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

Special Collection *This article is part of a joint ChemBioChem-ChemMedChem Special Collection on Chemical Epigenetics available at bit.ly/chemepi2021.*

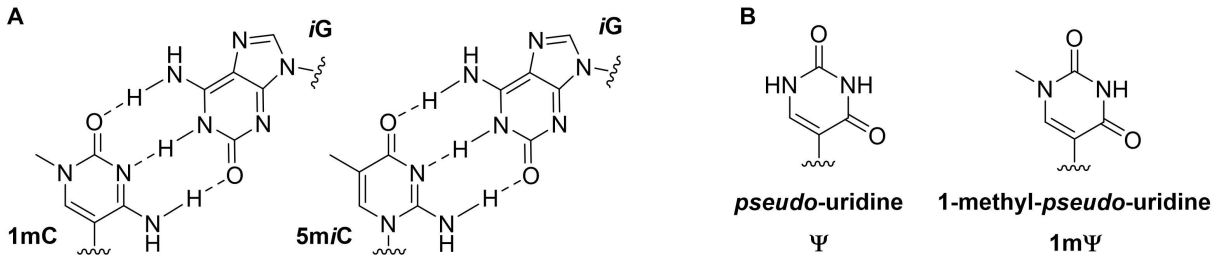


Figure 1. A) Base pairs 1-methylcytidine (1mC) – *iso*-guanosine (*i*G) and 5-methyl-*iso*-cytidine (5m*i*C) – *iso*-guanosine used in hachimoji DNA and other synthetic DNA applications.[5a,8] B) Nucleobase fragments in *pseudo*-uridine (Ψ) and 1-methyl-*pseudo*-uridine (1mΨ).

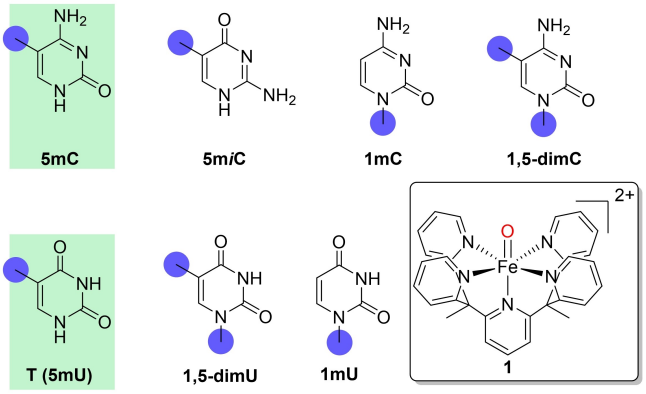


Figure 2. Methylated cytosine and uracil derivatives used in this work: 5-methylcytosine (5mC), 5-methyl-*iso*-cytosine (5miC), 1-methylcytosine (1mC), 1,5-dimethylcytosine (1,5dimC), thymine (T), 1-methyluracil (1mU), 1,5-dimethyluracil (1,5dimU). The studied methyl groups are marked with blue circles, 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

