J|A|C|S
ARTICLES
Published on Web 03/01/2003

Modeling H-Bonding and Solvent Effects in the Alkylation of Pyrimidine Bases by a Prototype Quinone Methide: A DFT Study

Mauro Freccero,* Cristiana Di Valentin, and Mirko Sarzi-Amadè

Contribution from the Dipartimento di Chimica Organica, Università di Pavia, V.le Taramelli 10, 27100 Pavia Italy

Received September 27, 2002 ; E-mail: freccero@chifis.unipv.it

Abstract: Nucleophilicity of NH_2, N3, and O^2 centers of cytosine toward a model quinone methide (*o*-QM) as alkylating agent has been studied using DFT computational analysis [at the B3LYP/6-311+G(d,p) level]. Specific and bulk effects of water (by C-PCM model) on the alkylation pathways have been evaluated by analyzing both unassisted and water-assisted reaction mechanisms. An ancillary water molecule, H-bonded to the alkylating agent, may interact monofunctionally with the *o*-QM oxygen atom (passive mechanisms) or may participate bifunctionally in cyclic hydrogen-bonded structures as a proton shuttle (active mechanisms). A comparison of the unassisted with the water-assisted reaction mechanisms has been made on the basis of activation Gibbs free energies ($\Delta G^{\ddagger}$). The gas-phase alkylation reaction at N3 does proceed through a passive mechanism that is preferred over both the active (by -6.3 kcal mol^{-1}) and the unassisted process. In contrast, in the gas phase, the active assisted processes at NH_2 and O^2 centers are both favored over their unassisted counterparts by -4.0 and -2.2 kcal mol^{-1}, respectively. The catalytic effect of a water molecule, in gas phase, reduces the gap between the TSs of the O^2 and NH_2 reaction pathways, but the former remains more stable. Water bulk effect significantly modifies the relative importance of the unassisted and water-assisted alkylation mechanisms, favoring the former, in comparison to the gas-phase reactions. In particular, the unassisted alkylation becomes the preferred mechanism for the reaction at both the exocyclic (NH_2) and the heterocyclic (N3) nitrogen atoms. By contrast, alkylation at the cytosine oxygen atom is a water-catalyzed process, since in water the active water-assisted mechanism is still favored. As far as competition, among all the possible mechanisms, our calculations unambiguously suggest that the most nucleophilic site both in gas phase (naked reagents: $N3 \gg O^2 \geq NH_2$) and in water solution (solvated reagents: $N3 \gg NH_2 \gg O^2$) is the heterocyclic nitrogen atom (N3) ($\Delta G^{\ddagger}_{gas} = +7.1$ kcal mol^{-1}, and $\Delta G^{\ddagger}_{solv} = +13.7$ kcal mol^{-1}). Our investigation explains the high reactivity and selectivity of the cytosine moiety toward *o*-QM-like structures both in deoxymononucleoside and in a single-stranded DNA, on the basis of strong H-bonding interactions between reactants and solvent bulk effect. It also offers two general reactivity models in water, uncatalyzed and active water-catalyzed mechanisms (for nitrogen and oxygen nucleophiles, respectively), which should provide a general tool for the planning of nucleic acid modification.

Introduction

Quinone methides (QMs) play a key role in the chemistry of several classes of antitumor compounds, and antibiotic drugs, including the commercial ones mitomycin C[1] and anthracycline antibiotics,[2] where they form covalent linkages with DNA bases. DNA cross-linking,[3] which is probably one of the most promising application of QMs reactivity, has been obtained as a result of two consecutive alkylating steps, both involving QMs.[4]

The reactivity of QMs is mainly due to their remarkable electrophilic nature, which is comparable to that one of stabilized carbocations.[5] In fact QMs are Michael acceptors adding nucleophiles at the exocyclic methylene group to form benzylic adducts. Interactions of quinone methides with simple sulfur-,[5b,c] nitrogen- and oxygen-centered nucleophiles have been experimentally[6] and computationally investigated.[7] Their reactivity has been also experimentally studied with biological

(1) (a) Tomasz, M.; Das, A.; Tang, K. S.; Ford, M. G. J.; Minnock, A.; Musser, S.; Waring, M. J. *J. Am. Chem. Soc.* **1998**, *120*, 11581. (b) Han, I.; Russel, J.; Kohn, H. *J. Org. Chem.* **1992**, *57*, 1799.

