

Computational Insights in DNA Methylation: Catalytic and Mechanistic Elucidations

Mansour H. Almatarneh ^{1,2,*}, Ghada G. Kayed ¹, Mohammednoor Altarawneh³, Yuming Zhao², Amita Verma⁴

Correspondence to: Mansour H. Almatarneh (E-mail: m.almatarneh@ju.edu.jo)

¹Department of Chemistry, University of Jordan, Amman 11942, Jordan.

²Department of Chemistry, Memorial University, St. John's, NL A1B 3X7, Canada.

³Department of Chemical and Petroleum Engineering, United Arab Emirates University, Al-Ain, 15551, UAE.

⁴Department of Pharmaceutical Sciences, Faculty of Health Sciences, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, 211007, India.

ABSTRACT

Methylation at C5 position of cytosine (5mC) is the most abundantly occurring methylation process at CpG island, which has been well-known as an epigenetic modification linked to many human's diseases. Recently, another methylation approach has been discovered to show that DNA methyltransferases (DNMTs) promote the addition of methyl group at position 3 to yield 3mC. The existence of 3mC can cause severe damages to the DNA strand, such as blocking its replication, repair, and transcription, affecting its stability, and initiating a double-strand DNA break. To gain a deeper insight into the formation of 3mC, we have performed density functional theory (DFT) modeling studies at different levels of theory to clearly map out the mechanistic details for this new methylation approach. Our computed results are in harmony with pertinent experimental observations and shed light on a crucial off-target activity of DNMTs.

KEYWORDS

Epigenetics, DNA Methylation, DNA Methyltransferases, 3-Methyl Cytosine, Catalysts, DNA Stand-Break

1. Introduction

Although methylation of cytosine at position 5 is a hallmark of epigenetic modifications in organisms produced and conserved by DNA methyl transferases (DNMTs), other methylation pathways still possibly exist under the catalysis of the same enzyme, while the resulting methylated products can pose profoundly

serious threats to the DNA helix. In 2018, a transmutative relationship between alkylation and DNMTs appeared to the surface, showing the methylation of cytosine can take place at position 3 as well to form 3-methyl cytosine (3mC) as the product [1]. DNMT plays a catalytic role **Figure 1** in this reaction, and this methylation pathway has been found to result in a spontaneous DNA mutation in the cell.

Experimental analyses using ultrasensitive liquid chromatography/mass spectrometry (LC/MS) have unambiguously confirmed that DNMT can catalyze the methylation at the N3 position of cytosine with S-adenosyl-L-methionine (AdoMet). This finding reflects that the generation of 3mC is a general reactivity enabled by cytosine methyltransferases [1].

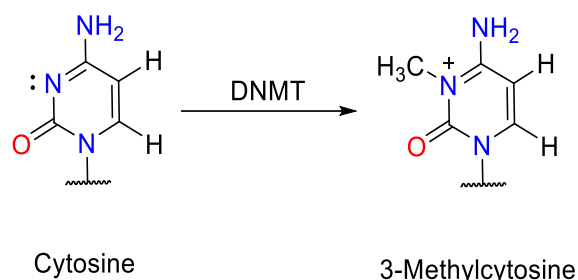


Figure 1: Reaction scheme of the generation of 3mC from cytosine under the catalysis of DNMT.

Mechanistic details of this newly observed methylation pathway, however, have not yet been disclosed. Duktaz *et al.* [2] recently conducted a mutational study of catalytically relevant amino acids in DNMT enzyme. Different amino acid residues were mutated in the active site of DNMT, and the catalytic roles that DNMT plays in methylation at C5 and N3 were assessed. Based on the experimental results, they proposed that the N3 methylation occurs through a mechanism of cytosine inversion step; that is, cytosine bound into the active-site pocket of DNMT is flipped into an inverted orientation. The flipping allows the N3 site of cytosine to be positioned in a close proximity to the methyl group of AdoMet to facilitate methyl group transfer, see **Figure 2** *vid infra*. The role of amino-acids present in the active-site in the formation of 3mC and 5mC was varied, unlike the N3 position, methylation on the C5 position was found to be attenuated, suggesting these amino acids play vital roles in the catalytic mechanisms. Cysteine (C710) has no catalytic role in the mechanism of N3 methylation, while mutation of Arg (R792) to Ala induces the formation of 3mC with a lower rate of formation, which was explained by the loss of H-bond of

R792 to the deoxyribose due to inversion of the base by 180°. Glutamic acid (E756) has a neutral effect on producing both 5mC and 3mC. The generation of both products was catalyzed through the formation of a H-bond with the exocyclic-amino group (N4), which stabilizes the base [2]. Despite the repair system displayed by the cell, alkylation repair protein B (ALKB2/3) can erase alkylation adducts as 1-methyladenine and 3-methylcytosine to help the survival of the genome [3]. However, when ALKB2 loses its functionality, the level of 3mC increases by around 15% [1]. It has also been reported that if the damage occurs during DNA replication and when the damage is beyond the limits of correction, ALKB is no longer capable of removing such a lesion [4].

On one side, it is well recognized that the occurrence of 5mC is mutagenic and can cause base mispairing G-T by the deamination of 5mC to thymine [5]. The deamination reactions of nucleic acid bases and related compounds have been extensively studied, the results of which are of great value in proposing the mechanisms investigated in this work[6–13]. On the other side, however, 3mC lesion affects the stability of single-strand (ss) and double-strand (ds) DNA [3]. Rošić *et al.* [1] reported that 3mC results in double-strand DNA (dsDNA) breaking.

The damage incurred in the presence of 3mC contradicts with replication, repair, and transcription of the DNA strand, in addition to spontaneous mutations of C to A and C to T [5]. In comparison to unmethylated cytosine, 3mC can accelerate deamination to form 3-methyluracil (3mU) by a factor of 4×10^3 at pH 7.4 (physiological conditions) [14].

Alkylation of cytosine at position 3 with carcinogenic alkylating agents, such as methanediazonium and ethanediazonium ions, has been studied theoretically through the B3LYP density functional method. Carcinogenic alkylating agents are produced from the metabolic activation of methyl and ethyl-

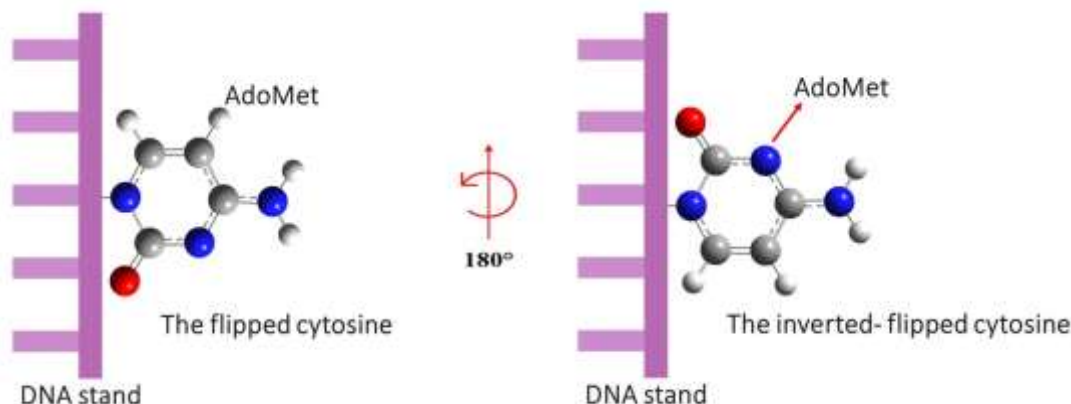


Figure 2: Schematic illustration of 180° inversion of flipped target base.

