

Unraveling the Reaction Mechanism of
Enzymatic C5-Cytosine Methylation of DNA. A
Combined Molecular Dynamics and QM/MM
Study of wild type and Gln119 variant

Juan Aranda¹, Kirill Zinovjev¹, Katarzyna Świderek^{1,2}, Maite
Roca^{1*}, Iñaki Tuñón^{1*}

¹Departament de Química Física, Universitat de València, 46100 Burjassot,
Spain

²Institute of Applied Radiation Chemistry, Lodz University of Technology, 90-
924 Lodz, (Poland)

Ignacio.tunon@uv.es

M.Teresa.Roca@uv.es

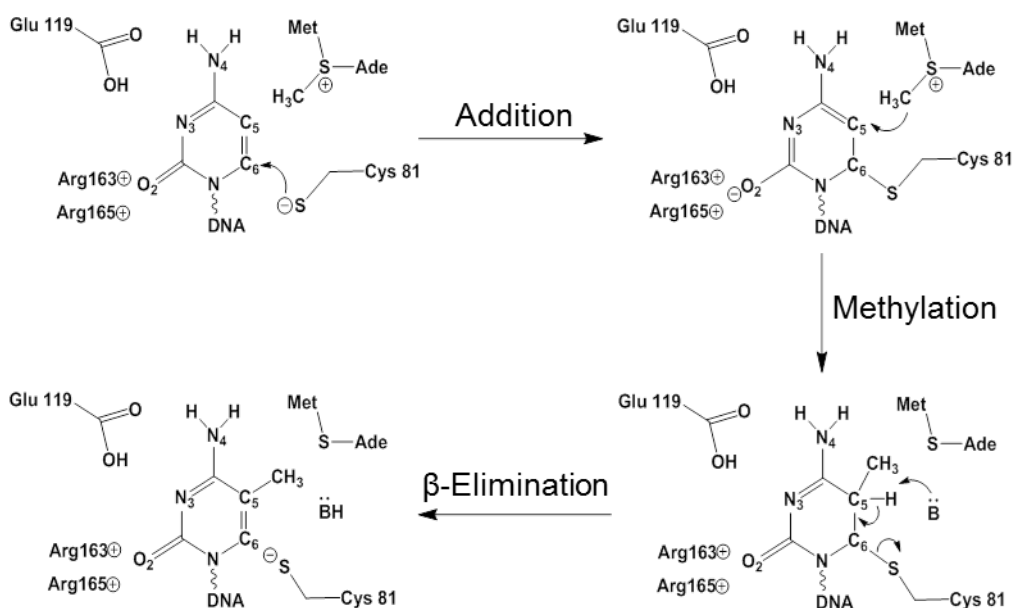
Abstract

M.HhaI is a DNA Methyltransferase from *Haemophilus haemolyticus* that catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the C₅ position of a cytosine. This enzyme is a paradigmatic model for C₅ DNA Methyltransferases due to its major homology to mammalian enzymes and to the availability of high resolution structures of the DNA-enzyme complex. In spite of the number of experimental and theoretical analysis carried out for this system many mechanistic details remain unraveled. We have used full atomistic classical Molecular Dynamics simulations to explore the protein-SAM-DNA ternary complex where the target cytosine base is flipped out into the active site for both the wild type and Glu119Gln mutant. The relaxed structure was used for a combined Quantum Mechanics/Molecular Mechanics exploration of the reaction mechanism using the string method. Exploration of multidimensional Free Energy Surfaces allowed determining a complete picture of the reaction mechanism in agreement with experimental observations. In our proposal Cys81 becomes unprotonated after formation of the ternary complex. This residue then attacks C₆ position of the target cytosine establishing a fast and reversible equilibrium. Methyl transfer from SAM to position C₅ takes place after the previous addition and the Transition State is stabilized by means of hydrogen bond interactions established between the cytosine base and Glu119. Although previous proposals suggest that a hydroxide anion could be the base abstracting the proton from the C₅ position, our calculations show that the free energy cost needed to have this anion in the active site is too high. Instead, a crystallographic water molecule can remove the excess proton once the cytosine has been activated by means of a proton transfer from Glu119 to the N₃ position. This observation explains the consequences of mutations of this residue on the pre-steady state and steady state rate constants. Our proposal then clarifies the role of Glu119 and identifies the nature of the base in charge of proton abstraction, two issues that have been the subject of a long debate in the literature.

1. Introduction

DNA Methyltransferases (DNA MTases) are the enzymes in charge of DNA methylation, an essential biological process in bacteria, plants and animal cells. In general, DNA methylation is a central mechanism in epigenetic inheritance because information can be encoded in a stable but reversible manner.¹ Epigenetic mechanisms comprise gene silencing, parental imprinting, maintenance of genome integrity, X chromosome inactivation, protection of the genome against selfish DNA and regulation of development.²⁻⁴ Autoimmune disorders, carcinogenesis and other diseases are often related to aberrant DNA methylation. Methylation patterns, for example, are disrupted in tumor cells with hypomethylation and region-specific hypermethylation.⁵

DNA MTases of prokaryotes catalyze the methyl transfer from S-adenosyl-L-methionine (SAM) to an adenine (N₆ position) or a cytosine nucleobase (N₄ or C₅ positions); while in mammals and other vertebrates methylation only takes place at the C₅ position of cytosine.⁶ The catalytic C-terminal part of C₅-MTases is conserved between prokaryotic and eukaryotic C₅-MTases.⁷ In addition these enzymes share a similar overall reaction mechanism; all DNA MTases flip the target nucleobase out of the DNA helix and place it inside the active site before transferring the methyl group from SAM to the nucleobase.⁸ The reaction mechanism in C₅-MTases (see Scheme I) involves the nucleophilic addition of a cysteine placed in the active site to the C₆ atom of the cytosine ring and the methyl transfer from the cofactor to the C₅ position.⁹ It has been suggested that the nucleophilic attack by cysteine could be favored by a transient protonation of the cytosine ring carried out by a conserved and protonated glutamic residue present in the active site. Afterwards (see Scheme I), the excess H₅ proton from the C₅ position must be abstracted and the covalent bond between the enzyme and the cytosine base cleaved to liberate the methylated DNA. Some questions concerning the mechanism have been controversial during the recent years:¹⁰⁻²³ the stepwise or concerted nature of each chemical step, the protonation state of the nucleophilic cysteine, the role of the conserved glutamic residue and, in particular, which is the base abstracting the proton from C₅ during the β -elimination step.



Scheme I. General reaction mechanism for DNA methylation in *M.HhaI*.

Development of inhibitors of the C_5 -MTases activity is an active field for cancer treatment.²⁴ C_5 -MTases inhibitors can be classified into two broad categories: nucleoside and non-nucleoside inhibitors.^{25,26} The most promising nucleoside based inhibitors with antiproliferative cancer activity are zebularine,²⁷ decitabine, 5-azacytidine and its 2' deoxy analog.²⁸ These compounds are incorporated to DNA and act forming covalent adducts with DNA MTases. Zebularine, for example, traps the enzyme through a reversible covalent complex.²⁹ Nevertheless, nucleoside inhibitors have to be improved in terms of toxicity and stability to decrease side effects. On the other side, non-nucleoside inhibitors generally act binding to DNA MTases and competing with natural substrates, SAM or DNA. Some examples of reversible DNA MTases inhibitors are known cardiovascular drugs as hydralazine or procainamide and natural compounds like polyphenols and antibiotics.²⁵ Sequence specific oligomers that bind to the DNA inhibiting enzymatic CpG methylation preventing the translation of the MTase are also under research.³⁰

In this work we focus on the C_5 -MTase from *Haemophilus haemolyticus* (*M.HhaI*), an excellent model to study DNA methylation due to its major homology to mammalian enzymes and to the availability of high resolution structures of the DNA-enzyme complex.³¹ This enzyme catalyzes the

methylation in the C₅ position of the underlined cytosine within the recognition sequence 5'-GCGC-3' in duplex DNA. This system has been the target of several experimental^{10,12,15-18,32} and theoretical²⁰⁻²³ studies because provides a paradigm for enzymes inducing conformational changes in DNA.^{10,12,31}

Previous computational studies on the reaction mechanism of *M.HhaI* have shown that the nucleophilic attack by Cys81 and the methyl transfer from SAM could take place either concertedly or sequentially.^{21,23} The role of Glu119 has been also a matter of controversy. It has been proposed that proton transfer from protonated Glu119 to the N₃ position of the target cytosine could activate the base for the nucleophilic addition of the thiolate group of Cys81 to the C₆ position of the cytosine.^{17,18} Kinetic measurements showed that the catalytic activity of Glu119 mutants is dramatically reduced, concluding that this residue has a crucial role in catalysis and not only a structural function.¹⁷ A theoretical analysis on a cluster model carried out by Zangi and coworkers showed that Glu119 does not protonate the substrate during the methylation step.²² Bruice and coworkers found no evidences for a proton transfer from Glu119 to the substrate during the methyl transfer step.²¹ Finally, using QM/MM modelling, Zhang and coworkers proposed that Glu119 may transiently transfer the proton to the N₃ position during the methyl transfer step.²³ Regarding the β -elimination step, Bruice and coworkers proposed that a hydroxide anion could reach the active site through a water channel observed in the X-ray structures.²¹ The free energy profile for the abstraction of the excess H₅ proton on C₅ position by a hydroxide anion was recently explored with QM/MM methods, displaying a reasonable free energy barrier.²³ However, the proposals in which the hydroxide anion acts as the base in charge of the C₅-proton abstraction has never considered the free energy cost of having the anion in the active site, which can be significantly different to the cost of creating the anion in bulk water.

In order to unravel the still controversial features of the reaction mechanism for C₅ methylation in *M.HhaI* we have performed classical molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) calculations to obtain the free energy profiles associated with the reaction mechanism. Because of the nature of the reaction, where several chemical steps can take place sequentially or concertedly, we used the string method³³ to explore many

different mechanistic variants and characterize the most favorable paths on a multidimensional free energy surface. According to these calculations the reaction proceeds through a reversible nucleophilic attack of a deprotonated Cys81 to the cytosine ring followed by a methyl transfer from SAM to the C₅ position. Proton transfer from Glu119 to the substrate is not needed at this stage. However, protonation of the N₃ position of the target cytosine by this residue is essential for the following step, the elimination of the excess proton from the C₅ atom. Our calculations show that a water molecule placed near the C₅ position, as seen in the X-ray structure,¹⁵ is the base that abstracts the proton, while formation of a hydroxide anion in the active site has a large free energy cost. Finally, in the last step, deprotonation of the N₃ atom of cytosine by Glu119 and the cleavage of the covalent bond formed with Cys81 leads to the reaction products. This mechanistic proposal allows a complete reconciliation between theoretical modelling and experimental data in *M.HhaI*, a C₅-Mtase enzyme that serves as a model for this relevant epigenetic process.