(2) (a) Angle, S. R.; Rainier, J. D.; Woytowicz C. *J. Org. Chem.* **1997**, *62*, 5884. (b) Gaudiano, G.; Frigerio, M.; Bravo, P.; Koch, T. H. *J. Am. Chem. Soc.* **1990**, *112*, 6704.

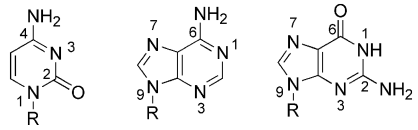
(3) (a) Rajski, S. R.; Williams, R. M. *Chem. Rev.* **1998**, *98*, 2723. (b) Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666. (c) Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856. (d) Zhou, A.; Duan, W.; Simmons, D.; Shayo, Y.; Raymond, M. A.; Dorr, R. T.; Hurley, L. H. *J. Am. Chem. Soc.* **2001**, *123*, 4865. (e) Bando, T.; Iida, H.; Saito, I.; Sugiyama, H. *J. Am. Chem. Soc.* **2001**, *123*, 5158. (f) Zhou, Q.; Pande, P.; Johnson, A. E.; Rokita, S. E. *Bioorg. Med. Chem.* **2001**, *9*, 2347.

(4) (a) Zeng, Q.; Rokita, S. E. *J. Org. Chem.* **1996**, *61*, 9080. (b) Nakatani, K.; Higashida, N.; Saito, I. *Tetrahedron Lett.* **1997**, *38*, 5005.

(5) (a) Richard, J. P.; Amyes, T. L.; Toteva, M. M. *Acc. Chem. Res.* **2001**, *34*, 981. (b) Richard, J. P.; Crugeiras, J. *J. Am. Chem. Soc.* **2000**, *122*, 1664. (c) Richard, J. P.; Amyes, T. L.; Bei, L.; Stubblefield, V. *J. Am. Chem. Soc.* **1990**, *112*, 9513. (d) McClelland, R. A.; Moriarty, M. M.; Chevalier, J. M. *J. Chem Soc., Perkin Trans. 2* **2001**, 2235.

10.1021/ja028732+ CCC: $25.00

Chart 1



nucleophiles such as free amino acids,[6] oligopeptides,[6,8] and DNA bases.[9−11]

Chemoselectivity in the alkylation processes of DNA bases is still a matter of intensive investigation and hot debate. Competition between various nucleophilic centers in DNA, as mentioned by Rokita, is controlled by (i) the arrangement of the nucleobases within the duplex DNA, (ii) noncovalent preassociation effects, between alkylating reactant or its precursors and DNA, and (iii) intrinsic reactivity.[11] Data on intrinsic reactivity of DNA bases seldom appear to be transferable among different alkylating reactants because the mechanism affecting product formation remains uncertain.

Although selectivity and reactivity in alkylation processes have been often extrapolated from the nucleophilicity of DNA bases (experimentally described by numerous researchers, using a wide variety of reagents, such as diazonium ions, carbocations,[12] benzyl halides,[13] epoxides,[14] and both *p*-QMs[10,15] and *o*-QMs),[9,12] it is important to be aware that nucleophilicity of a substrate is not just an intrinsic characteristic of the nucleophile itself but is also strongly dependent on the nature of the substrate undergoing nucleophilic addition and on solvent effects.[16] In light of this observation, it is not surprising that a generalization of the selectivity on a qualitative basis is an almost impossible task. The only general statement accepted by the experimental community is that hard electrophiles preferentially alkylate the oxygen atoms of cytosine (O^2) and guanine (O^6) and soft electrophiles preferentially modify N-nucleophiles, particularly adenine N1 and cytosine N3 (see Chart 1, for numbering). Unfortunately this statement is too general and vague, and it does not take into account any specific noncovalent interaction between the reactants, like H-bonding, which may heavily affect the selectivity under kinetic control.

The control of the selectivity in alkylation reactions of polyfunctional nucleophiles such as DNA bases, by H-bonding between substrate, alkylating agent, and protic solvent, seeks a rationalization. Although it is known from experimental and computational data that the reactivity of *o*-QM as alkylating agent is highly enhanced by hydrogen bonding with a protic solvent such as water[6,7] and by acid catalysis,[6,17] the foregoing aspect has not yet been subject of any computational investigation.

In the past decade, many theoretical papers tried to model physical and chemical properties of DNA bases (natural and modified)[18,19] and base pairs,[20] but the evaluation of models for DNA modification by alkylating agents has been to our knowledge sporadic.[21,22]

In addition, thermodynamic aspects [namely stability of alkylation adduct evaluated on the basis of reaction free energy values, both in gas phase (ΔG_{gas}) and in solution (ΔG_{sol})] which could partially control the profile of the experimentally detectable adducts, with only few recent exceptions,[11a,23,24] have often been overlooked.

