A Model Study of the Enzyme-Catalyzed Cytosine Methylation Using ab Initio Quantum Mechanical and Density Functional Theory Calculations: pK_a of the Cytosine N3 in the Intermediates and Transition States of the Reaction

Mikael Peräkylä*

Contribution from the Department of Chemistry, University of Kuopio, P.O. Box 1627, FIN-70211, Kuopio, Finland

Received April 24, 1998. Revised Manuscript Received September 28, 1998

Abstract: The reaction mechanism of the DNA (cytosine-5)-methyltransferase-catalyzed cytosine methylation was investigated using ab initio quantum mechanical (at the MP2/6-31+G*//HF/6-31+G* and MP2/6-31+G*//HF/3-21+G* levels) and density functional theory calculations (Becke3LYP/6-31+G*) in the gas phase and in solution. The effects of aqueous solvation on the reaction energies were included by using an isodensity surface-polarized continuum model. The quantum mechanical model consisted of 1-methylcytosine (the model of the target cytosine), methylthiolate (the model of the side chain of the catalytic cysteine), and trimethylsulfonium (the model of the methyl-donating AdoMet). In addition, an approach is presented to estimate the pK_a of the cytosine N3 in the reaction intermediates and transition states and on the calculated reaction profiles. The approach involves calculation of the gas-phase proton affinities and solvation energies of the neutral and protonated forms of the molecules using ab initio quantum mechanical and continuum solvation methods. The calculated aqueous-phase proton affinities were calibrated using a set of 13 nitrogen acids with pK_a values from 0.5 to 17.5. The correlation coefficient (r^2) between the calculated aqueous proton affinities and the experimental pK_a's was 0.988. During the attack of methylthiolate on C6 of the cytosine, the pK_a of the N3 atom of the cytosine was calculated to increase from 5 to 17. In the subsequent reaction step, where C5 of the cytosine is methylated, the pK_a of N3 drops from 17 to 5. The protonation and deprotonation of the N3 atom was calculated to catalyze the two reaction steps. It seems likely that in the DNA (cytosine-5)-methyltransferase-catalyzed cytosine methylation proton transfers take place between Glu119 of the active site and N3 of the cytosine. The implications of the model calculations for the DNA (cytosine-5)-methyltransferase-catalyzed reactions are discussed.

Introduction

Methylation of DNA is an essential element of genomic function in organisms ranging from bacteria to mammals.[1] Methylation of cytosine residues in specific DNA sequences is catalyzed by *S*-adenyl-L-methionine (AdoMet)-dependent DNA (cytosine-5)-methyltransferases (DCMtase). Biochemical experiments[2−5] and recent X-ray crystallographic studies[6−9] of the complexes between *Hha*I and *Hae*III DNA (cytosine-5)-methyltransferases, AdoMet, and oligomeric DNA have revealed the basic features of the catalytic events.[10] Interestingly, the X-ray structures showed that the target cytosine is flipped out of the DNA helix and positioned in the active site of the enzyme.[10,11] The DNA methylation reaction is initiated by a nucleophilic attack of a conserved cysteine residue on C6 of the target cytosine (Figure 1). The attack generates a covalent enzyme−DNA intermediate and activates C5 of the cytosine ring for methyl transfer from AdoMet. Activation of the ring carbon by a nucleophilic addition to the adjacent carbon is a common feature in the mechanism of several enzymes. For example, thymidylate synthase, dUMP and dCMP hydroxymethylases, certain RNA-modifying enzymes, and DNA (cytosine-5)-methylases utilize such activation.[12,13] In the case of cytosine methylases, the nucleophilic attack increases the pK_a of the N3 of the cytosine, which probably becomes protonated.[12,14,15] The protonation may catalyze the addition reaction by neutralizing the negative charge of the base generated upon adduct formation.[6] However, N3 protonation makes cytosine more susceptible to enzyme-mediated hydrolytic deamination and its con-

* (e-mail) mikael.perakyla@uku.fi.

(1) Adams, R. L. P. *Biochem. J.* **1990**, *265*, 309−320.
(2) Smith, S. S.; Kaplan, B. E.; Sowers, L. C.; Newman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4744−4748.
(3) Chen, L.; MacMillan, A. M.; Chang, W.; Ezaz-Nikpay, K.; Lane, W. S.; Verdine, G. L. *Biochemistry* **1991**, *30*, 11018−11025.
(4) Ho, D. K.; Wu, J. C.; Santi, D. V.; Floss, H. G. *Arch. Biochem. Biophys.* **1991**, *284*, 264−269.
(5) Wu, J. C.; Santi, D. V. *J. Biol. Chem.* **1987**, *262*, 4778−4786.
(6) O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X. *J. Mol. Biol.* **1996**, *261*, 634−645.
(7) Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. *Cell* **1995**, *82*, 143−153.
(8) Chen, X.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J. *Cell* **1993**, *74*, 299−307.
(9) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357−369.
(10) Verdine, G. L. *Cell* **1994**, *76*, 197−200.
(11) Roberts, R. J. *Cell* **1995**, *82*, 9−12.
(12) Gabbara, S.; Sheluho, D.; Bhagwat, A. S. *Biochemistry* **1995**, *34*, 8914−8923.
(13) Jones, P. A.; Taylor, S. M. *Cell* **1980**, *20*, 85−93.
(14) Erlanson, D. A.; Chen, L.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 12583−12584.
(15) Chen, L.; MacMillan, A. M.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 5318−5319.

10.1021/ja981405a CCC: $15.00
Published on Web 11/24/1998

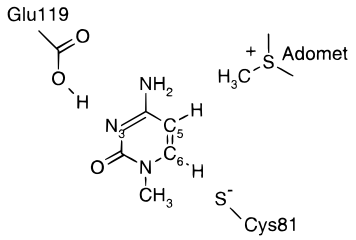


Figure 1. Schematic picture of the active site of DNA (cytosine-5)-methyltransferase.

version to thymine.[16] The last step in the enzyme mechanism is β-elimination of the C5-proton and C6-thiolate, resulting in a methylated cytosine. The elimination step can be blocked by using 5-fluorocytosine as a target base.[17] This feature has been used in trapping the covalent DNA−enzyme intermediate for structural and biochemical studies.