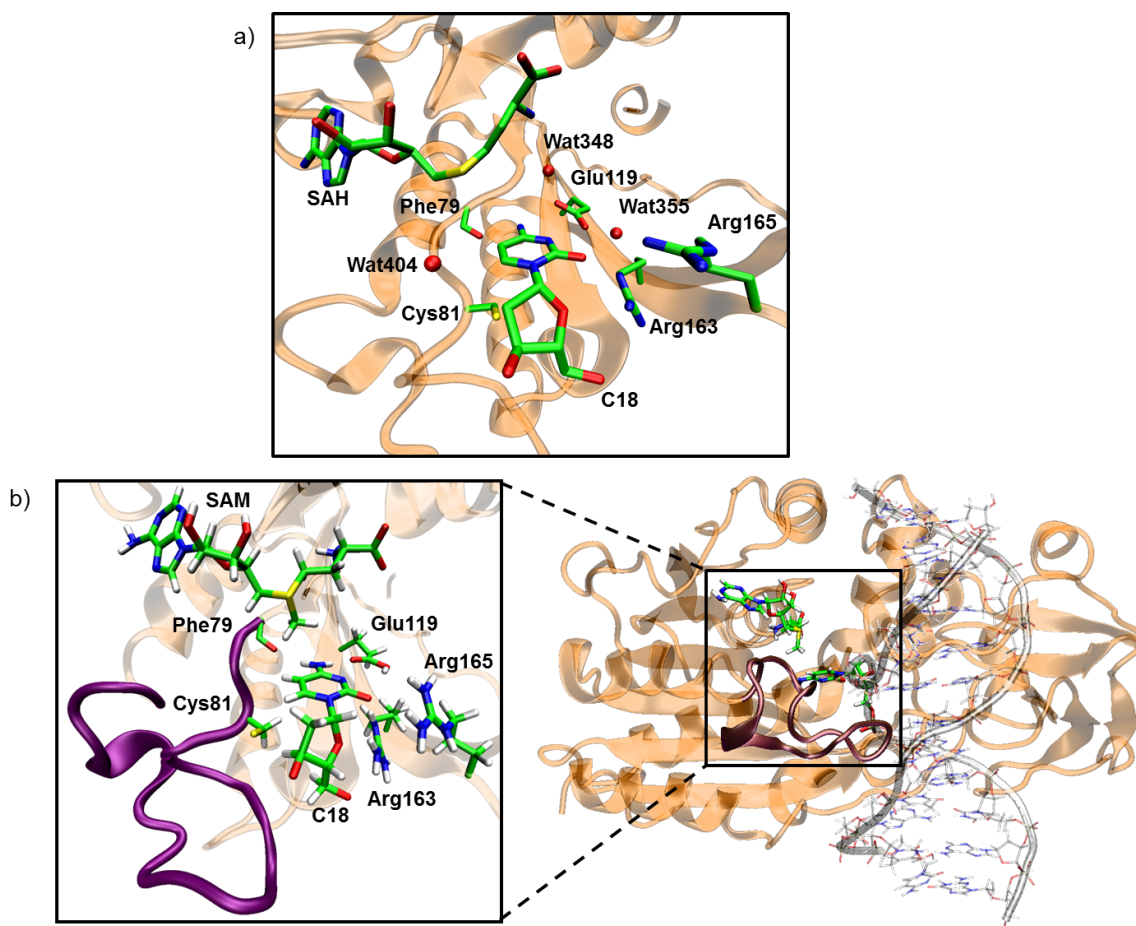


Figure 1. a) Active site insert from the X-ray structure of *M.HhaI* complexed with a dodecamer of DNA (PDB code 2HR1). b) Model system used in the simulations (solvent water molecules are omitted for clarity). The catalytic loop (residues 80-100) is drawn in purple. The insert shows an insight into the active site with cofactor SAM, the target cytosine base (C18) and key residues establishing hydrogen bond interactions with C18.

2. Methodology

Building the System

The initial coordinates for our study were taken from the crystallized structure of C_5 -MTase from *Haemophilus haemolyticus* (*M.HhaI*) in complex with a DNA 13-mer and with S-Adenosyl Homocysteine, SAH (PDB ID:2HR1). This structure was solved with a resolution of 1.96 Å.¹⁵ In the X-ray structure the target cytosine nucleobase is flipped out of the DNA helix and inserted into the catalytic pocket of the *M.HhaI* enzyme. We remodeled the crystalized SAH into

SAM by adding a methyl group while DNA was modelled as a dodecamer because of the absence of structural information for one of the terminal nucleosides in the X-ray structure. We calculated the pK_a of all titratable residues at pH=7 using Propka 3.0.³⁴⁻³⁷ Once pK_a values were determined, hydrogen atoms were added to the enzyme-SAM-DNA complex using the LEAP module of AMBER program.³⁸ The pK_a obtained for Cys81 was 10.6 in absence of DNA and about 14.0 when DNA was included. This change reflects the electrostatic influence on the pK_a . However, as these calculations do not consider the effect of the DNA ionic atmosphere we built up two different model systems: one where Cys81 was neutral and another one where it was deprotonated and we analyzed the deprotonation process. As discussed below, our calculations show that the free energy difference between the protonated and unprotonated forms of Cys81 in the ternary complex must be small. Both systems were solvated with a buffer of water molecules of at least 15 Å around the system in all directions, keeping crystallization water molecules. This procedure was carried out with LEAP module and resulted in simulation boxes of 88 x 88 x 99 Å of size with 19/20 sodium ions added to neutralize the system for the protonated/deprotonated Cys81 models, respectively. The final systems consisted in a total of 23523/23235 water molecules and a total of 76592/75728 atoms, for protonated and deprotonated Cys81 models, respectively. We used the TIP3P³⁹ and AMBER99SB force fields,⁴⁰ including the DNA-refined parameters parmbsc0,⁴¹ to carry out optimizations and molecular dynamics simulations with the NAMD program.⁴² Parameters for the cofactor SAM were taken from Stacklies et al.⁴³ These parameters were already successfully employed in simulations of another DNA MTase.⁴⁴ Analysis of the hydrogen bond interactions established between the cofactor SAM and the protein during our simulation confirmed the correctness of these parameters (see Figure S1 in the Supporting Information).

Molecular Dynamics Simulations

The same simulation protocol was applied to both model systems containing protonated and deprotonated Cys81. Classical MD simulations were performed with a time step of 2 fs applying the SHAKE algorithm⁴⁵ to bond lengths involving hydrogen atoms. Periodic boundary conditions were applied using the

particle mesh Ewald method with a cutoff of 12 Å in the real part of the sum.⁴⁶ Short range non-bonded interactions were calculated using a smooth switching function with inner and outer cutoff radii of 10 and 12 Å. In order to equilibrate the systems water molecules and counterions were first relaxed by means of 1 ns of MD simulation in the NPT ensemble followed by 1 ns of NVT simulation. Pressure was kept at 1 atm using the Langevin piston method with a period of 100 fs and a damping time constant of 50 fs. Langevin-Verlet dynamics were performed using a damping coefficient of 10 ps⁻¹ at a target temperature of 300 K, the same temperature than in previous theoretical studies.^{20,23,47} Subsequently, the enzyme and DNA atoms were gradually released starting with side chains and nucleotide bases atoms and finishing with backbone atoms. At each step the system was first minimized by means of 10000 steps of conjugate-gradient algorithm followed by MD simulations where the system was gradually heated from 0 to 300 K rescaling velocities during 500 ps. Additional 500 ps were performed at constant temperature. During this process harmonic restraints were applied to the hydrogen bond interactions observed between the flipped out cytosine base and active site residues (Phe79, Glu119, Arg163 and Arg165, see Figure 1), to the distance between the sulfur atom of Cys81 and the C₆ atom of cytosine and to the distance between the the carbon atom of SAM methyl group and the C₅ atom of cytosine. These harmonic restraints were gradually removed through several cycles of optimization and MD simulations. After equilibration, we performed 100 ns of fully unrestrained MD simulation in the NVT ensemble at 300 K for the two model systems.

QM/MM Calculations

The fDynamo program⁴⁸ was used to perform QM/MM-MD simulations, starting with a structure selected from the last 10 ns of the MD simulation. The QM subsystem was described at the AM1 semiempirical level of theory⁴⁹ and at the M06-2X level⁵⁰ (see below) while the MM system was described using the AMBER force field as implemented in fDynamo.^{44,51} The QM subsystem included SAM, the target cytosine base and the side chains of those residues participating in each of the steps of the chemical reaction. Thus, the QM region is different for each reaction step and we refer the reader to the Supporting Information for a detailed list of the side chains atoms incorporated into the QM

subsystem at each reaction step (see Schemes S1 to S5). To saturate the valence of the QM/MM frontier we used the link atom procedure.^{52,53} In the AMBER force field nucleotides and aminoacids side chains have no-integer charge. To overcome this issue we redistributed the small fractional charges remaining in the MM subsystem between C1' and H1' atoms of the sugar, in the case of the target cytosine, and between CA and HA atoms for those residues whose side chains were included in the QM subsystem.⁵⁴ For each QM/MM definition the system was prepared by means of a minimization and a 500 ps long NVT simulation with periodic boundary conditions where all the atoms of the system were free to move. Short range non-bonded interactions were calculated using a smooth switching function with inner and outer cutoff radii of 10 and 12 Å. We also explored the Potential Energy Surfaces (PES) at the M06-2X/6-31G*/MM level for the complete system as described above, studying different mechanistic options. **These computational levels have been shown to produce good results in the study of other methyltransferases.**^{44,55} These calculations were done with an interface between fDynamo and Gaussian09.⁵⁶

Free Energy Calculations

We performed QM/MM-MD simulations to obtain minimum free energy paths (MFEPs) by means of the on-the-fly string method.³³ After obtaining the MFEP for a given mechanistic step a path collective variable^{55,57} was defined to obtain the potential of mean force (PMF) using umbrella sampling.⁵⁸ This methodological combination offers an efficient approach for studying the multidimensional free energy landscape associated with complex reactions where the dependence of the reaction coordinate on several degrees of freedom has to be considered. Therefore, we were able to take into account multiple variables that conform the best global reaction coordinate at each step, as required in the analysis of the complex reaction mechanism of the C₅-MTase. The set of coordinates selected to trace the MFEP for each reaction step (the active spaces) are detailed in the Supporting Information (Schemes S1 to S5). A total of 30 string nodes were used to localize each MFEP and the values employed for the force constant and friction in the string method were $K = 6000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ and $\gamma = 9000 \text{ ps}^{-1}$, respectively. For each string we performed 30 ps of Langevin dynamics at 300 K with a time step of 1 fs or 0.5 fs

if a proton transfer was involved. The strings converged within the first 10 ps of dynamics and the averaged positions of the string nodes were determined over the last 20 ps. We used different initial guesses to explore all possible reaction mechanisms within the selected active spaces. In most cases different initial guesses converged to the same string indicating that among the proposed mechanisms only one corresponds to a MFEP on the free energy surface.

In order to obtain the PMF a set of 100 points were interpolated from the string nodes to be used in the definition of the collective path coordinate (the s coordinate)^{55,57} for each MFEP. This coordinate was parameterized with values ranging from 0 to L , where L corresponds to the total arc-length of the path. Negative values and values larger than L correspond to structures extrapolated beyond the reactant and product structures, respectively. This extrapolation procedure is used to ensure well-defined reactant and product valleys in the PMF. For each converged string, the force constants for 120 umbrella sampling windows were adjusted to flatten the free energy profiles obtained from the string method with values ranging between 2000 and 5000 $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$. Initial structures for each window were taken from the final snapshot of the closest string node. 10 ps of relaxation followed by 30 ps of production dynamics with a time step of 1.0 or 0.5 fs, if a proton transfer is involved, were performed for each simulation window. Then PMFs were integrated from the respective histograms of the s coordinate using WHAM.⁵⁹ **PMF statistical uncertainties were calculated using statistical bootstrap technique⁶⁰ with 1000 samples generated from the original data from Umbrella Sampling.** PMFs were obtained using interpolated corrections^{61,62} along the s coordinate:

$$E = E_{QM}^{LL} + E_{QM/MM}^{LL} + E_{MM} + \text{Spl}[\Delta E_{LL}^{HL}(s)]$$

where Spl denotes a one dimensional cubic spline function and its argument, ΔE_{LL}^{HL} , is a correction term obtained as the difference between the single-point high-level (HL) energy of the QM system and the low level one (LL=AM1). As HL theory we chose M06-2X/6-311+G**.⁵⁰ The structures were optimized at different values of the s coordinate at the low level. To confirm the conclusions obtained with our correction scheme we also explored the PES associated with the converged mechanisms at the M06-2X/6-31G**/MM level of theory.

