doi:10.1016/j.jmb.2010.05.021 J. Mol. Biol. (2010) 400, 632–644

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

# Mechanism of DNA Methylation: The Double Role of DNA as a Substrate and as a Cofactor

**Ronen Zangi[1,2]*, Ana Arrieta[1] and Fernando P. Cossío[1]**

[1]*Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018 San Sebastian, Spain*

[2]*IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain*

*Received 8 March 2010; received in revised form 29 April 2010; accepted 9 May 2010*
Available online
21 May 2010

Methylation of cytosine residues in the DNA is one of the most important epigenetic marks central to the control of differential expression of genes. We perform quantum mechanical calculations to investigate the catalytic mechanism of the bacterial *HhaI* DNA methyltransferase. We find that the enzyme nucleophile, Cys81, can attack C6 of cytosine only after it is deprotonated by the DNA phosphate group, a reaction facilitated by a bridging water molecule. This finding, which indicates that the DNA acts as both the substrate and the cofactor, can explain the total loss of activity observed in an analogous enzyme, thymidylate synthase, when the phosphate group of the substrate was removed. Furthermore, our results displaying the inability of the phosphate group to deprotonate the side chain of serine is in agreement with the total, or the large extent of, inactivity observed for the C81S mutant. In contrast to results from previous calculations, we find that the active site conserved residues, Glu119, Arg163, and Arg165, are crucial for catalysis. In addition, the enzyme–DNA adduct formation and the methyl transfer from the cofactor *S*-adenosyl-L-methionine are not concerted but proceed via stepwise mechanism. In many of the different steps of this methylation reaction, the transfer of a proton is found to be necessary. To render these processes possible, we find that several water molecules, found in the crystal structure, play an important role, acting as a bridge between the donating and accepting proton groups.



*Keywords:* DNA cytosine methyltransferase; methylation reaction mechanism; DNA as cofactor; DNA phosphate group; quantum mechanical calculations

***Edited by D. Case***

## Introduction

In eukaryotes, methylation of the DNA at cytosines is an epigenetic mark that plays an important role in embryonic development, regulation of gene expression, X-chromosome inactivation, and genomic imprinting.[1,2] The methylation reaction, catalyzed by DNA methyltransferases (DMTs), proceeds by a transfer of a methyl group from *S*-adenosyl-L-methionine (AdoMet) to position 5 of the cytosine ring. An outline of the catalytic mechanism (shown in Fig. 1) has been proposed based on analogy to the enzymatic methylation reaction of 2′-deoxyuridine monophosphate by thymidylate synthase (TS).[3,4] Later, the crystal structure (presented in Fig. 2) of *HhaI* methyltransferase (M.*HhaI*) bound to the substrate, DNA, and the cofactor AdoMet confirmed this mechanism and provided more details.[5,6] Nevertheless, many issues in regard to the way DMT executes the methyltransfer reaction remain unclear. A remarkable feature of this DNA–enzyme interaction is the flipping of the target cytosine base out of the DNA helix, which then can fit into the active site of the enzyme.[7,8] This is necessary because, in normal B-DNA structure, carbon-5 of cytosine is buried too deep inside the DNA helix, preventing the methylation reaction.

In general, the carbon at position 5 of the cytosine ring is not a strong nucleophile that can react with the electrophile methyl donor AdoMet. However, a nucleophile from the enzyme can attack the cytosine ring at position 6 (Michael or conjugated addition reaction) forming a covalent bond between the

*Corresponding author. E-mail address: r.zangi@ikerbasque.org.

Abbreviations used: DMT, DNA methyltransferase; AdoMet, *S*-adenosyl-L-methionine; TS, thymidylate synthase; M.*HhaI*, *HhaI* methyltransferase; dUMP, 2′-deoxyuridine 5′-monophosphate.

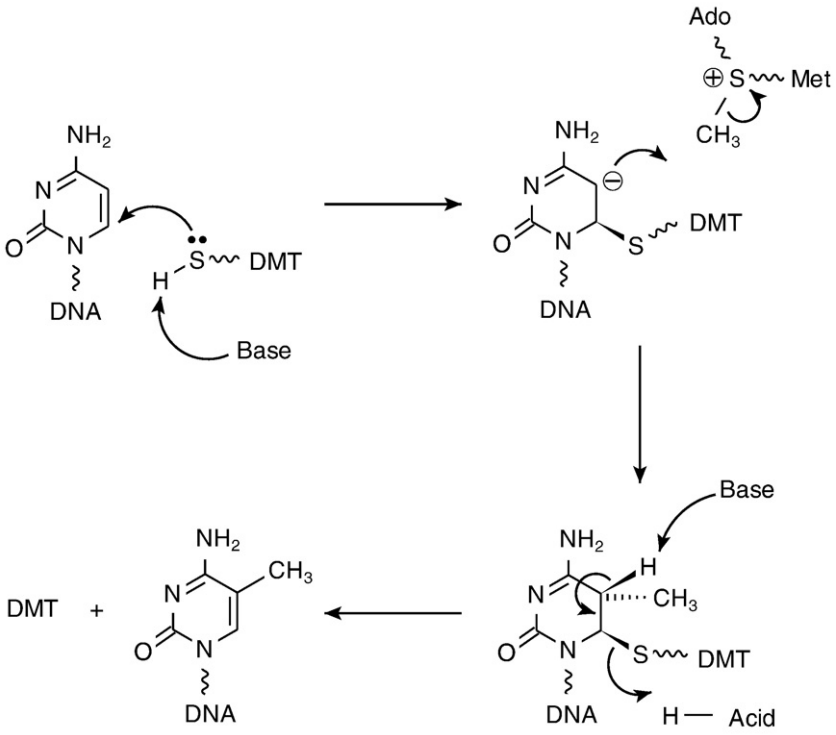


**Fig. 1.** Proposed schematic representation of the methylation reaction catalyzed by DMT. In the first step, a nucleophile from the enzyme (Cys81) attacks position 6 of cytosine. This activates position 5 for an electrophilic attack by the cofactor, AdoMet, where a methyl group is transferred to cytosine. Subsequent elimination of the enzyme nucleophile and abstraction of the proton at position 5 yield the methylated cytosine and the enzyme.

enzyme and the DNA. In this covalent adduct, the nucleophilic character of C5 of cytosine is activated for an attack on AdoMet and a methyl group is transferred. Subsequent elimination of the enzyme nucleophile and deprotonation at position 5 resolve the nucleotide–enzyme complex. Sequence alignment of DMT from 18 different sources found six highly conserved motifs.[9,10] These building blocks exhibit the same order, and between certain motifs even the same spacings, in the different DMTs. Five of these motifs are part of the active site and the AdoMet binding pocket, which suggests that all cytosine-5 DMTs share the same mechanism for catalyzing the methylation reaction.

Five amino acid residues that are part of the active site [Cys81, Glu119, Val121, Arg163, and Arg165 (sequence numbering of M.*Hha*I)] are completely conserved.[10] It is very well established that Cys81 is the enzyme nucleophile that attacks C6 of cytosine, forming a transient covalent complex with the DNA.[3,5,6,11–15] Mutation of Val121 to alanine suggests that the role of this valine residue is to stabilize the extrahelical cytosine that binds into the active site of the enzyme.[16] The other active site conserved residues, Glu119, Arg163, and Arg165, were suggested to facilitate the nucleophilic attack of Cys81 on C6 of cytosine. This is obtained by increasing the electrophilic character of C6 through electron delocalization (see Fig. 3). Arg163 and Arg165 were suggested to activate C6 by electrostatic interactions with O2 of cytosine,[15] while Glu119 does so by mediating protonation of N3 of cytosine.[6] Mutation studies of the enzyme active site concluded that Arg165 and Glu119 are important for catalysis, cytosine positioning, and DNA binding.[5,17] Nevertheless, a QM/MM study found that Glu119 has no catalytic function, while Arg163 and Arg165 displayed only a minor assistance, for the methylation reaction.[18] It was, therefore, concluded that the role of these residues is only to create the reactive initial state conformer. Earlier quantum mechanical calculations also investigated the catalytic steps of DMT methylation reaction.[19] However, in this study, small models of the reacting molecules were used without considering the effect of the enzyme active-site residues on catalyzing the different steps in the reaction.