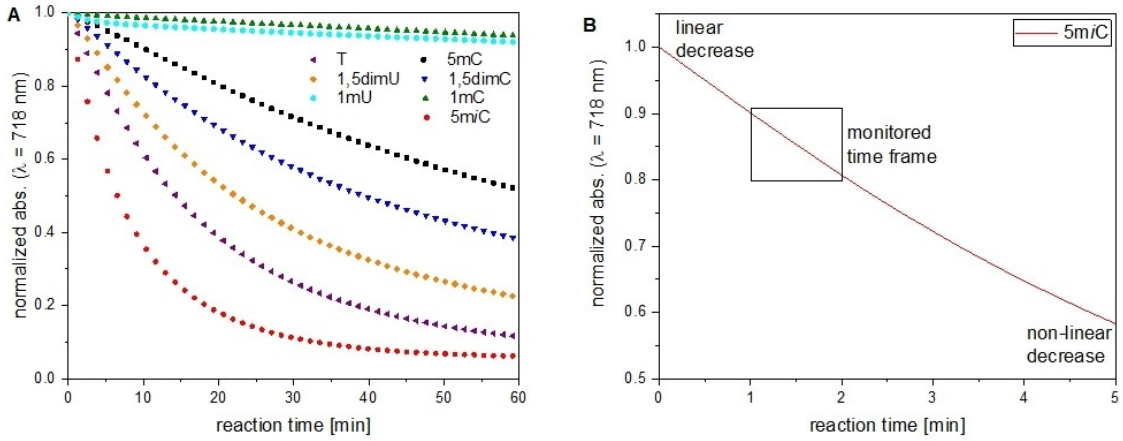


Figure 3. UV/Vis spectroscopy kinetics: A) Plot of the development of the absorbance of a series of reactions. B) Linear and non-linear decrease of the absorbance at λ = 718 nm. Conditions: [S] = 1 mM, [**1**] = 5 mM, H_2O, 30 °C.

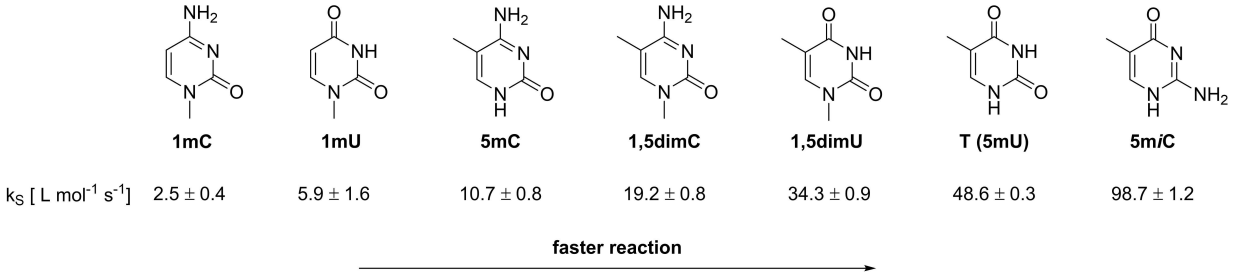


Figure 4. Observed rate constants k_S of the substrates at the following conditions: [S] = 1 mM, [**1**] = 5 mM, H_2O, 30 °C. Rate constants k_S were calculated using a second order rate equation (Eq. 1).

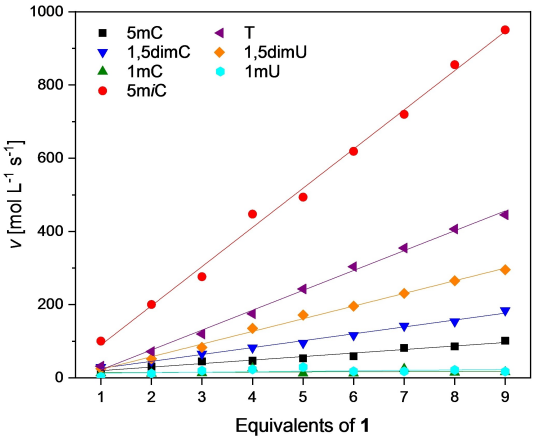


Figure 5. Plot of the measured reaction rates *v* of the reaction of **1** with nucleobase substrates. For a zoomed-in version of 5mC, 1mC, and 1,5dimC results see Supporting Information Figure S3[B]. 5mC (black squares), 5m*i*C (red dots), 1mC (green triangles), 1,5dimC (blue inverted triangles), T (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.

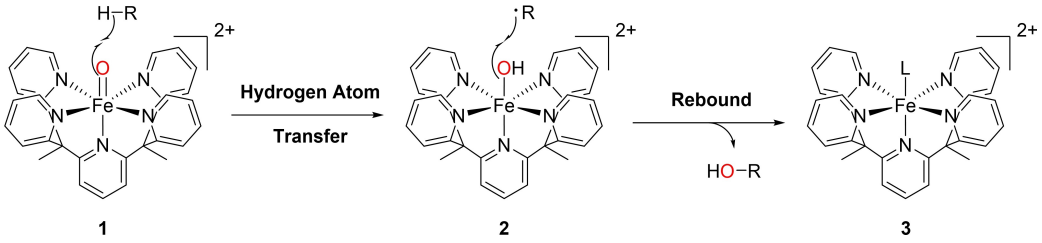
BDE calculation

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**.

