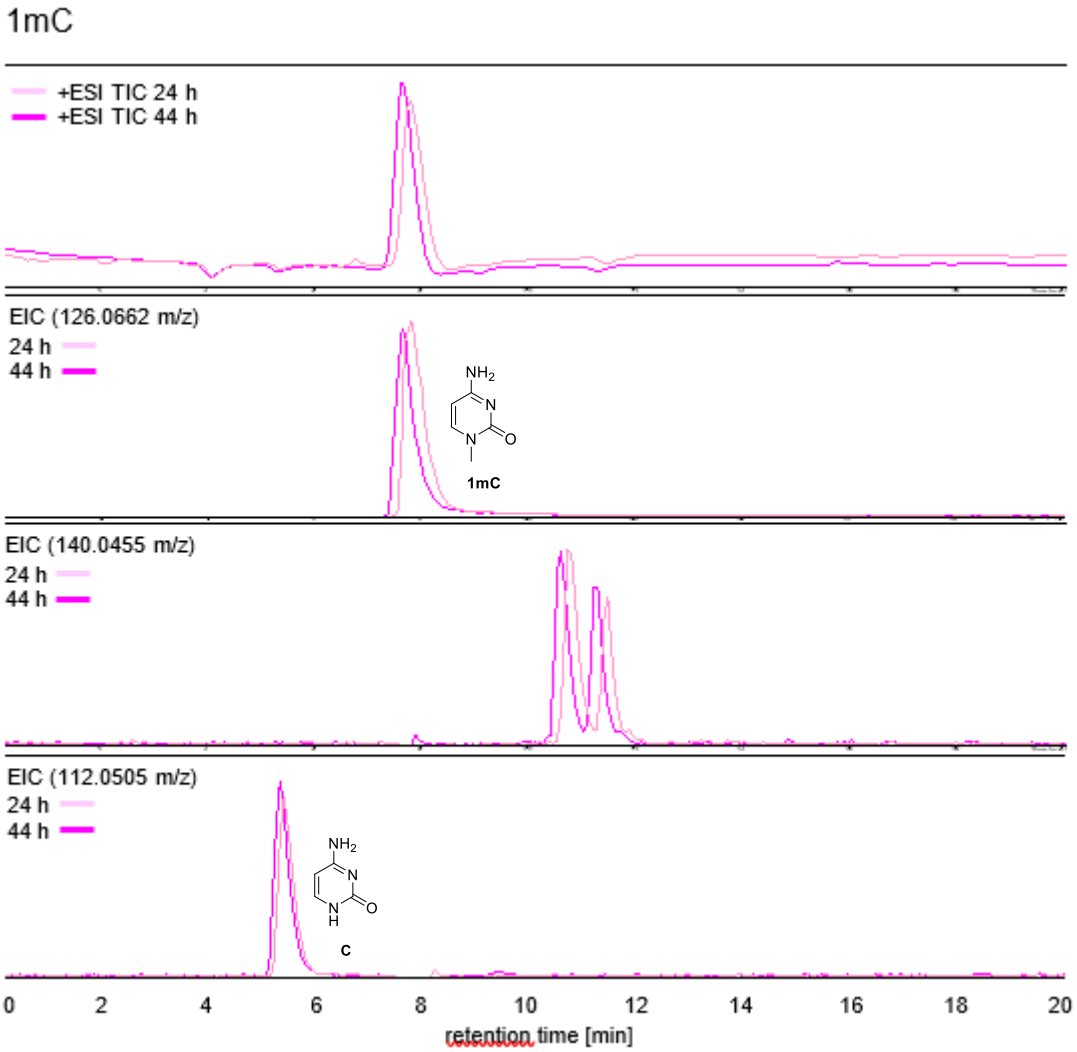


Figure S14: Displayed are the total ion chromatograms (TIC) of a control reaction of 1mC without **1** after 24 h and 44 h. Conditions: [S] = 1 mM, H_2O, T = 30 °C. Additionally, extracted ion chromatograms (EIC) of 1mC (126.0662 m/z), cytosine (112.0505 m/z) and 140.0455 m/z are shown. Cytosine was found as a minor impurity stemming from synthesis of 1mC, but no significant changes in intensity were observed. The EIC of 140.0455 m/z shows an impurity also observed in reaction with **1** (Figure S13) which is considered an impurity originating from the starting material.

GC-MS Traces

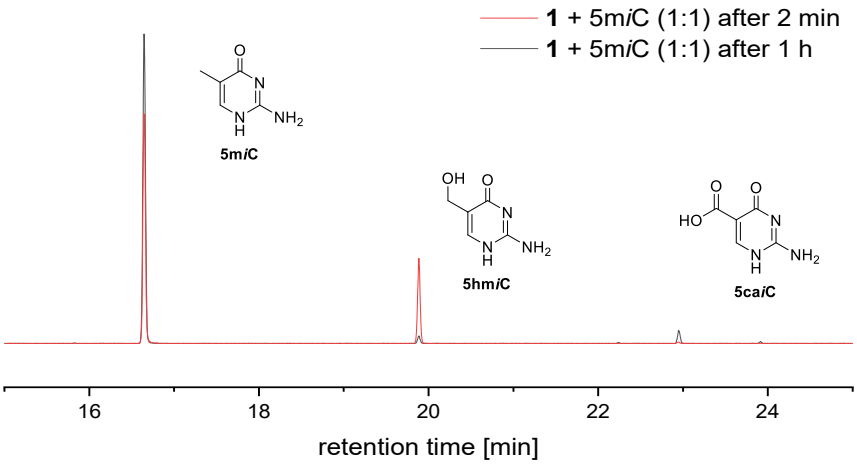


Figure S15: Excerpt of the GC-MS traces of reactions of **1** with 5miC after 1 h or 2 min. 5hmiC and 5caiC were observed in addition to the starting material. Conditions: [**1**] = 1 mM; [5miC] = 1 mM, H_2O, T = 30 °C.

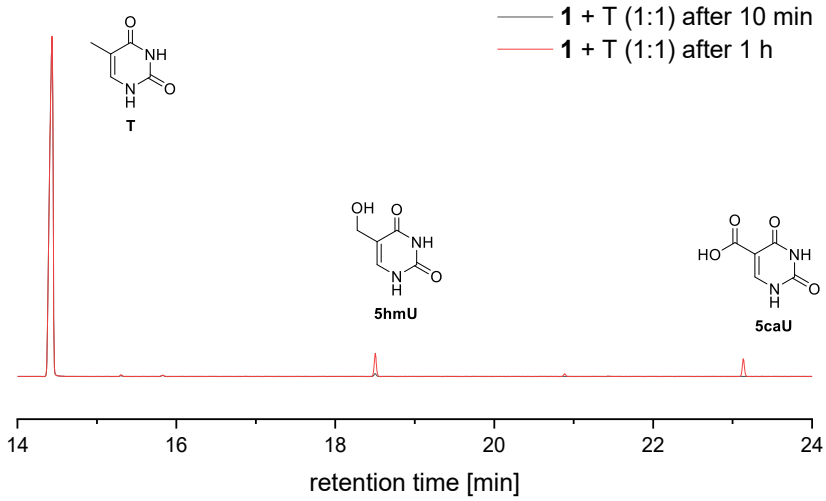


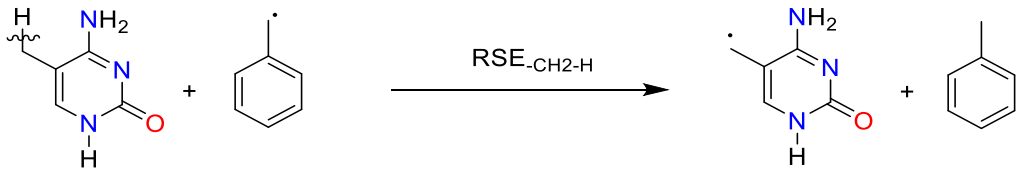
Figure S16: Excerpt of the GC-MS traces of reactions of **1** with T after 10 min or 1 h. 5hmU and 5caU were observed in addition to the starting material. Conditions: [**1**] = 1 mM; [T] = 1 mM, H_2O, T = 30 °C.

5.2 Bond Dissociation Energy Calculations

Methodology

As in previous studies on radical stabilities, [2-8] geometry optimizations have been performed with a combination of the (U)B3LYP hybrid functional [9] complemented by the D3 dispersion correction [10] and the 6-31+G(d,p) basis set [11-12] in the gas phase. Thermochemical corrections (corr. Δ*H* & Δ*G*) to 298.15 K have been calculated at the same level of theory using the rigid rotor/harmonic oscillator model. Enthalpies (ΔH_{298}) and Gibbs energies (ΔG_{298}) at (U)B3LYP-D3/6-31+G(d,p) level have been obtained through addition of corr. ΔG and corr. ΔH to ΔE_{tot}, respectively. Single point energies have subsequently been calculated with the DLPNO-CCSD(T) method [13-14] as implemented in ORCA 4.2.1 [15] in combination with the cc-pVTZ and cc-pVQZ basis sets, followed by extrapolation to the complete basis set (CBS) limit to DLPNO-CCSD(T)/CBS total energies. [16] Enthalpies at 298.15 K have subsequently been calculated through combination of the DLPNO-CCSD(T)/CBS total energies with thermochemical corrections (Δ*H*) calculated at the (U)B3LYP-D3/6-31+G(d,p) level before. In order to estimate the impact of aqueous solvation, single point energies have been calculated using the SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) level using the gas phase-optimized geometries obtained before [17]. Solvation free energies in water (ΔG_{solv}) have then been obtained as the difference of total energies obtained from SMD(H_2O)/(U)B3LYP-D3/6-31+G(d,p) and (U)B3LYP-D3/6-31+G(d,p) calculations. Combination of the solvation energies with gas phase enthalpies then yield enthalpy values in water.