Although there is no doubt in regard to the identity of the first nucleophile attacking C6 of cytosine, the protonation state of the sulfur atom of Cys81 during this attack is not established. In the majority of the studies, the nature of the attacking sulfur atom is ambiguous or not specified.[3,6,12,16,17,20] In other works, it is assumed that a thiol[4,15] (neutral cysteine), while in others a thiolate[13,18,19] (anionic cysteine), attacks position 6 of the cytosine ring. The p$K_a$ value of the side chain of cysteine in aqueous solution is about 8.4,[21] and inside the active site, away from bulk water molecules, the sulfur atom is expected to be protonated. Within this context, the questions we are interested to address are the following: Is the thiol group a strong enough nucleophile to attack the cytosine ring? Is there a potential base in the vicinity for deprotonating the thiol group before its attack on C6? What is the role of the conserved active site amino acid residues Glu119, Arg163, and Arg165 in catalyzing the reaction? Additional unresolved issue is the final elimination step of the methylation reaction. The enzyme nucleophile, Cys81, attacks the cytosine ring on the face that is opposite to that attacked by the cofactor,

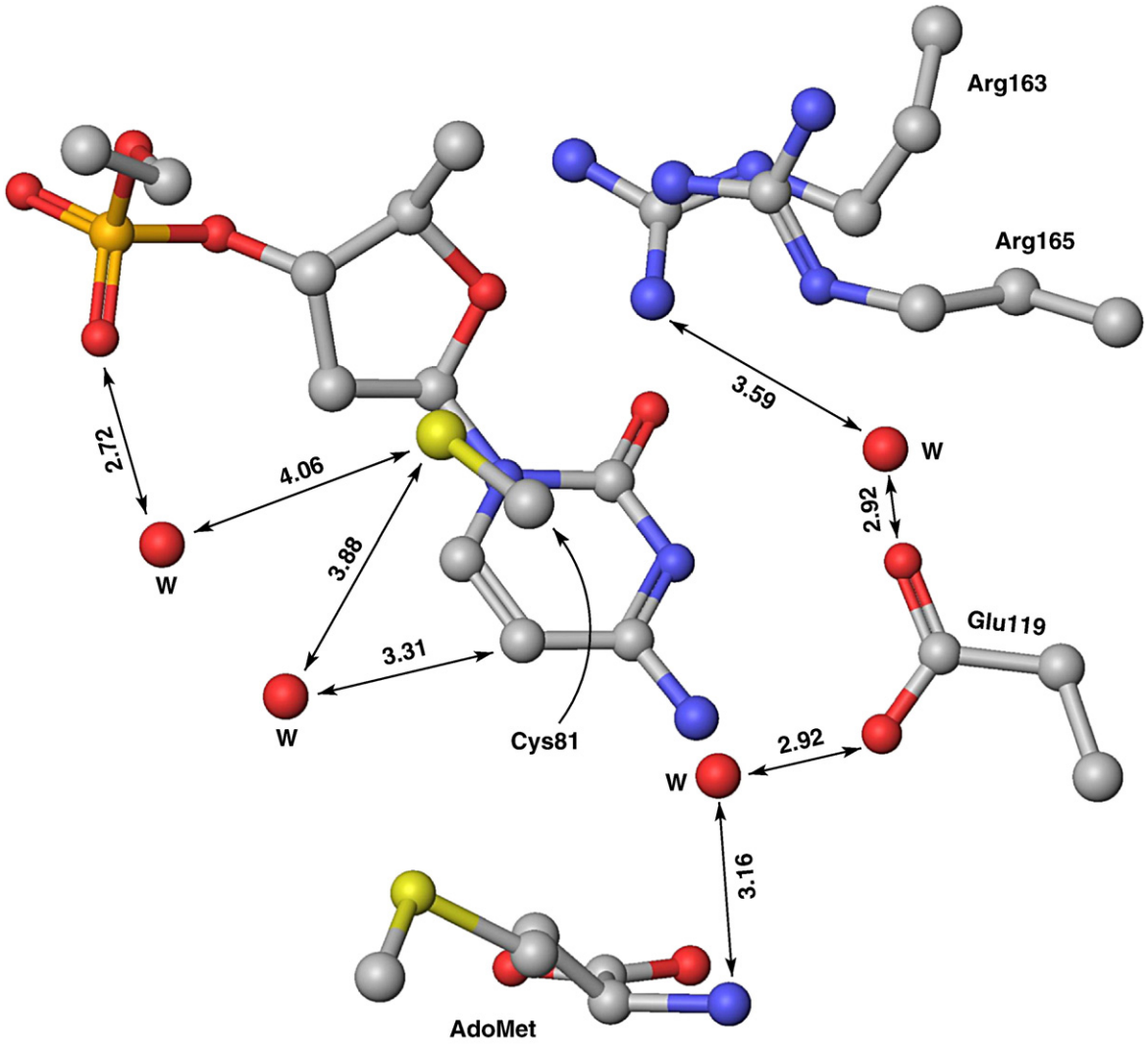


**Fig. 2.** Crystal structure of cytosine bound to the active site of M.*HhaI*.[5] Only the side chains of the conserved residues Cys81, Glu119, Arg163, and Arg165 and the methionine moiety of AdoMet are shown. We also present four of the active-site water molecules (W) that we find in this article to be important for catalysis, as well as their distances to relevant surrounding atoms. The distances indicated are in angstroms.

AdoMet.[5,6] This would require the reaction to proceed via the energetically unfavorable syn elimination. Inspection of the crystal structure of M.*HhaI* reveals no suitable base that can abstract the hydrogen atom at position 5 of the cytosine ring. Involvement of a phosphate group or a water molecule has been suggested,[6] as well as hydroxide (from autoionization of bulk water) that diffuses to the active site through a water channel.[18]

In this article, we perform quantum mechanical calculations to investigate the catalytic mechanism of M.*HhaI*. We find that in order to attack C6 of cytosine, the thiol group of the enzyme nucleophile, Cys81, has to be deprotonated. The negatively charged DNA phosphate group, mediated by a bridging water molecule, acts as the proton abstracting base (the proton transferred might then be donated to the solvent water molecules). This finding, which indicates that the DNA acts as both the substrate and the cofactor, can explain the total loss of activity observed in analogous enzyme, TS, when the phosphate group of the substrate was removed. Furthermore, our results displaying the inability of the phosphate group to deprotonate the side chain of serine is in agreement with the total, or the large extent of, inactivity observed for the C81S mutant. In contrast to results from previous calculations, we find that the formation of the enzyme–DNA covalent complex and the methyl addition from the cofactor, AdoMet, are not concerted but proceed via stepwise mechanism. In addition, we demonstrate that the active site conserved residues Glu119, Arg163, and Arg165 are crucial for catalysis. The activation of many of the different steps in this methylation reaction depends on the ability to transfer a proton. It is shown that water molecules in the active site, which are found in the X-ray structure of M.*HhaI* (Fig. 2), play an important role in facilitating several of these proton transfer reactions.

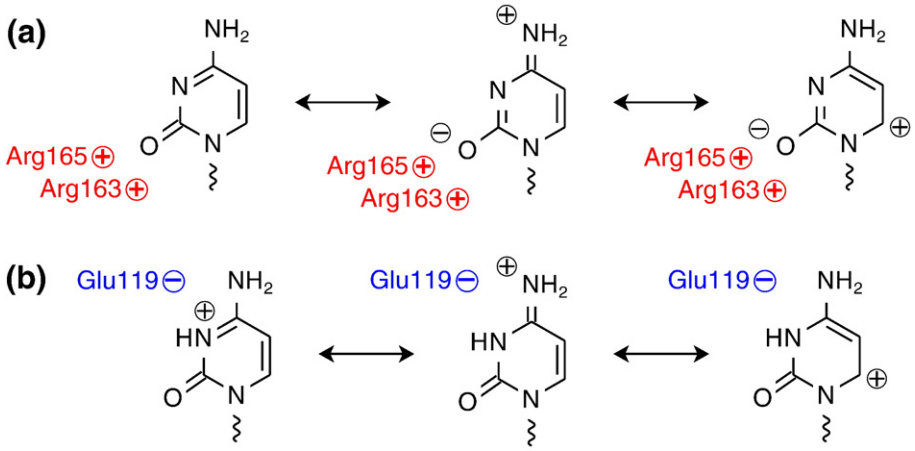


**Fig. 3.** Possible schemes for activating C6 of cytosine by different conserved residues in the active site of DMT. In both cases, the activation is a result of electron delocalization by (a) electrostatic interactions of O2 with Arg163 and Arg165 and (b) N3 protonation mediated by Glu119.

## Results and Discussion

### The enzyme nucleophilic attack

In many studies, it was shown that Cys81 is the enzyme nucleophile that attacks C6 of cytosine. Nevertheless, the protonation state of the attacking sulfur atom was not addressed. We performed many transition-state optimizations corresponding to this attack where different conserved residues were placed around the cytosine ring for its activation. However, except for one case (see below), no proper transition state was found. Therefore, we calculated two potential energy profiles for this reaction. In the first one, a thiol group (-SH) attacks position 6 of cytosine, while in the second, the attacking group was thiolate ($-S^-$). A thiolate group is a better nucleophile than a thiol group, and indeed, it was used as the enzyme nucleophile in the previous two theoretical studies on the mechanism of DNA methylation.[18,19] Therefore, in order to facilitate this reaction for a thiol group as much as possible, we included Glu119 (modeled as acetate) around the cytosine ring where N3 is protonated, as well as Arg163 and Arg165 (modeled as positively charged guanidines). Figure 3 displays the potential activation through electron delocalization by these residues. The energy profiles as a function of the distance between the sulfur atom of methyl thiol/thiolate and C6 of cytosine are shown in Fig. 4. As this distance decreases, the potential energy increases in the case of methyl thiol while it decreases for methyl thiolate. These optimizations were performed, probably, with maximum potential activation of C6. Therefore, it is very unlikely that a neutral thiol group can attack C6 of cytosine. We also tried to find a concerted transition state for this reaction in which the proton from the thiol group is transferred to a water molecule or to a phosphate group of the DNA but were unsuccessful. Therefore, we conclude that Cys81 in a deprotonated thiolate form is the enzyme nucleophile attacking C6 of cytosine.