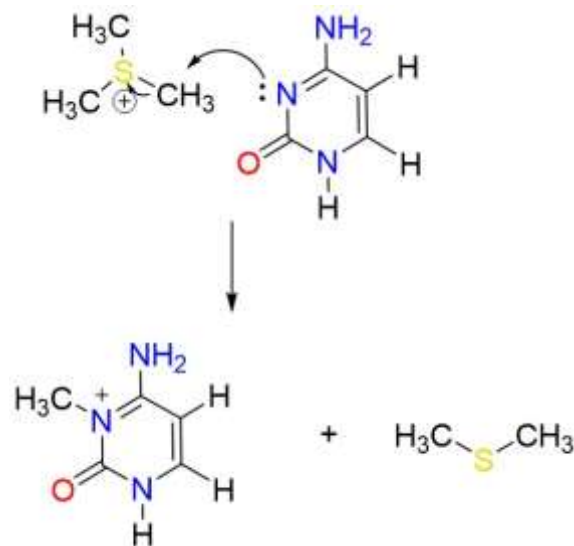
substituted nitrosoureas, and they tend to add a methyl group to N3 and O2 of cytosine; N7, N3 and O6 of guanine; N1 of adenine; and O2 and O4 of thymine via the S_N2 mechanism [5, 15].

Another exogenous strong alkylating agent is dimethyl sulfate (DMS). The results obtained from combustion analysis have been subjected to DFT theoretical investigations to understand the methylation of N3 of cytosine, N7 of guanine and N1, N3 and N7 of adenine at room temperature using M06-2X/6-31+G(d) and B3LYP/6-311+G(2df,2p) levels of theory. The solvation effects were examined using the conductor-like polarizable continuum model (CPCM). The obtained results indicated that the most reactive site towards methylation in the gas phase is N7 of adenine with an activation energy of 70.42 kJ.mol⁻¹, while the activation energy of cytosine is 96.52 kJ.mol⁻¹ [16].

To the best of our knowledge, the catalytic activity of DNMTs towards alkylation of cytosine at position 3 has not been investigated theoretically. Furthermore, the side reaction of 3mC promoted by DNMT warrant theoretical analysis to assess the validity of the proposed “inverted base flipping” mechanism. In this study, we performed a computational study on the most plausible reaction pathway starting from the general mechanism shown in **Scheme 1**. Based on calculations done at high-level theories, mechanistic details concerning the reactivity of DNMTs towards N3 of cytosine were

established, and the results were compared to the methylation reaction at conventional C5 position.

The catalytic roles of amino acids present in the active site such as cystine (Cys), glutamic acid (Glu), arginine (Arg), and alanine (Ala) in the gaseous phase were determined as well.



Scheme 1: The proposed one step mechanism of 3mC reaction (pathway A).

In vivo abundant potential catalytic species, such as phosphate and carbonate ions, have been known to play a vital role in the biological process. Most recently, Kato and co-workers studied the deamination of glutamine residues that were catalyzed by hydrogen carbonate and dihydrogen phosphate catalysts (HCO_3^- and H_2PO_4^-) [17]. In this study, the effects of such catalyst anions were also examined in order to understand their influence on the energy cost of methylation at various positions of cytosine ring, the structure of transition states, and potential energy barriers.

2. Computational Methods

We conducted geometry optimizations using Gaussian 16 (G16) quantum chemistry package. The resulting geometries were confirmed to be energy minima or transition states (TSs) on the potential energy surface by calculating their vibrational frequencies, where the TSs were confirmed by showing one and only imaginary frequency. Furthermore, TSs were subjected to intrinsic reaction coordinate (IRC) calculations at the B3LYP/6-31G(d) level of theory in order to connect the TSs with the local minima of the reactants and products on the potential energy surface. Geometry optimization was performed using the B3LYP/6-31G(d), B3LYP/6-31G(d,p), M06-2X/6-31G(d) and APFD/6-31G(d) methods. We also performed APFD/6-31+G(d) and APFD/6-31G(2df,2p) calculations in order to detect the effect of the polarization and diffuse functions on the system. The proposed pathways have been studied in the aqueous medium using the solvation model based on density (SMD) at APFD/6-31G(d) level of theory. Thermodynamic parameters: ΔS , ΔH , and ΔG , as well as the kinetic parameters: activation energy (E_a), enthalpy of activation (ΔH^\ddagger), and the Gibbs energies of activation (ΔG^\ddagger) have been also calculated for each proposed reaction mechanism.

3. Results and Discussion

In this thorough study, comprehensive computational quantum chemistry calculations for 10 reaction pathway mechanisms were proposed for the methylation reaction of cytosine at position 3. We designated different pathways in order to study and observe how each component will behave and affect this off-target methylation reaction. Pathways **A** represents the cytosine and AdoMet system as a starting point to study the methylation reaction on N3 position. As shown in **Figure 3** below.

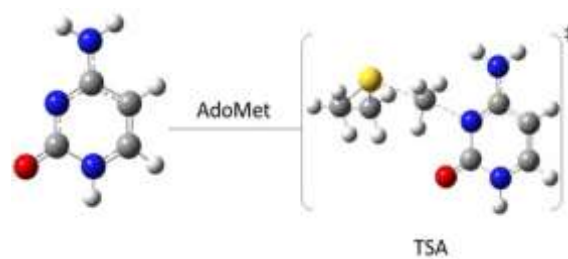


Figure 3: Optimized structure of reactant and transition state **TSA** of the proposed pathway **A**.

Pathways **B**→**D** include the participation of amino acids Cys, Glu and Ala respectively, as shown in **Figures 4**. Meanwhile, pathways **E** and **F** comprise the methylation reaction with our proposed catalytic anions such as hydrogen carbonate and dihydrogen phosphate respectively, as depicted in **Figure 5**.

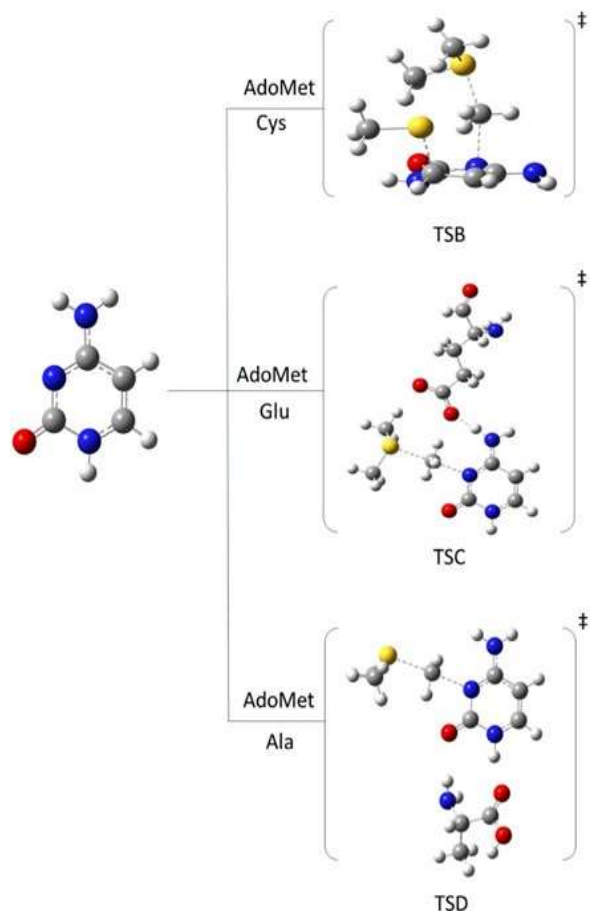


Figure 4: Optimized structure of reactant and transition state **TSB**, **TSC** and **TSD** of the proposed pathways **B**, **C** and **D**.