In this study, the catalytic steps of the DNA (cytosine-5)-methyltransferase-catalyzed cytosine methylation were investigated by using small models of the catalytic residues and ab initio quantum mechanical and density functional theory (DFT) calculations. The effects of aqueous solvation on the reaction energies were included by using a continuum solvation model. The quantum mechanical model of the active site species of the DCMtase included the target cytosine (as modeled by 1-methylcytosine), the side chain of the nucleophilic cysteine (methylthiolate), and the methyl-donating AdoMet (trimethylsulfonium). The purpose of the study was, first, to shed light on the intrinsic properties of the DNA methylation reaction in the gas phase and in aqueous solution, and second, to investigate the protonation state of the N3 of the cytosine ring in the reaction intermediates, transition states, and on the calculated reaction profiles. This allowed us to reach conclusions about the catalytic effects of the N3-protonation, which in the DCMtase-catalyzed reaction can take place between Glu119 and the N3-atom of the target cytosine (Figure 1). The computational approach applied to the calculation of the pK_a values is based on the quantum mechanically calculated gas-phase proton affinities (PA(gas)) and solvation energies (ΔG_{solv}). The calculated aqueous proton affinites (PA(aq) = PA(gas) + $\Delta\Delta G_{solv}$) were calibrated using a set of compounds with known pK_a values. Although the computational approach is not completely novel,[18−23] this is the first time pK_a values of reaction intermediates and transition states are calculated using purely ab initio quantum mechanical methods. The approach used here can be extended to study other pH-dependent reactions as well. Furthermore, the implications of the model calculations for the DNA (cytosine-5)-methyltransferase-catalyzed reaction are discussed.

Computational Details

All the ab initio quantum mechanical and DFT calculations reported in this work were performed with the Gaussian 94 program.[24] The geometries of the minimum and transition states were optimized at the HF/3-21+G*, HF/6-31+G*, and B3LYP/6-31+G* levels unless otherwise noted. Vibrational frequencies were calculated to confirm the nature of the transition states (one negative eigenvalue, at the HF/3-21+G* and HF/6-31+G* levels). Energies were further calculated at the MP2/6-31+G* level for the species optimized at the HF/3-21+G* (MP2/6-31+G*//HF/3-21+G*) and HF/6-31+G* (MP2/6-31+G*//HF/6-31+G*) levels. The effect of aqueous solvation on the relative energies was estimated by using the isodensity surface-polarized continuum model (IPCM option of Gaussian94).[25,26] A dielectric constant (ϵ) of 78.3 (water), the HF/6-31+G* level, and geometries optimized at the HF/6-31+G* level were used in the solvation calculations (IPCM-HF/6-31+G*//HF/6-31+G*). In the IPCM calculations, the solute cavity is determined from the electron density of the solute. A value of 0.0004 e/au^3 for the charge density was applied in the determination of the solute boundary. This value was observed in this and in our earlier studies[27,28] to give solvation energies which, when combined with the gas-phase energies, reproduce the experimental pK_a values and the reaction energies reasonably well.

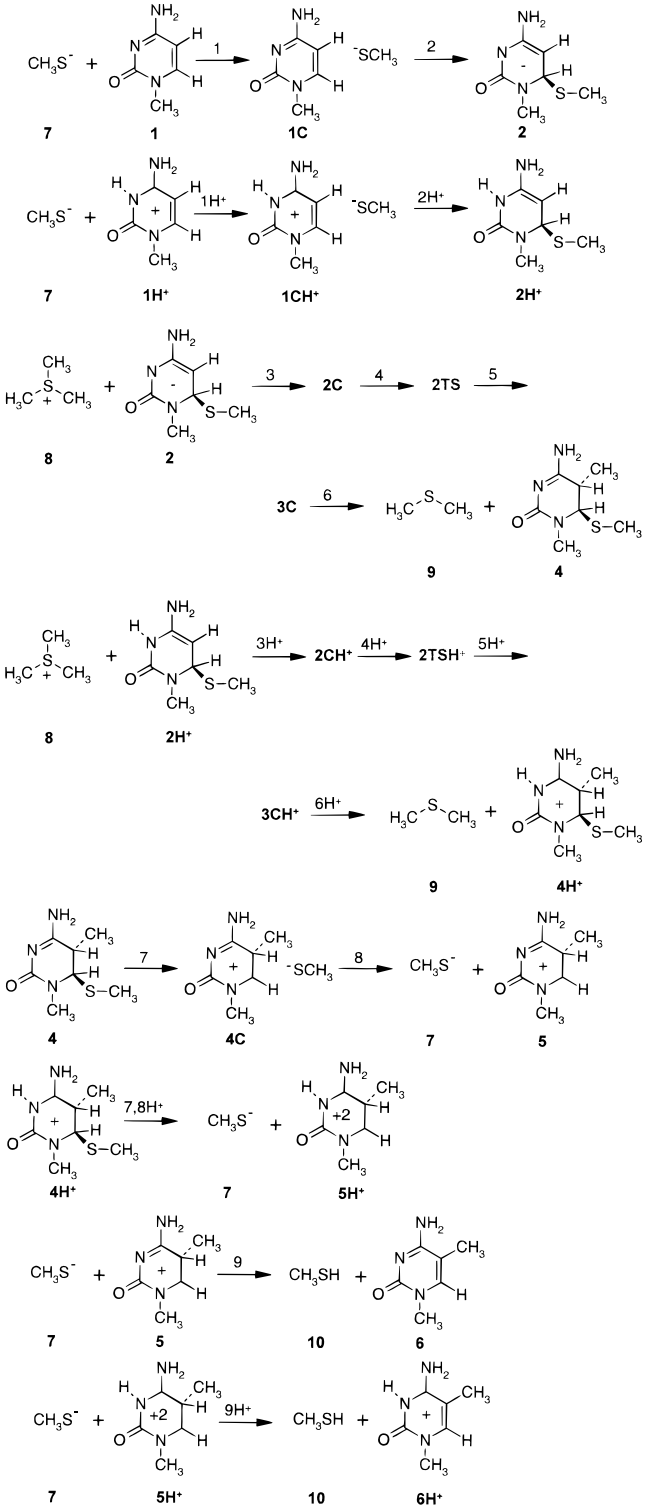
Results and Discussion

All the reaction steps (1−9 and 1H$^+$−9H$^+$) of the cytosine methylation reaction studied in this work are presented in Scheme 1. It also presents the structures of the isolated molecules, molecular complexes, and transition states calculated. Note that the reaction sequence is calculated for the neutral (**1**) and N3-protonated (**1**H$^+$) cytosine. Energies for the reaction steps at the MP2/6-31+G*//HF/3-21+G*, MP2/6-31+G*//HF/6-31+G*, and B3LYP/6-31+G* levels in the gas phase and at the B3LYP/6-31+G* level with solvation energies (ΔG_{solv}, IPCM-HF/6-31+G*) included are listed in Table 1.