Given the above finding, a suitable base that can abstract the proton from the thiol group of Cys81 needs to be identified. From inspection of the crystal structure of the ternary complex[5] of DMT, bound DNA, and *S*-adenosyl-L-homocysteine, we did not find such a group within a reasonable distance from the thiol group, which is part of the enzyme. However, the negatively charged phosphate group of the DNA from the same, or the next, nucleotide as the target cytosine can act as the base abstracting the thiol proton of Cys81. This proton can subsequently be transferred from the phosphate group to solvent water molecules. The distance in the crystal structure between the sulfur atom and the closest negatively charged oxygen of the phosphate group on the same nucleotide is 5.38 Å, and that on the next

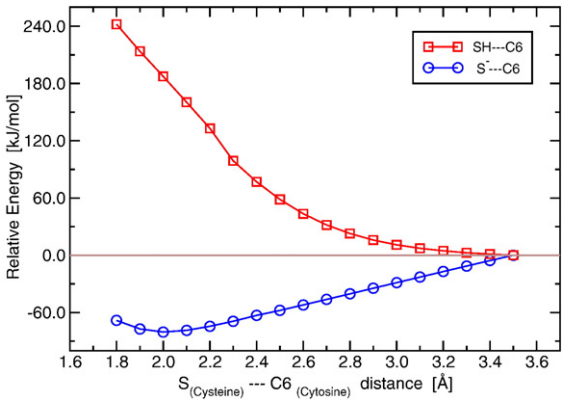


**Fig. 4.** Relative potential energy of the reaction coordinate of the attack of Cys81 via thiol (-SH), as well as, thiolate ($-S^-$), on the C6 of cytosine. In both cases, the cytosine is activated by (i) protonation at position N3, hydrogen bonded to the carboxylic group of Glu119, and (ii) strong electrostatic interactions at position O2 through the positively charged guanidinium groups of Arg163 and Arg165.

nucleotide is 4.65 Å. The existence of a water bridge between these atoms (Fig. 2) can, in principle, permit this proton transfer reaction. Therefore, we choose to model this reaction computationally. The model system includes the methyl thiol of Cys81, a water molecule, the nucleoside (sugar and base) of the target cytosine, the phosphate group of the next nucleotide, and methyl carboxylate representing Glu119 (see Fig. 5). In order to preserve the conformation found in the crystal structure, we constrained two distances: one between the sugar and the methyl attached to the phosphate group and the other between the methyl thiol and the cytosine ring. A transition state is obtained with a barrier of 31.4 kJ/mol. The structure is shown in Fig. 5. The product has an energy of 13.2 kJ/mol above the reactant. Including interactions with the environment through reaction field calculation increases the barrier to 40.8 kJ/mol. Therefore, a proton transfer from a thiol group of Cys81 to a phosphate group of the DNA is possible with hardly any need for a conformational change (relative to the X-ray structure) of the enzyme–DNA complex. A similar calculation but without the constraint on the distance between the sulfur atom of Cys81 and C6 of cytosine resulted in a lower energy barrier of 23.6 kJ/mol (37.4 kJ/mol with the solvent effect). However, the structure of the transition state exhibits a large deviation compared to the crystal structure.

One can argue that, alternatively, a water molecule in the active site can abstract the thiol group proton. To address whether this is possible, we performed a search for a transition state but without the sugar and the phosphate groups. The calculations met with no success. Therefore, we constructed a potential energy profile for the proton transfer reaction from a methyl thiol to a water molecule. This is shown in Fig. 6a as a function of the distance between the sulfur atom and the proton to be

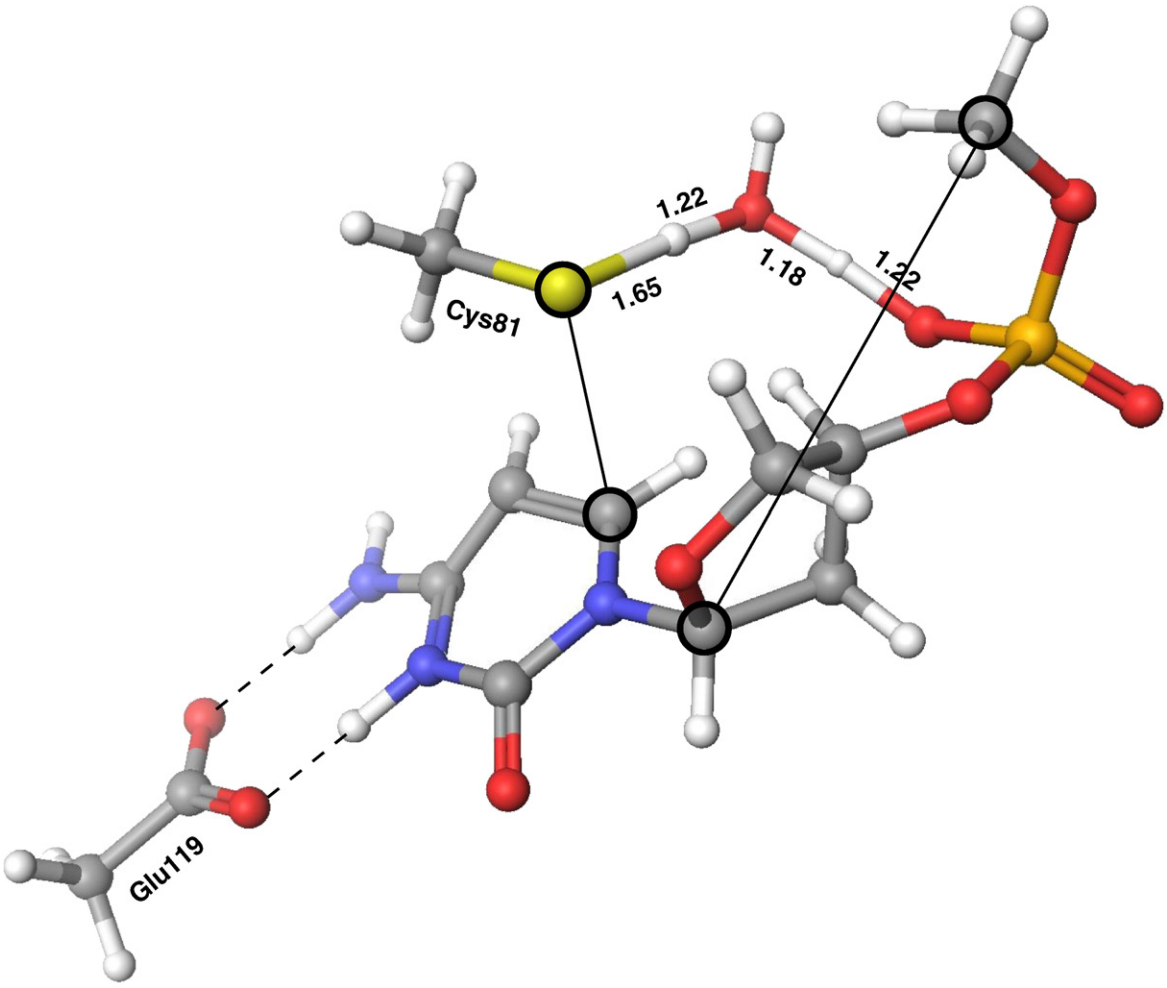


**Fig. 5.** Structure of the transition state ($\nu_{TS} = 1069i$ cm$^{-1}$) of the proton transfer from the methyl thiol (Cys81) to the DNA phosphate group through a water molecule. The distances of the two atom pairs, which are circled and connected by black lines, were constraint during the optimization search to the values obtained from the crystal structure. The corresponding barrier height of the reaction is 31.4 kJ/mol (40.8 kJ/mol when accounting for the environment). Removing the constraint between the sulfur atom and C6 of cytosine lowers the barrier height to 23.6 kJ/mol (37.4 kJ/mol with solvent effect). The distances indicated are in angstroms.

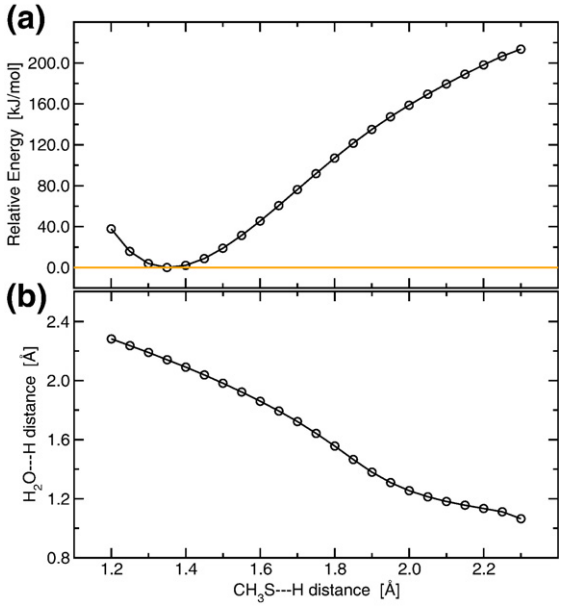


**Fig. 6.** A hypothetical proton transfer from methyl thiol to a water molecule (without the presence of a phosphate group). (a) Relative energy and (b) the H–OH[2] distance, as a function of the reaction coordinate, $CH_3S$–H distance.

transferred. The optimized distance between the proton to be transferred and the water molecule is shown in Fig. 6b, indicating that the water molecule accepted the proton and converted into hydronium. From these graphs, it is clear that such reaction is not feasible since it corresponds to an uphill energy curve with a very large activation energy (larger than 200 kJ/mol).