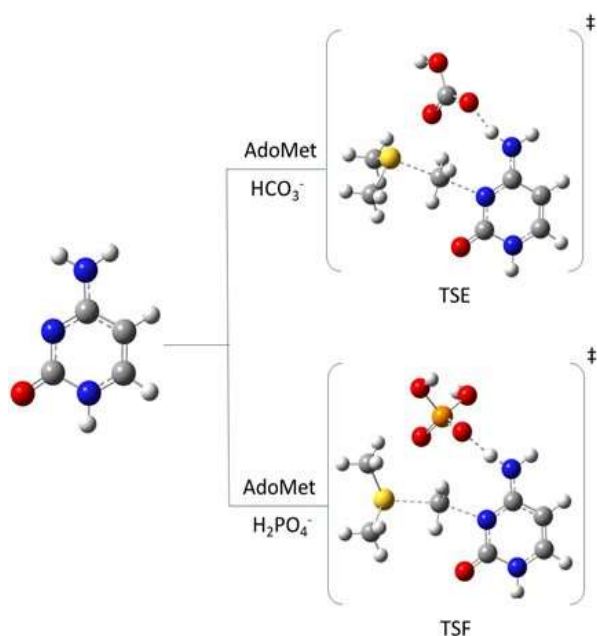


Figure 5: Optimized structure of reactant and transition state **TSE** and **TSF** of the proposed pathways **E** and **F**.

Furthermore, we studied the combined effect of amino acids and the proposed catalysis, in pathways **G** and **H**, as presented in **Figure 6**. It is worth observing that all pathways proceed through one step mechanism, except pathway **B**.

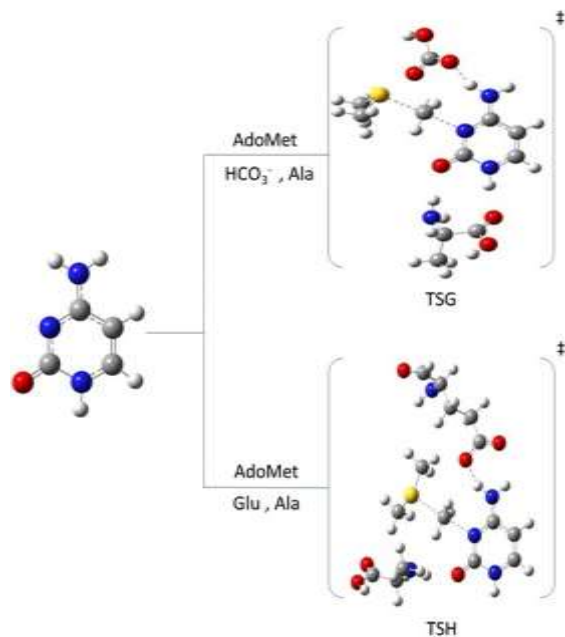


Figure 6: Optimized structure of reactant and transition state **TSG** and **TSH** of the proposed pathways **G** and **H**.

All investigated pathways are exothermic reactions.

The activation energies (E_a), enthalpies of activation (ΔH^\ddagger), and Gibbs energies of activation (ΔG^\ddagger) were calculated at different levels of theory **Tables 1** for all proposed pathways. Moreover, the potential energy diagrams (PEDs) were also constructed for studied pathways. While the thermodynamic parameters for the proposed pathways are presented in **Table 2**. The Potential energy diagram (PED) for pathway **A** calculated at different levels of theory in aqueous and gaseous phases (in kJ.mol^{-1} at 298.15 K) is depicted in **Figure 7**; while for pathways **B→J** were included in the Supporting Information (SI) in **Figures S1-S9**.

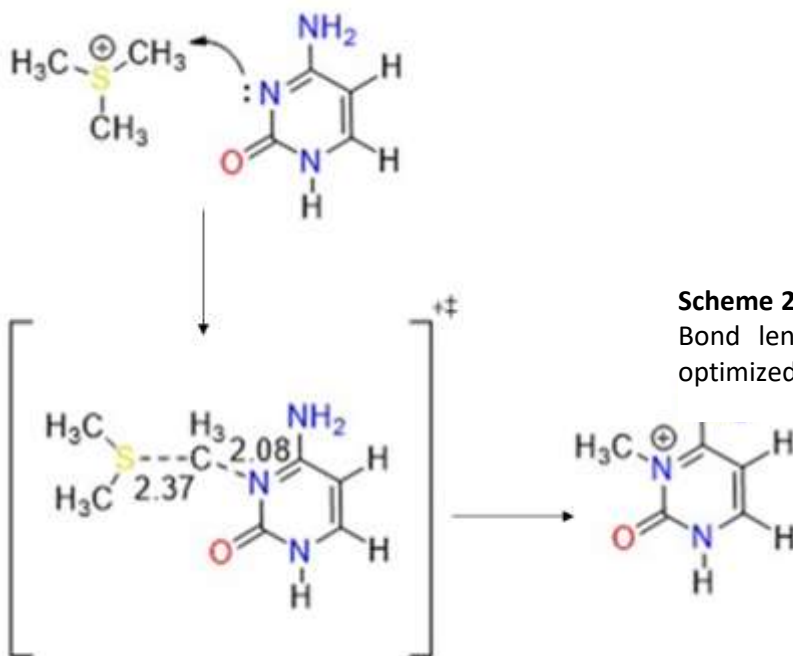
3.1 The Original Reaction: Pathway A

As mentioned above, DNMT can catalyze the substitution of a methyl group at the N3 position of cytosine ring through a straightforward electrophilic addition mechanism [15]. The

details of this one-step reaction are outlined as pathway **A**, **Figure 2**, and **Scheme 2**.

In this pathway, the reactants are placed to form electrostatic interaction between the positively charged AdoMet and the partially negative carbonyl oxygen of cytosine. For the **TSA**, the methyl group approaches the N3 atom at a semi-linear angle of 174.9° in a side-way manner to the ring. The methyl carbon shows a distance of 2.08 \AA to N3, which is indicative of a loosely bound $\text{S}_{\text{N}}2$ transition state, as shown in **Scheme 2** of pathway **A** reaction mechanism.