Radical stabilization energies (RSE) for C-centered radicals generated by C-H homolytic bond cleavage are measured with reference to toluene (Ph-CH_2-H) using the isodesmic hydrogen exchange reaction shown in eq. S1. R-CH_2-H bond dissociation energies (BDE) can then be derived from the calculated RSE values through the addition of the reference BDE(C-H) value in Ph-CH_2-H (375.5 ± 5 kJ/mol^{-1}) [18] as expressed in eq. S2. In the subsequent discussion we assume this value to be valid in the gas phase as well as in aqueous solution.



Eq. S1

$$\mathrm{BDE}_{\mathrm{RCH_2-H}} = \mathrm{RSE}_{\mathrm{RCH_2-H}} + \mathrm{BDE}^{\mathrm{Exp}}_{\mathrm{RCH_2-H}} \qquad \textbf{Eq. S2}$$

In Figure S17 the methylated nucleobases under investigation are depicted. The number indicates the location of the methyl group. The prefix “n” indicates a neutral molecule. For radicals, “r” precedes the methyl-location marker.

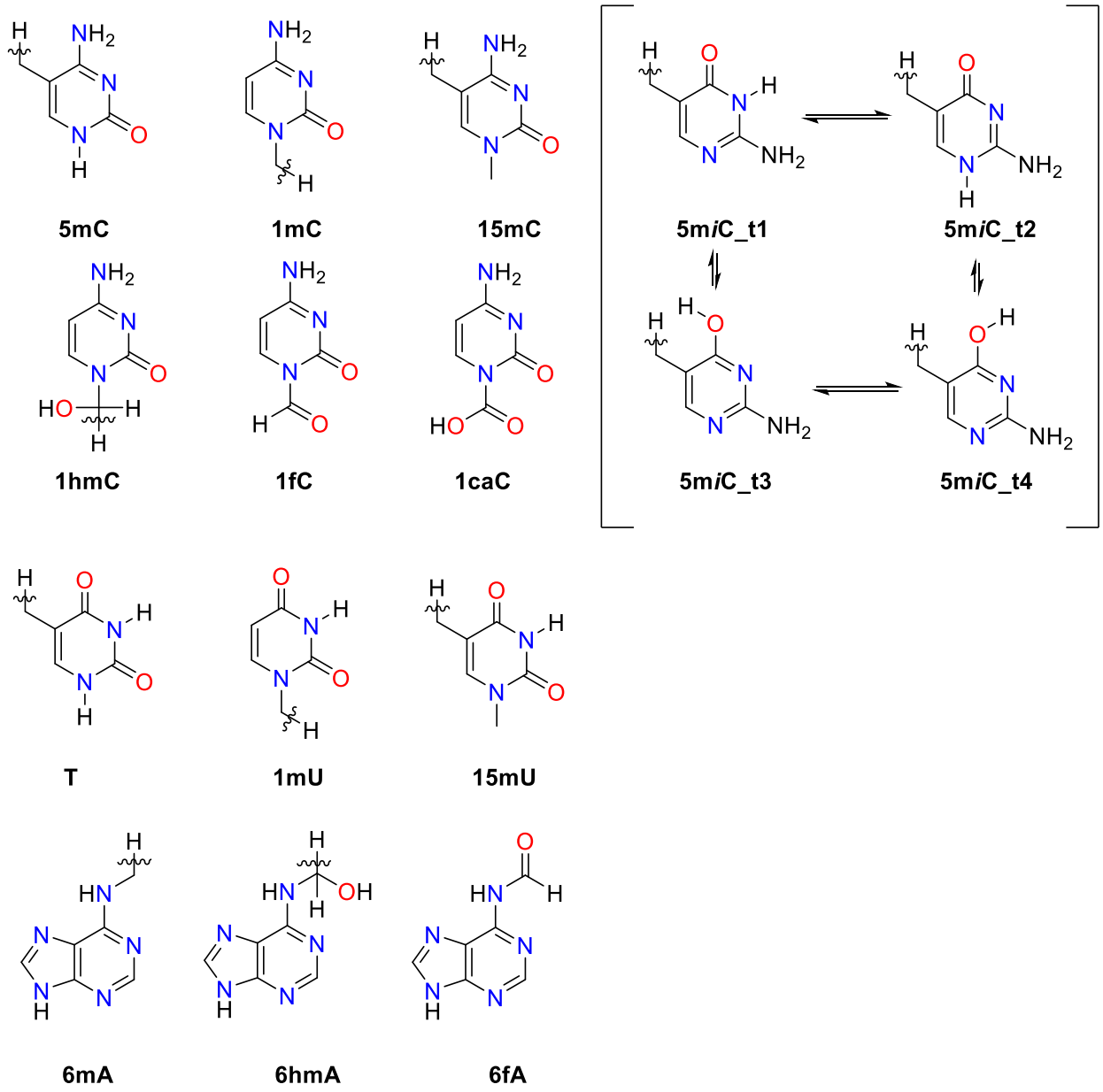


Figure S17: Naming of methylated nucleobases under investigation (n_ =neutral, r_ = radical).

Calculations of BDE and RSE values:

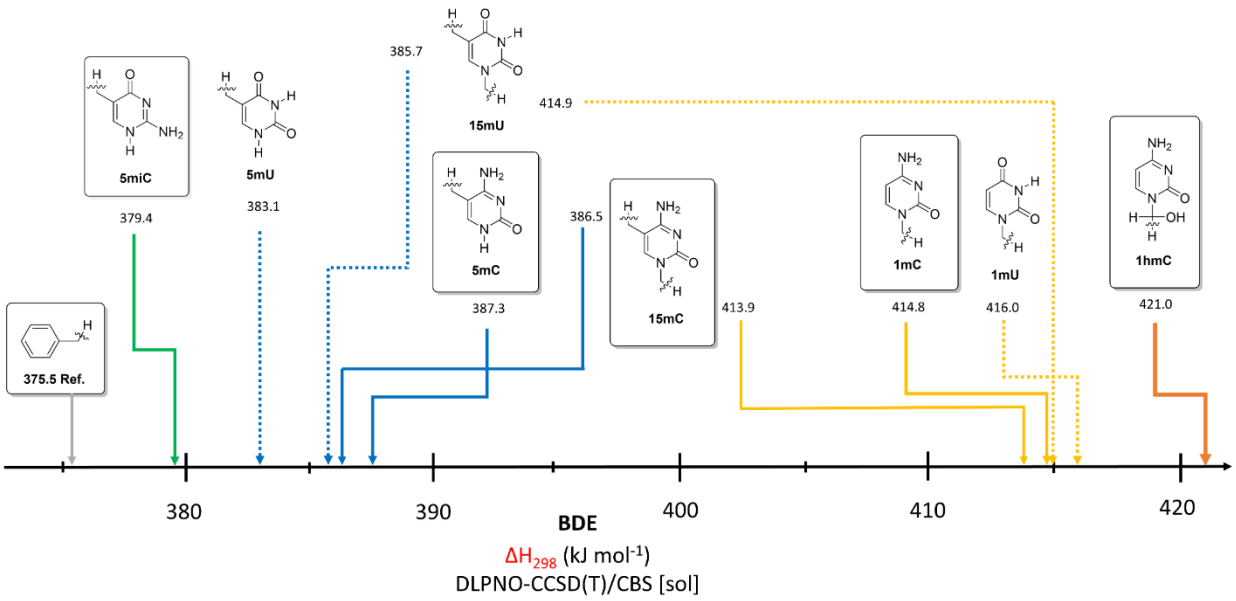


Figure S18: Aqueous phase ($\Delta H_{sol} = \Delta H_{298} + \Delta G_{solv}$) RCH_2-H bond dissociation energies (BDEs) calculated at the DLPNO-CCSD(T)/CBS level of theory.

In Figure S18 and Table S2, the RSE and BDE values of modified cytosine nucleobases in aqueous phase with toluene as a reference compound (with BDE(C-H) = +375.5 kJ mol^{-1}) can be seen. The lowest C-H bond energies can be found for the methyl groups attached to the C5 position, the value for isocytosine 5m*i*C at BDE(C-H) = +379.4 kJ mol^{-1} being slightly lower than for 5-methylcytosine or 1,5-dimethylcytosine (both around BDE(C-H) ≈ +387 kJ mol^{-1}). C-H Bond energies for the methyl group attached to the N1 position are approximately 24 kJ mol^{-1} stronger (as can be seen in 1-methylcytosine (BDE(C-H) = +414.8 kJ mol^{-1}) or 1,5-dimethylcytosine). C-H Bond energies for the uracil derivatives shown in Figure S16 are closely similar to those in the respective cytosine derivatives (see Figure S18). When moving from the gas phase to aqueous solution a general increase can be noted for all C-H bond energy values in Table S2, the effect being larger for the methyl substituents attached to the *N1*-position (with ΔBDE(C-H) ca. 11 kJ mol^{-1}) as compared to the *C5* methyl substituents (with only half this value). Analysis of the individual solvation free energies for all species in isodesmic equation (1) shows this to be the result of systematically larger solvation free energies of the closed shell parents as compared to the respective product radicals.

The values obtained here for 5mC are largely similar to those reported by Hu et al. at several different levels in the gas phase (BDE(5mC,CBS-QB3) = +375.5 kJ/mol and BDE(5mC,G3(MP2)B3 = +380.8 kJ/mol) as well as in aqueous solution ((BDE(5mC,CPCM/CBS-QB3) = +379.2 kJ/mol) [19].