Therefore, within this scenario, we considered worthwhile a computational study on the alkylation reactions of cytosine and 1-methylcytosine (**1-MeC**, as prototype of deoxycytidine in a single stranded DNA) by *o*-QM, to clarify the role of H-bonding on reactivity and selectivity toward DNA bases in water. The reported results should as well provide some more meaningful discussion regarding the mechanism of DNA base alkylation by QMs in condensed protic media. Concerning the *o*-QM reactivity as electrophiles, the interest of the scientific community is high, as proved by several papers in the recent literature,[4−11,15,17,18] because it is a prototype for soft alkylating agents, highly polarizable and showing well-defined hydrogen bond acceptor properties.[7] In summary, the aim of this paper is to computationally define cytosine and 1-methylcytosine nucleophilicity in alkylation reactions, as a prototype for that of a DNA base to clarify to what extent reactivity and selectivity of an alkylating agent is the result of combined effects, such as H-bonding interaction between reactants and protic solvent (which are operative under kinetic control), adduct stability (under thermodynamic control), and bulk solvent effects.

Computational Methods

The B3LYP method is now well established as a method that can compute potential energy surfaces (PES) for organic reactions described by a single electronic configuration. The suitability of DFT for reliably describing hydrogen-bonded systems has been the subject of many investigations,[25] and such methods have proved quite useful for studying

(6) Modica, E.; Zanaletti, R.; Freccero, M.; Mella, M. *J. Org. Chem.* **2001**, *66*, 41.
(7) Di Valentin, C.; Freccero, M.; Zanaletti, R. Sarzi-Amadè, M. *J. Am. Chem. Soc.* **2001**, *123*, 8366.
(8) (a) McCracken, P. G.; Bolton, J. L.; Thatcher, G. R. J. *J. Org. Chem.* **1997**, *62*, 1820. (b) Bolton, J. L.; Turnipseed, S. B.; Thompson, J. A. *Chem.-Biol. Interact*. **1997**, *107*, 185. (c) Bolton, J. L.; Valerio, L. G.; Thompson, J. A. *Chem. Res. Toxicol*. **1992**, *5*, 816. (d) Thompson, D. C.; Perera, K.; Krol, E. S.; Bolton, J. L. *Chem. Res. Toxicol*. **1995**, *8*, 323.
(9) (a) Rokita, S. E.; Yang, J.; Pande, P.; Shearer, J.; Greenberg, W. A. *J. Org. Chem.* **1997**, *62*, 3010. (b) Pande, P.; Shearer, J.; Yang, J.; Greenberg, W. A.; Rokita, S. E. *J. Am. Chem. Soc.* **1999**, *121*, 6773.
(10) Lewis, M. A.; Graff Yoerg, D.; Bolton, J. L. Thompson, J. A. *Chem. Res. Toxicol*. **1996**, *9*, 1368.
(11) (a)Veldhuyzen, W. F.; Shallop, A. J.; Jones, R. A.; Rokita, S. E. *J. Am. Chem. Soc.* **2001**, *123*, 11126. (b) Veldhuyzen, Lam, Y.-F.; Rokita, S. E. *Chem. Res. Toxicol.* **2001**, *14*, 1345.
(12) Blans, P.; Fishbein, J. C. *Chem. Res. Toxicol*. **2000**, *13*, 431.
(13) (a) Moschel, R. C.; Hudgins, R. W.; Dipple, A. *J. Org. Chem.* **1986**, *51*, 4180. (b) Moon, K.-Y.; Moschel, R. C. *Chem. Res. Toxicol*. **1998**, *11*, 696.
(14) (a) Nakatani, K.; Okamoto, A.; Matsuno, T.; Saito, I. *J. Am. Chem. Soc.* **1998**, *120*, 11219. (b) Barlow, T.; Dipple, A. *Chem. Res. Toxicol*. **1998**, *11*, 44.
(15) (a) Zhou, Q.; Turnbull, K. D. *J. Org. Chem.* **1999**, *64*, 2847. (b) Zhou, Q.; Turnbull, K. D. *J. Org. Chem.* **2001**, *66*, 7072.
(16) Lewis, E. D.; Douglas, T. A.; McLaughlin, M. L. A Quantitative Measure of Nucleophilic Character. In *Nucleophilicity*; Harris, J. M., McManus, S. P., Eds.; American Chemical Society: Washington, DC, 1987; Chapter 3, pp 35−49.
(17) (a) Chiang, Y. A.; Kresge, J.; Zhu, Y. *J. Am. Chem. Soc.* **2001**, *123*, 8089. (b) Chiang, Y. A.; Kresge, J.; Zhu, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9854. (c) Chiang, Y. A.; Kresge, J.; Zhu, Y. *J. Am. Chem. Soc.* **2002**, *123*, 717. (d) Wan, P.; Barker, B.; Diao, L.; Fisher, M.; Shi, Y.; Yang, C. *Can. J. Chem*. **1996**, *74*, 465. (e) Brousmiche, D.; Wan, P. *Chem. Commun.* **1998**, 491.
(18) (a)Stewart, E. L.; Foley, C. K.; Allinger, N. L.; Bowen, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 7282. (b) Del Bene, J. E. *J. Phys. Chem.* **1983**, *87*, 367. (c) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1983**, *87*, 367. (d) Roberts, C.; Bandaru, R.; Switzer, C. *J. Am. Chem. Soc.* **1997**, *119*, 4640.
(19) Nowak, M. J.; Lapinski, L.; Kwiatkowski J. S.; Leszczynsky, J. In *Computational Chemistry: Reviews of Current Trends*; Leszczynsky, J., Ed.; World Scientific Publishing: Singapore, 1997; Vol. 2, Chapter 4 and references therein.
(20) (a) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117. (b) Gould, I. R.; Kollman P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493. (c) Floriàn, J.; Leszczynski J. *J. Am. Chem. Soc.* **1996**, *118*, 3010.
(21) Ford, G. P.; Scribner, J. D. *Chem. Res. Toxicol*. **1990**, *3*, 219.
(22) Dannenberg J. J.; Tomasz, M. *J. Am. Chem. Soc.* **2000**, *122*, 2062.
(23) (a) Ouyang, A.; Skibo, E. B. *Biochemistry* **2000**, *39*, 5817. (b) Ouyang, A.; Skibo, E. B. *J. Org. Chem.* **1998**, *63*, 1893.
(24) Zewail-Foote, M.; Hurley, L. H. *J. Am. Chem. Soc.* **2001**, *123*, 6485.