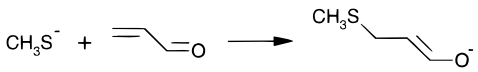
Nucleophilic Attack on C6. The cytosine methylation reaction is initiated by attack of the anionic methylthiolate on C6 of the cytosine ring. Formation of the methylthiolate adduct of the neutral cytosine (**2**) from the bimolecular cytosine−methylthiolate complex (**1**C) is endothermic in the gas phase and in solution (Scheme 1, reaction 2). At the B3LYP/6-31+G* level, the energy of the reaction is 12.9 kcal/mol, and at the MP2/6-31+G*//HF/3-21+G* and MP2/6-31+G*//HF/6-31+G* levels, 5.6 and 1.3 kcal/mol, respectively. Inclusion of solvation increases the reaction energies by 2.6 kcal/mol. In the case of the protonated cytosine (**1**H$^+$), formation of the cytosine−methylthiolate adduct (**2**H$^+$) from the bimolecular complex (**1**CH$^+$, reaction 2H$^+$) is strongly exothermic in the gas phase. Although aqueous solvation lowers the reaction energy by 25.5 kcal/mol, it is still −4.5 kcal/mol at the B3LYP level. The energies of these adduct formation reactions are calculated to be 7−18 kcal/mol more favorable at the MP2 than the B3LYP level. In addition to reactions 2 and 2H$^+$, this is seen in the energies of the reactions 7 and 7,8H$^+$, in which the energy differences between the two methods are even larger. Similar trends between the two computational methods are observed in the calculated reaction profiles of the addition reactions 2 and 2H$^+$ (Figures 2 and 3). Recently Zheng and Ornstein[29] supposed

(16) Yebra, M. J.; Bhagwat, A. S. *Biochemistry* **1995**, *34*, 14752−14757.
(17) Osterman, D. G.; DePillis, G. D.; Wu, J. C.; Matsuda, A.; Santi, D. V. *Biochemistry* **1988**, *27*, 5204−5210.
(18) Gao, J.; Pavelites, J. J. *J. Am. Chem. Soc.* **1992**, *114*, 1912−1914.
(19) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610−5620.
(20) Warshel, A. *Biochemistry* **1981**, *20*, 3167−3177.
(21) Jorgensen, W. L.; Briggs J. M. *J. Am. Chem. Soc.* **1989**, *111*, 4190−4197.
(22) Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, *109*, 6857−6858.
(23) Yang, B., Wright, J., Eldefrawi, M. E., Pou, S., MacKerell, A. *J. Am. Chem. Soc.* **1994**, *116*, 8722−8732. 1994.
(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Peterson, G. A.; Montgomery, J. A.; Rachavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzales, C.; Pople, J. A. *Gaussian 94* (Revision D.3); Gaussian, Inc.: Pittsburgh, PA, 1995.
(25) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.
(26) Wiberg, K. B.; Rablen, P. R.; Rush, D. J.; Keith, T. A. *J. Am. Chem. Soc.* **1995**, *117*, 4261−4270.
(27) Peräkylä, M. *J. Chem. Soc., Perkin Trans 2* **1997**, 2185−2199.
(28) Peräkylä, M. *J. Org. Chem.* **1996**, *61*, 7420−7425.

Scheme 1. Reaction Steps 1−9 and $1H^+$−$9H^+$ of the Cytosine Methylation Reaction Studied and Structures of the Isolated Molecules, Complexes, and Transition States



that MP2/6-31+G**//HF/6-31+G** level calculations overestimate the stability of the transition state for C−S bond formation. They studied the mechanism of nucleophilic aromatic substitution of 1-chloro-2,4-dinitrobenzene by glutathione as a model of the mode of action of glutathione *S*-transferases. In addition, the B3LYP method has been found to greatly underestimate the barrier for S_N2 reactions of Cl^- with CH_3Cl and CH_3Br.[30] Therefore, the reliability of the MP2/6-31+G*//HF/6-31+G* and B3LYP/6-31+G* calculations in predicting the C−S bond formation was investigated by calculating the energies of Michael addition of methylthiolate to acrolein (reaction 1).



(1)

Thomas and Kollman[31] have used this reaction to model the first step of the catalytic mechanism of thymidylate synthase. Reaction energies for (1) are −23.9, −23.7, and −17.9 kcal/mol at the MP2/6-31+G*//HF/6-31+G*, MP2/6-31+G*//MP2/6-31+G*, and B3LYP/6-31+G* levels, respectively. The B3LYP also predicts less negative formation energies than the MP2 method. Inclusion of more electron correlation at the MP4(SDQ)/6-31+G*//MP2/6-31+G* and MP4(SDTQ)/6-31+G*//MP2/6-31+G* levels results in reaction energies of −18.8 and −21.7 kcal/mol. Interestingly, inclusion of computationally intensive triple substitutions on the MP4 energy has a large effect (−2.9 kcal/mol). Enlargement of the basis set has practically no effect on the reaction energy. At the MP2/AUG-cc-pVDZ[32] level, the energy for methylthiolate addition to acrolein is −23.6 kcal/mol. Based on this comparison, we can conclude that the MP2 calculations probably slightly overestimate the stability of the C−S bond and the barriers for C−S bond formation and B3LYP calculations probably underestimate them.

In Figure 2, the reaction profiles are presented for the addition of methylthiolate to C6 of neutral cytosine, and in Figure 3 to C6 of N3-protonated cytosine. The reaction coordinate *r*(C−S) is the distance between S of the thiolate and C6 of the cytosine. In the gas-phase reaction profiles, the energies (B3LYP/6-31+G*) of the neutral (Figure 2) and protonated systems (Figure 3) increase and decrease monotonically, respectively, as the C−S distance decreases. In contrast to the B3LYP calculations, a TS was located in the gas phase at the HF/6-31+G* level (geometry of the TS was optimized) and can be seen to exist (Figure 2) at the MP2/6-31+G*//HF/6-31+G* level for the addition of methylthiolate on **1**H. In the TS optimized at the HF/6-31+G* level, the C−S distance is 2.22 Å and energy 2.1 kcal mol^{-1} at the MP2/6-31+G*//HF/6-31+G* level. This is only 0.2 kcal mol^{-1} larger than at the point *r*(C−S) = 2.2 Å obtained using the C−S distance as the reaction coordinate. Inclusion of solvation energies induces barriers on the profiles. The top of the barrier is located at a C−S distance of about 2.4 Å on the profile of neutral cytosine and 2.6 Å on the profile of N3-protonated cytosine. Together, the calculated reaction profiles and the reaction energies (reactions 1, 2, $1H^+$, and $2H^+$) show that adduct formation is facilitated by the N3-protonation in the gas phase and in solution. Furthermore, especially in the case of the N3-protonated cytosine, the rate of the addition reaction is considerably more favorable in the gas phase than in aqueous solution. It can be concluded that the enzyme may facilitate the addition reaction by protonating the N3-atom of the ring and by providing a nonaqueous (i.e., desolvated) active site environment.[33,34]

(29) Zheng, Y.-J.; Ornstein, R. L. *J. Am. Chem. Soc.* **1997**, *119*, 648−655.