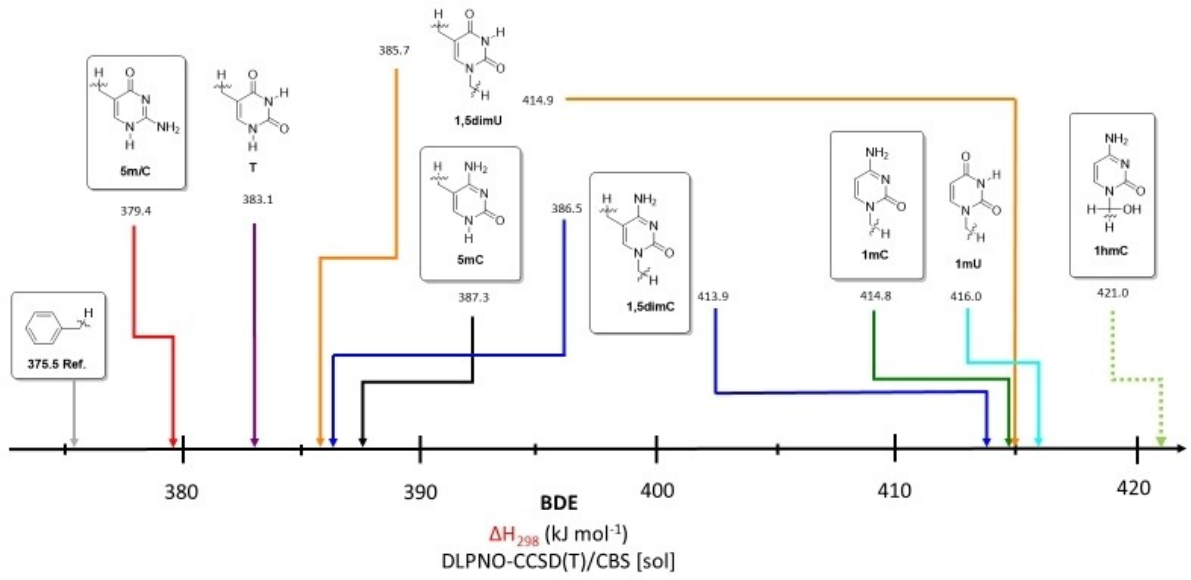


Figure 6. Aqueous phase ($\Delta H_{sol} = \Delta H_{298} + \Delta G_{solv}$) RCH_2–H bond dissociation energies (BDEs) of the relevant substrates calculated at the DLPNO-CCSD(T)/CBS level of theory. Dotted lines indicate the BDEs of molecules that were calculated but not included in experiments. For 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.

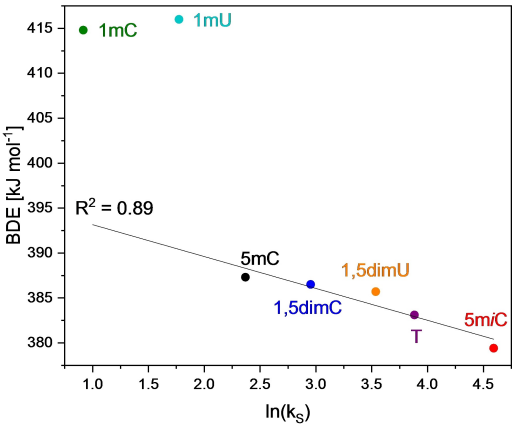


Figure 7. Plot of the calculated BDE values in aqueous phase against the observed rate constants k_s on a logarithmic scale ($R^2 = 0.891$) of the substrates 1mC, 1mU, 5mC, 1,5dimC, 1,5dimU, T, and 5m*i*C. In the case of the demethylated substrates 1,5dimC and 1,5dimU only the BDE of the carbon bound 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.