We also performed free energy perturbation (FEP) calculations in order to unravel which is the base abstracting the C₅-proton during the β-elimination step. Previous theoretical studies^{21,23} have considered that a hydroxide anion would be a suitable base to abstract this proton and to transfer it to the solvent. These proposals estimated the free energy cost of generating the hydroxide anion from its concentration in pure water (12 kcal·mol⁻¹).^{21,23} However, it should be taken into account that the free energy of generating such an anion in the catalytic pocket can be quite different from the value obtained in bulk water due to the very different environment. To investigate this issue we performed FEP calculations to estimate the free energy cost to bring the anion to the catalytic pocket from bulk. Two alchemical transformations were made for this purpose: a water molecule was transformed into a hydroxide ion in a periodic box of water molecules and the same calculation was performed for a water molecule located inside the active site. Finally, in order to estimate the free energy cost of having a hydroxide anion in the active site, the free energy difference between the two transformations was added to the free energy cost estimated from the self-ionization product of pure water. Computational details for these FEP calculations are provided as Supporting Information (see Free Energy Perturbation Protocol and Scheme S6). Using the same protocol, we performed FEP calculations to obtain the free energy associated to the delivery of the leftover proton from the active site to the bulk once it has been abstracted by a water molecule found in the active site (see Free Energy Perturbation Protocol and Scheme S7).

3. Results

MD Simulations

A previous analysis of 1 ns MD simulation on the same enzymatic system emphasized the role of some residues in catalysis (Cys81 and Glu119), and in the stabilization of the DNA chain (Gln237 and Ser87) and of the flipped out cytosine (Arg163, Arg165 and Glu119).²⁰ In the present work we performed 100 ns long NVT MD simulations for the two DNA-enzyme-SAM systems, containing a protonated or a deprotonated Cys81 residue, respectively. From these trajectories we analyzed DNA structure, DNA-enzyme interactions and active

site interactions established between the flipped cytosine base and the surrounding residues. Here below we only present the analysis for the Cys81 deprotonated system, which is the one undergoing the nucleophilic attack to the target cytosine base. Plots of the Root-Mean-Square Deviations (RMSDs) of enzyme and DNA backbone atoms in the Cys81 deprotonated system are presented as Supporting Information (Figure S2). The number of direct and water mediated hydrogen bonds between the DNA and the enzyme are provided as Supporting Information (Figure S3) as well as the analysis of the important protein-DNA hydrogen bond interactions (Figure S4). The average number of hydrogen bond interactions between nucleobase pairs is also shown as Supporting Information (Figure S5). Finally, a complete analysis of the simulation of the neutral Cys81 system is provided only in the Supporting Information (see Figures S6-S10).

Active Site Interactions

The flipped out cytosine base C18 is stabilized by several hydrogen bond contacts when inserted in the active site of *M.HhaI*. Figure 2 shows the average distances and standard deviations for the hydrogen bonds established between the target base C18 and the protein during the MD simulation. The exocyclic nitrogen atom of C18 establishes two hydrogen bonds, one with the O_{ε1} atom of Glu119 and another with the carbonyl oxygen of Phe79 backbone. The O_{ε2} atom of Glu119 acts as hydrogen bond donor to the N₃ atom of the C18 base. In some cases a water molecule can be placed between these donor and acceptor atoms. This water molecule (labelled as 355) is found in the X-ray structure at 2.9 Å of the Glu119 O_{ε2} atom.¹⁵ The O₂ oxygen atom of the cytosine base establishes up to three different hydrogen bonds with the surrounding arginines: Arg163 is hydrogen bonded through the N_{η1} atom while Arg165 can be simultaneously hydrogen bonded through N_{η2} and N_ε atoms. These arginine residues have been postulated to play an important role stabilizing the cytosine base in the active site and accommodating the negative charge developed on the cytosine ring after the nucleophilic attack by Cys81.¹⁵ In our MD simulation with Cys81 deprotonated the sulfur atom of this residue is found at an average distance of 3.46 ± 0.18 Å from the C₆ atom of the target C18 base. In general, the small values of the standard deviations associated with the distances

analyzed in this section indicate that the interactions between active site residues and the C18 base remain stable over the simulation with small fluctuations.

After a 100 ns simulation of the system where the Cys81 residue was modelled in its neutral form the average distance between the sulfur atom and the C₆ atom of the cytosine was found to be 3.36 ± 0.27 Å, very similar to the value found in the simulation of the unprotonated form. In order to elucidate the mechanism for Cys81 deprotonation we also monitored the interactions established between this residue and its surroundings. During this simulation the thiol group of Cys81 was found donating a hydrogen bond either to the O_γ atom of Ser85 (with an occupancy of 16.7 %), to a water molecule (with an occupancy of 10.9 %) or to the O_δ atom of Asn120 (7.5 % of occupancy). It has been previously proposed that the water molecule hydrogen bonded to Cys81 can assist its deprotonation transferring the proton to the nearest phosphate group of DNA.²² In our simulation, in most of the structures where the thiol group of a neutral Cys81 donates a hydrogen bond to the O_γ atom of Ser85, the hydroxyl group of this residue is, in turn, hydrogen bonded to the C18 phosphate group of DNA. Thus, an activated Ser85 could also act as the base in charge of Cys81 deprotonation. This possibility is also explored in the present work.

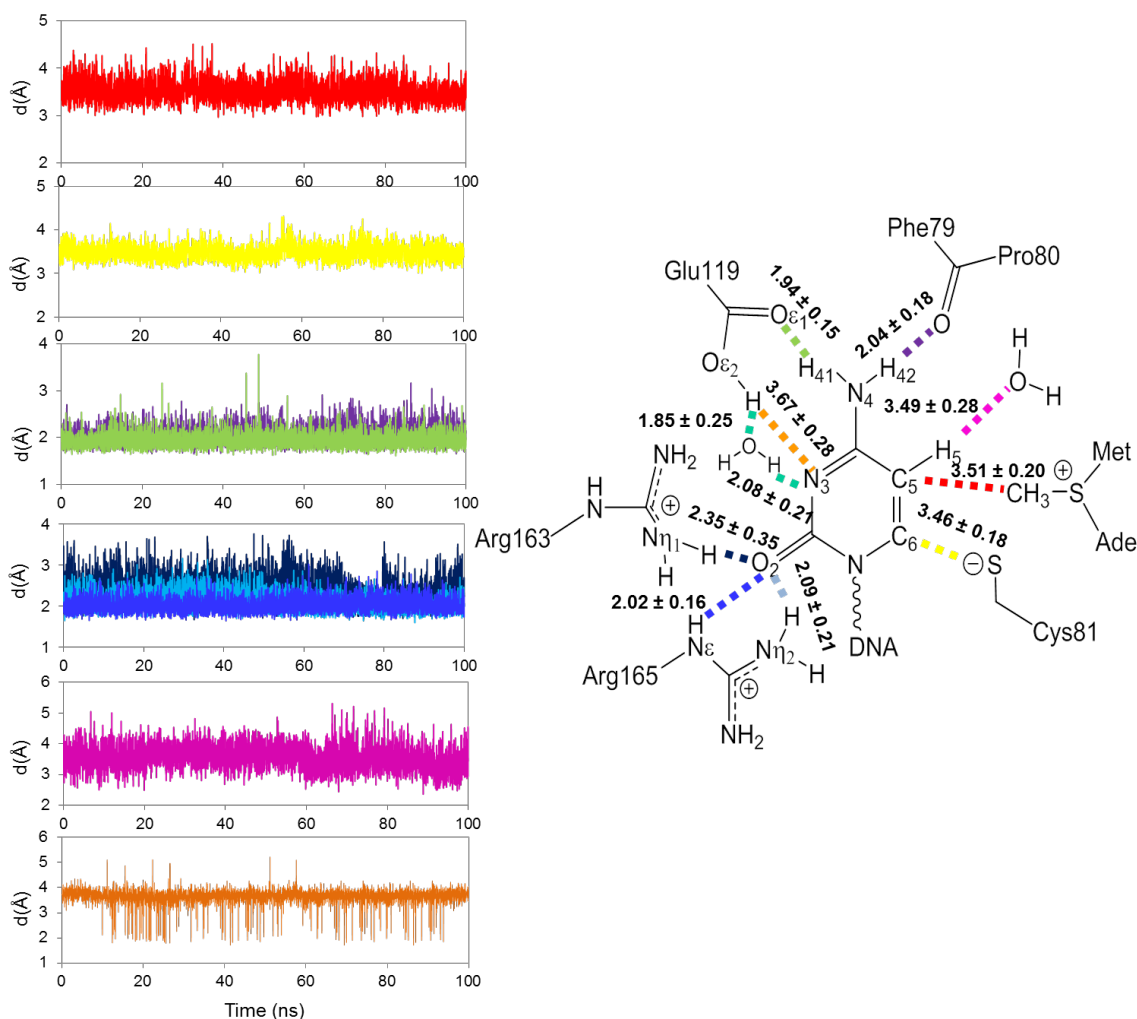


Figure 2. Interactions established between the C18 base and amino acids of the *M.HhaI* active site. The plots on the left show the instantaneous distances observed during the simulations. The values shown in the active site scheme displayed on the right correspond to the average values and their standard deviations. Distances are given in Å.

M.HhaI Reactivity

Cytosine methylation at C₅ position in *M.HhaI* may involve the participation of several residues present in the active site. Although extensive research has been carried out in this system, the role of some of these residues remain unclear.¹⁰⁻²³ Previous theoretical studies of the reaction mechanism have been carried out exploring the Potential Energy Surface (PES) and/or the Free Energy Surface (FES) associated with a reduced number of geometrical reaction coordinates (usually one or two). However, when the mechanistic step

under analysis involves different chemical events that can or cannot take place simultaneously the selection of a reduced set of distinguished reaction coordinates could inevitably bias the conclusions. This is the case of cytosine C₅-methylation in *M.HhaI*, a process that can take place concertedly or not with the nucleophilic addition of Cys81 to C₆ position and with a hypothetical proton transfer from Glu119 to N₃. Therefore a reliable approach for the analysis of this reaction must face the problem associated with the exploration of a multidimensional FES as a function of all the relevant geometrical coordinates and then determine the most reliable mechanism in terms of free energy barriers.

In the present work we have applied the on-the-fly string method to obtain the MFEP on a multidimensional FES. This approach has already shown to be successful for the analysis of several complex enzymatic reactions.^{44,55} In our calculations the coordinates used to trace the MFEP included all the interatomic distances associated with all possible bond breaking and forming events that could take place at each reaction step. Therefore no *a priori* assumptions have to be made about the reaction mechanism. For each possible mechanism an independent string calculation was performed starting from an adequate initial guess. Subsequently, PMF calculations with interpolated corrections were performed for each converged string using a path collective variable that maps the advance of the system along the MFEP. As a general scheme (Scheme 1) the reaction mechanism consists in three steps: i) deprotonation of the Cys81 residue; ii) methylation at the C₅ position and iii) β -elimination.

i) Cys81 deprotonation

Nucleophilic addition at C₆ position of the target cytosine is a common mechanistic step in all known C₅-MTases.³¹ In *M.HhaI* the nucleophile is Cys81. Some authors^{22,23} have proposed that deprotonation of this residue can take place through a nearby phosphate group of DNA. As discussed above, in our MD simulations with neutral Cys81 the thiol group of this residue remains hydrogen bonded to Ser85, which is a conserved amino acid in mammalian C₅-DNA MTases, or to a solvent molecule during a considerable fraction of the simulation (see Active Site interactions section). In turn, Ser85 and the water

molecule are placed next to C18 and G19 phosphate groups of DNA, respectively. Therefore we have studied two Cys81 deprotonation mechanisms, where the proton is transferred from the thiol moiety to a DNA phosphate group through a water molecule or through Ser85. Figure 3 shows the free energy profiles for the two possible mechanisms. The calculated activation free energy associated with the Ser85-mediated deprotonation of Cys81 (from state **1** to **2**, see Figure 3) is $14.3 \pm 0.3 \text{ kcal}\cdot\text{mol}^{-1}$, while the value obtained for the water-mediated deprotonation (**1** to **2'**) is $14.8 \pm 0.2 \text{ kcal}\cdot\text{mol}^{-1}$. In both cases the deprotonation takes place in a concerted asynchronous process (see Table S1). The PESs for the two reaction mechanisms and a table showing key distances for the stationary states appearing along the deprotonation processes are provided in Figure S11 and Table S1 of the Supporting Information. The proton transfer from Cys81 to DNA phosphate is largely endergonic (between 7.6 and 9.2 $\text{kcal}\cdot\text{mol}^{-1}$ depending on the considered mechanism). However, a protonated DNA is not the final state because the proton can be easily released from the phosphate group to the solvent. The C18 and G19 phosphate groups are exposed to the solvent and therefore can be easily deprotonated resulting in state **3** (see Figure 3). Considering that the pK_a of a DNA phosphate group must be about 1.5-2.0⁶³ the free energy variation associated with the transfer of the proton from these phosphate groups to bulk water at $\text{pH}=7$ must be about -7.5 $\text{kcal}\cdot\text{mol}^{-1}$. Thus, the global proton transfer from neutral Cys81 to bulk water in the Michaelis ternary complex must be a nearly thermoneutral process.