The activation energy was calculated to be $103.5 \text{ kJ.mol}^{-1}$ at the B3LYP/6-31G(d) level of theory which was highly correlated with the value obtained at B3LYP/6-31G(d,p) with $104.2 \text{ kJ.mol}^{-1}$ (see **Figure 7** and **Table 1**). We attributed this high activation energy to the presence of a positive charge on the cytosine ring. The effect of solvent (water) was studied using SMD solvation model. The activation energy of **TSA** was noticeably decreased in the aqueous environment with 92.2 kJ.mol^{-1} .



Scheme 2: Reaction mechanism of pathway **A**. Bond lengths are in \AA obtained from the optimized structures using B3LYP/6-31G(d).

Table 1: Kinetic parameters (E_a and ΔG^\ddagger) for the investigated pathways of 3mC (in $\text{kJ}\cdot\text{mol}^{-1}$) at 298.15 K.

Transition state	B3LYP/6-31G(d)		B3LYP/6-311G(d,p)		M06-2X/6-31G(d)		APFD/6-31G(d)		SMD*	
	E_a	ΔG^\ddagger	E_a	ΔG^\ddagger	E_a	ΔG^\ddagger	E_a	ΔG^\ddagger	E_a	ΔG^\ddagger
TSA	103.5	105.2	104.2	105.5	118.0	117.2	110.8	114.9	92.2	98.8
TSB	140.6	150.1	144.4	153.9	125.3	136.6	123.9	134.3	132.2	140.9
TSC	86.1	89.7	83.6	87.6	107.3	108.9	101.8	100.1	97.6	99.0
TSD	79.1	81.1	97.4	94.1	104.6	93.1	95.6	86.4	86.4	95.3
TSE	64.0	60.3	61.7	58.3	62.3	69.0	56.5	62.5	77.2	88.1
TSF	66.5	79.9	64.7	72.6	78.8	84.9	72.2	74.1	97.6	98.5
TSG	59.5	54.3	57.4	52.0	64.4	70.7	96.6	91.2	82.3	96.3
TSH	76.8	70.9	74.9	66.9	75.5	76.4	97.3	82.3	91.2	84.4

* Calculated at APFD/6-31G(d).

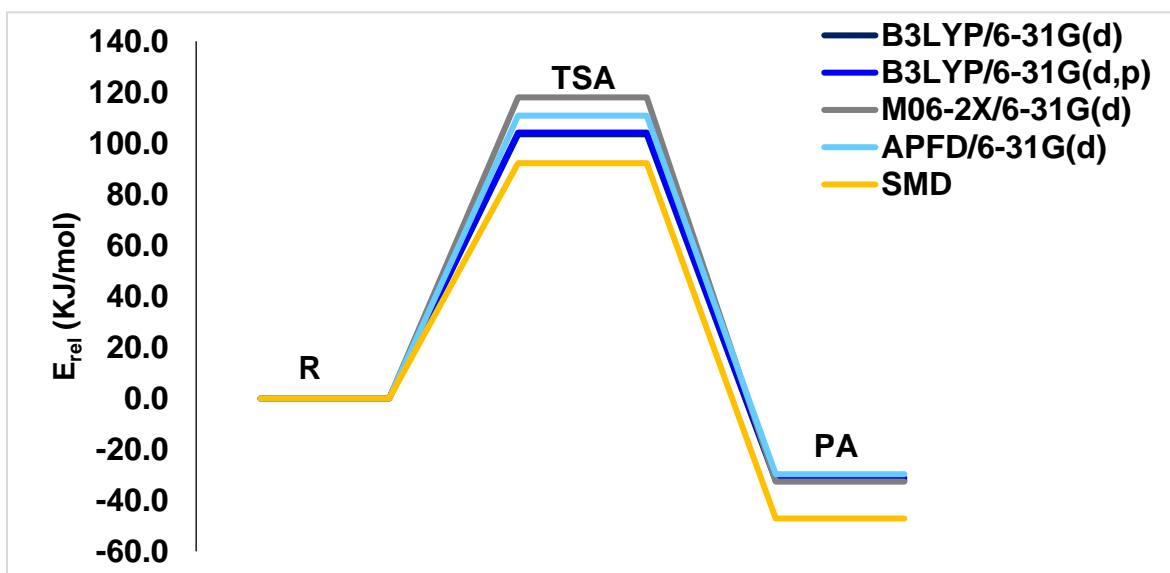


Figure 7: The PED of pathway **A** calculated at different levels of theory in aqueous and gaseous phases, in $\text{kJ}\cdot\text{mol}^{-1}$ at 298.15 K.

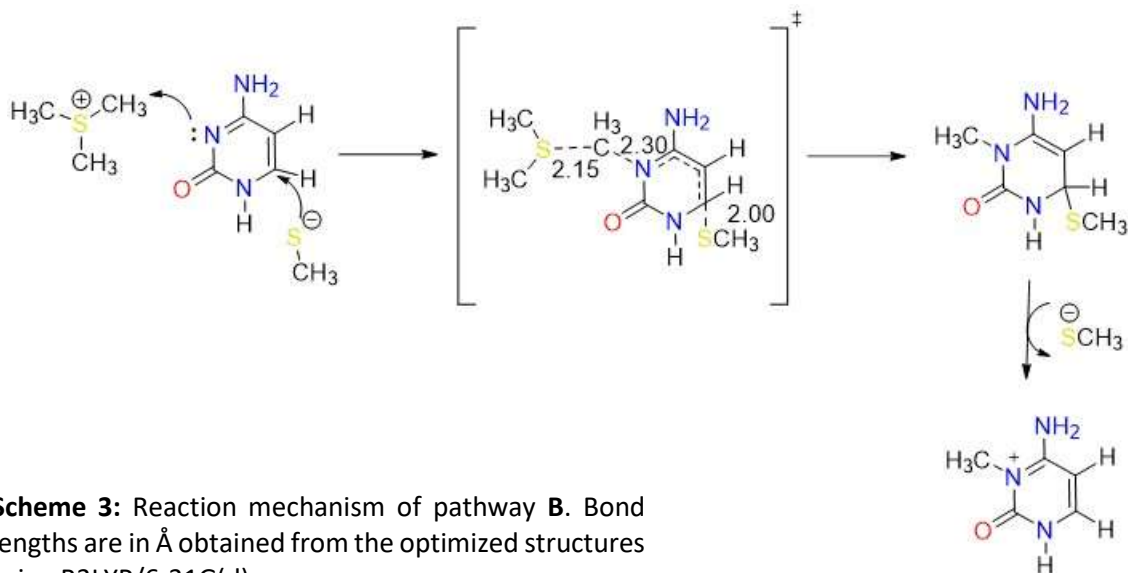
3.2 The Effect of Amino Acids Cys, Glu, Arg and Ala: Pathways B, C, and D

Herein, we performed a study that investigates the effect of adding amino acids (Cys, Glu, and Ala) on the activation energy. We started by mutating all amino acids and activated Cys (C710) as in pathway **B**. Then, we activated Glu (E756) in pathway **C** and then Ala (A791) in pathway **D**. We intended to compare the calculated activation energy of each pathway with the main reaction activation energies (pathway **A**).