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

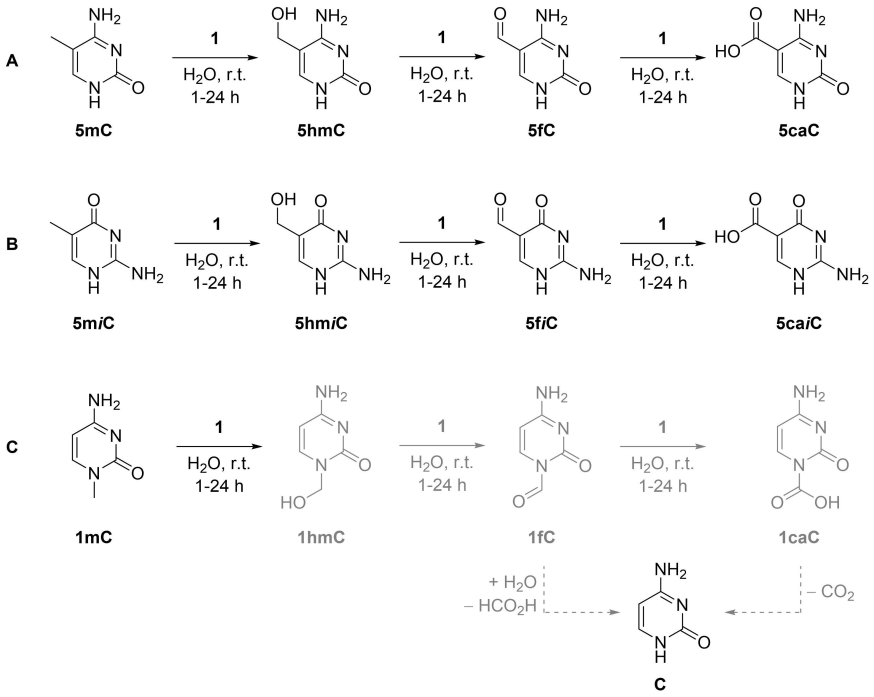
The presented data is a remarkable result, as it both provides evidence to confirm the previously postulated mechanism of **1** and shows that calculated BDEs can be reliably used to predict reaction rates for these types of substrates (in the absence of an enzyme's second coordination sphere). It is 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.

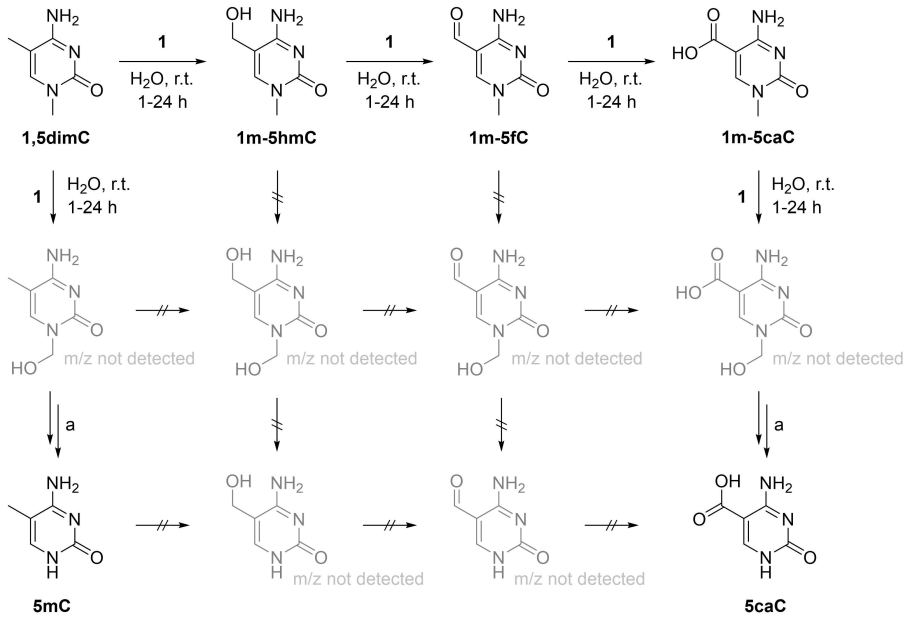
deformylation of 1fC is $-17.0\ kJ\,mol^{-1}$ (Figure S21[C]), for the decarboxylation of 5caC it is $-44.9\ kJ\,mol^{-1}$ (Figure S21[D]).

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