It must be taken into account that in our model Cys81 deprotonation is facilitated after DNA binding. The kinetic scheme of the conformational transitions consists in the association of the DNA and the C5-MTase enzyme, the binding of the cofactor, the base flipping and loop locking of the target base with the bound cofactor inside the catalytic active site.¹⁸ In this sense our results would be in agreement with direct optical observation of cytosine flipping¹⁸ showing a tight coupling between base flipping and closure of the catalytic loop where Cys81 is placed. This study showed that the rate of covalent binding is considerably slower than cytosine flipping and according to our results Cys81 deprotonation could be the rate-limiting step for the nucleophilic attack of this residue to the target base (see below).

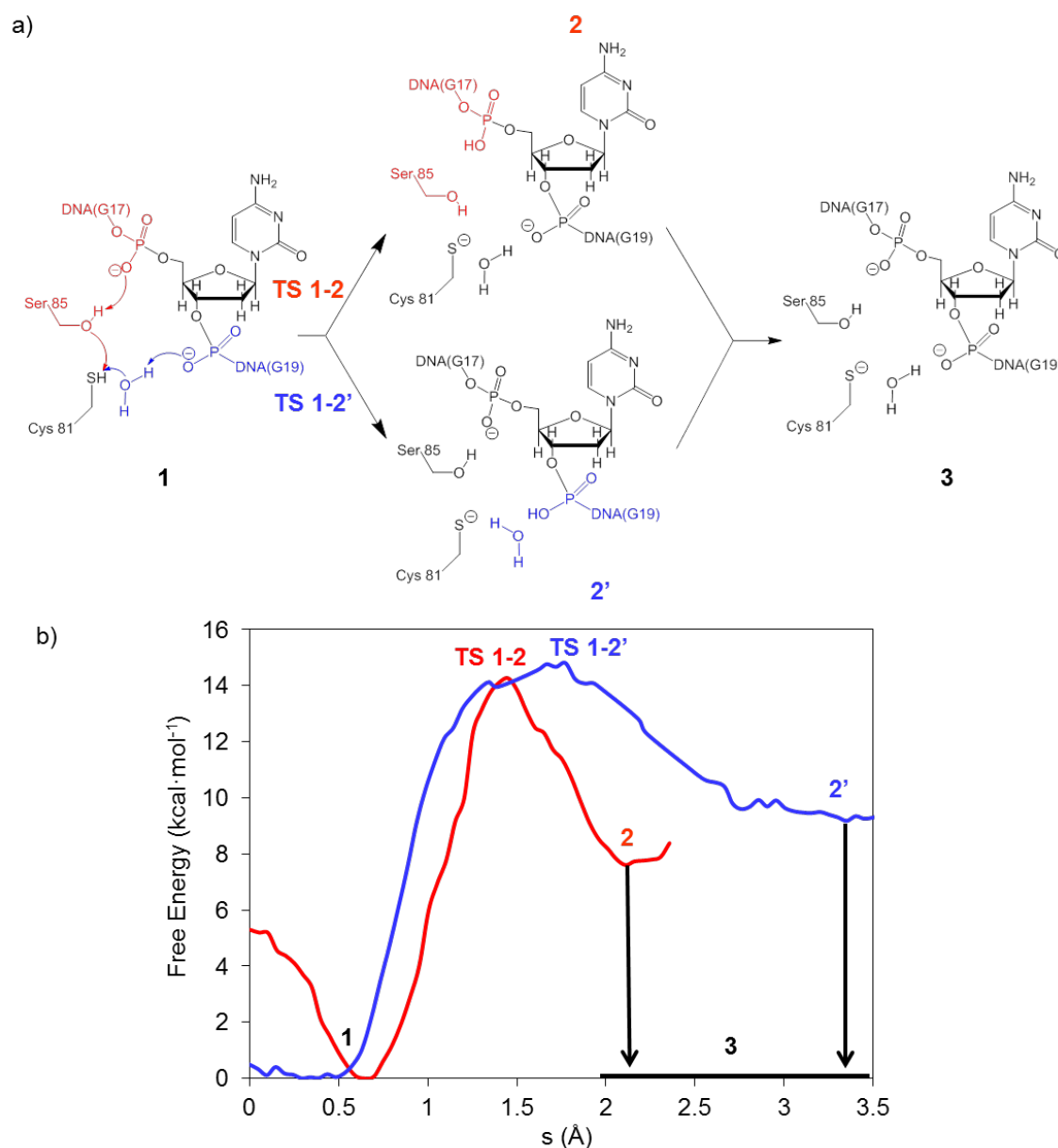


Figure 3. a) Scheme for two reaction mechanisms analyzed for Cys81 deprotonation. b) AM1/MM PMFs in ***M.HhaI* corrected** at the M06-2X/6-311+G** level, obtained for the Ser85-mediated proton abstraction (profile in red) and the water-mediated proton abstraction (in blue) as a function of the collective path coordinate s .

ii) C_5 methylation

It is widely accepted that C_5 -methylation involves the nucleophilic addition of a deprotonated Cys81 residue to the C_6 carbon atom of cytosine and the methyl transfer from the cofactor SAM to the C_5 position of the nucleobase.^{1,6} X-ray

structural analysis has shown that a covalent adduct between Cys81 and the cytosine nucleobase can be formed.^{9,64} In addition, it has been shown that activation of the target nucleobase through the nucleophilic attack is fast and reversible.¹² However, while some computational studies concluded that the nucleophilic addition and the methyl transfer take place concertedly;^{21,23} another study proposed a stepwise mechanism.²² In addition, the role of Glu119 in this step is also under debate. It was proposed that activation of the nucleobase for nucleophilic addition required protonation of the N₃ position by Glu119.⁹ Kinetic studies on Glu119 *M.HhaI* mutants have shown that this residue is essential for catalysis.¹⁷ Regarding theoretical analysis, a study in aqueous solution showed that the N₃ atom could be protonated during the nucleophilic addition,¹⁹ while more recent theoretical calculations on reduced active site models have shown that this protonation could be not necessary^{21,22} and that the main role of Glu119 would be to contribute to cytosine binding in the active site. QM/MM calculations agreed with this last picture, although transient protonation of the N₃ position from Glu119 was observed at advanced stages of the methylation step.²³

In order to solve these questions we explored the multidimensional FES associated with the methylation step, considering the bond distances associated with the proton transfer from Glu119 to N₃, the nucleophilic addition of Cys81 to C₆ and the methyl transfer from SAM to C₅ (a total of five interatomic distances, see Supporting Information, Scheme S3). The string method was employed using different initial guesses that correspond to different mechanistic proposals (stepwise and concerted mechanisms for the addition and methyl transfer events; with and without proton transfer from Glu119). After several trials we obtained a single converged MFEP corresponding to the reaction mechanism depicted in Figure 4 together with the corresponding PMF corrected at the M06-2X level.

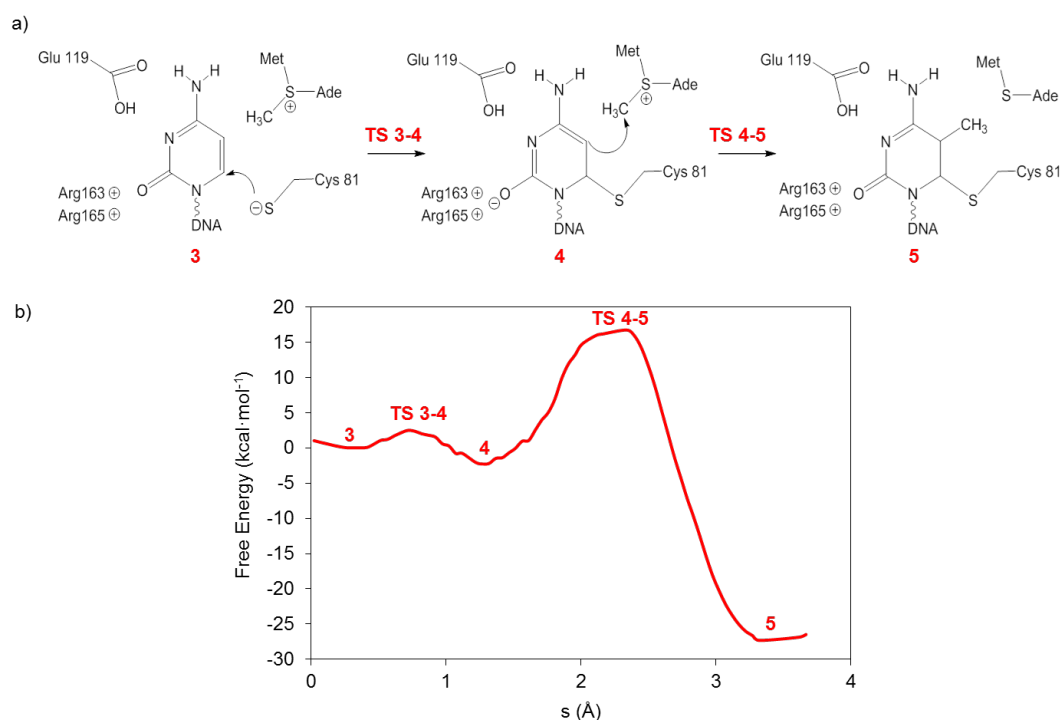


Figure 4. a) Scheme for the cytosine methylation step consisting in the nucleophilic attack by Cys81 and the methyl transfer from SAM. b) AM1/MM PMFs, corrected at the M06-2X/6-311+G** level, obtained for the Cys81 addition (**TS 3-4**) and methyl transfer (**TS 4-5**) in *M.HhaI* as a function of the collective path coordinate s .