The results supporting the involvement of the phosphate group in deprotonating the enzyme nucleophile are intriguing since they demonstrate for the first time that the DNA is not only the substrate (through the cytosine base) but also a cofactor (via the phosphate group). This means that by removing the phosphate moiety from the DNA, the methylation reaction by DMT cannot take place. Since such a function of the phosphate group as a cofactor had not been suggested in the past, no experimental studies were designed to test this double role of the DNA in any enzymatic mechanism. Nevertheless, support for this mechanism is obtained from an X-ray crystallography study[22] that addressed the additivity of substrate fragments in the binding of 2′-deoxyuridine 5′-monophosphate (dUMP) to TS. In analogy to the covalent bond between C6 of cytosine and the sulfur atom of Cys81 in M.*HhaI*, C6 of dUMP forms a covalent bond with the sulfur atom of Cys146 in TS (in the presence of the cofactor CB3717). Modification of dUMP to two compounds not containing the phosphate group (2′-deoxyuridine and to 2′,5′-dideoxyuridine) resulted in substrates that can bind in a very similar manner to the active site of TS. However, the covalent bond, due to the Michael addition reaction, is not formed. This is evident by the fact that in the case of dUMP, the distance between its C6 and $S^{\gamma}$ of Cys146 was 1.8 Å and the pyrimidine ring was not planar. On the other hand, the corresponding distance with either of the modified substrates was 3.8 Å, and the pyrimidine ring was, unambiguously, planar.[22] It is clear from these experiments that the phosphate moiety is essential for forming the dUMP–TS covalent adduct and it plays a fundamental role in the catalytic mechanism. The interpretation given by the authors for these observations is that the phosphate group in dUMP introduces substantial amount of stereochemical strain that activates C6 of the pyrimidine ring. However, we argue that these observations can be readily explained by our results showing that the phosphate group is the conjugate base deprotonating the thiol group of cysteine to form a thiolate, and only in this anionic state can cysteine attack C6 of pyrimidine and form a covalent bond.

The involvement of the DNA as a cofactor can also rationalize the experimentally observed results of the C81S mutation. The mutation to serine, which preserves the geometry of the side chain and contains a nucleophilic hydroxyl group, exhibits a total, or almost a complete, loss of activity.[12,13,15] The reason for this drastic effect is the inability of the phosphate group of the DNA to abstract a proton from the hydroxyl group of serine. A detailed study of this mutation is presented in the Supplementary Information.

## Activation of C6 of cytosine

As mentioned in the Introduction and shown in Fig. 3, three of the conserved residues in the active site, Glu119, Arg163, and Arg165, can potentially

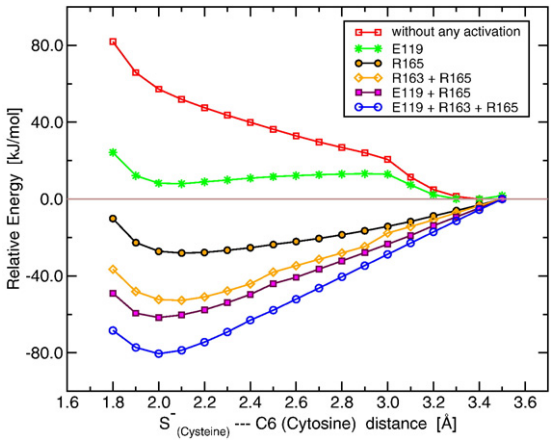


**Fig. 7.** The role of the different active site conserved residues in activating the nucleophilic attack of methyl thiolate (Cys81) on C6 of cytosine. The graph exhibits the relative energy as a function of the reaction coordinate, the distance between $S^-$ and C6. Activation by arginine (Arg163/Arg165) is modeled by guanidinium cation, and that of glutamate (Glu119) is modeled by acetic acid. In the latter, N3 of cytosine is protonated and the acid is deprotonated, while in the absence of Glu119, N3 of cytosine is deprotonated.

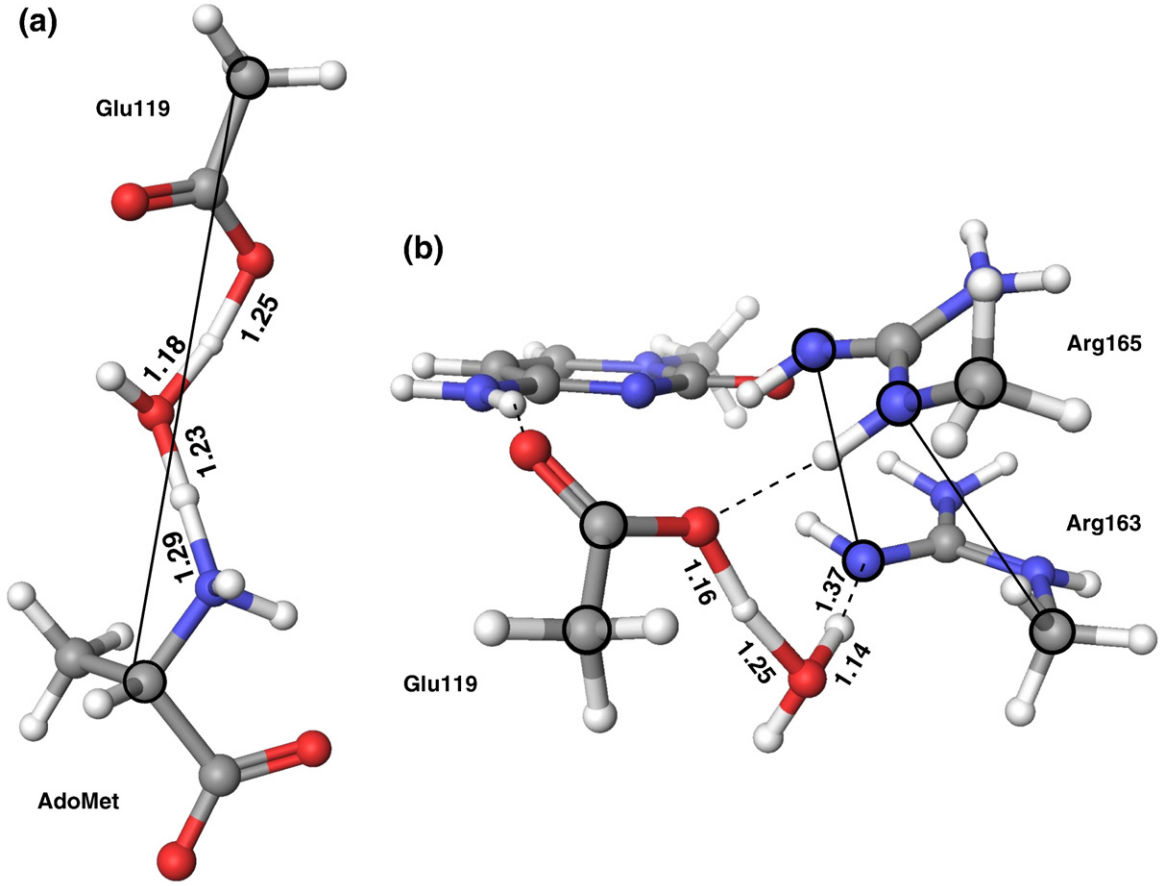


**Fig. 8.** Transition-state structure of two possible protonation reactions of Glu119. (a) From the methionine moiety of AdoMet ($\nu_{TS}=986i$ cm$^{-1}$). The barrier for the reaction is 13.5 kJ/mol (21.6 kJ/mol when the environment is implicitly taken into account); however, the carboxyl group of Glu119 experiences large motion relative to the cytosine ring. (b) From the guanidino group of Arg163 ($\nu_{TS}=733i$ cm$^{-1}$). The barrier for the reaction is 45.4 kJ/mol (54.0 kJ/mol with the solvent effect). A similar computation, however, with freezing the position of $C^{\zeta}$ (the central carbon of the guanidino group) or $N^{\eta}$ (the one closer to the cytosine ring), instead of $C^{\delta}$, of Arg165 resulted with a barrier height of 15.8 kJ/mol (with $U_{TS}=878i$ cm$^{-1}$) or 23.7 kJ/mol (with $\nu_{TS}=965i$ cm$^{-1}$), respectively. These values for the barrier increase to 35.3 and 45.5 kJ/mol, respectively, when accounting for the environment. In the drawing of both transition states, the position of the atoms circled by black lines and the distances between atom pairs circled and connected by black continuous lines were kept frozen to the values in the crystal structure.

promote the electrophilic character of C6 of cytosine through electron delocalization. The arginine residues can do so by electrostatic interaction with O2, while Glu119 does it by mediating the protonation of N3, of cytosine.[23–25] To address the contribution of each of these residues to the activation of C6, we calculated the potential energy profile of the enzyme nucleophilic attack for different combinations of these residues around the cytosine ring. This is shown in Fig. 7. In all cases, the reaction coordinate is the distance between the sulfur of methyl thiolate and C6 of cytosine. Since these residues are charged, the position of their non-carbon heavy atoms was frozen. When none of these residues is activating the cytosine ring, the attack of thiolate is uphill and the reaction is not likely to take place. With cytosine protonated at position N3 and coordinated by Glu119, there is a maximum in the energy profile at approximately 3.0 Å with a height of about 15 kJ/mol. Indeed, transition-state optimization of this system found a structure with one imaginary frequency with the same sulfur–C6 distance. A stronger activation is obtained by the interaction of O2 with an arginine residue. When only Arg165 is present, the curve is downhill and the product is substantially stabilized and is lower in energy compared to the reactant. Further inclusion of Arg163 or Glu119 resulted in a larger stabilization of the product. The biggest stabilization was obtained with all three residues surrounding the cytosine ring. However, it is not clear whether there is a need for such strong activation by all three residues, as the energy profile is already downhill. In fact, it seems that activation by Arg165 and either Arg163 or Glu119 is sufficient. The importance of Glu119 was demonstrated experimentally by mutation to glutamine.[17] In this case, the enzyme loses its activity probably because glutamine cannot protonate, or mediate the protonation of, N3 of cytosine. The sufficiency to activate C6 by only one of the arginines is important because in the conformation of the active site, the guanidino groups of Arg163

and Arg165 are very close to each other, and if they are both positively charged, there is probably an energetic penalty to maintain such conformation (see below).