Previous studies showed that cysteine (C710) residue significantly influences the cytosine ring activation for conventional 5mC generation, in which Cys initiate a nucleophilic attack at C6 to activate the C5 for methyl group transfe.[18–21] In the case of 3mC generation, however, its catalytic effects are insignificant. This was observed to noticeably increase the activation barrier for methyl addition, which was consistent with the results obtained experimentally by Dukatz and co-workers. Methylation of N3 in the presence of C710 denoted as pathway **B** was initiated by nucleophilic addition of Cys-S⁻ residue at C6 of the ring in concert with the methyl addition to the target N3 followed by the departure of Cys-S⁻.

Unlike the mechanism in pathway **A**, pathway **B** proceeds through a nearly perpendicular attack at 113.1° and syn-addition of AdoMet and C710, concerning the ring. In **TSB**, the distance between Cys-S⁻ and C6 was 2.00 Å, and the methyl group of AdoMet was 2.30 Å from N3 (see **Scheme 3**).

The activation energy of **TSB** was calculated as 140.6 kJ.mol⁻¹ using the B3LYP/6-31G(d) and was 144.4 kJ.mol⁻¹ at B3LYP/6-311G(d,p) level of theory. It has also been found that the M06-2X/6-31G(d) level of theory is much more sensitive toward this step and produces good results, with the activation energy being 125.3 kJ.mol⁻¹ comparable results were obtained using APFD/6-31G(d) level of theory (see **Figure S1** and **Table1**). Solvation model (SMD) showed an increase in the activation energy of **TSB** to 130.5 kJ.mol⁻¹. The increase in energy was attributed to the loss of π -conjugation over the ring. In comparison to pathway **A**, pathway **B** is unfavorable, and C710 has no catalytic effect. Considering pathways with lower values are taken as the most plausible mechanism.

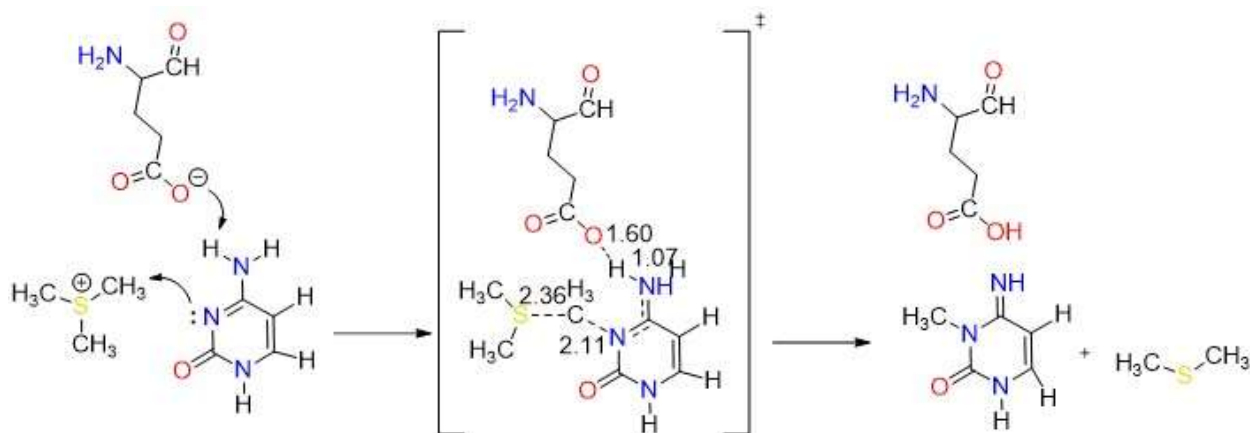


The catalytic role of glutamic acid residue (E756) was observed to be unique, although its role in conventional C5 methylation is still debatable.[22-23] Herein it enhances the reaction through a new approach, in which a proton (H4) abstraction occurs on the exocyclic amine group in concert with the methyl group addition. Removing H4 enhances the nucleophilicity of the ring to easily undergo the methylation reaction.

In **TSC**, H4 approaches the Glu acidic oxygen at 1.60 Å and C4-N4 become 1.33 Å for the double bond to be formed. Meanwhile, the methyl group is transferred at a distance of 2.11 Å from N3, as displayed in **Scheme 4** below.

The activation energy of **TSC** is lower than those of the previous pathways, with a value of 86.1 and 83.6 kJ.mol⁻¹ at the B3LYP/6-31G(d) and B3LYP/6-31G(d,p) level of theory, respectively (**Figure S2** and **Table 1**). The energy of the resulting imino-tautomeric product **PC** was calculated as -77.3 kJ.mol⁻¹ at the M06-2X/6-31G(d) level, which is much more stable than **PA** (-32.7 kJ.mol⁻¹ calculated at the same level of theory).

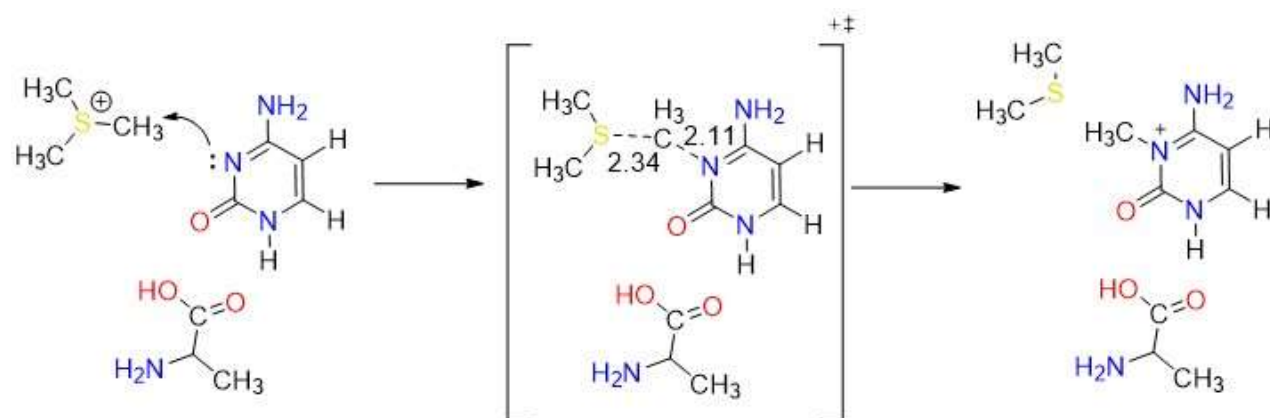
The **PC** is neutral product which provides extra stability, unlike the positively charged product of pathway **A**. This result is comparable with the data obtained from double proton transfer between DNA base pair (guanine-cytosine), which led to a stable but rare imino tautomer. Further, this has been suggested to cause a spontaneous mutation [24– 26].



Scheme 4: Reaction mechanism of pathway **C**. Bond lengths are in Å obtained from the optimized structures using B3LYP/6-31G(d).

The presence of Arg (R792) hinders the methylation of N3 and offers no assistance in 3mC production, despite the catalytic role and the stabilization effect that it can provide for the 5mC generation.[18] On the other hand, mutation of Arg to Ala was seen to boost the reaction dramatically, particularly, pathway **D**. This pathway proceeded as pathway **A** in one step. In **TSD** (**Scheme 5**), Ala provided extra stabilization for the cytosine to react through forming H-bonding with H1 of cytosine at 1.80 Å and at 2.24 Å from O2.