Our results show that methylation is a stepwise mechanism where Cys81 addition to cytosine precedes the methyl transfer from SAM to C₅. Both processes take place without requiring the proton transfer from Glu119 to the N₃ position of cytosine. While it has been argued that this proton transfer could activate the nucleobase for the nucleophilic attack, **our calculations show that this addition takes place with a very low free energy barrier ($2.5 \pm 0.3 \text{ kcal}\cdot\text{mol}^{-1}$)** even if the cytosine base remains unprotonated. According to our results, the nucleophilic addition of Cys81 to cytosine C18 can be described as a rapid equilibrium (see states **3** and **4** in Figure 4), in agreement with experiments.^{12,13} **Afterwards, the methyl transfer takes place with a free energy barrier of $19.1 \pm 0.4 \text{ kcal}\cdot\text{mol}^{-1}$ from **4**.** This calculated free energy barrier is in good agreement with the value derived from the application of Transition State Theory to the pre-steady-state rate constants at 310 K that range, depending on the DNA substrate, from 0.12 to 0.039 s^{-1} ,^{12,17,65} resulting in free energy barriers from

19.5 to 20.2 kcal·mol⁻¹. The reaction free energy for the methyl transfer step is -25.1 kcal·mol⁻¹, indicating that formation of state **5** is an irreversible step, also in agreement with experimental observations.⁶⁶ A previous QM/MM analysis²³ based on B3LYP calculations found that the nucleophilic addition took place simultaneously to the methyl transfer. In order to check the robustness of our mechanistic proposal we traced M06-2X/MM and B3LYP/MM PESs as a function of the C₆-S(Cys81) distance and the antisymmetric combination of distances defining the methyl transfer from SAM to C₅ ((S(SAM) - CH₃) - (CH₃ - C₅)). These results, presented as Supporting Information (Figure S12), confirm that while at the B3LYP level the reaction can be described as concerted but asynchronous (with Cys81 addition preceding methyl transfer) the reaction mechanism is clearly stepwise at the M06-2X level. **The fact that the B3LYP treatment does not incorporate dispersion contributions to stabilize the unbounded Cys81 could explain the failure of this method to describe the rapid equilibrium between **3** and **4** and could be alleviated including dispersion correction to the B3LYP functional.**

Table 1 provides average values for key distances relevant for the C₅-methylation step. In our QM/MM description the reactant state (**3**) has the methyl group of SAM placed 3.27 ± 0.09 Å from the C₅ carbon atom of the target cytosine while the sulfur atom of the Cys81 residue is 3.15 ± 0.11 Å from the C₆ atom. The nucleobase establishes five hydrogen bond interactions with Phe79, Glu119, Arg163 and Arg165 (see Table 1). The first transition state (**TS 3-4**) consists in the addition of the sulfur atom of Cys81 to C₆, resulting in an adduct with an average S-C₆ distance of 2.12 ± 0.08 Å. Two noticeable changes are observed at this intermediate structure: as a result of the development of a negative charge on cytosine C18, both the methyl group of SAM and the Glu119 residue approach to the nucleobase (see Table 1). From the resulting covalent adduct the reaction mechanism continues with the methyl transfer from SAM to the C₅ position. At the TS (**TS 4-5**) the distance between the sulfur atom of SAM and the carbon atom of the methyl group is 2.10 ± 0.05 Å while the distance between this methyl group and the C₅ atom of cytosine is 2.21 ± 0.05 Å and the average S-CH₃-C₅ angle is close to linearity (170 ± 4 degrees). The positioning of the transferred methyl group between the donor and acceptor

atoms is stabilized by means of C-H \cdots O hydrogen bond interactions established between the methyl group and the carbonyl oxygen atoms of Asn304 and Gly78 (average distances of 2.62 ± 0.09 and 2.71 ± 0.12 Å, see Figure 5). Similar stabilizing interactions have been identified in other MTases.⁶⁷ **TS 4-5** is also stabilized by the hydrogen bond established between the N₃ atom of cytosine and Glu119 that displays a shorter average distance at this TS. The reinforcement of the hydrogen bond formed between the N₃ atom of the nucleobase and Glu119 indicates a preferential stabilization for the intermediate species **4** and the subsequent TS (**TS 4-5**), which is the dynamical bottleneck for methylation. The hydrogen bond established between this residue and the exocyclic amine group of the C18 nucleobase is also reinforced during the methylation step (see Table 1). These observations agree with experimental analysis of Glu119 mutants that concluded that this residue is essential not only for the stabilization of the extrahelical cytosine but also for catalysis.¹⁷ However, according to our results, proton transfer from this residue to the N₃ position is not required at this stage.

Table 1. Key average distances (Å) obtained for reactant, intermediate, product and transition states of the C₅-cytosine methylation step in *M.HhaI*. Standard deviations are included. Free energies obtained after M06-2X/6-311+G**/MM corrections are given in kcal·mol⁻¹.

	3	TS 3-4	4	TS 4-5	5
d(S(SAM) - CH ₃ (SAM))	1.83 ± 0.05	1.82 ± 0.04	1.83 ± 0.04	2.10 ± 0.05	2.96 ± 0.03
d(CH ₃ (SAM) - C ₅ (C18))	3.27 ± 0.09	3.09 ± 0.14	2.93 ± 0.05	2.21 ± 0.05	1.52 ± 0.03
d(S(Cys81) - C ₆ (C18))	3.15 ± 0.11	2.54 ± 0.06	2.12 ± 0.08	2.06 ± 0.06	1.96 ± 0.05
d(O ϵ_2 (Glu119) - H ϵ_2 (Glu119))	0.98 ± 0.02	0.98 ± 0.02	0.97 ± 0.02	0.97 ± 0.03	0.98 ± 0.03
d(HO ϵ_2 (Glu119) - N ₃ (C18))	3.14 ± 0.34	3.25 ± 0.29	2.72 ± 0.56	2.17 ± 0.22	3.45 ± 0.34
d(O ϵ_1 (Glu119) - H ₄₁ (C18))	2.24 ± 0.32	2.08 ± 0.19	2.34 ± 0.33	2.19 ± 0.25	2.02 ± 0.22
d(O(Phe79) - H ₄₂ (C18))	2.16 ± 0.17	2.09 ± 0.18	2.12 ± 0.18	2.05 ± 0.15	2.04 ± 0.17
d(HN ϵ (Arg165) - O ₂ (C18))	1.91 ± 0.09	1.95 ± 0.13	1.92 ± 0.11	1.89 ± 0.09	1.96 ± 0.13
d(HN η_2 (Arg165) - O ₂ (C18))	2.16 ± 0.16	2.04 ± 0.15	2.02 ± 0.13	2.09 ± 0.17	2.05 ± 0.14
d(HN η_1 (Arg163) - O ₂ (C18))	1.99 ± 0.09	1.99 ± 0.14	1.96 ± 0.14	2.11 ± 0.15	2.17 ± 0.21
ΔG	0.0	2.5 ± 0.3	-2.3	16.8 ± 0.4	-27.4

Nevertheless we have thoroughly explored the possibility of forming a protonated cytosine intermediate by computing the PES and the free energy profile associated with the process where Cys81 attacks the C₆ atom of cytosine and Glu119 donates a proton to the N₃ atom of cytosine (see Supporting Information Figure S13-S15 and Table S2). These results confirm that protonation of the N₃ atom of cytosine by Glu119 is not needed during the nucleophilic attack of Cys81 to C₆. However, once intermediate **4** is formed the extrahelical cytosine can be easily protonated at position N₃ by Glu119 with a low energy barrier. The resulting species, where the N₃ atom of the cytosine is protonated and the Cys81 is bonded to the cytosine, does not easily undergo methylation (the energy barrier is significantly higher than for the unprotonated species, see Figure S15 in Supporting Information) and thus can be considered as a side-reaction in equilibrium with intermediate **4**, which is the species that can reach the methylated product through **TS 4-5**.

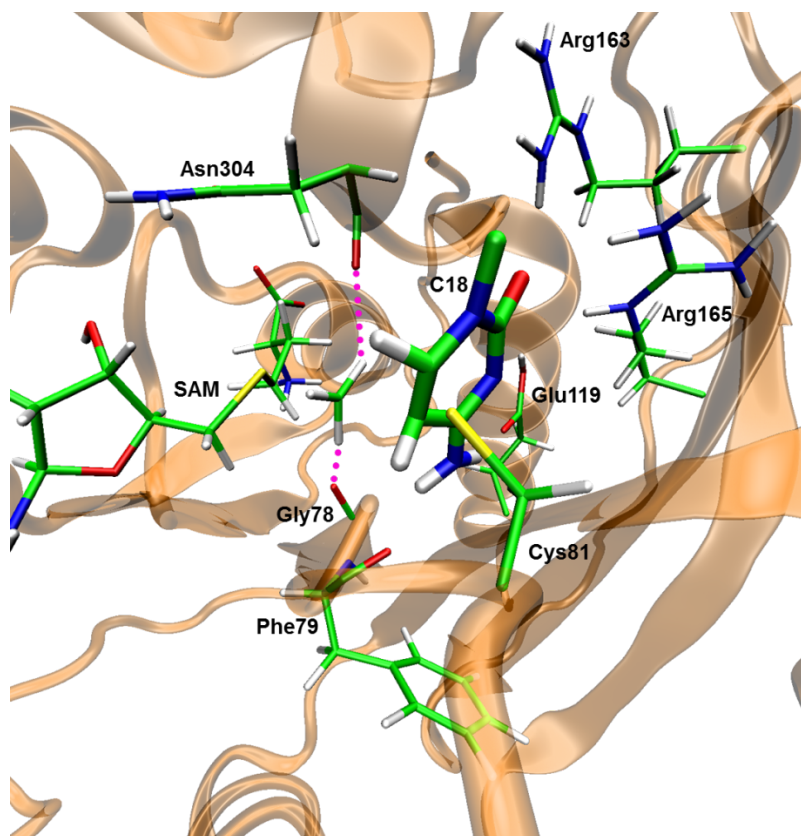


Figure 5. TS structure for the rate limiting step of the methylation process corresponding to the methyl transfer from SAM to the C₅ atom of cytosine (**TS 4-5**). Dotted lines indicate the stabilizing interactions established between the methyl group and the carbonyl oxygen atoms of Asn304 and Gly78.

Our results show that the main role of Glu119 in the methylation step would be attained by means of hydrogen bonding to the substrate. Substrate protonation would take place in a subsequent step (see below). However, these results are confronted to the experimental observation showing that the Gln119 variant is at least 10^6 fold less active than the wild type enzyme and that also the global and active site structures of the Gln119 variant and WT are similar.¹⁷ This result could indicate that hydrogen bonding is not enough at this stage of the reaction and that proton transfer to N_3 plays an essential role not captured in our QM/MM calculations. However, alternative explanations are also feasible for the experimental result of the Gln119 variant, considering that the mutation can perturb the active site and its catalytic properties. To investigate further this point we carried out 60 ns of MM MD simulations of the ternary complex in the Gln119 variant (see Figure S16 in Supporting Information). Analysis of the simulation shows that while Gln is able to keep the hydrogen bond to N_3 , other changes introduced in the active site after the mutation could explain the low activity of the mutant. For example, hydrogen bond interactions of Arg163 and Arg165 with the O_2 atom of the substrate, which are important for catalysis, are substantially weaker in the mutant than in the wild type enzyme. The water molecule hydrogen bonded to N_3 is also found at larger distances in the mutant. The hydrogen bond of the mutated residue to the exocyclic amino group of the target base is also weakened with respect to that reported for the wild type enzyme (Figure 4). These changes could also explain the low catalytic efficiency of the mutant without invoking proton transfer to N_3 .

iii) β -elimination

In order to form the final reaction product, intermediate **5** formed after the methylation step has to eliminate the excess proton located at position C_5 of the extrahelical cytosine and the covalent bond formed with Cys81 must be broken. It should be taken into account that the C_5 -H and the C_6 -S bonds are found in the same side of the cytosine ring and then the elimination reaction would be a *syn* elimination that requires larger activation energies.²² In addition, there is little experimental evidence about the nature of groups that could be involved in acid or base-assisted catalysis during this β -elimination process.¹⁷

Several theoretical studies have tried to characterize which is the residue in charge of the abstraction of the excess proton from C₅ position. It was suggested that a DNA phosphate group could act as a base for proton abstraction, assisted by a bridging water molecule,⁹ although this possibility seems difficult to accomplish after inspection of the distances from DNA phosphate groups to C₅. It was also proposed that a hydroxide anion could have access to the active site through a water channel observed in X-ray structures^{15,20} and then act as the reacting base.²¹ This possibility was reconsidered in a recent QM/MM exploration of the reaction mechanism.²³ The free energy cost of generating a hydroxide anion was estimated from the concentration of this ion in pure water at pH=7. Once the hydroxide anion is generated (at a free energy cost of 12 kcal·mol⁻¹), the free energy barrier for proton abstraction was estimated to be 8.7 kcal·mol⁻¹, resulting in a total free energy barrier of 20.7 kcal·mol⁻¹, a value compatible with the free energy barriers derived from the observed steady-state rate constants at 37 °C (0.011-0.085 s⁻¹),^{12,17,65} this is between 19.7 and 21.0 kcal·mol⁻¹. However, these proposals have never considered the free energy cost of generating the hydroxide anion in the active site, an environment which is quite different from that found for the anion in bulk water. We thus decide to critically evaluate this proposal considering the free energy cost of generating the hydroxide anion in the active site ($\Delta G_{OH^-,as}$) as the sum of two contributions, the free energy cost of having the anion in pure water ($\Delta G_{OH^-,aq}$) plus the free energy cost of moving this anion from bulk water to the active site:

$$\Delta G_{OH^-,as} = \Delta G_{OH^-,aq} + \Delta G_{aq \rightarrow as} \quad (1)$$