(30) Glukhovtsev, M. N.; Bach, R. D.; Pross, A.; Radom, L. *Chem. Phys. Lett.* **1996**, *260*, 558−565.

(31) Thomas, B. E., IV; Kollman, P. A. *J. Org.Chem.* **1995**, *60*, 8375−8381.

(32) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(33) Zheng, Y.-J.; Bruice, T. C. *J. Am. Chem. Soc.* **1997**, *119*, 3868−3877.

Table 1. Energies (kcal/mol) of the Reaction Steps 1−9 and $1H^+$−$9H^+$ (Scheme 1)[a]

reaction	MP2/6-31+G*// HF/3-21+G* (gas)[b]	MP2/6-31+G*// HF/6-31+G* (gas)	B3LYP/ 6-31+G*(gas)	B3LYP/ 6-31+G*(aq)
1	−22.8	−23.2	−19.4	−0.6
2	5.6	1.3	12.9	15.5
$1H^+$	−89.6	−88.0	−88.3	−8.1
$2H^+$	−43.8	−47.6	−30.0	−4.5
3	−87.7 (−99.9)	−75.9	−71.1	−1.1
4	5.4 (17.6)	−1.8	−3.4	9.9
5	−72.0 (−74.5)	−69.1	−63.8	−57.2
6	6.1 (8.5)	2.9	−1.6	−3.0
$3H^+$	−5.7 (−22.9)	−6.9	−0.3	6.8
$4H^+$	14.3 (31.6)	17.2	9.0	10.7
$5H^+$	−40.8 (−51.8)	−41.1	−32.9	−33.0
$6H^+$	−0.1 (10.9)	0.6	−3.5	−10.2
7	55.3	58.3	35.4	6.4
8	110.2	107.8	111.6	36.0
$7,8H^+$	273.5	273.7	255.9	61.9
9	−149.8	−151.1	−148.4	−91.6
$9H^+$	−259.7	−260.4	−259.7	−68.8

[a] Calculated at the MP2/6-31+G*//HF/3-21+G*, MP2/6-31+G*//HF/6-31+G*, and B3LYP/6-31+G* levels in the gas phase and at the B3LYP/6-31+G* level with solvation energies (ΔG_{solv}, IPCM-HF/6-31+G*) included. [b] Reaction energies calculated using geometries of the complexes fully optimized at the HF/3-21+G* level are in parentheses.

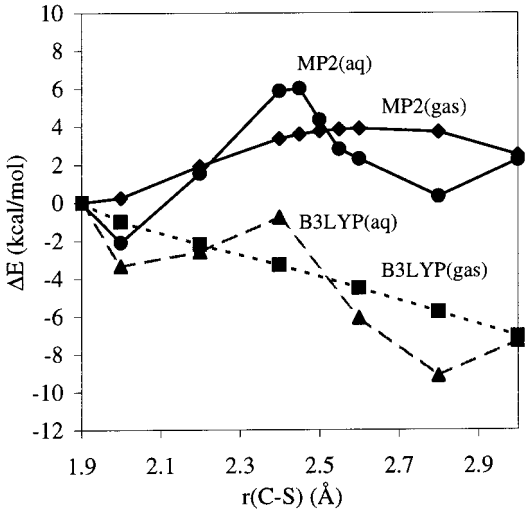


Figure 2. Reaction profile for the addition of methylthiolate (**7**) on C6 of neutral cytosine (**1**) ring in the gas phase at the MP2/6-31+G*//HF/6-31+G* (◆, MP2(gas)) and B3LYP/6-31+G* (■, B3LYP(gas)) levels and in aqueous solution at the MP2/6-31+G*//HF/6-31+G* + $\Delta\Delta G_{solv}$ (●, MP2(aq)) and B3LYP/6-31+G* + $\Delta\Delta G_{solv}$ (▲, B3LYP-(aq)) levels. *r*(C−S) is the distance between S of the attacking thiolate and C6 of cytosine.

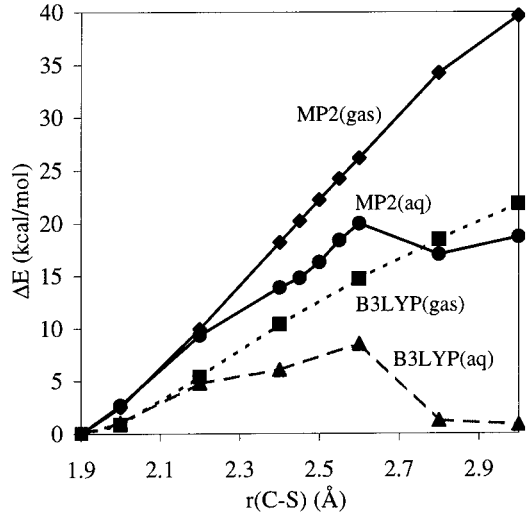


Figure 3. Reaction profile for the addition of methylthiolate (**7**) on C6 of the N3-protonated cytosine ($\mathbf{1}H^+$) ring in the gas phase at the MP2/6-31+G*//HF/6-31+G* (◆, MP2(gas)) and B3LYP/6-31+G* (■, B3LYP(gas)) levels and in aqueous solution at the MP2/6-31+G*//HF/6-31+G* + $\Delta\Delta G_{solv}$ (●, MP2(aq)) and B3LYP/6-31+G* + $\Delta\Delta G_{solv}$ (▲,B3LYP(aq)) levels. *r*(C−S) is the distance between S of the attacking thiolate and C6 of cytosine.

Methyl Transfer to C5. The next step in the reaction sequence is transfer of the methyl group from trimethylsulfonium, which mimics the methyl group-donating part of the AdoMet, to the activated C5 of the covalent C6-thiolate adduct of cytosine. Energies were calculated for formation of the bimolecular cytosine adduct−trimethylsulfonium complex (reaction 3, **2**C) from the isolated reactants, C6-thiolate adduct of cytosine (**2**) and trimethylsulfonium (**8**), for the TS of the methyl-transfer reaction (reaction 4, **2**TS), for formation of the bimolecular complex **3**C from **2**TS (reaction 5), and for formation of products, the isolated **4** and **9**, from the bimolecular complex **3**C (reaction 6). The same reaction steps were calculated for N3-protonated cytosine as well (reactions $3H^+$−$6H^+$). The geometries of all of the species were fully optimized excluding the bimolecular complexes **2**C, $\mathbf{2}CH^+$, **3**C, and $\mathbf{3}CH^+$. The energies of these species were estimated using a simplified approach. The structures of fully optimized C6-thiolate adducts (**2**, $\mathbf{2}H^+$, **4**, $\mathbf{4}H^+$), trimethylsulfonium (or dimethyl sulfide), and the corresponding TS structures were used. The cytosine adduct and trimethylsulfonium (or dimethyl sulfide) were least-squares fitted on the TS structure using the heavy atoms of the cytosine ring and C12, S11, and C13 of trimethylsulfonium (or dimethyl sulfide, the numbering is shown in Figure 4). After that the distance between the cytosine adduct and the trimethylsulfonium (or dimethyl sulfide) was varied by changing the C5−C10 (**2**C

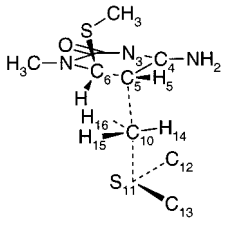


Figure 4. Atomic numbering of the transition states **2**TS and $\mathbf{2}TSH^+$.