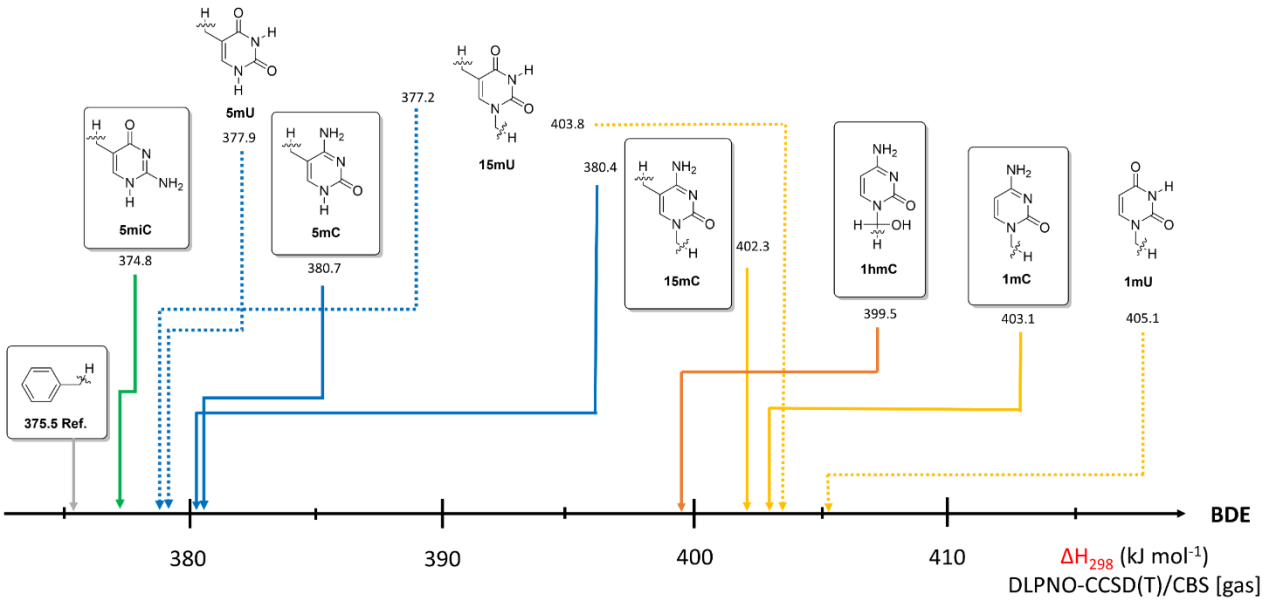


Figure S19: Gas phase (ΔH_{298}) R-CH_2-H bond dissociation energies (BDEs) calculated at the DLPNO-CCSD(T)/CBS level of theory.

Table S2: Gas- and solution phase RSE und BDE values with toluene as reference for the systems shown in Figure S17.

compound	gas phase				water (SMD model)				$\Delta\Delta G_{solv}$ [kJ/mol]
	(U)B3LYP-D3/ 6-31+G(d,p)[a]		DLPNO-CCSD(T)/ CBS[a]		(U)B3LYP-D3/ 6-31+G(d,p)[a]		DLPNO-CCSD(T)/ CBS[a]		
	RSE - ΔH_{298} [kJ/mol]	BDE - ΔH_{298} [kJ/mol]	RSE - ΔH_{298} [kJ/mol]	BDE - ΔH_{298} [kJ/mol]	RSE - ΔH_{298} [kJ/mol]	BDE - ΔH_{298} [kJ/mol]	RSE - ΔH_{298} [kJ/mol]	BDE - ΔH_{298} [kJ/mol]	
toluene		**375.5**[b]							
5m*i*C_t1	-6.3	369.2	-0.7	374.8	-1.7	373.8	3.9	**379.4**	4.6
5m*i*C_t2[c]	-2.0	373.5	4.4	379.9	1.0	376.5	7.4	**382.9**	3.0
T	-2.5	373.0	2.4	377.9	2.8	378.3	7.6	**383.1**	5.3
1,5dimU@5	-3.5	372.0	1.7	377.2	5.0	380.5	10.2	**385.7**	8.4
1,5dimC@5	-1.2	374.3	4.9	380.4	4.9	380.4	11.0	**386.5**	6.1
5mC	-0.5	375.0	5.2	380.7	6.1	381.6	11.8	**387.3**	6.6
1,5dimC@1	19.0	394.5	26.8	402.3	30.6	406.1	38.4	**413.9**	11.6
1mC	20.0	395.5	27.6	**403.1**	31.7	407.2	39.3	**414.8**	11.7
1,5dimU@1	24.0	399.5	28.3	403.8	35.1	410.6	39.4	**414.9**	11.1
1mU	25.5	401.0	29.6	405.1	36.4	411.9	40.5	**416.0**	11.0
1hmC	3.8	379.26	24.0	**399.5**	25.3	400.8	45.5	**421.0**	21.5
6mA	6.5	381.98	13.9	389.4	13.5	389.0	20.9	**396.4**	7.0
6hmA	2.3	377.76	19.1	394.6	16.3	391.8	33.2	**408.7**	14.1

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries; [b] Reference value: BDE(Ph-CH_2-H) = 375.5 ± 5 kJ/mol^{-1}; [c] the population of tautomer n_5miC_t2 amounts to 17% (solution free energy Boltzmann-weighted, see Table S9); [d] most stable Boltzmann-weighted conformer (see Table S10).

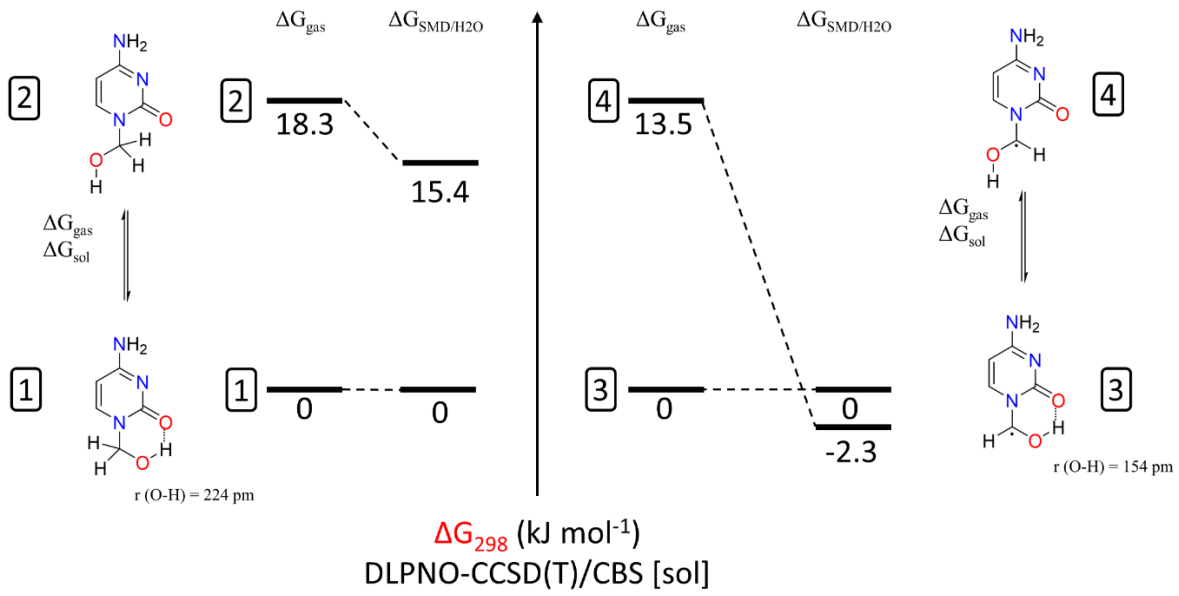


Figure S20: Comparison of solvation effects of closed shell (left) and open shell (right) conformers of 1hmC calculated at the DLPNO-CCSD(T)/CBS level of theory.

When looking at the bond dissociation energies of 1hmC (Table S2) a significant increase can be noted comparing gas to solution phase models (BDE(C-H, DLPNO) = +399.5 kJ mol^{-1} vs. BDE(C-H, DLPNO, H_2O) = +421.0 kJ mol^{-1}). This change in energies can be attributed to the differences in free energy of solvation ΔG_{solv} of the lowest closed shell (Figure S20, [1] + [2]) as well as open shell (Figure S20, [3] + [4]) conformers. This may lead to a higher predictive quality of the gas phase BDE values as compared to the solution phase energies obtained with aid of the SMD model.

For *N6*-methylated adenine the BDE values increase when going from 6mA (BDE(C-H) = +396.4 kJ mol^{-1}) to 6hmA (BDE(C-H) = +408.7.4 kJ mol^{-1}). The BDE value for 6mA is approximately 20 kJ mol^{-1} lower than for 1mC (see Table S2).

Reactivity of the N-methylated compounds

In Figure S21 [A], the deformylation reaction of *N1*-hydroxymethylcytosine (1hmC) to yield cytosine and formaldehyde is depicted. Systematic studies of the conformational space of 1hmC identified a single low energy conformation characterized by an internal hydrogen bond between the hydroxyl group hydrogen atom and the neighboring carbonyl oxygen atom. Formaldehyde generated in water will, of course, react further to the respective hydrate. Combining both steps into a single process as shown in Figure S21 [B] yields slightly different energetics (ΔG_{298}(Figure S21 [B]) = +24.5 kJ/mol) as compared to the deformylation alone (ΔG_{298}(Figure S21 [A]) = +24.9 kJ/mol, Table S3). Both variants thus indicate that the deformylation process is endergonic. Under oxidative conditions 1hmC may react further to 1-formylcytosine (fC). The hydrolytic deformylation of this latter species as shown in Figure S21 [C] has therefore also been studied. This reaction is actually exergonic with ΔG_{298}(Figure S21 [C]) = - 17.0 kJ/mol. The most exergonic reaction with - 44.9 kJ/mol is the decarboxylation of 1caC towards cytosine and carbon dioxide shown in Figure S21 [D].