The first term of the right hand-side in equation (1) was considered to be 12 kcal·mol⁻¹ as in previous works^{21,23} while the second term was obtained by means of a thermodynamic cycle that considered the transformation of a water molecule into a hydroxide anion in solution and in the active site of *M.HhaI* using FEP techniques (details are provided as Supporting Information in Free Energy Perturbation Protocol section). Our calculations show that the contribution due to the transportation of the hydroxide anion from the bulk to the active site of *M.HhaI* is 12.2 ± 0.2 kcal·mol⁻¹. **Note that this free energy**

difference could roughly correspond to the formation of an additional moderately strong hydrogen bond in solution. Then our estimation of the free energy cost of generating the hydroxide anion *in the active site* is $\Delta G_{OH^-, as} = 24.2 \text{ kcal}\cdot\text{mol}^{-1}$. This result clearly shows that the possibility of a hydroxide anion acting as the base in charge of C₅-proton abstraction must be disregarded. If we add this value to the free energy barrier for the proton abstraction estimated in the aforementioned QM/MM calculations (8.7 kcal·mol⁻¹) the total free energy barrier would be 32.9 kcal·mol⁻¹, too high for an enzymatic reaction. The reason for this large increase is that the interactions established by a hypothetical hydroxide anion in the active site of *M.HhaI* would be much less favorable than in bulk water, where strong hydrogen bonds can be established between the anion and many available water molecules. Obviously, the opposite situation, where the generation of a hydroxide anion is easier in an enzymatic active site than in solution, would be possible in an active site containing a positively charged center, for example a metal cation, that could stabilize the hydroxide anion.⁶⁸

Once the presence of a hydroxide anion as the base deprotonating C₅ position has been ruled out, we reevaluated the possibility of a water molecule acting as a base. The energy barrier calculated for the proton abstraction from C₅ by a water molecule and the Cys81 expulsion was found too high to be compatible with experimental rate constants in previous studies.²¹ For this reason we explored the possibility that the process could take place after activation of intermediate **5** by means of N₃ protonation. According to this proposal, Glu119 would protonate the N₃ atom of the cytosine base at this step and not during the methylation. Experimental studies have showed that the catalytic activity of *M.HhaI* is highly sensitive to mutations of this residue. Glu119Ala, Glu119Gln and Glu119Asp mutants show a considerable reduction not only in the rate-constant of the methylation step (which, as discussed above, could be explained by the stronger hydrogen bonds formed between this residue and the nucleobase at **TS 4-5**) but also in the steady-state rate constant, that corresponds to a slower step occurring after methylation.¹⁷ In the X-ray structure¹⁵ a water molecule labelled as 404 is located 3.31 Å from the C₅ atom of the target cytosine while in our MD simulations we observe an average

distance of 3.49 ± 0.28 Å between the oxygen atom of the closer water molecule and the excess proton (H_5). We then explored the possibility that intermediate **5**, activated after protonation at the N_3 position by Glu119, undergoes the otherwise difficult β -elimination with this water molecule acting as the abstracting base. Once the proton is abstracted by this water molecule it can be easily transferred to the bulk by means of a water channel.²¹

In Figure 6 we present the mechanism found in this work for the β -elimination step. The coordinates considered for the exploration of the multidimensional FES include the distances associated with the proton transfer from Glu119 to N_3 , the C_6 -S(Cys81) distance and the distances describing the proton transfer from C_5 to the water molecule. In our QM/MM description the reaction proceeds from the methylated intermediate **5** to the first TS (**TS 5-6**) where the proton from Glu119 is transferred to the N_3 atom of the substrate leading to intermediate **6**. This step has a free energy barrier of 4.8 ± 0.3 kcal·mol⁻¹ and leads to a positively charged intermediate **6** which is 3.4 kcal·mol⁻¹ above **5**. In this intermediate the charge of the excess proton increases from **5** to **6** (0.244 and 0.327 a.u., respectively, according to CHELP calculations)⁶⁹ facilitating the proton abstraction by a water molecule. The free energy barrier for this step is of 15.3 ± 0.5 kcal·mol⁻¹. Therefore the overall free energy cost to perform the proton abstraction from structure **5** by a water molecule is 18.7 kcal·mol⁻¹ (see Figure 6). This activation free energy is in reasonable agreement with the values derived from the steady-state rate constants, 19.7-21.0 kcal·mol⁻¹ (see above). It must be noticed that the high free energy obtained for the product of this proton abstraction step (state **7**) is due to the presence of a hydronium ion in the active site (see below). The mechanism proposed in this work for C_5 proton abstraction has been further confirmed analyzing the PESs obtained at the M06-2X/MM level (see Figure S17 provided as Supporting Information).

Key averaged distances for the stationary states found in this step are provided in Table 2. In the first TS, associated with the proton transfer from Glu119 to cytosine (**TS 5-6**), the transferred proton is found at 1.29 ± 0.03 Å of the O_ϵ atom of Glu119, while the distance to the N_3 atom of cytosine is 1.27 ± 0.03 Å. In intermediate **6** the negatively charged Glu119 residue is hydrogen bonded to the protonated N_3 atom (the O_ϵ -H distance being 1.89 ± 0.10 Å). As a

consequence of the delocalization of the positive charge on the cytosine ring, in intermediate **6** the water molecule approaches the H₅ proton, displaying an average O_w-H₅ distance of only 2.11 ± 0.04 Å. Afterwards the reaction proceeds through **TS 6-7** where the water molecule abstracts the H₅ proton from the cytosine nucleobase. In this TS the bond between the oxygen atom of the water molecule and the H₅ proton is almost completely formed displaying an average distance of 1.19 ± 0.04 Å; while the distance between the C₅ atom and the H₅ proton is 1.48 ± 0.04 Å. Finally a hydronium ion and a neutral intermediate (**7**) are formed, where the distance between C₅ and H₅ atoms is lengthened to 2.33 ± 0.09 Å. The hydronium ion is hydrogen bonded to a water molecule which is part of a water channel that leads to the bulk (see Figure S18 in Supporting Information).

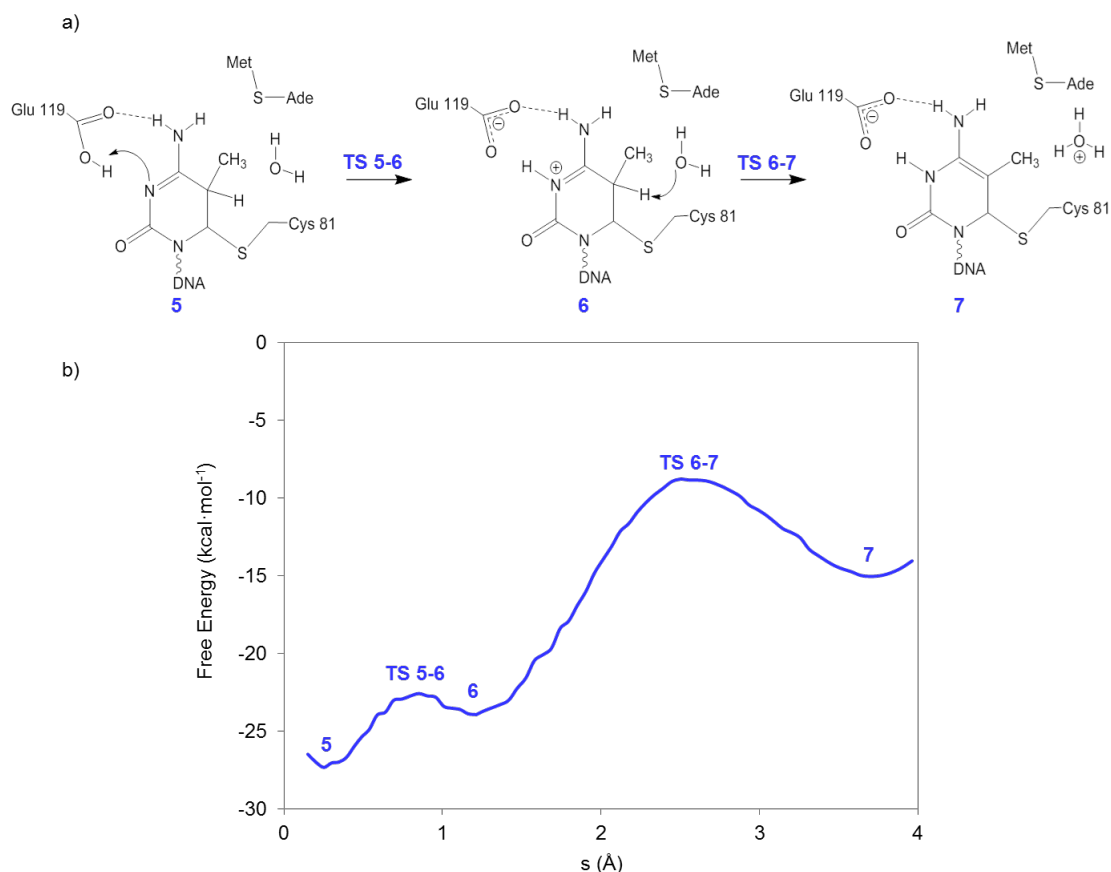


Figure 6. a) Scheme for the β -elimination step in *M.HhaI*. b) AM1/MM PMFs, corrected at the M06-2X/6-311+G** level, obtained for the protonation of N₃ atom of cytosine by Glu119 (**TS 5-6**) and H₅ proton abstraction by a water molecule (**TS 6-7**), as a function of the collective path coordinate *s*.

Table 2. Key average distances (in Å) obtained for reactant, intermediate, product and transition states for the H₅ abstraction in *M.HhaI*. Standard deviations are included. Free energies obtained after M06-2X/6-311+G**/MM corrections are given in kcal·mol⁻¹.

	TS 5-6	6	TS 6-7	7
d(C ₅ (C18) - H ₅ (C18))	1.13 ± 0.03	1.14 ± 0.03	1.48 ± 0.04	2.33 ± 0.09
d(H ₅ (C18) - O(WAT))	3.43 ± 0.25	2.11 ± 0.04	1.19 ± 0.04	1.00 ± 0.03
d(S(Cys81) - C ₆ (C18))	1.92 ± 0.05	1.93 ± 0.05	1.94 ± 0.05	1.96 ± 0.05
d(O _{ε2} (Glu119) - H _{ε2} (Glu119))	1.29 ± 0.03	1.89 ± 0.10	3.02 ± 0.27	3.31 ± 0.23
d(HO _{ε2} (Glu119) - N ₃ (C18))	1.27 ± 0.03	1.06 ± 0.03	1.01 ± 0.03	1.01 ± 0.03
d(O _{ε1} (Glu119) - H ₄₁ (C18))	2.18 ± 0.16	1.84 ± 0.11	2.30 ± 0.27	2.25 ± 0.30
d(O(Phe79) - H ₄₂ (C18))	1.95 ± 0.14	3.25 ± 0.48	4.02 ± 0.30	4.39 ± 0.35
d(HN _ε (Arg165) – O ₂ (C18))	1.94 ± 0.11	2.02 ± 0.14	2.04 ± 0.13	1.97 ± 0.11
d(HN _{η2} (Arg165) – O ₂ (C18))	2.11 ± 0.15	2.11 ± 0.17	2.11 ± 0.17	2.07 ± 0.14
d(HN _{η1} (Arg163) – O ₂ (C18))	2.46 ± 0.26	2.62 ± 0.25	2.26 ± 0.22	2.08 ± 0.17
ΔG	-22.6 ± 0.3	-24.0	-8.7 ± 0.5	-15.1

For the subsequent step we considered that this hydronium can be easily removed from the active site transferring the proton through the water channel to the bulk, reaching intermediate **7'**. In order to estimate the free energy difference between the intermediate **7** and the intermediate **7'**, where the hydronium ion formed from the previous step has already transferred the excess proton to the solvent, we have considered the free energy cost of generating the hydronium cation in the active site and in pure water following the same scheme as for the formation of the hydroxide (see FEP protocol in Supporting Information). In this case the formation of the hydronium cation was calculated to be 5.6 ± 0.1 kcal·mol⁻¹ more stable in pure water than in the active site. It could be somewhat surprising that the same active site (although at different stages of the catalytic cycle) destabilizes both the hydroxide and the

hydronium ions. This suggests that ion desolvation (including both electrostatic and hydrogen bonding contributions) is the main reason for the two cases. The free energy differences observed for the two ions (ranging between 5 and 12 kcal·mol⁻¹) are of the order of magnitude of a moderately strong hydrogen bond.