(34) Lightstone, F. C.; Zheng, Y.-J.; Maulitz, A. H.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8417−8420.

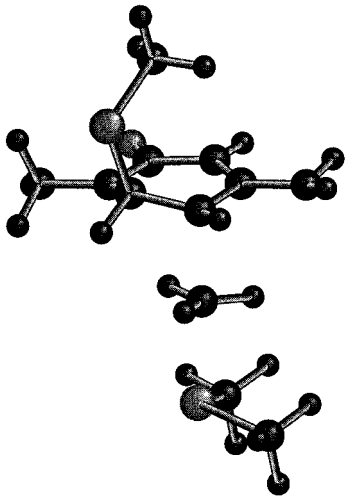


Figure 5. Structure of **2**TSH$^+$ optimized at the B3LYP/6-31+G* level.

and **2**CH$^+$) or C10−S11 (**3**C and **3**CH$^+$) distance in small steps (0.1−0.2 Å) while keeping the two parts rigid. This was done in order to be sure that there are no bad nonbonding interactions between the two parts. It was found that the original fitting of the molecules on the TSs provided acceptable structures. Thus, the energies of the bimolecular reactant and product complexes are rough estimates of the energies of the reactive conformations. In these structures, the arrangement of the reactive species is close to the one they have in the active site of DCMtase. In addition to this the geometries of the bimolecular complexes **2**C, **2**CH$^+$, **3**C, and **3**CH$^+$ were fully optimized at the HF/3-21+G* level and energies calculated at the MP2/6-31+G*//HF/3-21+G* level. These calculations provided energies for the fully relaxed ground-state conformations of the molecular complexes. Comparison of the TS energies obtained with the two approaches provides an estimate on how much rate acceleration can (at the most) be gained by positioning the reactive groups in a conformation where the reaction can take place without much structural reorganization.[35,36]

The structure of the transition state **2**TSH$^+$ optimized at the B3LYP/HF/6-31+G* level is presented in Figure 5. Selected geometric parameters of **2**TS and **2**TSH$^+$ optimized at the HF/3-21+G*, HF/6-31+G*, and B3LYP/6-31+G* are listed Table 2. From the table one can see that the geometries of the TSs are remarkably similar at the three computational levels employed.

The methyl group, which is transferred in the reaction, is almost planar at the TS. The H14−C10−H15−H16 angle, which measures the planarity of the methyl group (it is 180.0° for the planar CH_3), is 152−161° for the neutral cytosine species, and 168−173° for the species with protonated cytosine. In both cases, the top of the CH_3 umbrella points toward the sulfur of the methyl donor. In the case of the N3-protonated cytosine, the C5 of the cytosine ring, which accepts the methyl group in the reaction, is slightly more tetrahedral (H5−C5−C4−C6 angle in Table 2). In addition, in the TS of the N3-protonated species, the methyl group-accepting C5, the C10 of the transferring methyl, and the donor S11 are arranged almost linearly: the C5−C10−S11 angle is 175−176° for the protonated system compared to 158−166° for the neutral one. Thus, as indicated by the more favorable reaction energy of neutral cytosine, the TS of the methyl-transfer reaction is slightly earlier for neutral cytosine than for the N3-protonated cytosine. The TS of the N3-protonated cytosine closely resembles the corresponding TS calculated for the model of the catechol *O*-methyltransferase (COMT)-catalyzed methyl transfer.[37] In the model of COMT, the methyl transfer take place between catecholate anion and trimethylsulfonium. In both cases, the transferring methyl group is close to being planar. Similarly to the TS of this study, the attacking oxygen, the transferring methyl carbon, and the leaving sulfur atom are in an almost linear arrangement in the transition state of the catecholate−trimethylsulfonium system. In the case of catecholate, the angle was 177° as compared to 175−176° from the present work.

The overall methyl-transfer reactions from the isolated reactants to the isolated products of the neutral and N3-protonated cytosine adducts are highly exothermic in the gas phase and in solution. In the case of neutral cytosine, the methyl-transfer step (reaction 4) has a TS barrier of −3.4 kcal/mol in the gas phase and 9.9 kcal/mol in solution (B3LYP/6-31+G*) when the geometry of the reactant complex was modeled using the TS structure. The negative barrier of reaction 4 indicates that the methyl transfer to the neutral cytosine is highly favorable in the gas phase when the reactants are placed in a conformation where the reaction can readily take place. For N3-protonated cytosine, the corresponding barrier (reaction 4H$^+$) is 9.0 kcal/mol in the gas phase and 10.7 kcal/mol in solution. Full optimization of the reactant complex increases the TS energies in the gas phase by 12.2 kcal/mol for the neutral system and 17.3 kcal/mol for the N3-protonated system. In the optimized complexes, the trimethylsulfonium is located in the plane of the ring and the methyl hydrogens make hydrogen bonds with the N3 and carbonyl oxygen of the ring. The calculations clearly indicate that methyl transfer is more favorable for neutral cytosine than for N3-protonated cytosine. In addition, the rate of methyl transfer can be greatly accelerated by placing the transferring methyl group in a position that more closely resembles the transition state of the reaction.

The low TS energies, highly favorable energies of the overall methyl-transfer reactions, and the steps from the TSs to the product complexes (reactions 5 and 5H$^+$) of the model calculations indicate that the methylthiolate adduct (**2** or **2**H$^+$) is a transient intermediate which is rapidly methylated by the enzyme. The short lifetime of the methylthiolate adduct would be beneficial for the proper function of DCMTase. Namely, it would lower the probability of the methylthiolate adduct, which exists in an N3-protonated form (see pK_a calculations below), to undergo hydrolytic deamination and mutation to thymine.[16]

Elimination of the C5-Proton and C6-Thiolate. From the C6-thiolate−C5-methyl adduct of cytosine (**4** or **4**H$^+$) the reaction can occur by (i) elimination of the C6-thiolate, (ii) elimination of the C5-proton, or (iii) concerted elimination of the C6-thiolate and C5-proton. The elimination of methylthiolate from the neutral cytosine adduct (reactions 7 and 8) and the N3-protonated cytosine (reaction 7,8H$^+$) is highly endothermic in the gas phase. At the B3LYP/6-31+G* level, the energies are 147.0 and 255.9 kcal/mol for the neutral and the N3-protonated systems, respectively. The corresponding energies are 18−19 kcal/mol larger at the MP2/6-31+G*//HF/6-31+G* level. The inclusion of solvation energies lowers the reaction energies considerably, but the reactions are still highly unfavorable. At the B3LYP/6-31+G* + $\Delta\Delta G_{solv}$ level, the energies are 42.4 kcal/mol for neutral cytosine adduct and 61.9 kcal/mol for the N3-protonated species. Since the elimination

(35) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595−2605.