hydrogen-bonded complexes.[26] The B3LYP functional in particular has proven highly effective, at least as long as an appropriate basis set is used.[27] Basis set extensions with polarization function also for hydrogen [i.e., B3LYP/6-311G(d,p)] as well as introduction of diffuse functions [i.e., B3LYP/6-311+G(d,p)] are certainly useful to properly describe lone pairs and hydrogen-bonding interactions,[7] which are very important in controlling the reactivity and selectivity of *o*-QM in the alkylation reactions of nucleobases. The usual dilemma between high computational cost and low-level calculations can be satisfactorily solved for the systems under study by carrying out geometry optimization at the B3LYP/6-31G(d) level and improving the energy description by single point calculations with more extended basis sets. In fact, we have shown (by studying the *o*-QM alkylation reaction of ammonia, water, and hydrogen sulfide) that optimized TS geometries do not change on going from B3LYP/6-31G(d) to B3LYP/6-311+G(d,p) methods.[7] However, not only absolute but also the relative energies of stationary points can change appreciably.[7] In particular, a significant variation in absolute and relative energies takes place when the diffuse and polarization functions are introduced [i.e., on going from the B3LYP/6-31G(d) to B3LYP/6-311+G(d,p) method]. Fortunately, probably as a result of the remarkable geometry constancy when basis set is changed, higher level single-point calculations on B3LYP/6-31G(d) optimized geometries very closely reproduce the relative energies obtained by the corresponding higher level full optimization procedures (as clearly documented by our previous investigation on *o*-QM and three prototype nucleophiles NH_3, H_2O, and H_2S).[7] For reactive systems of larger size (more than 18 heavy atoms), we report B3LYP/6-31G(d) fully optimized geometry of stationary points and related energies as well as B3LYP/6-311+G(d,p)//B3LYP/ 6-31G(d) energies. The discussion on reaction energetics will be based on the latter data.

All calculations were carried out using the Gaussian 94[28] and Gaussian 98[29] program packages.