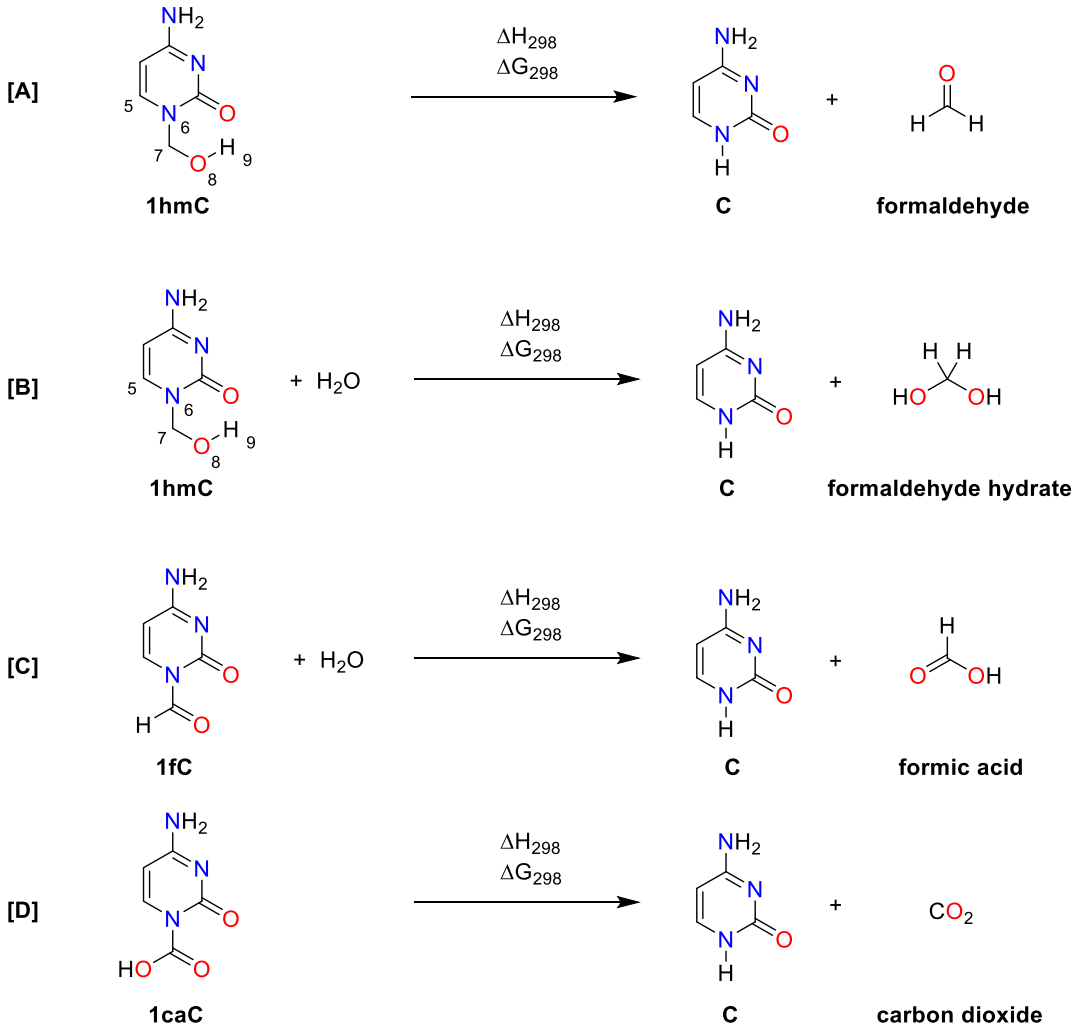


Figure S21: Reaction energies ΔH_{298} and ΔG_{298} for the deformylation of N1-hydroxymethyl cytosine to cytosine ([A]: direct, [B]: water-mediated), [C] for the water-mediated decarboxylation of N1-formyl cytosine to cytosine and formic acid and [D] for the direct decarboxylation of 1caC to cytosine and CO_2.

Table S3: Gas- and solution phase reaction energies ΔE_{tot}, ΔH_{298} and ΔG_{298} for the deformylation and decarboxylation of N1-modified cytosine derivatives shown in Figure S21.

	gas phase		**water (SMD model)**	
Figure 21 [A]				
	B3LYP-D3/ 6-31+G(d,p)[a]	**DLPNO-CCSD(T)/ CBS[a]**	**B3LYP-D3/ 6-31+G(d,p)[a]**	**DLPNO-CCSD(T)/ CBS[a]**
ΔE_{tot}	+85.3	+93.9	+75.6 +73.8[c]	
ΔH_{298}	+71.2	+79.8	+69.4[b]	**+78.0[b]**
ΔG_{298}	+18.1	+26.7	+16.3[b] +16.3[b,c]	**+24.9[b]**
Figure 21 [B]				
ΔE_{tot}	+22.5	+32.9	+18.8	
ΔH_{298}	+25.3	+35.8	+21.6[b]	**+32.1[b]**
ΔG_{298}	+17.7	+28.2	+14.0[b]	**+24.5[b]**
Figure 21 [C]				
ΔE_{tot}	-18.9	-11.3	-23.5	-15.9
ΔH_{298}	-14.4	-6.8	-19.0[b]	**-11.4[b]**
ΔG_{298}	-20.1	-12.4	-24.6[b]	**-17.0[b]**
Figure 21 [D]				
ΔE_{298}	6.6	3.9	10.4	
ΔH_{298}	3.3	0.6	7.1[b]	**4.3[b]**
ΔG_{298}	-45.9	-48.7	-42.2[b]	**-44.9[b]**

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries; [b] including standard state correction $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M}$ = +7.91 kJ mol^{-1}. [c] Using geometries optimized at the SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) level of theory.

While no experimental results appear to exist for the deformylation of 1-hydroxymethylcytosine, this is different for the conceptually similar deformylation of *N6*-hydroxymethyladenine (6hmA). This reaction may again be described as the deformylation alone (Figure S22 [A]) or as the combined deformlyation/ hydration reaction of formaldehyde (Figure S22 [B]). In both variants the reaction in water is slightly endergonic (⍰G_{298}(Figure S22 [A]) = +16.9 kJ/mol vs. ⍰G_{298}(Figure S22 [B]) = +16.5 kJ/mol, Table S4). The hydrolysis reaction of *N6*-formyladenine (6fA) formed as a possible by-product under oxidative conditions is, in contrast, practically thermoneutral (Figure S22 [C], Table S4).

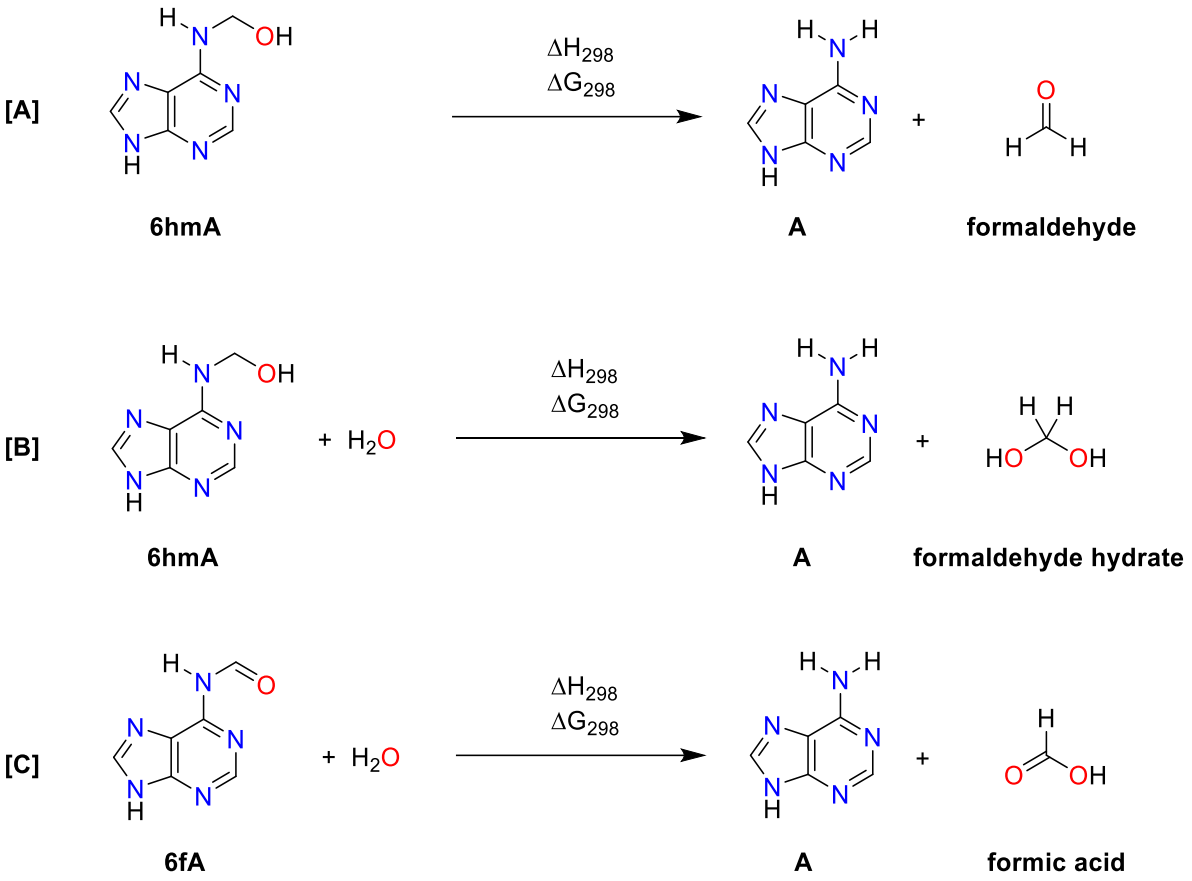


Figure S22: Reaction energies ΔH_{298} and ΔG_{298} for the deformylation of N6-hydroxymethyladenine to adenine ([A]: direct, [B]: water-mediated), and [C] for the water-mediated decarboxylation of N6-formyladenine to adenine and formic acid.