(36) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1997**, 91033-9113.

(37) Zheng, Y.-J.; Bruice, T. C. *J. Am. Chem. Soc.* **1997**, *119*, 8138−8145.

Table 2. Selected Geometric Parameters of the Transition State of the Methyl-Transfer Reaction of the Neutral (**2**TS) and N3-Protonated Methylthiolate Adducts (**2**TSH$^+$)[a]

	HF/3-21+G*		HF/6-31+G*		B3LYP/6-31+G*	
parameter	neutral	protonated	neutral	protonated	neutral	protonated
C4−C5	1.388	1.371	1.391	1.374	1.402	1.390
C5−C6	1.504	1.511	1.493	1.508	1.486	1.504
C5−H5	1.075	1.074	1.076	1.075	1.086	1.085
C5−C10	2.476	2.304	2.496	2.319	2.568	2.324
C10−S11	2.250	2.339	2.204	2.361	2.184	2.313
C10−H14	1.065	1.068	1.066	1.067	1.079	1.082
C10−H15	1.070	1.066	1.070	1.067	1.085	1.080
C10−H16	1.061	1.066	1.065	1.067	1.080	1.080
S11−C12	1.811	1.809	1.812	1.809	1.827	1.824
S11−C13	1.811	1.809	1.807	1.809	1.823	1.824
C4−C5−H5	120.29	118.49	119.53	117.67	120.62	117.81
C6−C5−H5	116.61	116.30	117.12	116.22	117.46	116.51
C4−C5−C10	85.01	97.36	84.94	97.43	83.84	98.56
C6−C5−C10	100.02	100.25	102.51	103.15	102.49	102.99
C5−C10−S11	158.93	175.25	165.51	175.53	158.03	176.36
C10−S11−C12	95.65	101.35	96.55	102.91	94.29	100.50
C10−S11−C13	98.61	102.14	101.68	103.74	99.55	101.36
H14−C10−H15−H16	160.9	171.3	158.5	173.2	151.7	168.2
H5−C5−C4−C6	−149.5	−150.4	−149.0	−147.7	−153.1	−147.7

[a] Optimized at the HF/3-21+G*, HF/6-31+G*, and B3LYP/6-31+G* levels. Distances in angstroms, angles and torsion angles in degrees.

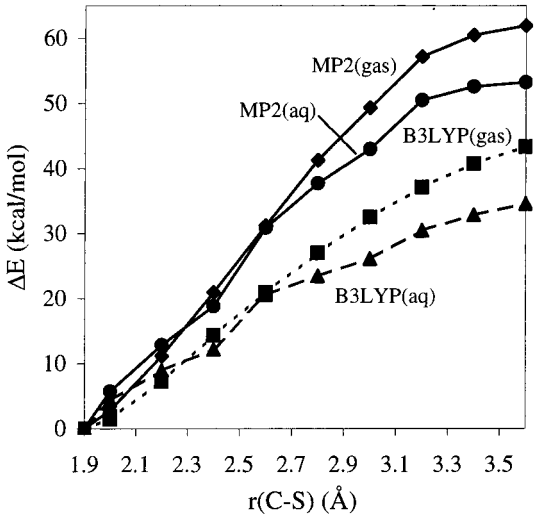


Figure 6. Reaction profile for the elimination of methylthiolate (**7**) from the C6-thiolate−C6-proton intermediate **4** in the gas phase at the MP2/6-31+G*//HF/6-31+G* (◆, MP2(gas)) and B3LYP/6-31+G* (■, B3LYP(gas)) levels and in aqueous solution at the MP2/6-31+G*//HF/6-31+G* + $\Delta\Delta G_{solv}$ (●, MP2(aq)) and B3LYP/6-31+G* + $\Delta\Delta G_{solv}$ (▲, B3LYP(aq)) levels. r(C−S) is the distance between S of the leaving thiolate and C6 of cytosine.

reaction is clearly more favorable for neutral cytosine, the reaction profile for methylthiolate elimination was calculated only for this case (Figure 6). The reaction was followed on the basis of the C−S distance of the optimized covalent intermediate (1.9 Å) to a C−S distance of 3.6 Å. The energy of the system increased without a barrier when the C−S distance was increased. In the gas phase, the point C−S = 3.6 Å was calculated to be 62.0 and 43.3 kcal/mol higher in energy than the stable intermediate at the MP2/6-31+G*//HF/6-31+G* and B3LYP/6-31+G* levels, respectively. Solvation energies stabilized the point C−S = 3.6 Å by 8.7 kcal/mol as compared to the covalent structure. The geometry of the bimolecular product complex, **4**C, was optimized as well. In the gas phase, this complex was calculated at the MP2/6-31+G*//HF/6-31+G level to be 58.3 kcal/mol and at the B3LYP/6-31+G* level to be 35.4 kcal/mol less stable than the covalent intermediate. Solvation stabilized the complexes by 29.0 kcal/mol. Thus, the reaction energies and the calculated reaction profile allowed us to conclude that methylthiolate elimination has a large reaction barrier.

The pK_a value of the C5-proton of the thiolate−methyl adducts were roughly estimated to be about 18 for the neutral species and 11 for the N3-protonated adduct (see below the estimates of pK_a values). Due to the high pK_a values of the C5-proton, the lack of an active site base that could abstract the proton,[5,9] and the large barrier for the C6-thiolate elimination, a concerted elimination of C5-proton and C6-thiolate seems the most plausible alternative for the mechanism of product formation. In the concerted reaction, elimination of the C5-proton would lower the barrier of C6-thiolate elimination (deduced from the reaction profiles of Figures 2 and 3). On the other hand, elimination of the C6-thiolate would lower the barrier for proton elimination. The latter effect is seen from the highly favorable energies for abstraction of the C5-proton from the product of the C6-thiolate elimination reaction (reactions 9 and 9H$^+$) by methylthiolate. Also, the total energy change in the C6-thiolate and C5-proton elimination steps is clearly favorable. The anionic sulfur of the leaving C6-thiolate could function as a proton-abstracting base in the elimination step. In this case, the reaction would take place via syn-elimination which, based on the structures of energy profile of C6-elimination, seems to be a structurally unfavorable alternative (Figure 7, point r(C−S) = 2.6 Å). However, a change in the ring conformation would arrange the C6−S and C5−H5 bonds into a synperiplanar position, suitable for syn-elimination. Analyses of X-ray structures have not identified an active site base that is located suitably to abstract the C5-proton. It has been proposed that the base would be a negatively charged phosphoryl oxygen of the flipped nucleotide.[6] In this case, the proton abstraction would be mediated by a water molecule located between the phosphoryl oxygen and the C5-proton. More detailed experimental studies or simulations of this reaction step in the enzyme−DNA environment are needed to resolve this question.