To confirm the nature of the stationary points and to produce theoretical activation parameters, vibrational frequencies (in the harmonic approximation) were calculated for all the optimized structures and used, unscaled, to compute the zero point energies, their thermal corrections, the vibrational entropies, and their contributions to activation enthalpies, entropies, and activation Gibbs free energies (simply called in the paper activation free energies). The computed relative (to reactants) electronic energies for transition structures and products with the thermodynamic activation parameters [at the B3LYP/6-31G(d) level], obtained from gas-phase vibrational frequencies, are listed in Tables 1−3 for the cytosine, adenine, and guanine alkylation processes, respectively.

The computed enthalpy, entropy, and free energy were converted from the 1 atm standard state into the standard state of molar concentration (ideal mixture at 1 mol L^{-1} and 1 atm)[30] to allow a direct comparison with the experimental result in water solution.[6]

The contributions of bulk solvent effects to the activation free energy of the reactions under study were calculated via the self-consistent reaction field (SCRF) method using the conductor version of PCM (C-PCM)[31] employing the HF parametrization of Barone's united atom topological model (UAHF),[32] as implemented in Gaussian 98. Such a model includes the nonelectrostatic terms (cavitation, dispersion, and repulsion energy) in addition to the classical electrostatic contribution. The former terms, unlike the latter one, always give a positive contribution to the solvation energy (δG_{sol}: effect of the solvation on stationary points), but it is much less important than the electrostatic term. For all PCM-UAHF calculations, the number of initial tesserae/atomic sphere was set to 60 as in the default. For comparative purposes, C-PCM calculations of the solvation energies with 60 initial tesserae were also performed for the TSs on the oxygen alkylation pathway, using Bondi's and Pauling's set of atomic radii (options Radii=Pauling and Radii=Bondi in the PCM version implemented in Gaussian 98). The resulting absolute free activation energies in water solution ($\Delta G^{\ddagger}_{sol}$) are in both cases slightly higher (from +1 to +4 kcal mol^{-1} for Bondi and Pauling options, respectively) in comparison to the values obtained with the default UAHF procedure. Relative free activation energies among TSs remain substantially unaltered. Few tests have been performed on the convergence of the results increasing the tessellation up to 100. The difference with the result obtained from the default settings is negligible (less than 2% on solvation energy). Solvation effect has been evaluated in water solution by single point calculation (i.e., with unrelaxed gas-phase reactant and TS geometries) at the B3LYP/6-31G(d) level and used to evaluate both the B3LYP/6-31G(d) and B3LYP/6-311+G(d,p)//B3LYP/6-31G(d) free energies in aqueous solution.

To evaluate the importance of specific solvent effects in a competing "water-assisted mechanism" we systematically investigated the role of a water molecule added to the cytosine alkylation reaction both in gas phase and in water bulk.

Results and Discussion

Alkylation under Thermodynamic Control. Exploring the potential energy surface (PES) of the alkylation reaction of cytosine (**C**) by *o*-QM in gas phase, we located a prereaction-complex **IC** (Scheme 1), where cytosine is involved in a H-bonding (hydrogen bond length H---O of 1.91 Å and N−H---O angle 173.1°) to *o*-QM, and three alkylation adducts **P0**, **P1**, and **P2**.

The intermediate (**IC**), when free energies are considered, is slightly more stable than free reactants, in gas phase (−0.7 kcal mol^{-1}). **IC** may easily evolve to the oxygen alkylation adduct **P2** via TS **S2**, but the fast equilibration between **IC** and reactants allows one to neglect it in the discussion of product formation.

(25) Guo, H.; Sirois, S.; Proynov, E. I.; Salaub, D. R. In *Theoretical Treatment of Hydrogen Bonding*; Hadzy, D., Ed.; Wiley: New York, 1997.
(26) Rablen, P. R.; Lockman, J. W.; Jorgensen W. L. *J. Phys. Chem.* **1998**, *102*, 3782.
(27) (a) Del Bene, J. E.; Person, W. B. Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705. (b) Kim, K.; Jordan, K. D. *J. Phys. Chem.* **1994**, *98*, 10089.
(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, 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. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision E.3; Gaussian, Inc.: Pittsburgh, PA, 1995.
(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.
(30) For conversion from 1 atm standard state to 1 mol/L standard state, the following contribution need to be added to standard enthalpy, entropy, and free energy: $-RT$, $-R - R \ln R'T$, and $RT \ln R'T$, where R' is the value of R in L × atm/mol × K.[43] For a reaction with A + B = C stoichiometry (such as the unassisted alkylation mechanism), the corrections for $\Delta H'$, $\Delta S'$, and $\Delta G'$ are RT, $R + R \ln R'T$, and $-RT \ln R'T$. At 298 K the corrections amount to 0.59 and −1.90 kcal mol^{-1} for $\Delta H'$ and $\Delta G'$ and +8.34 eu for $\Delta S'$.[44] For a reaction with A + B + C = D stoichiometry (such as the water-assisted alkylation mechanism), the corrections for $\Delta H'$, $\Delta S'$, and $\Delta G'$ are $2RT$, $2(R + R \ln R'T)$, and $-2RT \ln R'T$. At 298 K the corrections amount to 1.18 and −3.79 kcal mol^{-1} for $\Delta H'$ and $\Delta G'$ and +16.68 eu for $\Delta S'$.
(31) (a) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995. For the original COSMO model, see: (a) Klamt, A.; Schüürmann. *J. Chem. Soc., Perkins Trans.* **1993**, 2. (b) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074. (b) For a more comprehensive treatment of solvation models, see: (c) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(32) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