The distances, observed in the crystal structure of M.*HhaI*, between the carboxylate oxygen atoms of Glu119 and the nitrogen atom at position 3 of cytosine strongly suggest the presence of a proton between the two groups. It is hard to determine what the protonation state of Glu119 is when the cytosine binds to the active site of DMT. If it exists in the ionized form, then a proton transfer from a potential acid should occur. It was proposed that the source of this proton might be the ammonium group ($-NH_3^+$) of the methionine moiety of AdoMet. A water molecule found in the crystal structure (Fig. 2) between these donating and accepting groups can mediate the transfer of this proton.[6] We investigated this proton transfer reaction from AdoMet to Glu119 via a bridging water molecule. When the cytosine ring was included, the calculations did not find a proper transition state and, therefore, it was removed. A transition-state structure is shown in Fig. 8a, and the corresponding barrier height is 13.5 kJ/mol (21.6 kJ/mol with the effect of the environment). In this structure, the carboxy group of Glu119 is oriented almost perpendicular to the plane of the cytosine ring (if included in the calculations), whereas in the crystal structure, it was almost in the same plane. If such rotation inside the active site is possible is yet to be studied.

We also investigated a proton transfer reaction where the acid was the positively charged Arg163. The p$K_a$ value of the guanidino group of arginine is high, 12.48, and in aqueous solution, it is protonated. However, inside the active site and when the accepting base is negatively charged, this value is expected to be lower. In addition, the close proximity of another positively charged arginine residue should greatly reduce this value even further. In the crystal structure, few of the nitrogen–nitrogen or the nitrogen–carbon distances between the two guanidino groups range between 3 and 4 Å. These distances are about the same, or just above, their van der Waals contact diameter. If these guanidino groups are positively charged, there should be a relatively large energetic penalty to permit such conformation. As shown in Fig. 7, if N3 of cytosine is protonated via Glu119, the existence of the second arginine does not provide a significant activation for the nucleophilic attack on C6 of cytosine. Therefore, from the enzyme evolutionary point of view, there is no necessity to maintain the energetic penalty associated with the close proximity of the two charged arginines, unless it plays a role in a different aspect of the methylation reaction. For example, it might be possible that one of the arginines donates its proton to Glu119. One of the transition states of this reaction is shown in Fig. 8b. Depending on the atoms frozen (in order to keep the two positively charged arginines next to each other), the barrier height is found to be between 15.8 and 45.4 kJ/mol (with continuum solvent calculations the range of the barriers increased to 35.3–54.0 kJ/mol). Also in this case, a water molecule between the oxygen of Glu119 and the nitrogen Arg163 is present in the crystal structure (Fig. 2). In summary, both the protonated amino group of the methionine moiety of AdoMet as well as the protonated guanidino group of Arg163 can act as potential conjugated acid to donate a proton to Glu119, via a mediating water molecule. The barrier height in the former is lower, while in the latter, smaller relative motion of the participating residues is necessary.

### The AdoMet electrophilic attack

The addition of the thiolate group to C6 activates the nucleophilic character of C5 of the cytosine ring. This is necessary for the transfer of the methyl group from AdoMet to C5. Previous calculations suggest that the reaction is more likely to take place when N3 of cytosine is deprotonated.[19] This is very reasonable since, in this case, there is more negative charge on the cytosine ring, and with electron delocalization, it can also concentrate on C5. An equilibrium reaction of the proton transfer between Glu119 and N3 can be established. We calculated the potential energy profile of this reaction, where the thiolate group is covalently bound to C6 and the three active site conserved residues Glu119, Arg163, and Arg165 activating the cytosine ring. We found two minima corresponding to the reactants and products separated by a maximum with a height of 3.8 kJ/mol. The energy of the products (deprotonated N3 of cytosine and protonated glutamate) is lower by 18.0 kJ/mol than that of the reactants (figure not shown).

Considering N3 of cytosine in its deprotonated form, we calculated the reaction profile of the AdoMet (modeled as trimethylsulfonium) electrophilic attack. A transition-state structure is shown in Fig. 9 with a barrier height of 8.8 kJ/mol. With reaction field calculation to account for the environment, this barrier increases to 44.8 kJ/mol. When both Glu119 and Arg165 were not taken into account in the calculation (and N3 was still deprotonated), the barrier was much higher, 76.8 kJ/mol (77.9 kJ/mol with solvent effect). Electrostatic interactions of the guanidino group with O2 and hydrogen-bond formation of the carboxyl group with N3 will withdraw negative charge from C5. This effect is likely to reduce the nucleophilic character of C5 and, therefore, increase the reaction barrier. On the other hand, hydrogen-bond donation of N4 to the carboxy group will allow to push negative charge towards C5 and, thus, reduce the reaction barrier. The results suggest that for this reaction, the latter effect is more important than the former. In addition, the ability of the carboxy group to push negative charge towards C5 explains why with reaction field calculations the barrier height increases, when Glu119 and Arg165 are taken into account, while it hardly changes when these residues are not surrounding the cytosine ring. In the former case, the Mulliken charge of C5 is $-0.23\,e$ for the reactant and $-0.055\,e$ for the transition state (i.e., when included in

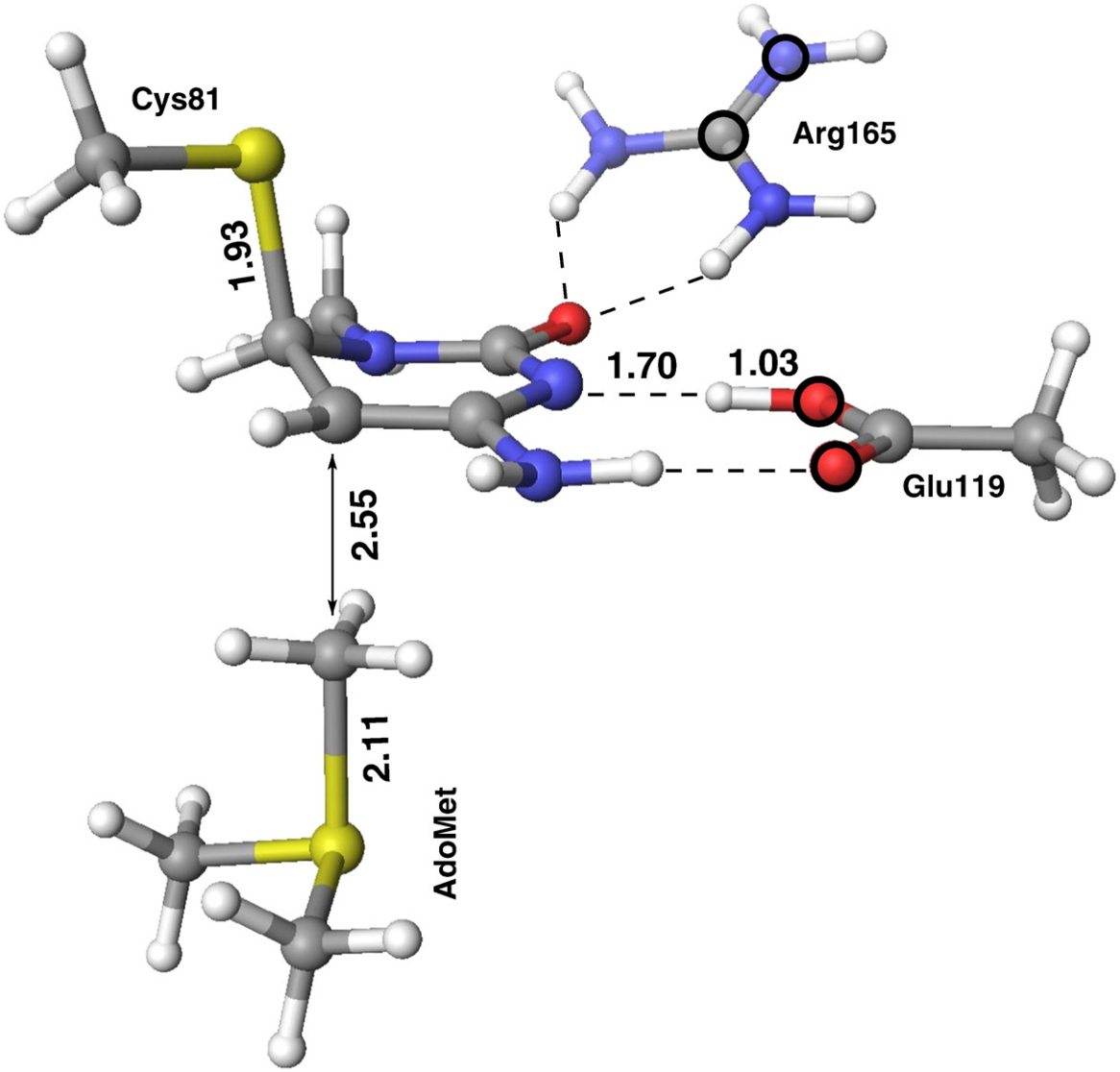


**Fig. 9.** Structure of the transition state ($\nu_{TS}=353i$ cm$^{-1}$) of the attack of trimethylsulfonium (AdoMet) onto C5 of cytosine. The corresponding barrier height of the reaction is 8.8 kJ/mol (44.8 kJ/mol with solvent effect). A similar transition-state structure, however, without the surrounding Glu119 (N3 deprotonated) and Arg165 residues, yielded a much higher barrier height of 76.8 kJ/mol (77.9 kJ/mol with solvent effect). Atoms circled by black lines, as well as the angle $C4_{cyt}$–$C5_{cyt}$–$S_{AdoMet}$, were kept frozen (the positions to the values in the crystal structure and the angle to the value found in the transition state without Glu119 and Arg165) during the optimization search.