Table S4: Gas- and solution phase reaction energies ΔE_{tot}, ΔH_{298} and ΔG_{298} for the deformylation of N6-hydroxymethyladenine shown in Figure S22.

	gas phase		**water (SMD model)**	
Figure S22 [A]				
	(U)B3LYP-D3/ 6-31+G(d,p)[a]	**DLPNO-CCSD(T)/ CBS[a]**	**(U)B3LYP-D3/ 6-31+G(d,p)[a]**	**DLPNO-CCSD(T)/ CBS[a]**
ΔE_{tot}	+76.1	+83.8	+82.0	
ΔH_{298}	+60.6	+68.3	+66.5	**+74.1**
ΔG_{298}	+3.4	+11.0	+9.2	**+16.9**
Figure S22[B]				
ΔE_{tot}	+13.2	+22.8	+17.2	
ΔH_{298}	+14.7	+24.3	+18.7	**+28.3**
ΔG_{298}	+3.0	+12.6	+6.9	**+16.5**
Figure S22[C]				
ΔE_{tot}	-9.5	-4.1	+4.2	
ΔH_{298}	-7.1	-1.8	+6.6	**+11.9**
ΔG_{298}	-17.9	-12.6	-4.2	**+1.1**
[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries				

Thermodynamic Data of Calculated Compounds

Table S5: Gas- and solution phase RSE values with various methods with toluene as the reference for the systems shown in Figure S17. All gas phase optimized conformers were calculated at the (U)B3LYP-D3\6-31+G(d,p) and DLPNO-CCSD(T) level of theory.

compound	gasphase				water (SMD model)			
	RSE - ΔH_{298} [kJ/mol]				RSE - ΔH_{298} [kJ/mol]			
	(U)B3LYP-D3/ 6-31+G(d,p)[a]	DLPNO-CCSD(T)/ cc-pVTZ[a]	DLPNO-CCSD(T)/ cc-pVQZ[a]	DLPNO-CCSD(T)/ CBS[a]	(U)B3LYP-D3/ 6-31+G(d,p)[a]	DLPNO-CCSD(T)/ cc-pVTZ[a]	DLPNO-CCSD(T)/ cc-pVQZ[a]	DLPNO-CCSD(T)/ CBS[a]
toluene	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
5mC	-0.5	4.4	4.9	5.2	6.1	11.0	11.5	11.8
1mC	20.0	26.6	27.2	27.6	31.7	38.3	38.9	39.3
5m*i*C_t1	-6.3	-1.3	-0.9	-0.7	-1.7	3.4	3.7	3.9
5m*i*C_t2	-2.0	3.6	4.1	4.4	1.0	6.7	7.1	7.4
1,5dimC@5	-1.2	4.4	4.7	4.9	4.9	10.4	10.7	11.0
1,5dimC@1	19.0	25.9	26.4	26.8	30.6	37.5	38.0	38.4
T	-2.5	1.6	2.1	2.4	2.8	6.8	7.4	7.6
1mU	25.5	28.5	29.1	29.6	36.4	39.5	40.1	40.5
1,5dimU@5	-3.5	1.1	1.6	1.7	5.0	9.6	10.0	10.2
1,5dimU@1	24.0	27.6	28.0	28.3	35.1	38.6	39.1	39.4
1hmC	3.8	21.0	22.8	24.0	25.3	42.5	44.3	45.5
1fC	44.3	43.2	48.5	45.8	48.1	47.0	48.5	49.6
6mA	6.5	12.8	13.4	13.9	13.5	19.8	20.5	20.9
6hmA	2.3	17.9	18.5	19.1	16.3	31.9	32.5	33.2
6fA[c]	31.8	32.2	33.8	35.0	41.3	41.6	43.3	44.5

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries; [b] Reference value: BDE(Ph-CH_2-H) = 375.5 ± 5 kJ/mol^{-1}; [c] most stable Boltzmann-weighted conformer (see Table 9).

Table S6: Gas- and solution phase BDE values with various methods with toluene as reference for the systems shown in Figure S17. All gas phase optimized conformers were calculated at the (U)B3LYP-D3\6-31G+(d,p) and DLPNO-CCSD(T) level of theory.

compound	gasphase				water (SMD model)			
	BDE - ΔH_{298} [kJ/mol]				BDE - ΔH_{298} [kJ/mol]			
	(U)B3LYP-D3/ 6-31+G(d,p)[a]	DLPNO-CCSD(T)/ cc-pVTZ[a]	DLPNO-CCSD(T)/ cc-pVQZ[a]	DLPNO-CCSD(T)/ CBS[a]	(U)B3LYP-D3/ 6-31+G(d,p)[a]	DLPNO-CCSD(T)/ cc-pVTZ[a]	DLPNO-CCSD(T)/ cc-pVQZ[a]	DLPNO-CCSD(T)/ CBS[a]
toluene	**375.5[b]**	**375.5[b]**	**375.5[b]**	**375.5[b]**				
5mC	375.0	379.9	380.4	380.7	381.6	386.5	387.0	387.3
1mC	395.5	402.1	402.7	403.1	407.2	413.8	414.4	414.8
5m*i*C_t1	369.2	374.2	374.6	374.8	373.8	378.9	379.2	379.4
5m*i*C_t2	373.5	379.1	379.6	379.9	376.5	382.2	382.6	382.9
1,5dimC@5	374.3	379.9	380.2	380.4	380.4	385.9	386.2	386.5
1,5dimC@1	394.5	401.4	401.9	402.3	406.1	413.0	413.5	413.9
T	372.97	377.07	377.60	377.9	378.3	382.3	382.9	383.1
1mU	400.98	404.00	404.62	405.1	411.9	415.0	415.6	416.0
1,5dimU@5	372.05	376.64	377.10	377.2	380.5	385.1	385.5	385.7

1,5dimU@1	399.54	403.05	403.51	403.8	410.6	414.1	414.6	414.9
1hmC	379.26	396.47	398.27	399.5	400.8	418.0	419.8	421.0
1fC	419.82	418.70	424.00	421.3	423.6	422.5	424.0	425.1
6mA	381.98	388.27	388.91	389.4	389.0	395.3	396.0	396.4
6hmA	377.76	393.39	393.99	394.6	391.8	407.4	408.0	408.7
6fA[c]	407.34	407.66	409.32	410.5	416.8	417.1	418.8	420.0

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries; [b] Reference value: BDE(Ph-CH_2-H) = 375.5 ± 5 kJ/mol^{-1}; [c] most stable Boltzmann-weighted conformer (see Table 9).

Table S7: QM properties for gas phase optimized conformers shown in Figure S17 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory (n = neutral molecule, r = radical, position of r indicates location of radical, t = tautomer, c = conformer).

compound	**E_{tot}[a]** (UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (SMD(H2O)/ UB3LYP-D3/ 6-31+G(d,p))	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVTZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ cc-pVQZ)	**E_{tot}[a]** (DLPNO-CCSD(T)/ CBS)	**corr. ΔH[a]** (UB3LYP-D3/ 6-31+G(d,p))	**corr. ΔG[a,b]** (UB3LYP-D3/ 6-31+G(d,p))
	[Hartree]					**[Hartree]**	
Tol	-271.597084	-271.597992	-271.046063	-271.123446	-271.171726	0.134933	0.097272
r_tol	-270.945089	-270.945771	-270.392891	-270.469013	-270.516453	0.121248	0.084945
5m_C	-434.294327	-434.327258	-433.536803	-433.667983	-433.748906	0.135443	0.093389
r5m_C	-433.642504	-433.672686	-432.881950	-433.011683	-433.091648	0.121745	0.079639
n_1m_C	-434.285803	-434.315437	-433.528525	-433.659206	-433.739823	0.135374	0.093144
n_r1m_C	-433.625393	-433.650346	-432.864432	-432.993636	-433.073263	0.120909	0.078726
n_15m_C	-473.609537	-473.638972	-472.771605	-472.913874	-473.001785	0.164885	0.119476
n_1r5m_C	-472.957993	-472.984891	-472.116776	-472.257665	-472.344651	0.151202	0.105495
n_r15m_C	-472.949558	-472.974344	-472.107810	-472.248624	-472.335557	0.150450	0.105252
n_5m_U	-454.180586	-454.203334	-453.415551	-453.552361	-453.636830	0.123522	0.082190
n_r5m_U	-453.529634	-453.550145	-452.761863	-452.897208	-452.980739	0.109918	0.068962
n_1m_U	-454.169547	-454.190793	-453.404337	-453.540848	-453.625093	0.123555	0.081618
n_r1m_U	-453.507105	-453.523948	-452.739567	-452.874580	-452.957818	0.109128	0.067870
n_15m_U	-493.495205	-493.515149	-492.649360	-492.797353	-492.888854	0.152978	0.106590
n_1r5m_U	-492.844598	-492.861109	-491.995827	-492.142384	-492.232990	0.139366	0.093614
n_r15m_U	-492.833316	-492.848818	-491.984957	-492.131512	-492.222046	0.138555	0.093558
n_1f_C_2	-508.305717	-508.332655	-507.4569442	-507.609235	-507.7030437	0.117101	0.073646
n_r1f_C_2	-507.637693	-507.662950	-506.7881728	-506.938636	-507.031181	0.104269	0.060095
n_1hmC	-509.514881	-509.547802	-508.661390	-508.816566	-508.912162	0.141795	0.097717
n_r1hm_C	-508.859249	-508.883761	-507.998030	-508.151257	-508.245546	0.125907	0.081323
5m*i*C							
n_5m_iC_t4	-434.296548	-434.314366	-433.542192	-433.672556	-433.753129	0.135393	0.093646
n_5m_iC_t1	-434.296536	-434.322439	-433.539711	-433.670291	-433.750971	0.135566	0.093797