Scheme 1

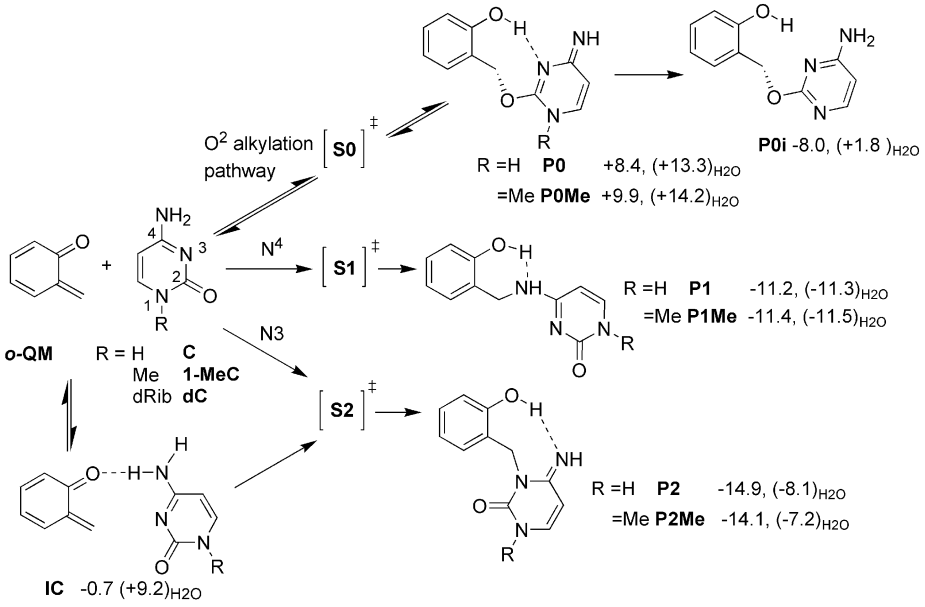


Table 1. B3LYP/6-31G(d) and B3LYP/6-311G+(d,p)//B3LYP/6-31G(d) Electronic Energy (ΔE, in kcal mol^{-1}), Enthalpy (ΔH), Entropy (ΔS), Free Energy in the Gas Phase (ΔG_{gas}), Solvation Free Energy (δG_{sol}), Solvent Effect on Free Energy ($\Delta\Delta G_{sol}$), and Free Energy in Solution (ΔG_{sol}) for the Alkylation Adducts Arising from Reaction of **C** and **1-MeC** with *o*-QM[a]

struct	ΔE	ΔH	ΔS	ΔG_{gas}	δG_{sol}[b]	$\delta\Delta G_{sol}$[c]	ΔG_{sol}[d]	μ_{gas}	μ_{sol}
P0	−9.0[e]	−5.9	−35.8	+4.8	−14.6	+4.9	+9.8	−6.8	−9.4
	−5.5[f]			+8.4			+13.3		
P0Me	−7.7[e]	−4.36	−35.60	+6.26	−11.9	+5.35	+11.61	−7.2	−9.8
	−4.1[f]			+9.89			+14.24		
P1	−27.5[e]	−23.7	−34.2	−13.6	−19.6	−0.0	−13.6	−5.8	−8.1
	−25.1[f]			−11.2			−11.3		
P1Me	−27.8[e]	−24.1	−35.7	−13.4	−17.3	−0.1	−13.6	−5.7	−7.9
	−25.4[f]			−11.4			−11.5		
P2	−33.4[e]	−29.6	−38.8	−18.0	−12.8	+6.8	−11.2	−6.8	−8.5
	−30.3[f]			−14.9			−8.1		
P2Me	−32.8[e]	−28.9	−31.9	−17.2	−10.4	+6.8	−10.3	−7.1	−8.8
	−29.7[f]			−14.1			−7.2		