the calculations the dielectric property of the solvent stabilizes the reactants relative to the transition-state), while in the latter case, it is $-0.025$ *e* for both the reactant and the transition state ( i.e., when taken into account the dielectric property of the solvent does not change the stability of the reactants relative to the transition-state).

## The elimination reaction

The last step of the methylation reaction is the elimination of the enzyme nucleophile, Cys81, at C6 and the abstraction of the proton at C5 of the cytosine ring. Earlier calculations concluded that stepwise elimination, where either the nucleophile or the proton leaves first, is energetically unfavorable.[19] In addition, the fact that no suitable base can be identified to act as the proton abstracting agent in the crystal structure also favors the concerted elimination scenario. It was suggested that a negatively charged phosphoryl oxygen of the DNA can act as the base assisted by a bridging water molecule.[6] However, from the X-ray positions of the groups involved, such reaction is hard to envision (the distances between C5 and the two closest phosphate groups are 7.9 and 10.6 Å), unless significant conformational change takes place.

The addition of the methyl group from AdoMet occurs on the face opposite to the one of the nucleophilic attack by Cys81. This means that the two leaving groups in the elimination reaction leave from the same face of the cytosine ring; that is, it is a syn elimination, which requires larger activation energy. In the crystal structures solved,[5,6] few water

molecules are observed nearby the hydrophobic edge (C5 and C6) of the flipped cytosine (Fig. 2). Therefore, we assumed that the abstracting base in the elimination reaction is the anionic sulfur of the leaving thiolate bridged by a water molecule. We obtained a transition-state structure of such reaction; however, the barrier height was relatively high (157.8 kJ/mol). Therefore, we considered an alternative elimination step. The cytosine base can exist in different tautomeric forms.[26,27] We consider here the imino hydroxy tautomer where an amine hydrogen at position 4 is transferred to the carbonyl oxygen at position 2. The conserved residues, Glu119, Arg163, and Arg165, can facilitate such proton transfer reaction.

In Fig. 10 we display the transition-state structure found for the elimination reaction. In this case, only one of the arginine residues is taken into consideration. Note that many attempts to find a transition structure with both Arg163 and Arg165, or with one arginine and Glu119, did not succeed. The height of the barrier of the transition state is 125.1 kJ/mol, and the products are more stable than the reactants by 20.6 kJ/mol. By including the effect arising from the dielectric constant of the environment, the transition-state barrier is reduced to 111.0 kJ/mol. In the transition structure, we also included an acetone molecule, representing the backbone carbonyl of Phe79, which is hydrogen bonded to the mediating water molecule. In the crystal structure, the distance between the oxygen atom of this carbonyl and C5 cytosine is 3.66 Å and is likely to be able to form a hydrogen bond with a nearby water molecule. By forming this hydrogen bond, the barrier of the transition state is reduced by 10.4 kJ/mol. This is probably due to the ability to stabilize better the partial positive charge that is developed on the water molecule in the transition state.

## Conclusions

In this article, the different steps of cytosine methylation catalyzed by DMT were investigated by quantum calculations. We find the reaction to be strongly catalyzed by the active site conserved

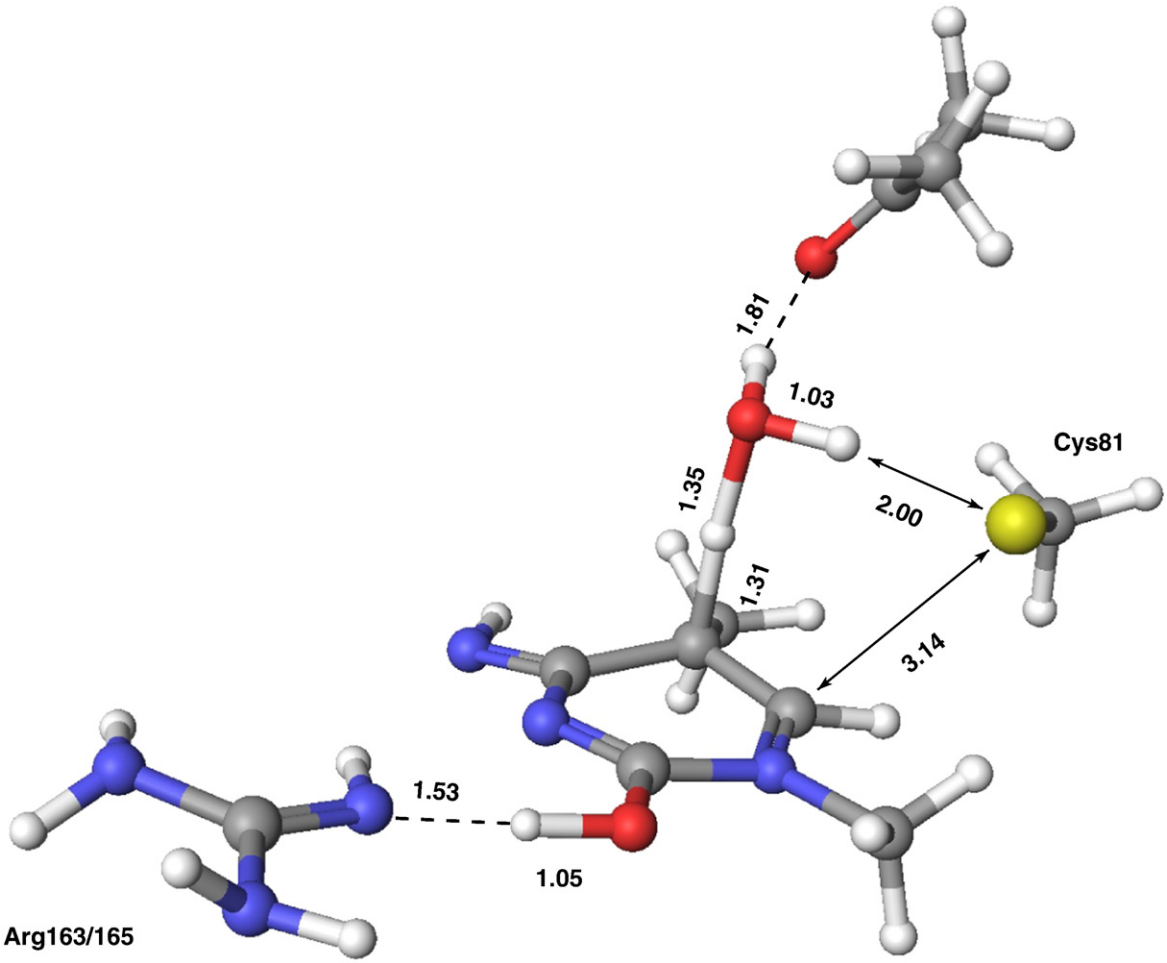


**Fig. 10.** Structure of the transition state ($\nu_{TS} = 898i\ cm^{-1}$) of the elimination reaction. The barrier height is 125.1 kJ/mol and the product state is −20.6 kJ/mol below the energy level of the reactant. With reaction field calculations, the barrier height reduces to 111.0 kJ/mol. The acetone molecule hydrogen bonded to the bridging water molecule, representing the carbonyl backbone of Phe79, helps to stabilize the transition structure by 10.4 kJ/mol relative to the reactant.

residues, Glu119, Arg163, and Arg165, making the enzyme nucleophilic attack as well as the AdoMet electrophilic attack feasible. In particular, Glu119 mediates the protonation of N3 of cytosine from either the amino group of the methionine moiety of AdoMet or the positively charged guanidino group of Arg163. In addition, Arg165 and, possibly, Arg163 withdraw negative charge from the cytosine ring via electrostatic interactions with O2. These results are in sharp contrast to the findings obtained by a quantum mechanics/molecular mechanics calculations[18] where it was argued that (i) the residues Glu119, Arg163, and Arg165 play no catalytic role, and (ii) the nucleophilic attack by Cys81 and the electrophilic attack by AdoMet proceed in a single step.