n_5m_iC_t3	-434.288426	-434.313036	-433.534860	-433.665431	-433.746058	0.135135	0.093434
n_5m_iC_t2	-434.281176	-434.318129	-433.5261508	-433.657121	-433.737909	0.135337	0.093294
r5m*i*C							
n_r5m_iC_t1	-433.647007	-433.670924	-432.887099	-433.016270	-433.096041	0.121957	0.080559
n_r5m_iC_t2	-433.629947	-433.665523	-432.871587	-433.001112	-433.080959	0.121647	0.079768
1hmC							
1hmC_C1	-509.514881	-509.547802 -509.549955[c]	-508.661390	-508.816566	-508.912162	0.141795 0.142116[c]	0.097717 0.098326[c]
1hmC_C2	-509.505305	-509.539323 -509.541509[c]	/	/	/	0.141277	0.095281
1hmC_C3	-	-509.545282[c]	/	/	/	0.141906[c]	0.096865[c]
n_6mA							
n_6mA_C2	-506.6721266	-506.6967031	-505.768023	-505.9169157	-506.0090917	0.1500520	0.1054460
n_6mA_C1	-506.6743398	-506.6994632	-505.770119	-505.9192773	-506.0115603	0.1501280	0.1053670
n_r6mA_C2	-506.0175025	-506.0391695	-505.109827	-505.2572139	-505.3483743	0.1362070	0.0930140
n_r6mA_C1	-506.0182143	-506.0408057	-505.110575	-505.2581748	-505.3494131	0.1362500	0.0928040
r_6hmA							
n_r6hmA_C1	-581.2496770	-581.2743833	-580.242637	-580.4146234	-580.5207140	0.1423400	0.0967280
n_r6hmA_C6	-581.241774	-581.270106	-580.236138	-580.4082863	-580.5144529	0.142478	0.096222
n_r6hmA_C4	-581.249206	-581.274336	-580.241469	-580.4133400	-580.5193843	0.142379	0.097263
6hmA							
n_6hmA_C1	-581.9030375	-581.9333208	-580.903123	-581.0766011	-581.1837729	0.1565280	0.1110180
n_6hmA_C3	-581.8967977	-581.9324302	-580.897064	-581.0706722	-581.1778838	0.1562520	0.1099090
n_6hmA_C4	-581.8960418	-581.9310750	-580.896397	-581.0698447	-581.1769964	0.1562610	0.1102790
n_6hmA_C6	-581.9040620	-581.9340549	-580.903808	-581.0770723	-581.1841647	0.1565160	0.1116760
n_6fA							
n_6fA_C1	-580.7009735	-580.7293299	-579.704925	-579.8757152	-579.9812086	0.1321100	0.0876180
n_6fA_C2	-580.6911846	-580.7254693	-579.695525	-579.8662439	-579.9716500	0.1318830	0.0869790
n_6fA_C3	-580.6898778	-580.7234146	-579.694851	-579.8656710	-579.9711580	0.1316320	0.0871860
n_6fA_C4	-580.7006313	-580.7287276	-579.704634	-579.8754173	-579.9809128	0.1320360	0.0876750
n_r6fA							
n_r6fA_C1	-580.0376449	-580.0621630	-579.040295	-579.2091939	-579.3133838	0.1192170	0.0742340
n_r6fA_C2	-580.0354243	-580.0606139	-579.038109	-579.2069539	-579.3111341	0.1195480	0.0744560
n_r6fA_C3	-580.0319291	-580.0567300	-579.035283	-579.2040424	-579.3081988	0.1193200	0.0740190
n_r6fA_C4	-580.0368960	-580.0617069	-579.039692	-579.2085560	-579.3127421	0.1190540	0.0739940
n_formic_acid	-189.7765265	-189.7862545	-189.500591	-189.5600909	-189.5963896	0.0378790	0.0096730
n_A	-467.3621144	-467.3894777	-466.539260	-466.6771507	-466.7623499	0.1201880	0.0783630
form	-114.511930	-114.515630 -114.515846[c]	-114.333204	-114.368249	-114.389516	0.030433 0.030694[c]	0.004951 0.005856[c]
C	-394.970446	-395.003367 -395.005991[c]	-394.293673	-394.413263	-394.486893	0.105988 0.105841[c]	0.067168 0.067544[c]
form_hyd	-190.969940	-190.985315	-190.687746	-190.750415	-190.788675	0.061973	0.032383
water	-76.434057	-76.448021	-76.331927	-76.359479	-76.375952	0.025064	0.003636
1caC							

n_1caC_C1	-583.5635409	-583.5912411	-582.620793	-582.7971325	-582.9057379	0.1223900	0.0780290
n_1caC_C2	-583.5441490	-583.5785648	-582.600915	-582.7774501	-582.8861539	0.1230600	0.0765800
n_1caC_C3	-583.5400894	-583.5775050	-582.597192	-582.7736452	-582.8822771	0.1229260	0.0764940
carbondioxide	-188.5905823	-188.5869314	-188.325595	-188.3826376	-188.4173712	0.0151420	-0.0091530

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries. [b] Standard state correction of $\Delta G_{0K\rightarrow 298K}^{1atm\rightarrow 1M}$ = +7.91 kJ mol^{-1} not yet applied. [c] Using geometries optimized in water at the SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p) level of theory.

Table S8: QM properties for gas phase optimized conformers shown in Figure S17 calculated at the (U)B3LYP-D3\6-31G+(d,p) level of theory.

compound	H(298, B3LYP-D3) [Hartree]	H(298, B3LYP-D3, H2O) [Hartree]	H(298, DLPNO(TZ) [Hartree]	H(298, DLPNO(TZ), H2O) [Hartree]	ΔG_{solv} [kJ/mol]
tol	-271.462151	-271.463059	-270.911130	-270.912037	-2.4
r_tol	-270.823841	-270.824523	-270.271643	-270.272325	-1.8
n_5m_C	-434.158884	-434.191815	-433.401360	-433.434292	-86.5
n_r5m_C	-433.520759	-433.550941	-432.760205	-432.790387	-79.2
n_1m_C	-434.150429	-434.180063	-433.393151	-433.422785	-77.8
n_r1m_C	-433.504484	-433.529437	-432.743523	-432.768476	-65.5
n_15m_C	-473.444652	-473.474087	-472.606720	-472.636154	-77.3
n_1r5m_C	-472.806791	-472.833689	-471.965574	-471.992472	-70.6
n_r15m_C	-472.799108	-472.823894	-471.957360	-471.982146	-65.1
n_5m_U	-454.057064	-454.079812	-453.292029	-453.314777	-59.7
n_r5m_U	-453.419716	-453.440227	-452.651945	-452.672456	-53.9
n_1m_U	-454.045992	-454.067238	-453.280782	-453.302028	-55.8
n_r1m_U	-453.397977	-453.414820	-452.630439	-452.647282	-44.2
n_15m_U	-493.342227	-493.362171	-492.496382	-492.516326	-52.4
n_1r5m_U	-492.705232	-492.721743	-491.856461	-491.872972	-43.4
n_r15m_U	-492.694761	-492.710263	-491.846402	-491.861904	-40.7
n_1f_C_2	-508.188616	-508.215554	-507.339843	-507.366781	-70.7
n_r1f_C_2	-507.533424	-507.558681	-506.683904	-506.709161	-66.3
1hmC	-509.373086	-509.406007	-508.519595	-508.552516	-86.4
n_r1hm_C	-508.733342	-508.757854	-507.872123	-507.896635	-64.4
n_6mA_C2	-506.522075	-506.546651	-505.642548	-505.617971	-64.5
n_6mA_C1	-506.524212	-506.549335	-505.645115	-505.619991	-66.0
n_r6mA_C2	-505.881296	-505.902963	-504.995287	-504.973620	-56.9
n_r6mA_C1	-505.881964	-505.904556	-504.996917	-504.974325	-59.3
n_r6hmA_C1	-581.107337	-581.132043	-580.125003	-580.100297	-64.9
n_6hmA_C1	-581.746510	-581.776793	-580.776879	-580.746595	-79.5
n_6fA_C1	-580.568864	-580.597220	-579.601171	-579.572815	-74.4
n_6fA_C2	-580.559302	-580.593586	-579.597927	-579.563642	-90.0
n_6fA_C3	-580.558246	-580.591783	-579.596756	-579.563219	-88.1
n_6fA_C4	-580.568595	-580.596692	-579.600694	-579.572598	-73.8
n_r6fA_C1	-579.918428	-579.942946	-578.945596	-578.921078	-64.4
n_r6fA_C2	-579.915876	-579.941066	-578.943751	-578.918561	-66.1
n_r6fA_C3	-579.912609	-579.937410	-578.940764	-578.915963	-65.1
n_r6fA_C4	-579.917842	-579.942653	-578.945449	-578.920638	-65.1