The activation barrier is lowered to 79.1 kJ.mol⁻¹ and 97.4 kJ.mol⁻¹ at B3LYP/6-31G(d) and B3LYP/6-31G(d,p), respectively (**Figure S3** and **Table 1**). These results were consistent with the experimental findings observed by Dukatz et al. after the substantial increase in 3mC/5mC ratio upon mutating Arg to Ala. Fortunately, this can support the proposed “inverted flipped base” approach, as the role of Arg is lost in stabilizing the ring and catalyzing the reaction. In light of this, the pronounced stability of the ring is attributed to Ala amino acid. The activation energy of **TSD** is lower than those calculated for **TSA** and **TSB**, then the existence of Ala is favorable for this reaction much more than the Glu (and certainly Cys).



Scheme 5: Reaction mechanism of pathway **C**. Bond lengths are in Å obtained from the optimized structures using B3LYP/6-31G(d).

3.3 The Catalytic Effect of Hydrogen Carbonate and Dihydrogen Phosphate Anions: Pathways E and F

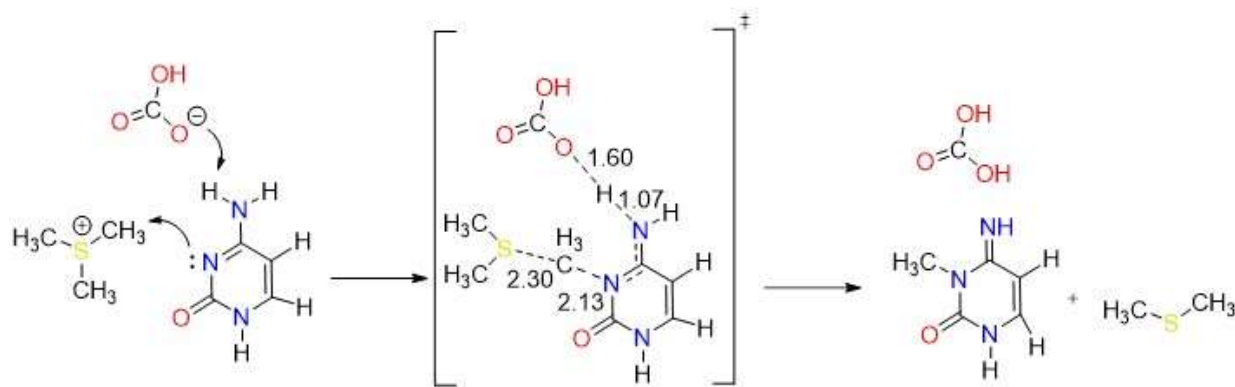
The role of potential catalyst molecules in the methylation of N3 was investigated; meanwhile, other amino acids are mutated, which are abundantly present *in vivo*. Hydrogen carbonate and dihydrogen phosphate ions can act as either acids or bases depending on the pH. At biological pH, phosphate can exist as dihydrogen phosphate ion (H_2PO_4^- , pKa 2.15), which is known to speed up the aging process, while carbonates exist as HCO_3^- (pKa 10.25). [17] These ions tend to follow the same proton transfer approach that has been described in pathway C (with E756) by forming the stable imino-tautomer concertedly with methylation at N3 (pathways E and F for carbonate and phosphate, respectively). **Scheme 6** demonstrate the mechanism of pathways E. Structurally, both **TSE** and **TSF** were remarkably similar. The reaction mechanism of pathway F has been presented in the supplementary information as **Scheme S1**.

Although pathways E and F proceeded as pathway C mechanism, but the ability of these proposed catalysts in lowering the activation barrier of the reaction is more significant and even much better than E756. Besides the lower bulkiness provided by hydrogen carbonate than E756. Hydrogen carbonate is the most basic (pKa=10.25); therefore, its ability to abstract H4 is higher than E756 (pKa around 5) and H_2PO_4^- (pKa=2.15).

On the other hand, the polarizability of the phosphorus of H_2PO_4^- enhances its ability to abstract the proton than E756.

At the APFD/6-31G(d) level of theory, the activation energy of **TSE** was calculated as 56.5 kJ.mol⁻¹ and reduced to 61.7 kJ.mol⁻¹ at B3LYP/6-31G(d,p) (see **Table 1**). In contrast, the activation energies of **TSF** are 66.5 and 64.7 kJ.mol⁻¹ at B3LYP/6-31G(d) and B3LYP/6-31G(d,p), respectively, see **Figure S4** and **S5**. It is worth noting that the products **PE** and **PF** are very stable with relative energies of -113.7 and -71.4 kJ.mol⁻¹, respectively, at APFD/6-31G(d) (which are the methylated rare tautomer).

It is important to address that the proton transfer mechanism is affected by the environment. [24] Despite many tries, using SMD model **TSC**, **TSE** and **TSF** did not show proton transfer. Taking the effect of solvation into account, the activation energies of **TSE** and **TSF** using SMD model were 77.2 and 97.6 kJ.mol⁻¹, respectively. This may be reflected in the higher activation energy of these transition states solvation models.



Scheme 6: Reaction mechanism of pathway E. Bond lengths are in Å obtained from the optimized structures using B3LYP/6-31G(d).

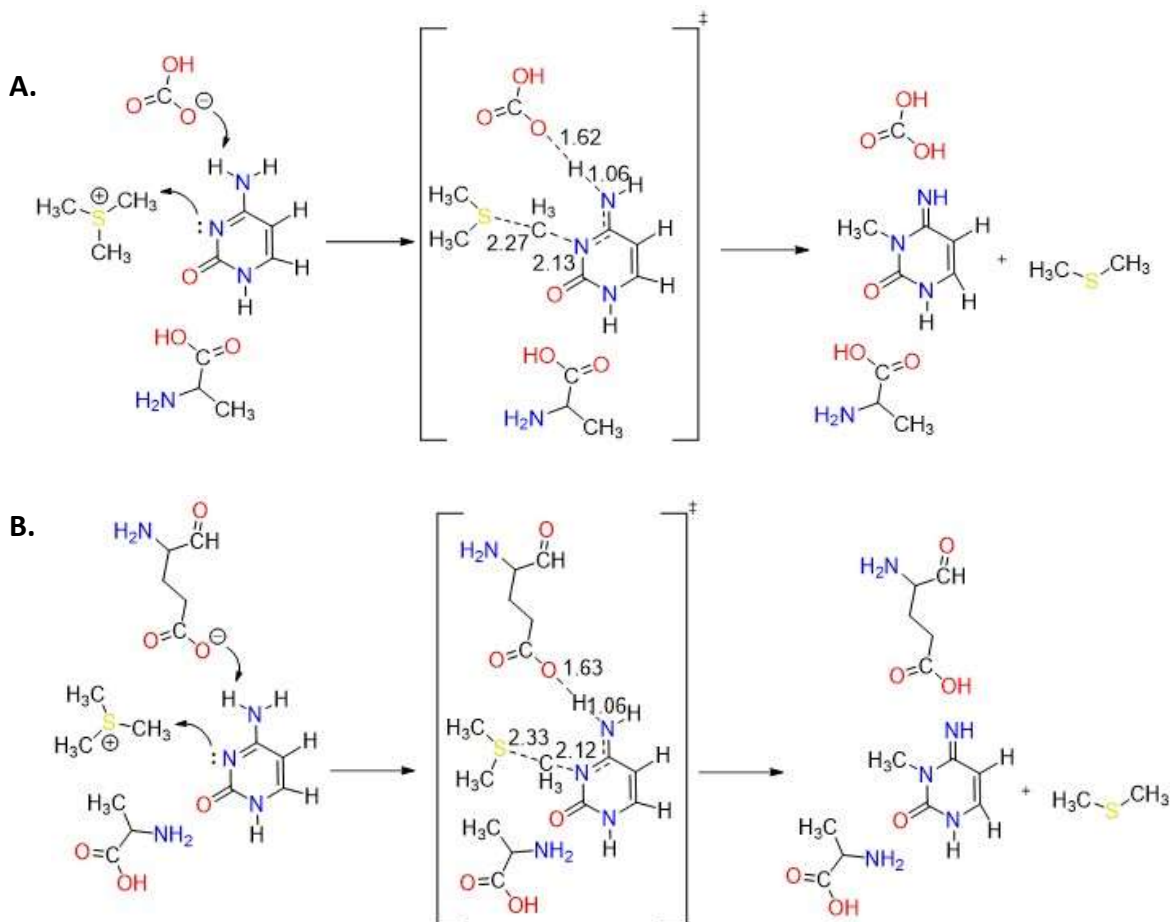
3.4 The Collaborative Role of Amino Acids and Catalysts: Pathways G→J