Arguably, the most intriguing result obtained from our study is the double role of the DNA, acting as the substrate and as a cofactor. The enzyme nucleophile can attack C6 of cytosine only in its anionic deprotonated form. However, within the protein, there is no suitable base in the vicinity of Cys81 that can protonate its thiol group. In this article, we have shown that the negatively charged phosphate group of the DNA adjacent to the target cytosine can form an equilibrium reaction, where it acts as a conjugate base, abstracting the thiol proton, via a mediating water molecule. Depending on its p$K_a$, this neutral protonated phosphate can donate this proton to solvent water molecules. With the activation of the conserved residues in the active site, this enzyme nucleophilic attack on C6 of cytosine is downhill in energy. The subsequent step of the AdoMet electrophilic attack is also strongly catalyzed by these conserved residues. In fact, only in the last step of the enzymatic mechanism did we find a substantial energy barrier (of 111.0 kJ/mol) that is associated with a syn elimination reaction. In some proton transfer steps studied in this work, we found that a bridging water molecule was necessary. Evidence of such involvement was obtained by recent X-ray crystallography study demonstrating that several active-site water molecules are important for catalysis.[28] Note, however, that the involvement of a bridging active-site water molecule will add an entropic penalty to the free-energy barrier of the reaction and reduce the rate of the reaction. In general, water molecules that are present in the active site are characterized by a very low entropy compared to bulk water and often referred to as structured water. This loss of entropy is compensated by the enthalpic gain from the interactions with active-site residues. The additional restriction of the translation and rotations of an active site water, so that it can participate in the transition-state complex, would require a much smaller entropic penalty than the freezing of a water molecule in the bulk. It is, however, difficult to provide an estimate for such penalty since the density of states of the water molecule inside the active site would need to be determined first.

That the enzymatic activity of the cysteine to serine mutant (C81S) is reduced is not surprising. However, it is remarkable that there was a total loss, or reduction by a factor of $10^4$, of the measured enzymatic activity.[12,13,15] Indeed, sulfur atom is known to be a better nucleophile than oxygen due to its larger size and polarizability. However, the difference in this nucleophilic ability is not strong enough to explain such large difference in catalytic efficiency. Our findings from the calculations addressing the cysteine-to-serine mutation indicate that this selectivity can be largely enhanced inside the protein active site. If the nucleophilic reaction proceeds only via the anionic form of the nucleophile, then in the absence of a strong base, cysteine (thiol) can undergo deprotonation while serine (alcohol) cannot, and thereby, only in the former case can the reaction takes place. In line with this argument, it is interesting to note that the difference in p$K_a$ between the hydroxy and mercapto groups is ca 6.5, which would translate to a difference by a factor of $10^6$–$10^7$ in enzymatic activity.

## Methods

The quantum mechanical calculations were performed using the Gaussian 03 program[29] with the polarized 6-31G** basis set. The energy resulting from the exchange and correlation of the electrons was approximated by the hybrid functional, B3LYP, of Hartree–Fock exchange with Density Functional exchange-correlation terms. Vibrational frequencies were calculated to confirm the existence of a transition state, which exhibits one, and only one, imaginary frequency. In addition, for all transition states identified, Intrinsic Reaction Coordinate calculations in the forward/reverse direction led to either the product or reactant complexes. However, in most cases, full convergence of the reactant/product state was achieved by an additional optimization step.

The effect of the environment (which is composed of the protein, DNA, and solvent) on the height of the barrier of the different steps was approximated by performing a Self-Consistent Reaction Field calculation using the Polarizable Continuum Model.

For various proteins, the dielectric constant $\varepsilon$ was estimated in the literature to be between 11 and 35.[30] The value of $\varepsilon$ is not uniform inside the protein, and estimation of its value locally yields a distribution in the range 1–20.[31] It is important to note that since the flexibility of an active site is usually smaller than that of the rest of the enzyme, it is argued that in this region, the dielectric constant should also be smaller.[32] However, it is further argued that water molecules that are present inside the active site strongly increase the value of $\varepsilon$.[32] Based on this and on the crystal structure of DMT indicating a large number of water molecules in the active site,[5] we chose a dielectric constant of $\varepsilon = 20.7$, which corresponds to solvating the model systems in acetone.

All starting coordinates in the calculations were extracted from the crystal structure of wild-type *HhaI* (from *Haemophilus haemolyticus*) C5-cytosine

DMT with unmodified DNA and *S*-adenosyl-L-homocysteine (Protein Data Bank ID: 2HR1[5]), which is shown in Fig. 2. Addition of hydrogen atoms and modifications, necessary to create the system model for the calculation of each step in the catalytic mechanism, were performed using the Molden processing program.[33]

The conserved amino acids in the active site proposed to catalyze the reaction (Glu119, Arg163, and Arg165) as well as the cofactor AdoMet are likely to be charged in the active site of the enzyme. Therefore, in order to avoid the collapse of two oppositely charged residues or the breakup of like-charge residues, certain constraints on atomic positions and/or angles had to be imposed. These constraints are described in the text and in the figures of the transition-state structure found. Of course, within the protein, the integrity of the active site is kept due to the stability of the native folded conformation. A disadvantage of applying these constraints is the possibility of observing imaginary vibrational modes associated only with employing the constraints. Nevertheless, in all the transition states presented, the constraints were chosen such that these additional imaginary vibrational modes are not present and only one imaginary frequency is observed. This procedure did not succeed for all the reaction steps investigated, and therefore, in these cases, we calculated the potential energy profile as an alternative. The search for a transition state with only one imaginary frequency while imposing the constraints on charged residues (which are necessary) also prevented us from applying the same model system for all elementary steps analyzed.

## Acknowledgements

We thank the Spanish MICINN (grants CTQ 2007-67528 and Ingenio Consolider CSD 2007-00006) and the UPV/EHU (grant IT-324-07) for financial support. Technical and human support provided by SGIker (USED SERVICES) (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.05.021

## References

1. Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem*, **3**, 274–293.
2. Szyf, M. (2003). Targeting DNA methylation in cancer. *Ageing Res. Rev.* **2**, 299–328.
3. Santi, D. V., Norment, A. & Garrett, C. E. (1984). Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc. Natl Acad. Sci. USA*, **81**, 6993–6997.
4. Wu, J. C. & Santi, D. V. (1987). Kinetic and catalytic mechanism of *Hha*I methyltransferase. *J. Biol. Chem.* **262**, 4778–4786.
5. Shieh, F.-K., Youngblood, B. & Reich, N. O. (2006). The role of Arg165 toward base flipping, base stabilization and catalysis in M.HhaI. *J. Mol. Biol.* **362**, 516–527.
6. O'Gara, M., Klimašauskas, S., Roberts, R. J. & Cheng, X. (1996). Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for *Hha*I methyltransferase–DNA–AdoHcy complexes. *J. Mol. Biol.* **261**, 634–645.
7. Kumar, S., Cheng, X., Pflugrath, J. W. & Roberts, R. J. (1992). Purification, crystallization, and preliminary x-ray diffraction analysis of an M.HhaI–AdoMet complex. *Biochemistry*, **31**, 8648–8653.
8. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
9. Pósfai, J., Bhagwat, A. S., Pósfai, G. & Roberts, R. J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17**, 2421–2435.
10. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J. & Wilson, G. G. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **22**, 1–10.
11. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. & Trautner, T. A. (1988). Sequential order of target-recognizing domains in multispecific DNA–methyltransferases. *EMBO J.* **7**, 2601–2609.
12. Wyszynski, M. W., Gabbara, S. & Bhagwat, A. S. (1992). Substitutions of a cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. *Nucleic Acids Res.* **20**, 319–326.
13. Chen, L., MacMillan, A. M. & Verdine, G. L. (1993). Mutational separation of DNA binding from catalysis in a DNA cytosine methyltransferase. *J. Am. Chem. Soc.* **115**, 5318–5319.
14. Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. (1995). The crystal structure of Haelll methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.
15. Gabbara, S., Sheluho, D. & Bhagwat, A. S. (1995). Cytosine methyltransferase from *Escherichia coli* in which active site cysteine is replaced with serine is partially active. *Biochemistry*, **34**, 8914–8923.
16. Estabrook, R. A., Lipson, R., Hopkins, B. & Reich, N. O. (2004). The coupling of tight DNA binding and base flipping. *J. Biol. Chem.* **279**, 31419–31428.
17. Shieh, F.-K. & Reich, N. O. (2007). AdoMet-dependent methyl-transfer: Glu119 is essential for DNA C5-cytosine methyltransferase M.HhaI. *J. Mol. Biol.* **373**, 1157–1168.
18. Zhang, X. & Bruice, T. C. (2006). The mechanism of M. HhaI DNA C5 cytosine methyltransferase enzyme: a quantum mechanics/molecular mechanics approach. *Proc. Natl Acad. Sci. USA*, **103**, 6148–6153.
19. Peräkylä, M. (1998). A model study of the enzyme-catalyzed cytosine methylation using ab initio quantum mechanical and density functional theory calculations: p$K_a$ of the cytosine N3 in the intermediates and transition states of the reaction. *J. Am. Chem. Soc.* **120**, 12895–12902.