After C₅ deprotonation, the reaction mechanism concludes breaking the bond between cytosine and Cys81 nucleophile and recovering the protonation state of Glu119. Figure 7 illustrates the mechanism found and the corresponding PMF corrected at the M06-2X level. This final step consists in a stepwise C₆-S(Cys81) bond breaking followed by proton transfer from N₃ back to Glu119. Starting from structure **7'** the free energy barrier for the C₆-S bond breaking process (**TS 7'-8**) is 4.8 ± 0.4 kcal·mol⁻¹, leading to a positively charged intermediate **8**. This intermediate has a free energy of 3.7 kcal·mol⁻¹ above **7'**. After this intermediate the reaction proceeds through a second TS (**TS 8-9**), in which the initial protonation state of Glu119 is recovered after a direct transfer from N₃. This last TS (**TS 8-9**) is found 3.5 ± 0.2 kcal·mol⁻¹ above intermediate **8** (7.2 kcal·mol⁻¹ above **7'**) and leads to the final reaction product where the active site is regenerated and ready for the next catalytic turnover. The proposed mechanism for this last step has been further tested analyzing the PESs at the M06-2X/MM level (see Figure S17 in Supporting Information).

Relevant distances associated with the stationary states of this final step are provided in Table 3. In **TS 7'-8** the distance between the S(Cys81) atom and the C₆ atom of cytosine is 2.56 ± 0.09 Å, significantly larger than the distance found at intermediate **7**, 1.96 ± 0.05 Å. After this TS an intermediate (**8**) is reached where the Cys81 residue is completely separated from the cytosine ring (the distance between the S(Cys81) atom and the C₆ atom of cytosine is 3.17 ± 0.06 Å). In this intermediate state the negatively charged Glu119 residue is doubly hydrogen bonded to the cytosine through the N₃ position and the exocyclic amino group (see Table 3). Finally, **TS 8-9** leads to the final reaction product (**9**) where Glu119 receives back its proton from N₃. At this TS, the distance between the O_ε atom of Glu119 and the proton bonded to the N₃ position is 1.33 ± 0.04 Å, while the distance between the N₃ atom and the proton to be transferred is of 1.23 ± 0.04 Å. In the final state, the target cytosine C18 has

been methylated at position C₅ and the active site is prepared for the next catalytic cycle.

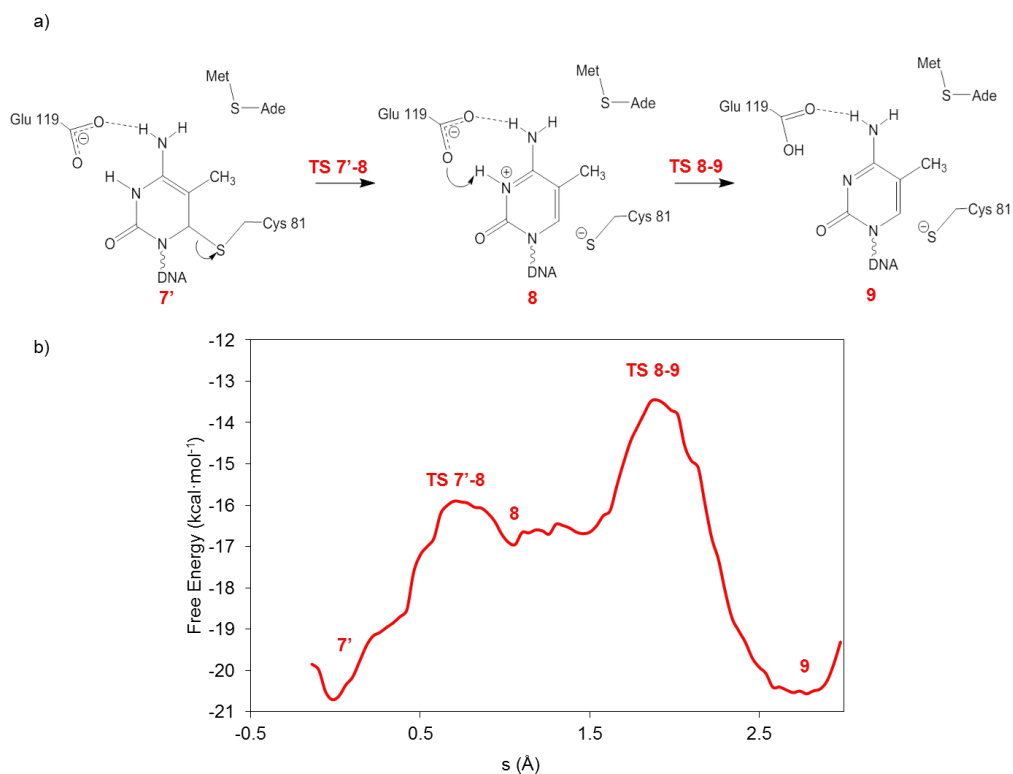


Figure 7. a) Scheme for the breaking of the enzyme-DNA covalent adduct and restoration of Glu119 protonation state. b) AM1/MM PMFs, corrected at the M06-2X/6-311+G** level, obtained for the process of breaking the covalent cytosine-Cys81 adduct (**TS 7'-8**) and reprotonation of the Glu119 residue (**TS 8-9**) in *M.HhaI* as a function of the collective path coordinate s .

Table 3. Key average distances (in Å) obtained for reactant, intermediate, product and transition states for the process of breaking the covalent cytosine-Cys81 adduct and reprotonation of the Glu119 residue in *M.HhaI*. Standard deviations are included. Free energies obtained after M06-2X/6-311+G**/MM corrections are given in kcal·mol⁻¹.

	TS 7'-8	8	TS 8-9	9
d(S(Cys81) - C ₆ (C18))	2.56 ± 0.04	3.17 ± 0.06	3.77 ± 0.23	3.31 ± 0.13
d(O _{ε2} (Glu119) - H _{ε2} (Glu119))	1.89 ± 0.10	1.96 ± 0.14	1.33 ± 0.04	0.98 ± 0.03
d(HO _{ε2} (Glu119) - N ₃ (C18))	1.03 ± 0.03	1.03 ± 0.03	1.23 ± 0.04	2.20 ± 0.07
d(O _{ε1} (Glu119) - H ₄₁ (C18))	2.00 ± 0.14	2.14 ± 0.21	2.29 ± 0.24	2.16 ± 0.16
d(O(Phe79) - H ₄₂ (C18))	2.70 ± 0.40	2.13 ± 0.32	1.92 ± 0.10	2.99 ± 0.35
d(HN _ε (Arg165) - O ₂ (C18))	2.01 ± 0.12	1.97 ± 0.12	1.92 ± 0.10	1.91 ± 0.09
d(HN _{η2} (Arg165) - O ₂ (C18))	2.09 ± 0.16	2.08 ± 0.17	2.09 ± 0.15	2.06 ± 0.14
d(HN _{η1} (Arg163) - O ₂ (C18))	2.19 ± 0.19	2.18 ± 0.17	2.39 ± 0.27	2.03 ± 0.14
ΔG	-20.7 ± 0.4	-15.9	-13.5 ± 0.2	-20.6

4. Conclusions

M.HhaI is a DNA Methyltransferase from *Haemophilus haemolyticus* that catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the C₅ position of a cytosine and that has been thoroughly used as a model for mammalian C₅-DNA MTases. In this work 100 ns long MD simulation has been carried out for the ternary complex formed by the cofactor, the protein and a DNA dodecamer that contains the recognition sequence. The starting point for our study was the X-ray structure found at the PDB with code 2HR1 that shows the target cytosine (C18) flipped out in the active site.

An equilibrated structure taken from the classical MD simulation was used as starting point for the exploration of the multidimensional free energy landscape associated with the reaction catalyzed by *M.HhaI*. QM/MM calculations were performed using the AM1 Hamiltonian, corrected by means of single-point calculations at the M06-2X level. The string method was used to identify the MFEPs associated with different mechanistic proposals. The reaction mechanism involves several chemical events: the nucleophilic addition of an unprotonated Cys81 to the C₆ position, the methyl transfer to the C₅ position and the abstraction of the excess proton from this atom. Some of these events can take place concertedly or in a stepwise manner. Usually exploration of free

energy landscapes associated with chemical reactions are made selecting one or two distinguished reaction coordinates, which inevitably bias the outcome of the simulations. In this work we have used the string method in a multidimensional space where different chemical events can be treated on equal-foot basis. Our results clarify most of the current debates about the reaction mechanism in this enzyme and the role of key residues present in the active site.

The free energy profile of the global process is shown in Figure 8. In agreement with experimental observations we found that the nucleophilic addition of an unprotonated Cys81 to the target nucleobase is a fast and reversible equilibrium preceding the methylation step. Deprotonation of Cys81 residue can be reached after the formation of the Michaelis complex by means of a proton transfer to the bulk solvent mediated by DNA phosphate groups. Methyl transfer is the rate-limiting step during the process of the formation of the methylated adduct, and our free energy barrier is in excellent agreement with the value derived from the experimentally determined pre-steady state rate constant. During this step proton transfer from Glu119 to the N₃ position is not required. However, this residue catalyzes the process by means of two hydrogen bonds formed with the extrahelical base that are strengthened during the process. This residue plays also a determinant role during the β -elimination step. FEP calculations have been used to rule out the participation of a hydroxide anion as the base in charge of the elimination of the excess proton from position C₅, as proposed in previous works. Instead, a water molecule could be in charge of this abstraction if the extrahelical cytosine is previously protonated at position N₃ by Glu119. The free energy barrier obtained with this proposal is also in very good agreement with the value derived from the steady-state rate constants. The reaction is completed breaking the adduct formed between the extrahelical cytosine and Cys81 and restoring the protonation state of Glu119.

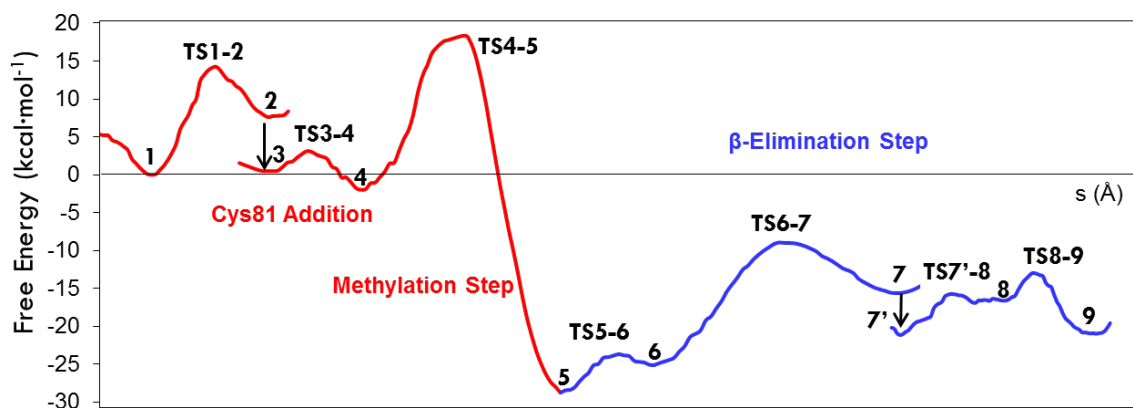


Figure 8. AM1/MM PMFs, corrected at the M06-2X/6-311+G** level, obtained for the addition of the Cys81 residue, the methylation transfer and the β -elimination step as a function of the collective path coordinate s .