So far, we have studied the impact of subunits available in the active site and mutate others to the methylation reaction to obtain the complete picture of the N3 methylation. According to the results obtained from the previously discussed pathways (A→F), we have considered the calculations including these units as an entire system to elucidate the properties of the active sites. Therefore, pathways **G** and **H** were assigned to mutate Arg to Ala simultaneously with activating HCO_3^- and Glu, respectively. While pathways **I** and **J** were designated by activating Arg with HCO_3^- and Glu. The reaction mechanism of pathways **G** and **H** are shown in **Schemes 7A** and **7B**, respectively.

The PED for pathways **G** and **H** are reported in **Figure S6** and **S7**. The data for pathways **I** and **J** are included in the Supplementary Information in **Table S1**. Consistent with our previous results, the presence of Ala stabilizes the cytosine ring and catalyzes the reaction to significantly drop down the activation energy barrier compared to that with R792. In **TSG**, Ala formed H-bonding with H1 at 1.87 Å and the carbonyl carbon O2 of cytosine at 2.04 Å, offering further stabilization to the transition state.

In **TSH**, the amine hydrogen from Ala side approaches the O2 of the cytosine ring with 2.01 Å as shown in **Scheme 7B**.

It was also found that HCO_3^- can reduce the energy values more than E756. As E756 is less basic than HCO_3^- and adds extra bulkiness around the ring.



Scheme 7: Reaction mechanism of: **A.** pathway **G** and **B.** pathway **H**. Bond lengths are in Å obtained from the optimized structures using B3LYP/6-31G(d).

The lowest activation energies obtained for **TSG** and **TSH**, are 59.5 and 76.8 kJ.mol⁻¹ at the B3LYP/6-31G(d) level of theory, and 64.4 and 75.5 kJ.mol⁻¹ at M06-2X/6-31G(d) level of theory, respectively. The results calculated using B3LYP/6-31G(d) and M06-2X/6-31G(d) are highly correlated for these pathways. However, utilizing APFD/6-31G(d) shows an increase in the activation energy with no reducing effect when a diffuse or polarization functions are considered (See **Table 1**, **Figures S6 and S7**). Taking the effect of solvation using SMD model, the barrier for **TSG** was dramatically lowered to -63.3 kJ.mol⁻¹ while found at 91.2 kJ.mol⁻¹ for **TSH**. Consequently, Ala and HCO₃⁻ are the most suitable candidates for catalyzing the reaction and lowering the activation energy.

Despite the occurrence of the proton transfer during the methyl transfer in pathways **I** and **J**, the effect of Arg was much more pronounced in increasing the activation energy of **TSI** and **TSJ** and diminishing the catalytic role of HCO₃⁻ and Glu. By the withdrawal effect of positively charged Arg making the ring more electrophilic. The activation energy of **TSI** and **TSJ** were studied using B3LYP/6-31G(d) and M06-2X/6-31G(d) as displayed in **Table S1** and **Figures S8 and S9**. (supplementary)

Using B3LYP/6-31G(d) the activation energies were 183.2 and 175.2 kJ.mol⁻¹ of **TSI** and **TSJ**, respectively. In comparison, the calculated values at M06-2X/6-31G(d) were 162.9 and 187.6 kJ.mol⁻¹. The obtained results were highly correlated and significantly higher than those observed while mutating Arg to Ala. Therefore, we found that it would be a computational and time waste to investigate these pathways further.

According to the experimental results obtained for 3mC generation, it was observed that the rate of 3mC generation was decreased when the impact of Arg is considered and further explained by the loss of the stabilizing H-bond between this amino acid and the ring [2].

Regarding the effect of polarization and diffuse functions, we have performed geometry optimization using APFD/6-31+G(d) and APFD/6-31G(2df,2p) levels of theory. According to our

results, these functions did not decrease the activation barriers for the studied pathways (see **Table S2** in the Supplementary Information (SI)).

3.5 Thermodynamic Parameters for Methylation of Cytosine at N3

The thermodynamic parameters (DH and DG) for the methylation of cytosine at position 3 along with its proposed pathways are given in **Table 2**, at all studied levels of theory.

The methylation reaction of cytosine at position 3 (pathway **A**) and in the presence of amino acids (Glu and Ala) as well as when catalytic anions (hydrogen carbonate and dihydrogen phosphate) are utilized in pathways (**A→H**) are found to be exothermic and exergonic at all levels of theory.

However, the cooperation of Arg in pathway **I**, yielded endothermic and endergonic results at B3LYP/6-31G(d) level of theory. While pathway **J**, it results in exothermic and exergonic reactions at studied levels of theory but with higher energy than other pathways. As a result, both pathways are not favored for the 3mC production direction.

According to the results obtained in **Table 2**, pathways **E** and **G** had the lowest thermodynamic parameter values, therefore, they are more spontaneous and plausible reactions to occur.

The solvation process, modeled by SMD, did not decrease but increase the thermodynamical parameters of all pathways.