In the X-ray crystal structure of *Hha*I methyltransferase complexed with S-adenosyl-L-homocysteine and DNA methylated at the target cytosine (reaction product), it was observed that the catalytic sulfur of Cys81 interacts strongly with C6 (C−S distance is 2.77 Å), and most interestingly, the C5 methyl

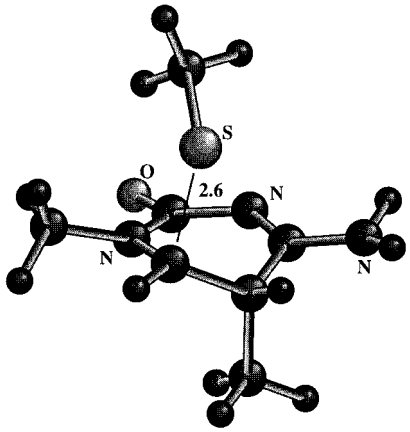


Figure 7. Structure of intermediate **5** at the point r(C−S) = 2.6 Å of the reaction profile of the methylthiolate elimination (Figure 6).

is bent ∼50° out of the plane of the cytosine ring.[6] It was proposed that the unusual position of the C5 methyl is probably due to partial sp^3 character at C5 and C6 and to the steric effects of the conserved amino acid residues Pro80 and Cys81. Comparison of the crystal structures and geometries calculated in this work led us to propose that a significant population of species corresponding to the C6-thiolate−C5-proton intermediate **4** exists in the X-ray structures. Thus, bending of the C5-methyl and sp^3 character at C5 of the cytosine observed in the crystal structure would be due to the protonation of C5. It seems unlikely, that the steric effects of the active site residues would be strong enough to bend the C5-methyl out of plane of the cytosine ring.[38] Furthermore, it has been suggested, that a slowly reversible covalent bond is formed between cysteine's S and C6 of the cytosine in the DCMtase−5-fluorocytosine−*S*-adenosyl-L-homocysteine complex.[39] Here the stabilizing effect of the fluoro substituent resembles that provided by the fluoro substitution (inductive effect) in the catalytic mechanism of a β-glycosidase.[40−43]

pK_a of the Cytosine N3. The protonation state of N3 of the target cytosine probably changes during the catalytic cycle of DNA methylation.[2,6] The N3-protonation has been proposed to facilitate the attack of cysteine by increasing the charge on C6[44] and by neutralizing the negative charge of the DNA−cytosine adduct.[6,15] Therefore, we estimated the pK_a of the cytosine N3 in the intermediates and the transition state of the methyl-transfer reaction and during the attack of methylthiolate on C6. This was done by calculating the aqueous proton affinity (PA(aq)) of the cytosine N3 (eq 2) at different stages of the

$$\mathrm{PA(aq)} = \mathrm{PA(gas)} + \Delta\Delta G_{\mathrm{solv}} \qquad (2)$$

reaction and calibrating the calculated values using a set of acids with known pK_a values. Thus, the calculated PA(aq) was equal to the sum of the gas-phase proton affinity (PA(gas)) of the acid and the solvation energy difference ($\Delta\Delta G_{\mathrm{solv}}$) between the acid and conjugate base. PA(gas) was calculated from the gas-phase energies of the acid and conjugate base without any thermodynamic corrections[45] at the MP2/6-31+G*//HF/6-31+G* level. The ΔG_{solv} energies were calculated at the HF/6-31+G* level using the isodensity surface-polarized continuum model (IPCM-HF/6-31+G*). To relate the calculated PA(aq)'s to the corresponding pK_a values, the correspondence between the two values was calibrated by calculating the PA(aq)'s of 13 heterocyclic and aliphatic compounds having pK_a's between 0.5 and 17.5. The molecules used in the calibration are listed in the figure caption of Figure 8. The figure also shows that there is good correlation between the calculated PA(aq)'s and known pK_a values;[46,47] the correlation coefficient (r^2) is 0.988 (standard deviation s = 0.606; correlation equation pK_a(calc) = 0.4137PA(aq) - 109.372).

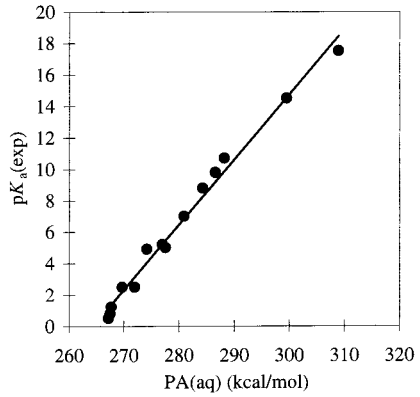


Figure 8. Correlation between the calculated aqueous-phase proton affinities (PA(aq) = PA(gas) + $\Delta\Delta G_{\mathrm{solv}}$) and the corresponding experimental pK_a values for nitrogen atoms of a set of cyclic and aliphatic compounds. Compounds and pK_a's: pyridine, 5.2; oxazole, 0.8; pyrrole, 17.5; pyrazole, 2.5; thiazole, 2.5; imidazole, 7 and 14.5; quinoline, 4.9; pyrimidine, 1.2; pyrazine, 0.5; methylamine, 10.7; trimethylamine, 9.8; allyldimethylamine, 8.8.

Table 3. Calculated Gas-Phase (PA(gas), kcal/mol) and Aqueous-Phase (PA(aq)) Proton Affinities and pK_a Values of Reaction Intermediates and Transition-State Species

molecule	PA(gas)	PA(aq)	pK_a(calc)
1H$^+$	232.8	275.2	4.5
1CH$^+$	297.6	278.7	5.9
2H$^+$	346.5	304.6	16.6
2CH$^+$	277.5	298.5	14.1
2TSH$^+$	258.6	291.2	11.1
3CH$^+$	230.6	269.9	2.3
4H$^+$	232.9	277.5	5.4
5H$^+$	125.3	259.3	−2.1
6H$^+$	234.6	275.9	4.8

Predicted pK_a values for the intermediates and TS of the studied reaction sequence are listed in Table 3. The calculations predict, in agreement with experiment, that the pK_a of N3 of an isolated cytosine is 4.5 in aqueous solution.[48] Complexation of cytosine with methylthiolate increases the pK_a to 5.9 due to the negative charge of methylthiolate, which stabilizes the protonated form of cytosine. In the complexes **1**C and **1**CH$^+$, the negatively charged sulfur of methylthiolate is located in the plane of the cytosine and is hydrogen bonded to H5 of the ring.