The mechanistic description described in the present study agrees with most of the experimental observations made on this system, rationalizes the role played by the conserved Glu119 residue and identifies the nature of the base in charge of proton abstraction, issues that have been the subject of a long debate in the literature. The robustness of our results have been confirmed a posteriori obtaining relevant PES at the M06-2X/MM level, which allowed us also to rationalize some differences observed with respect to previous theoretical analyses. The picture obtained in this work could be useful to design future mutational studies to confirm the nature of the reaction mechanism and as a guide for the design on new and more efficient inhibitors of C₅-DNA MTases. Mutations of Ser86 could affect Cys81 deprotonation while Pro80 mutations could block the access of water to the active site, maybe preventing the β -elimination. The use of a thioacid analog of Glu119 could be useful to clarify the role of this residue, as far as our simulations show that the Gln119 variant could provoke unexpected changes in the active site properties.

5. Acknowledgements

The authors gratefully acknowledge financial support from FEDER funds and the Ministerio de Economía y Competitividad (project CTQ2012-36253-C03-03) and Generalitat Valenciana (ACOMP/2015/239). J. A. thanks Ministerio de Economía y Competitividad for FPI fellowship and K. Z. acknowledges a FPU fellowship of the Ministerio de Educación. K. S. thanks to the Polish Ministry of Science and Higher Education for "Iuventus Plus" program, project no. 0478/IP3/2015/73, (2015-2016), Polish National Center for Science (NCN)

(grant 2011/02/A/ ST4/00246, 2012–2017) and the USA National Institute of Health (ref. NIH R01 GM065368). The authors also acknowledge the computational facilities of the Servei d'Informàtica de la Universitat de València in the "Tirant" Supercomputer.

Supporting Information Contents

A representation of the interactions established between cofactor SAM and the surrounding amino acids during the MD simulation is provided as Supporting Information in Figure S1. Schemes S1 to S5 show the different QM subsystems employed at every step of the reaction mechanism **and the convergence of the string method for each of them.** The free energy perturbation protocol and Schemes S6 and S7 illustrating the thermodynamic cycle employed to calculate the free energy cost of moving a hydroxide anion or a hydronium ion from the solution to the active site are provided. Plots of the Root-Mean-Squared-Deviations (RMSD) for the enzyme and DNA backbone atoms for the Cys81 deprotonated system are shown in Figure S2. Discussion of the Protein-DNA interactions and plots for the number of direct and water mediated hydrogen bonds between the enzyme and the DNA are shown in Figure S3. The most important protein-DNA hydrogen bond interactions are discussed and depicted in Figure S4. Discussion of the DNA structure and a plot showing the average number of hydrogen bonds between the DNA base pairs for the deprotonated system are provided in Figure S5. Analysis for the neutral Cys81 system is provided in Figures S6 to S10. Figure S11 shows the Potential Energy Surfaces calculated at the M06-2X/6-31G**/MM level for the deprotonation of Cys81 by a DNA phosphate group mediated by Ser85 and by a water molecule. In Table S1 we show the key averaged distances and free energies for the aforementioned processes. Figure S10 shows the Potential Energy Surface calculated at the M06-2X/6-31G**/MM and B3LYP/6-31G**/MM levels for the methyl transfer and the nucleophilic attack of Cys81. Figure S13 shows the side reaction where Glu119 protonates the cytosine ring and the AM1/MM PMFs corrected at the M06-2X/6-31G** level for this process. Table S2 provides the most important average distances and free energies for the aforementioned process. Figure S14 shows the Potential Energy Surface calculated at the M06-2X/6-31G**/MM level for the addition of Cys81 to the cytosine ring and the proton transfer from Glu119 to the N₃ atom of cytosine. Figure S15 shows the Potential Energy Surface calculated at the M06-2X/6-31G**/MM level for the proton transfer from Glu119 to the N₃ atom of cytosine and the methyl transfer from SAM to C5 position. Figure S16 shows the most important interactions between the cytosine base and the amino acids in the active site of the Glu119Gln mutated enzyme. Figure S17 shows the Potential Energy Surfaces calculated at the M06-2X/6-31G**/MM level for the two processes involved in the β -elimination step. Finally in Figure S18 a representation of the water channel connecting the

hydronium ion formed in the active site with the bulk, once the proton abstraction from the C₅ atom of the cytosine has taken place, is shown. Complete references 38 and 55 are also provided.

REFERENCES

- (1) Cheng, X., Hashimoto H, Horton JR, Zhang X In *Handbook of Epigenetics: The New Molecular and Medical Genetics*; Tollefsbol, T., Ed.; Oxford: Academic Press: 2010, p 9.
- (2) Okano, M.; Bell, D. W.; Haber, D. A.; Li, E. *Cell* **1999**, *99*, 247.
- (3) E, L.; Beard C Fau - Jaenisch, R.; R, J. *Nature* **1993**, *366*, 362.
- (4) Baylin, S. B. *Nature clinical practice. Oncology* **2005**, *2 Suppl 1*, S4.
- (5) Robertson, K. D.; Jones, P. A. *Carcinogenesis* **2000**, *21*, 461.
- (6) Jeltsch, A. *ChemBiochem* **2002**, *3*, 274.
- (7) Jurkowska, R. Z.; Jurkowski, T. P.; Jeltsch, A. *ChemBiochem* **2011**, *12*, 206.
- (8) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357.
- (9) O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X. *J Mol Biol* **1996**, *261*, 634.
- (10) Vilkaitis, G.; Merkiene, E.; Serva, S.; Weinhold, E.; Klimašauskas, S. *Journal of Biological Chemistry* **2001**, *276*, 20924.
- (11) Reither, S.; Li, F.; Gowher, H.; Jeltsch, A. *Journal of Molecular Biology* **2003**, *329*, 675.
- (12) Svedružić, Ž. M.; Reich, N. O. *Biochemistry* **2004**, *43*, 11460.
- (13) Svedružić, Ž. M.; Reich, N. O. *Biochemistry* **2005**, *44*, 9472.
- (14) Svedružić, Ž. M.; Reich, N. O. *Biochemistry* **2005**, *44*, 14977.
- (15) Shieh, F. K.; Youngblood, B.; Reich, N. O. *J Mol Biol* **2006**, *362*, 516.
- (16) Youngblood, B.; Shieh, F.-K.; Buller, F.; Bullock, T.; Reich, N. O. *Biochemistry* **2007**, *46*, 8766.
- (17) Shieh, F. K.; Reich, N. O. *J Mol Biol* **2007**, *373*, 1157.
- (18) Gerasimaitė, R.; Merkienė, E.; Klimašauskas, S. *Nucleic acids research* **2011**, *39*, 3771.
- (19) Peräkylä, M. *Journal of the American Chemical Society* **1998**, *120*, 12895.
- (20) Lau, E. Y.; Bruice, T. C. *Journal of Molecular Biology* **1999**, *293*, 9.
- (21) Zhang, X.; Bruice, T. C. *Proc Natl Acad Sci U S A* **2006**, *103*, 6148.
- (22) Zangi, R.; Arrieta, A.; Cossio, F. P. *J Mol Biol* **2010**, *400*, 632.
- (23) Yang, J.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y.; Broyde, S. *Biochemistry* **2013**, *52*, 2828.
- (24) Foulks, J. M.; Parnell, K. M.; Nix, R. N.; Chau, S.; Swierczek, K.; Saunders, M.; Wright, K.; Hendrickson, T. F.; Ho, K. K.; McCullar, M. V.; Kanner, S. B. *J Biomol Screen* **2012**, *17*, 2.
- (25) Medina-Franco, J. L.; Caulfield, T. *Drug Discov Today* **2011**, *16*, 418.
- (26) Fahy, J.; Jeltsch, A.; Arimondo, P. B. *Expert Opin Ther Pat* **2012**, *22*, 1427.
- (27) Zhou, L.; Cheng, X.; Connolly, B. A.; Dickman, M. J.; Hurd, P. J.; Hornby, D. P. *J Mol Biol* **2002**, *321*, 591.
- (28) Martinet, N.; Michel, B. Y.; Bertrand, P.; Benhida, R. *MedChemComm* **2012**, *3*, 263.
- (29) Champion, C.; Guianvarc'h, D.; Sénamaud-Beaufort, C.; Jurkowska, R. Z.; Jeltsch, A.; Ponger, L.; Arimondo, P. B.; Guieysse-Peugeot, A.-L. *PLoS ONE* **2010**, *5*, e12388.
- (30) Kang, J. S.; Meier, J. L.; Dervan, P. B. *Journal of the American Chemical Society* **2014**, *136*, 3687.
- (31) Cheng, X. D.; Roberts, R. J. *Nucleic Acid Res.* **2001**, *29*, 3784.
- (32) Jeltsch, A. In *DNA Methylation: Basic Mechanisms*; Doerfler, W., Böhm, P., Eds.; Springer Berlin Heidelberg: 2006; Vol. 301, p 203.
- (33) Maragliano, L.; Vanden-Eijnden, E. *Chemical Physics Letters* **2007**, *446*, 182.
- (34) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 704.

- (35) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73*, 765.
- (36) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 525.
- (37) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 2284.
- (38) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; University of California, San Francisco: 2012.
- (39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.
- (40) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
- (41) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. *Biophys J* **2007**, *92*, 3817.
- (42) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781.
- (43) Stacklies, W.; Xia, F.; Gräter, F. *PLoS Comput Biol* **2009**, *5*, e1000574.
- (44) Aranda, J.; Zinovjev, K.; Roca, M.; Tuñón, I. *Journal of the American Chemical Society* **2014**, *136*, 16227.
- (45) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684.
- (46) Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089.
- (47) Huang, N.; MacKerell Jr, A. D. *Journal of Molecular Biology* **2005**, *345*, 265.
- (48) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088.
- (49) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- (50) Zhao, Y.; Truhlar, D. *Theoretical Chemistry Accounts* **2008**, *120*, 215.
- (51) Krzemińska, A.; Paneth, P.; Moliner, V.; Świderek, K. *The Journal of Physical Chemistry B* **2015**, *119*, 917.
- (52) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
- (53) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
- (54) The new charges (q_i') of the boundary atoms ($i=CA$, H. f. a. a. i. C., H1' for sugars) were obtained redistributing the excess charge (Q) according to the charges of these atoms in the force field (q_i): .
- (55) Zinovjev, K.; Ruiz-Pernía, J. J.; Tuñón, I. *Journal of Chemical Theory and Computation* **2013**, *9*, 3740.
- (56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.;

Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Wallingford CT, 2009.

(57) Zinovjev, K.; Martí, S.; Tuñón, I. *Journal of Chemical Theory and Computation* **2012**, *8*, 1795.

(58) Torrie, G. M.; Valleau, J. P. *Journal of Computational Physics* **1977**, *23*, 187.

(59) Roux, B. *Computer Physics Communications* **1995**, *91*, 275.

(60) Hub, J. S.; de Groot, B. L.; van der Spoel, D. *Journal of Chemical Theory and Computation* **2010**, *6*, 3713.

(61) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. *The Journal of Physical Chemistry B* **2004**, *108*, 8427.

(62) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. *The Journal of Physical Chemistry B* **2006**, *110*, 17663.

(63) Kumler, W. D.; Eiler, J. J. *Journal of the American Chemical Society* **1943**, *65*, 2355.

(64) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. D. *Cell* **1994**, *76*, 357.

(65) Youngblood, B.; Buller, F.; Reich, N. O. *Biochemistry* **2006**, *45*, 15563.

(66) Wu, J. C.; Santi, D. V. *Journal of Biological Chemistry* **1987**, *262*, 4778.

(67) Horowitz, S.; Dirk, L. M. A.; Yesselman, J. D.; Nimtz, J. S.; Adhikari, U.; Mehl, R. A.; Scheiner, S.; Houtz, R. L.; Al-Hashimi, H. M.; Trievel, R. C. *Journal of the American Chemical Society* **2013**, *135*, 15536.

(68) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. *Journal of the American Chemical Society* **2012**, *134*, 13436.

(69) Chirlan, L. E.; Francl, M. M. *Journal of Computational Chemistry* **1987**, *8*, 894.

Table of Contents Graphic