[a] With respect to reactants (cytosine and *o*-QM). Kinetic contributions [nonpotential energy terms at B3LYP/6-31G(d) level] to molar entropy ($T\delta$S, at 298.15 K), enthalpy (δH), and free energy (δG) are reported in the Supporting Information (Table 1S). For conversion from 1 atm standard state to 1 mol/L standard state, see ref 30. [b] Solvent effect (δG) on stationary points by C-PCM single point calculations on gas-phase geometries B3LYP-C-PCM/6-31G(d)//B3LYP/6-31G(d). [c] Solvent effect on reaction free energy, calculated as $\delta\Delta G_{sol} = \delta G_{sol} - \delta G_{reactants}$ ($\delta G_{reactants}$ = sum of the solvent effect on each reactant). [d] Free energy in water solution calculated as $\Delta G_{solv} = \Delta G_{gas} + \delta\Delta G_{solv}$. [e] B3LYP/6-31G(d). [f] B3LYP/6-311G+(d,p)//B3LYP/6-31G(d).

Nitrogen alkylation adducts **P1** and **P2**, arising respectively from NH_2 (N^4) and N3 alkylation processes (Scheme 1) are both thermodynamically stable, lying −11.2 and −14.9 kcal mol^{-1} below reactants, respectively (Table 1). **P2** is the most stable in gas phase, and therefore, it is the resulting adduct from a thermodynamic controlled alkylation process. **P0** is the adduct of the oxygen (O^2) covalent modification of cytosine, and it is the only alkylation product unstable in gas phase, by +8.4 kcal mol^{-1}, in comparison to free reactants (Scheme 1). **P0** is the less stable among two tautomeric compounds both resulting from cytosine oxygen alkylation. Unlike **P0**, tautomer **P0i** is more stable than free reactants in the gas phase by −8.0 kcal mol^{-1}. Since our aim is to predict QM reactivity toward **dC** (modeled by **1-MeC**; see Scheme 2), where the above tautomerization pathway is not a feasible process, such a tautomer will not be discussed any further.

Replacement of the H atom at the cytosine N1 center with a methyl group generates 1-methylcytosine (**1-MeC**), which is better than cytosine as a model substrate for deoxycytidine (**dC**) in a single strand of DNA. Therefore, to evaluate *o*-QM-DNA adducts stability we will take in consideration free energy data for 1-methylcytosine alkylation adducts. No significant effect on product stability in gas phase is introduced by methyl substitution at N1, since **P0Me**, **P1Me**, and **P2Me** free energies values (+9.9, −11.4, and −14.1 kcal mol^{-1}, respectively), relative to free reactants, are comparable or slightly higher (in the case of **P0Me** and **P2Me**) than the unsubstituted counterparts (see free energy data reported underneath each structure in Scheme 1).

Under thermodynamic control in water, the selectivity should be reversed in favor of the NH_2 (N^4) group since the product **P1** more stable in water than **P2** by −3.2 kcal mol^{-1} (Scheme 1). Similar considerations hold for 1-methylcytosine, since methyl substitution at N1 introduces minor change in the free energy values relative to free reactants (see data in Table 1 and Scheme 1). *Oxygen alkylation* of **1-MeC** by *o*-QM, although feasible, actually cannot be an operative chemical pathway, since the resulting adduct (**P0Me**) is much less stable than free reactants both in gas phase and water solution ($\Delta G^{\ddagger}$ = +9.9 and +14.2 kcal mol^{-1}, respectively).

Under experimental conditions deoxycytidine undergoes alkylation at the N3 center to afford selectively **P2Me**-like