(38) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.
(39) Wyszynski, M.; Gabbara, S.; Kubareva, E. A.; Romanova, E. A.; Oretskaya, T. S.; Gromova, E. S.; Shabarova, Z. A.; Bhagwat, A. S. *Nucleic Acids Res.* **1993**, *21*, 295−301.
(40) Withers, S. G.; Street, I. P. *J. Am. Chem. Soc.* **1988**, *110*, 8551−8553.
(41) Withers, S. G.; Warren, R. A. J.; Street, I. P.; Rupitz, K.; Kempton, J. B.; Aebersold, R. *J. Am. Chem. Soc.* **1990**, *112*, 5887−5889.
(42) Namchuk, M. N.; Withers, S. G. *Biochemistry* **1995**, *34*, 16194−16202.
(43) Notenboom, V.; Birsan, C.; Nitz, M.; Rose, D. R.; Warren, R. A. J.; Withers, S. G. *Nature Struct. Biol.* **1998**, *5*, 812−818.
(44) Baker, D. J.; Kan, J. L. C.; Smith, S. S. *Gene* **1988**, *74*, 207−210.
(45) Hehre, W. J.; Radom, L.; von Ragué Schleyer, P.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(46) *Lange's Handbook of Chemistry*; McGraw-Hill: London, 1979.
(47) Davies, D. T. *Aromatic Heterocyclic Chemistry*; Oxford University Press: Oxford, UK, 1994.
(48) Mergny, J.-L., Lacroix, L., Han, X., Leroy, J.-L., Hélène, C. *J. Am. Chem. Soc.* **1995**, *117*, 8887−8898.

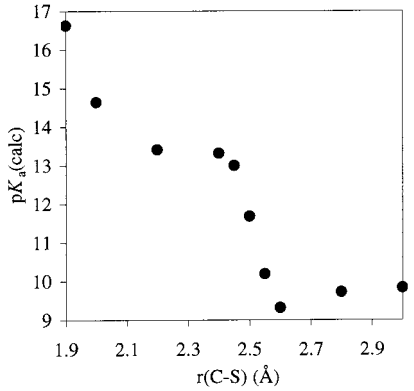


Figure 9. pK_a of N3 of cytosine during the attack of methylthiolate on C6. r(C−S) is the distance between S of the attacking thiolate and C6 of cytosine.

Upon the nucleophilic attack of methylthiolate at C6, the pK_a of the **2**H$^+$ increases to 16.6. The change in the pK_a as the addition reaction progresses is shown in Figure 9. When the distance between the nucleophilic S of methylthiolate and C6 of the ring is 3.0 Å, the pK_a of N3 in the bimolecular complex (**1**CH$^+$) has increased from 5.9 to 9.8. This increase is due to the favorable electrostatic interaction between the positively charged proton on N3 and the negative charge on thiolate. As the C−S distance decreases to 2.6 Å, the pK_a decreases to 9.3, probably because the favorable electrostatic interaction between the charges has decreased. In addition, the negative charge has not spread from the thiolate to the ring system, so the positive charge on the ring is not stabilized. When the addition reaction progresses and the C−S distance decreases to 2.4 Å, the pK_a of N3 increases rapidly to 13.3. Further decrease in the C−S distance increases the pK_a of the covalent adduct steadily to the final value 16.6. Interestingly, rapid change of the pK_a occurs and the TSs of the methylthiolate attacks are located at similar distances (Figures 2 and 3). In the active site of DNA (cytosine-5)-methyltransferase, a glutamic acid (Glu119) hydrogen is bonded to the N3 atom of the target cytosine (Figure 1) and is probably donated to N3 during catalysis.

The pK_a of the isolated covalent adduct **2**CH$^+$ is 16.6. Therefore, the cytosine is presumably in a protonated form at physiological pH and in the enzyme environment. The next step in the reaction sequence is transfer of methyl from trimethylsulfonium to the activated C5. The pK_a of N3 at the (gas-phase) transition state (**2**TSH$^+$) of the methyl-transfer reaction was calculated to be 11.1. That of the product complex, **3**CH$^+$, is 2.3, and that of the isolated C6-thiolate−C5-methyl adduct, **4**H$^+$, is 5.4. These calculations indicate that N3 of cytosine accepts a proton during the nucleophilic attack on C6 in aqueous solution and is deprotonated during transfer of the methyl group to C5. It is possible that protonation/deprotonation steps take place in the enzyme-catalyzed reaction as well. Both proton transfers were calculated to catalyze the corresponding reaction steps (Figures 2 and 3).

Conclusions

In this work, cytosine methylation reaction was investigated in the gas phase and in solution using ab initio quantum mechanical (up to the MP2/6-31+G*//HF/6-31+G* level) and density functional theory calculations (Becke3LYP/6-31+G*) in combination with an isodensity surface-polarized continuum model (IPCM-HF/6-31+G*). A complete reaction sequence, similar to the catalytic mechanism proposed for the DCMtase-catalyzed reaction was studied. In addition, a computational approach was presented to calculate pK_a values of reaction intermediates and transition-state species on the reaction profiles.

The pK_a calculations showed that, during the attack of methyl thiolate on C6 of the cytosine, the pK_a of N3 of the ring increases from 5 to 17. In the subsequent reaction step, where C5 of the cytosine is methylated, the pK_a of N3 drops from 17 to 5. These proton transfers were calculated to catalyze the two reaction steps. It is probable that, in the DCMtase-catalyzed cytosine methylation, proton transfers take place between Glu119 of the active site and N3 of the cytosine. Calculations also indicated that methyl transfer from AdoMet to the activated C5 of the C6-thiolate adduct of cytosine is a fast, highly exothermic reaction. According to the calculations and the recent X-ray crystal structures, DCMtase may accelerate the methyl-transfer step by positioning the reactive groups in a conformation where the reaction can take place without much structural reorganization. In the X-ray structure of *Hha*I methyltransferase complexed with *S*-adenosyl-L-homocysteine and DNA methylated at the target cytosine, it was observed that the C5-methyl of the ligand is bent ∼50° out of plane of the cytosine plane. It is possible that a significant population of species corresponding to the methylated C6-thiolate−C5-proton intermediate **4**, which was calculated to be the most stable intermediate of the reaction sequence studied, exists in the X-ray structure. This would explain the bent C5-methyl of the target cytosine observed experimentally.

Acknowledgment. This work was supported by the Academy of Finland. I am grateful to the Computer Science Center (Espoo, Finland) for computational resources. I thank Peter Kollman (UCSF) for valuable discussions.

JA981405A