20. Vilkaitis, G., Merkienė, E., Serva, S., Weinhold, E. & Klimašauskas, S. (2001). The mechanism of DNA cytosine-5 methylation. *J. Biol. Chem.* **276**, 20924–20934.
21. Henchoz, Y., Schappler, J., Geiser, L., Prat, J., Carrupt, P.-A. & Veuthey, J.-L. (2007). Rapid determination of p$K_a$ values of 20 amino acids by CZE with UV and capacitively coupled contactless conductivity detections. *Anal. Bioanal. Chem.* **389**, 1869–1878.
22. Stout, T. J., Sage, C. R. & Stroud, R. M. (1998). The additivity of substrate fragments in enzyme–ligand binding. *Structure*, **6**, 839–848.
23. Jordan, F. & Sostman, H. D. (1973). Molecular orbital (CNDO2 and MINDO) calculations on protonated deoxyribonucleic acid bases. The effects of base protonation on intermolecular interactions. *J. Am. Chem. Soc.* **95**, 6544–6554.
24. Hayatsu, H. (1976). Bisulfite modification of nucleic acids and their constituents. *Prog. Nucleic Acid Res. Mol. Biol.* **16**, 75–124.
25. Baker, D. J., Kan, J. L. C. & Smith, S. S. (1988). Recognition of structural perturbations in DNA by human (cytosine-5)methyltransferase. *Gene*, **74**, 205–210.
26. Scanlan, M. J. & Hillier, I. H. (1984). An ab initio study of tautomerism of uracil, thymine, 5-fluorouracil, and cytosine. *J. Am. Chem. Soc.* **106**, 3737–3745.
27. Sambrano, J. R., de Souza, A. R., Queralt, J. J. & Andrés, J. (2000). A theoretical study on cytosine tautomers in aqueous media by using continuum models. *Chem. Phys. Lett.* **317**, 437–443.
28. Youngblood, B., Shieh, F.-K., Buller, F., Bullock, T. & Reich, N. O. (2007). *S*-Adenosyl-L-methionine-dependent methyl transfer: observable precatalytic intermediates during DNA cytosine methylation. *Biochemistry*, **46**, 8766–8775.
29. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Rob, M. A., Cheeseman, J. R. *et al.* (2003). Gaussian 03, Revision A.1, Gaussian, Inc., Wallingford, CT.
30. Simonson, T. (1998). Dielectric constant of cytochrome *c* from simulations in a water droplet including all electrostatic interactions. *J. Am. Chem. Soc.* **120**, 4875–4876.
31. Nakamura, H., Sakamoto, T. & Wada, A. (1988). A theoretical study of the dielectric constant of protein. *Protein Eng.* **2**, 177–183.
32. Mertz, E. L. & Krishtalik, L. I. (2000). Low dielectric response in enzyme active site. *Proc. Natl Acad. Sci. USA*, **97**, 2081–2086.
33. Schaftenaar, G. & Noordik, J. H. (2000). Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput.-Aided Mol. Des.* **14**, 123–134.

# Supplementary Information:

# Mechanism of DNA Methylation: The Double Role of DNA as a Substrate and a Cofactor

Ronen Zangi[*,a,b], Ana Arrieta[a], Fernando P. Cossío[a]

[a]*Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain*
[b]*IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

## *I. Cys81 to Ser Mutation*

The enzyme nucleophile, Cys81 in M.*HhaI*, had been the subject of several mutational (to Gly, Ala, Val, Trp, and Ser) analysis that resulted in, either, complete loss or substantially reduced activity [1,2]. Indeed, any mutation to an amino acid without a good nucleophile is expected to hamper the methylation reaction. However, it is surprising that mutating cysteine to serine, which preserves the geometry of the side chain and contains a potentially nucleophilic hydroxyl group exhibits at least $10^4$-fold reduction in activity [1,2]. In another study it was found that the mutation to serine yielded no covalent complex between the enzyme and the DNA [3]. We performed additional calculations in order to understand the results of this mutation. As in the case of the thiol group (cysteine), the hydroxyl group of serine can attack C6 only when it is deprotonated to its anionic form. This is shown in Fig. 7a, where the attack by methanol is an uphill process, while that by methoxide exhibits a downhill curve. The slope of the energy profile of this nucleophilic reaction for methoxide is larger than that for thiolate by more than a factor of 2 (see also Fig. 4). Consequently, the stability of the product, relative to the reactants, is also larger. Thus, the methoxide can potentially attack C6 of cytosine at least as good as the thiolate. It is, however, questionable whether the negatively charged phosphate group can act as a conjugate base abstracting a proton from the more electronegative oxygen atom. For example, the pKa of mercaptoethanol is 9.5 [4] whereas the corresponding value for ethanol is 16 [5]. Starting from the transition state structure that we found for the proton transfer from the thiol to the phosphate, we replaced the sulfur atom by oxygen. Several optimizations did not locate any transition state for the proton transfer from methanol to the phosphate. We further searched for a concerted attack of the hydroxyl on C6 and a proton transfer to the phosphate, however, as with the case of the thiol, no transition state was found. We therefore, performed a potential energy profile, shown in Fig. 1b, for the proton transfer reaction as a function of the $CH_3O$–H distance of the methanol hydroxyl group. To confirm the expected structure of the products we display in Fig. 1c the optimized distances, involved in this proton transfer reaction, between the bridging water molecule and the methanol hydroxyl group, as well as, the negatively charged oxygen of the phosphate group. Fig. 1b

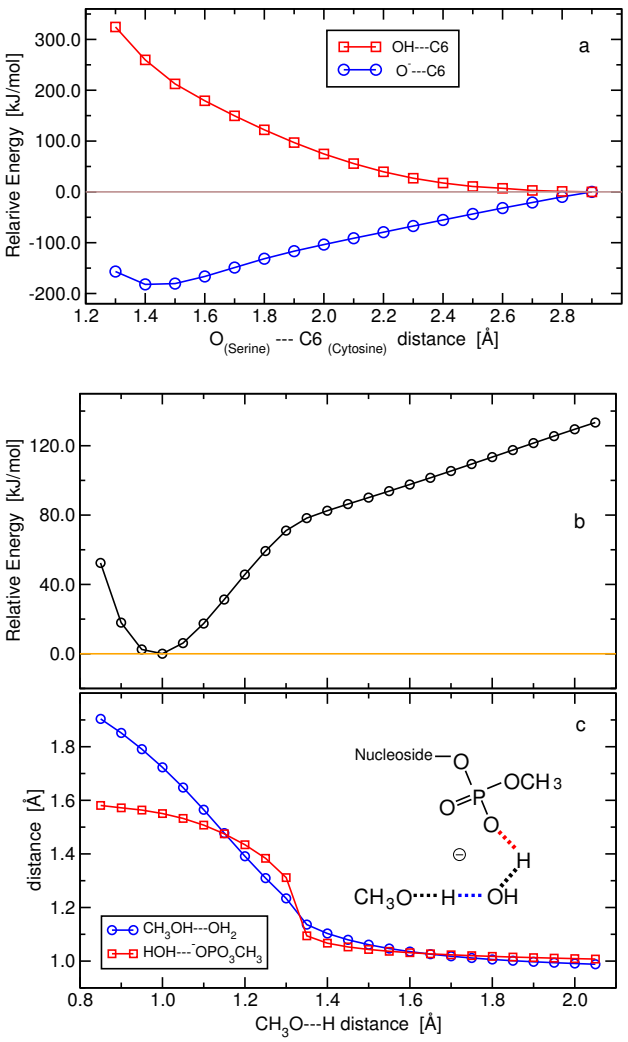


Figure 1: Modeling the C81S mutation. (a) The relative energy of the attack of methanol and methoxide (representing protonated and deprotonated serine, respectively) on C6 of cytosine as a function of the O—C6 distance. As in the case of Fig. 4, the cytosine ring is activated by Glu119, Arg163, and Arg165, where N3 is protonated. (b) The relative energy of deprotonating the methanol by the phosphate group through a bridging water molecule. (c) The distance between the proton transferred from methanol and the oxygen atom of the mediating water molecule (blue curve), as well as, the distance between the proton transferred from the mediating water molecule and the oxygen atom of the receiving phosphate group (red curve). In (b) and (c), the curves are plotted as a function of the O—H distance of the methanol molecule.

*Corresponding author: r.zangi@ikerbasque.org Tel: +34-943018112 Fax: +34-943015270

indicates that as the reaction proceeds there is only an increase in the potential energy of the system, with substantial activation barrier. This suggests that for the C81S mutant a proton transfer from the hydroxyl group to phosphate is not likely, and therefore, the methylation reaction cannot proceed by the enzymatic mechanism that we propose in this paper.

## References


[1] M. W. Wyszynski, S. Gabbara, A. S. Bhagwat, Substitutions of a cysteine conserved among DNA cytosine methylases result in a variety of phenotypes, Nucleic Acids Res. 20 (1992) 319–326.

[2] S. Gabbara, D. Sheluho, A. S. Bhagwat, Cytosine Methyltransferase from *Escherichia coli* in Which Active Site Cysteine Is Replaced with Serine Is Partially Active, Biochemistry 34 (1995) 8914–8923.

[3] L. Chen, A. M. MacMillan, G. L. Verdine, Mutational Separation of DNA Binding from Catalysis in a DNA Cytosine Methyltransferase, J. Am. Chem. Soc. 115 (1993) 5318–5319.

[4] J. T. Edsall, J. Wyman, Biophysical Chemistry Vol 1, Academic Press, New York, 1958.

[5] P. Ballinger, F. A. Long, Acid Ionization Constants of Alcohols. II. Acidities of Some Substituted Methanols and Related Compounds, J. Am. Chem. Soc. 82 (1960) 795–798.