n_formic_acid	-189.738648	-189.748376	-189.472440	-189.462712	-25.5
n_A	-467.241926	-467.269290	-466.446436	-466.419072	-71.8
n_5miC	For Boltzmann-weighting of tautomers see Table 6.				
r_t1	-433.525050	-433.548967	-432.765142	-432.789059	-62.8
t1	-434.160970	-434.186873	-433.404145	-433.430049	-68.0
r_t2	-433.508301	-433.543877	-432.749940	-432.785516	-93.4
t2	-434.145839	-434.182792	-433.390814	-433.427767	-97.0
t3	-434.1611547	-434.1789728	-433.4067992	-433.4246173	/
t4	-434.1458391	-434.1827921	-433.3908138	-433.4277668	/
compound	H(298, DLPNO(QZ) [Hartree]	H(298, DLPNO(QZ, H2O) [Hartree]	H(298, DLPNO(CBS) [Hartree]	H(298, DLPNO(CBS), H2O) [Hartree]	
tol	-270.988513	-270.989421	-271.036793	-271.037701	/
r_tol	-270.347765	-270.348447	-270.395205	-270.395887	/
n_5m_C	-433.532540	-433.565472	-433.613463	-433.646395	/
n_r5m_C	-432.889938	-432.920120	-432.969903	-433.000086	/
n_1m_C	-433.523832	-433.553466	-433.604449	-433.634083	/
n_r1m_C	-432.872727	-432.897680	-432.952354	-432.977307	/
n_15m_C	-472.748989	-472.778424	-472.836900	-472.866335	/
n_1r5m_C	-472.106463	-472.133361	-472.193449	-472.220346	/
n_r15m_C	-472.098174	-472.122961	-472.185107	-472.209893	/
n_5m_U	-453.428839	-453.451587	-453.513308	-453.536057	/
n_r5m_U	-452.787290	-452.807801	-452.870821	-452.891332	/
n_1m_U	-453.417293	-453.438539	-453.501538	-453.522784	/
n_r1m_U	-452.765452	-452.782295	-452.848690	-452.865532	/
n_15m_U	-492.644375	-492.664320	-492.735876	-492.755820	/
n_1r5m_U	-492.003018	-492.019530	-492.093624	-492.110136	/
n_r15m_U	-491.992957	-492.008459	-492.083491	-492.098993	/
n_1f_C_2	-507.492134	-507.519072	-507.585943	-507.612881	/
n_r1f_C_2	-506.834367	-506.859624	-506.926912	-506.952169	/
1hmC	-508.674771	-508.707692	-508.770367	-508.803289	/
n_r1hm_C	-508.025350	-508.049861	-508.119639	-508.144151	/
n_5miC	For Boltzmann-weighting of tautomers see Table 6.				
r_t1	-432.894313	-432.918230	-432.974084	-432.998002	/
t1	-433.534725	-433.560629	-433.615405	-433.641308	/
r_t2	-432.8794655	-432.915041	-432.9593129	-432.9948886	/
t2	-433.521784	-433.558737	-433.6025721	-433.6395251	/
t3	-433.530038	-433.547856	-433.610665	-433.628483	/
t4	-433.537219	-433.574172	-433.617792	-433.654745	/
n_6mA_C2	-505.766864	-505.791440	-505.859040	-505.883616	/
n_6mA_C1	-505.769149	-505.794273	-505.861432	-505.886556	/
n_r6mA_C2	-505.121007	-505.142674	-505.212167	-505.233834	/
n_r6mA_C1	-505.121925	-505.144516	-505.213163	-505.235755	/
n_r6hmA_C1	-580.272283	-580.296990	-580.378374	-580.403080	/
n_6hmA_C1	-580.920073	-580.950356	-581.027245	-581.057528	/
n_6fA_C1	-579.743605	-579.771962	-579.849099	-579.877455	/

n_6fA_C2	-579.734361	-579.768646	-579.839767	-579.874052	/
n_6fA_C3	-579.734039	-579.767576	-579.839526	-579.873063	/
n_6fA_C4	-579.743381	-579.771478	-579.848877	-579.876973	/
n_r6fA_C1	-579.089977	-579.114495	-579.194167	-579.218685	/
n_r6fA_C2	-579.087406	-579.112596	-579.191586	-579.216776	/
n_r6fA_C3	-579.084722	-579.109523	-579.188879	-579.213680	/
n_r6fA_C4	-579.089502	-579.114313	-579.193688	-579.218499	/
formic_acid	-189.522212	-189.531940	-189.558511	-189.568239	/
n_A	-466.556963	-466.584326	-466.642162	-466.669525	/

In case of 5m*i*C two relevant tautomers were found and Boltzmann weighted as depicted in Table S9 (see structures in Figure S17). Since it has been found that 5m*i*C crystalizes in a ratio of 1:1 between **5mic_t1** and **5miC_t2**, these results agree with experimental observation. [20]

Table S9: QM properties for gas phase optimized tautomers of isocytosine shown in Figure S17 calculated at the DLPNO-CCSD(T)/CBS level of theory.

	SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p)// DLPNO-CCSD(T)/CBS[a]			Relative Energy			Boltzmann Distribution			% Population (/100)		
Molecule	E_{tot}(DPLNO-CCSD(T)/CBS)	H(298, DLPNO-CCSD(T)/CBS, H2O)	G(298, DLPNO-CCSD(T)/CBS), H2O)	Rel. ΔE	Rel. ΔH	Rel. ΔG	E	H	G	E	H	G
5m*i*C_t1	-434.2965357	-433.641308	-433.683077	0.0	0.0	0.0	0.99	1.00	1.00	0.50	0.87	0.83
5m*i*C_t2	-434.2811761	-433.639525	-433.681568	40.4	4.7	4.0	0.00	0.15	0.20	0.00	0.13	0.17
5m*i*C_t4	-434.2965477	-433.635554	-433.677301	0.0	15.1	15.2	1.00	0.00	0.00	0.50	0.00	0.00
5m*i*C_t3	-434.2884264	-433.635532	-433.677233	21.3	15.2	15.3	0.00	0.00	0.00	0.00	0.00	0.00

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries.

Table S10: QM properties for gas phase optimized conformers of the systems shown in Figure S17 calculated at the DLPNO-CCSD(T)/CBS level of theory.

	SMD(H2O)/(U)B3LYP-D3/6-31+G(d,p)// DLPNO-CCSD(T)/CBS[a]			Relative Energy			Boltzmann Distribution			% Population (/100)		
Molecule	E_{tot}(DPLNO-CCSD(T)/CBS)	H(298, DLPNO-CCSD(T)/CBS, H2O)	G(298, DLPNO-CCSD(T)/CBS), H2O)	Rel. ΔE	Rel. ΔH	Rel. ΔG	E	H	G	E	H	G
n_6mA_1	-506.011560	-505.886556	-505.931317	0.0	0.0	0.0	1.00	1.00	1.00	0.93	0.96	0.96
n_6mA_2	-506.009092	-505.883616	-505.928222	6.5	7.7	8.1	0.07	0.04	0.04	0.07	0.04	0.04
n_6fA_1	-579.981209	-579.877455	-579.921947	0.0	0.0	0.0	1.00	1.00	1.00	0.58	0.61	0.64
n_6fA_4	-579.980913	-579.876973	-579.921334	0.8	1.3	1.6	0.73	0.60	0.52	0.42	0.37	0.33
n_6fA_2	-579.971650	-579.874052	-579.918956	25.1	8.9	7.9	0.00	0.03	0.04	0.00	0.02	0.03
n_6fA_3	-579.971158	-579.873063	-579.917509	26.4	11.5	11.7	0.00	0.01	0.01	0.00	0.01	0.01
n_6hmA_1_scan2_3	-581.177884	-581.057264	-581.1036073	16.5	1.0	0.0	0.00	0.67	1.00	0.00	0.25	0.50
n_6hmA_1_scan2_1	-581.183773	-581.057528	-581.1030382	1.0	0.3	1.5	0.66	0.89	0.55	0.40	0.33	0.27

n_6hmA_1_scan2_6	-581.184165	-581.057642	-581.1024816	0.0	0.0	3.0	1.00	1.00	0.30	0.60	0.37	0.15
n_6hmA_1_scan2_4	-581.176996	-581.055769	-581.1017506	18.8	4.9	4.9	0.00	0.14	0.14	0.00	0.05	0.07
n_1f_C_2	-507.703044	-507.612881	-507.656336	0.0	0.0	0.0	1.00	1.00	1.00	1.00	0.99	0.95
n_1f_C_1	-507.687306	-507.608090	-507.653588	41.3	12.6	7.2	0.00	0.01	0.05	0.00	0.01	0.05
n_1caC_1	-582.905738	-582.811048	-582.8554091	0.0	0.0	0.0	1.00	1.00	1.00	1.00	1.00	1.00
n_1caC_2	-582.886154	-582.797510	-582.8439897	51.4	35.5	30.0	0.00	0.00	0.00	0.00	0.00	0.00
n_1caC_3	-582.882277	-582.796767	-582.8431987	61.6	37.5	32.1	0.00	0.00	0.00	0.00	0.00	0.00

[a] Using gas phase optimized (U)B3LYP-D3/6-31+G(d,p) geometries.

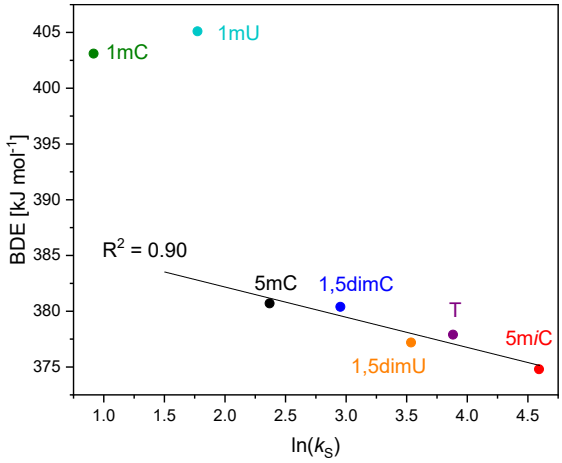


Figure S23: Plot of the calculated gas phase BDE values against the observed reaction rates on a logarithmic scale (R^2= 0.901).