Table 2: Thermodynamic parameters for methylation of cytosine at N3 position for all proposed pathways. Energies are in kJ.mol ⁻¹ .										
Pathway	B3LYP/6-31G(d)		B3LYP/6-31G(d,p)		M06-2X/6-31G(d)		APFD/6-31G(d)		SMD	
	ΔH	ΔG	ΔH	ΔG	ΔH	ΔG	ΔH	ΔG	ΔH	ΔG
A	-32	-35	-31	-34	-33	-35	-30	-30	-47	-48
B	-97	-99	-91	-93	-121	-117	-107	-106	-65	-60
C	-79	-97	-82	-98	-77	-83	-84	-77	-37	-43
D	-50	-61	-31	-44	-41	-52	-31	-49	-47	-50
E	-114	-133	-117	-134	-124	-127	-114	-120	-49	-38
F	-83	-86	-89	-96	-82	-91	-71	-82	-9	-18
G	-124	-131	-128	-134	-125	-133	-77	-86	-44	-43
H	-89	-97	-91	-101	-126	-121	-84	-89	-51	-54

4. Conclusions

In this work, we have performed an extensive study to investigate and provide mechanistic details concerning the methylation reaction of cytosine on position 3 utilizing accurate quantum chemical DFT calculations. The role of conserved residues and abundant anionic particles (hydrogen carbonate and dihydrogen phosphate) were investigated through designing pathways **A** \rightarrow **J** for 3mC generation. The potential energy diagram (PED) for each pathway was constructed employing the B3LYP/6-31G(d), M06-2X, APFD methods in the gaseous phase besides considering the aqueous environment SMD model. The thermodynamic functions (ΔH and ΔG) and kinetic parameters (E_a , ΔH^\ddagger , and ΔG^\ddagger) were calculated, using these DFT methods, for each proposed pathway. The connections of the TS's with the I's, R's, and P's of every pathway have been confirmed using the intrinsic reaction coordinate (IRC) calculations.

In this work, we have conducted a mechanistic exploration of the methylation of N3 position at the cytosine ring. Our results suggest that the factors that influence the N3 methylation play different roles in C5 methylation. For instance, for 3mC formation Ala and Glu showed catalytic activity toward the N3 reaction while Cys and Arg did not show any catalytic activity in the formation of 3mC. On the other hand, Cys and Arg play a key role in 5mC formation reaction. Methylation of N3 occur through electrophilic addition mechanism concerted with proton H4 transfer to the surrounding employed residues, which were Glu, HCO_3^- and H_2PO_4^- . The calculated results showed that HCO_3^- and Ala residue proved a powerful catalytic activity in lowering the activation energy to 59.5 kJ.mol⁻¹ employing B3LYP/6-31G(d) level of theory. The integrated proton transfer mechanism between cytosine and the surrounding molecules and methylation of N3 can double the trouble through introducing a serious mutagenic lesion. The agreement of our results with the experimental findings obtained by Dukatz and co-workers gives credence to our framework, which will pave the road to control these reactions and help drug design processes.

Disclosure of the mechanistic details of N3 methylation of cytosine opens the door to clearly understand the recovery process by ALKB2 enzyme.

Data Availability

The optimized structures of all pathways for DNA- Methylation are available at <https://github.com/matarneh/DNA-ethylation> to facilitate reproducibility of the results.

Conflicts of Interest

The authors declare no competing financial interests.

Acknowledgments

Mansour H. Almatarneh is grateful to the Deanship of Academic Research at the University of Jordan for the grant. We also gratefully acknowledge Compute Canada for the computer time.

Supporting Information

Potential energy diagrams for pathways **B→J** at different levels of theory. Kinetic parameters for pathways **A→H** calculated at B3LYP/6-31G(d) and M06-2X/6-31G(d) levels of theory. The cartesian coordinates for all optimized structures of all pathways to facilitate reproducibility of the results.

References

- [1] S. Rošić, R. Amouroux, C. E. Requena, A. Gomes, M. Emperle, T. Beltran, J. K. Rane, S. Linnett, M. E. Selkirk, P. H. Schiffer, A. J. Bancroft, R. K. Grencis, A. Jeltsch, P. Hajkova, P. Sarkies, *Nature Genetics*, **2018**.
- [2] M. Dukatz, C. E. Requena, M. Emperle, P. Hajkova, P. Sarkies, A. Jeltsch, *Journal of Molecular Biology*, **2019**.
- [3] S. L. Nay, D.-H. Lee, S. E. Bates, T. R. O'Connor, *DNA Repair*, **2012**.
- [4] A. Furrer, B. van Loon, *Nucleic Acids Research*, **2014**.
- [5] M. V. Yermolina, G. A. Papadantonakis, *Chemical Physics Letters*, **2020**.
- [6] M. H. Almatarneh, C. G. Flinn, R. A. Poirier, W. A. Sokalski, *J. Phys. Chem. A*, **2006**.
- [7] M. H. Almatarneh, C. G. Flinn, R. A. Poirier, *J. Chem. Inf. Model.*, **2008**.
- [8] K. M. Uddin, M. H. Almatarneh, D. M. Shaw, R. A. Poirier, *J. Phys. Chem. A*, **2011**.
- [9] A. I. Alrawashdeh, M. H. Almatarneh, R. A. Poirier, *Can. J. Chem.*, **2013**.
- [10] M. A. Halim, M. H. Almatarneh, R. A. Poirier, *J. Phys. Chem. B*, **2014**.
- [11] M. H. Almatarneh, A. A.-A. A. Abu-Saleh, I. A. Elayan, *Computational and Theoretical Chemistry*, **2017**.
- [12] M. H. Almatarneh, A. A.-A. A. Abu-Saleh, K. M. Uddin, R. A. Poirier, P. L. Warburton, *International Journal of Quantum Chemistry*, **2017**.

- [13] M. H. Almatarneh, R. A. Omeir, S. AL Demour, I. A. Elayan, S. Islam, R. A. Poirier, *Computational and Theoretical Chemistry*, **2020**.
- [14] L. C. Sowers, W. D. Sedwick, B. R. Shaw, *Mutation Research*, **1989**, 131–138.
- [15] K. S. Ekanayake, P. R. Lebreton, *Journal of Computational Chemistry*, **2006**.
- [16] D. R. Eichler, G. A. Papadantonakis, *Chemical Physics Letters*, **2017**.
- [17] K. Kato, T. Nakayoshi, E. Kurimoto, A. Oda, *ACS Omega*, **2019**.
- [18] X. Zhang, T. C. Bruice, *Proc Natl Acad Sci USA*, **2006**.
- [19] M. Peräkylä, *J. Am. Chem. Soc.*, **1998**.
- [20] R. Zangi, A. Arrieta, F. P. Cossío, *Journal of Molecular Biology*, **2010**, DOI:10.1016/j.jmb.2010.05.021.
- [21] Q. Du, Z. Wang, V. L. Schramm, *Proc Natl Acad Sci USA*, **2016**.
- [22] J. Aranda, K. Zinovjev, K. Świderek, M. Roca, I. Tuñón, *ACS Catal.*, **2016**.
- [23] F.-K. Shieh, N. O. Reich, *Journal of Molecular Biology*, **2007**.
- [24] J. P. Cerón-Carrasco, A. Requena, J. Zúñiga, C. Michaux, E. A. Perpète, D. Jacquemin, *J. Phys. Chem. A*, **2009**.
- [25] P.-O. Löwdin, Quantum Genetics and the Aperiodic Solid: Some Aspects on the Biological Problems of Heredity, Mutations, Aging, and Tumors in View of the Quantum Theory of the DNA Molecule**Sponsored in part by the King Gustaf VI Adolf's 70-Years Fund for Swedish Culture, Knut and Alice Wallenberg's Foundation, and in part by the Cancer Institute of National Institutes of Health, under Contract CA 06850–01 with Uppsala University., *Advances in Quantum Chemistry*, 2. Academic Press, , 213–360, **1966**.
- [26] P.-O. Löwdin, *Rev. Mod. Phys.*, **1963**.