5.3 SYNTHETIC PROCEDURES

Synthesis of 5-methylisocytosine

According to a modified literature procedure. [21] Ethyl propionate (4.50 mL, 3.98 g, 39.0 mmol, 1.5 equiv.) was added to a solution of sodium methoxide (2.82 g, 52.2 mmol, 2.0 equiv.) and dimethylformamide (4 mL) under nitrogen atmosphere. Methyl formate (1.60 mL, 1.57 g, 26.1 mmol, 1.0 equiv.) was added to the mixture over a period of one hour, resulting in a slight foaming. The solution was stirred for 30 min at room temperature. A solution of guanidine hydrochloride (2.49 g, 26.1 mmol, 1.0 equiv.) in methanol (9.2 mL) was added rapidly and the mixture was refluxed (bath temperature 95 °C) for 2 h. The suspension was allowed to cool and when it reached 30 °C it was filtered through a sintered glass frit. The filter cake was discarded, and the pH of the filtrate was adjusted to 6 with concentrated hydrochloric acid, resulting in a colorless precipitate. The mixture was kept at 4 °C for 30 min, filtered off through a sintered glass frit, washed with methanol and recrystallized in water. The precipitate was filtered, washed with water and dried *in vacuo* to yield 5-methylisocytosine as a colorless solid (684.4 mg, 5.47 mmol, 21%)

^{1}H NMR (400 MHz, DMSO-*d6*): δ [ppm] = 10.90 (br. s, 1H), 7.39 (q, *J* = 1.4 Hz, 1H), 6.33 (s, 2H), 1.74 (d, *J* = 1.1 Hz, 3H).

Elemental Analysis: calc. for $C_5H_7N_3O$: C, 47.99; H, 5.64; N, 33.58. Found: C, 47.93; H, 5.86; N, 33.48.

HR-MS (ESI): calc. for $C_5H_9N_3O^+$ $[M+H^+]^+$ 126.06619; found 126.06631 m/z.

Synthesis of 1-methyluracil

According to a modified literature procedure. [22] Uracil (5.00 g, 44.6 mmol, 1.0 equiv.) was placed in a flame dried flask and suspended in hexamethyldisilazide (HMDS, 25.2 g, 32.0 mL, 156 mmol, 3.5 equiv.). TMSCl (2.70 g, 3.14 mL, 25.0 mmol, 0.56 equiv.) was added and the mixture was refluxed under nitrogen atmosphere at an oil bath temperature of 130 °C until the suspension turned clear, about 2 h. After cooling to room temperature, the solution was concentrated to a volume of about 10 mL. After addition of iodomethane (27.2 g, 11.9 mL, 194 mmol, 4.34 equiv.) the mixture was refluxed for a further 18 hours at an oil bath temperature of 60 °C. Volatile reagents were evaporated and the residue was suspended in isopropanol (25 mL). It was then filtered off, washed with isopropanol and recrystallized in water to yield 1-methyluracil (4.03 g, 31.9 mmol, 73%) as colorless crystals.

^{1}H NMR (400 MHz, DMSO-*d6*): δ [ppm] = 11.22 (s, 1H), 7.61 (d, *J* = 7.8 Hz, 1H), 5.51 (d, *J* = 7.8 Hz, 1H), 3.21 (s, 3H).

^{13}C-NMR (100 MHz, DMSO-*d6*): δ [ppm] = 164.0, 151.3, 146.5, 100.5, 35.2.

Elemental Analysis: calc. for $C_5H_6N_2O_2$: C, 47.62; H, 4.80; N, 22.21. Found: C, 46.69; H, 4.68; N, 22.16.

HR-MS (EI): calc. for $C_5H_7N_2O_2^+$ $[M+H^+]^+$ 126.0429; found 126.0423 m/z.

Synthesis of 1,5-dimethyluracil

According to a modified literature procedure. [23] In an oven dried Schlenk flask under nitrogen, a suspension of thymine (1.90 g, 15.0 mmol), HMDS (35 mL) and Me_3SiCl (0.75 mL) was refluxed for 5h. After cooling down to 40 °C iodomethane (15 mL) was added to the reaction mixture and the suspension was refluxed overnight. All volatile compounds were removed under high vacuum with the addition of a cooling trap, to yield a brown solid. Now 35 mL of 6N acetic acid were added, the reaction mixture was stirred for 20 min at room temperature and the solvent was removed under reduced pressure. The crude product was recrystallized in water to yield 1,5-dimethyluracil (1.47 g, 10.5 mmol, 70%) as a slightly brown solid.

^{1}H-NMR (400 MHz, DMSO-*d6*): δ [ppm] = 12.21 (br. s, 1H), 7.50 (s, 1H), 3.19 (s, 3H), 1.73 (s, 3H).

^{13}C-NMR (100 MHz, DMSO-*d6*): δ [ppm] = 164.5, 151.2, 142.4, 108.1, 35.0, 11.9.

Synthesis of 1-methylcytosine

According to literature procedure. [24]

^{1}H-NMR (400 MHz, DMSO-*d6*): δ [ppm] = 7.55 (d, ^{3}J = 7.1 Hz, 1H), 6.91 (s, 2H), 5.60 (d, ^{3}J = 7.1 Hz, 1H), 3.19 (s, 3H).

^{13}C-NMR (100 MHz, DMSO-*d6*): δ [ppm] = 166.1, 156.3, 146.6, 92.9, 36.6.

Elemental Analysis: calc. for $C_5H_7N_3O$: C, 47.99; H, 5.64; N, 33.58. Found: C, 47.73; H, 5.64; N, 32.58.

HR-MS (EI): calc. for $C_5H_7N_3O^{\cdot+}$ $[M]^{\cdot+}$ 125.0584; found 125.0585 m/z.

5.4 REFERENCES

[1] N. S. W. Jonasson, L. J. Daumann, *Chem. Eur. J.* **2019**, 25, 12091-12097.

[2] J. Hioe, M. Mosch, D. M. Smith, H. Zipse, *RSC Adv.* **2013***, 3,* 12403-12408.

[3] J. Hioe, G. Savasci, H. Brand, H. Zipse, *Chem. Eur. J.* **2011**, 17, 3781-3789.

[4] J. Hioe, H. Zipse, *Org. Biomol. Chem.* **2010**, 8, 3609-3617.

[5] J. Hioe, H. Zipse, *Faraday Discuss.* **2010**, 145, 301-313.

[6] J. Hioe, H. Zipse, *Chem. Eur. J.* **2012**, 18, 16463-16472.

[7] D. Šakić, H. Zipse, *Adv. Synth. Catal.* **2016**, 358, 3909-3909.

[8] H. Zipse, *Radicals in Synthesis I*, 1 ed., Springer-Verlag Berlin Heidelberg, **2006**.

[9] A. D. Becke, *J. Chem. Phys.* **1993**, 98, 5648-5652.

[10] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, 132, 154104.

[11] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, 54, 724.

[12] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, 72, 650.

[13] A. Altun, F. Neese, G. Bistoni, *Beilstein J. Org. Chem.* **2018**, 14, 919-929.

[14] M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2017**, 146, 164105.

[15] F. Neese, *WIREs Comput. Mol. Sci.* **2018**, 8, e1327.

[16] T. H. D. Jr., *J. Chem. Phys.* **1989**, 90, 1007-1023.

[17] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, 113, 6378-6396.

[18] V. S. F. Muralha, R. M. Borges dos Santos, J. A. Martinho Simões, *J. Phys. Chem. A* **2004**, 108, 936-942.

[19] 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.

[20] G. Portalone, M. Colapietro, *Acta Crystallogr. E* **2007**, 63.

[21] P. Stoss, E. Kaes, G. Eibel, U. Thewalt, *J. Heterocyclic Chem.* **1991**, 28, 231-236.

[22] J. R. Bartels-Keith, J. B. Mahoney, A. J. Puttick, *J. Org. Chem.* **1985**, 50, 980-987.

[23] T. T. Sakai, A. L. Pogolotti Jr, D. V. Santi, *J. Heterocycl. Chem.* **1968**, 5, 849-851.

[24] E. Greco, A. E. Aliev, V. G. Lafitte, K. Bala, D. Duncan, L. Pilon, P. Golding, H. C. Hailes, *New J. Chem.* **2010**, 34, 2634-2642.

6 LIST OF ABBREVIATIONS

1mC	1-Methylcytosine
5mC	5-Methylcytosine
AID	activation-induced cytidine deaminase
APCI	atmospheric-pressure chemical ionization
APOBEC	apolipo-protein B mRNA-editing enzyme complex
Asp	asparagine
BDE	bond dissociation energy
BER	base excision repair
CoPaSi	"Complex Pathway Simulator" program
d	doublet
DCM	dichloromethane
DFG	Deutsche Forschungsgemeinschaft
DEP	direct exposure probe
DFT	density functional theory
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyl transferases
Dnmt1	DNA (cytosine-5)-methyltransferase 1
DPG	DNA-protein crosslink
dsDNA	double stranded deoxyribonucleic acid
E_{tot}	total energy
FWHM	full width at half maximum
G	free energy
GC	gas chromatography
GLU-COOH	glutamate
H	enthalpy
HAT	hydrogen atom abstraction
His	histidine
HPLC	high-performance liquid chromatography
K	equilibrium constant

k_{cat}	catalytic rate constant
K_{hyd}	equilibrium constant for hydration reaction
K_m	Michaelis constant
LC	liquid phase chromatography
LOD	limit of detection
LOQ	limit of quantification
MAE	mean absolute error
MeOH	methanol
MHz	Megahertz
mM	millimolar
MnO_2	manganese dioxide
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MUE	mean unsigned error
NH_4OH	ammonium hydroxyde
NMR	nuclear magnetic resonance spectroscopy
pK_a	acid dissociation constant
PLC	preparative layer chromatography
R	gas constant
R_f	retardation factor
R^2	coefficient of determination
RSE	radical stabilization energy
s	singlet
S/N	signal to noise
SAM	S-adenosyl-L-methionine
SMD	solvation model based on density
SMUG1	single-strand selective monofunctional uracil DNA glycosylase
sol	solution
solv	solvation
t	triplet
T/5mU	Thymine/5-Methyluracil
TDG	thymidine DANN glycosylase
TET	ten-eleven translocation methylcytosine dioxygenases

THase	thymine-7-hydroxylase
TLC	thin layer chromatography
δ	chemical shift [ppm]

The End