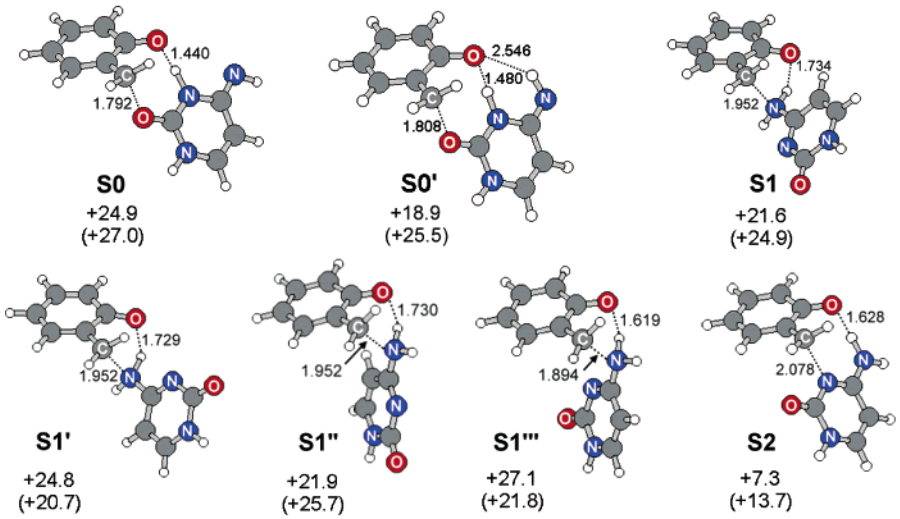


Figure 1. Optimized TS geometries (**S0**−**S2**) of the cytosine alkylation reaction by *o*-QM, without water assistance. Bond lengths (in Å) and activation Gibbs free energies (in kcal mol^{-1}) in the gas phase and water solution (in parentheses) at the B3LYP/6-311+G(d,p)//B3LYP/6-31G(d) level of theory are reported.

Table 2. B3LYP/6-31G(d) and B3LYP/6-311G+(d,p)//B3LYP/6-31G(d) Electronic Activation Energy ($\Delta E^{\ddagger}$, in kcal mol^{-1}), Enthalpy ($\Delta H^{\ddagger}$), Entropy ($\Delta S^{\ddagger}$), Free Energy in the Gas Phase ($\Delta G^{\ddagger}_{gas}$), Solvation Free Energy (δG_{sol}), Solvent Effect on Free Energy ($\Delta\Delta G_{sol}$), and Free Activation Energy in Solution ($\Delta G^{\ddagger}_{sol}$) for the Alkylation of **C** and **1-MeC** by *o*-QM[a]

struct	$\Delta E^{\ddagger}$	$\Delta H^{\ddagger}$	$\Delta S^{\ddagger}$	$\Delta G^{\ddagger}_{gas}$	δG_{sol}[b]	$\delta\Delta G_{sol}$[c]	$\Delta G^{\ddagger}_{sol}$[d]	μ_{gas}	μ_{sol}
				O Alkylation TSs					
S0	7.8[e]	11.0	−36.3	21.6	−17.5	+2.1	23.7	4.6	10.8
	11.0[f]			24.9			27.0		
S0Me	9.0[e]	9.4	−36.2	20.2	−15.2	+2.0	22.2	8.4	11.4
	12.2[f]			23.4			25.4		
S0′	3.1[e]	4.1	−37.1	15.2	−13.0	+6.6	21.8	6.5	9.4
	6.8[f]			18.9			25.5		
S0′Me	4.3[e]	5.4	−36.3	16.3	−10.6	+6.6	22.8	7.1	10.1
	7.9[f]			19.9			26.4		
				NH_2 (N^4) Alkylation TSs					
S1	5.3[e]	7.3	−35.5	17.8	−16.2	+3.4	21.2	3.0	3.1
	9.0[f]			21.6			24.9		
S1Me	4.8[e]	6.8	−37.1	17.8	−13.6	+3.6	21.2	3.3	4.2
	8.4[f]			21.4			25.0		
S1′	9.1[e]	10.6	−35.1	21.1	−23.8	−4.2	16.9	8.9	9.3
	12.9[f]			24.8			20.7		
S1′Me	8.6[e]	10.2	−36.7	21.1	−21.2	−4.0	17.1	8.9	13.0
	12.2[f]			24.7			20.7		
S1″	6.0[e]	7.9	−35.5	18.5	−15.8	+3.8	22.3	3.0	3.7
	9.4[f]			21.9			25.7		
S1‴	11.8[e]	13.1	−35.5	23.7	−24.9	−5.3	18.4	9.1	13.8
	15.2[f]			27.1			21.8		
				N3 Alkylation TSs					
S2	−8.0[e]	−6.5	−36.2	4.3	−13.2	+6.4	10.7	7.3	7.4
	−5.0[f]			7.3			13.7		
S2Me	−8.0[e]	−6.5	−36.6	4.6	−10.6	+6.6	11.2	7.5	9.7
	−5.0[f]			7.6			14.2		

[a] See footnote *a* in Table 1, considering also that symmetry numbers used to calculate entropy are $\sigma = 1$ for **1** and **C** and $\sigma = 2$ for H_2O. A correction of R ln 2 to ΔS has been added for the alkylation reactions, as the nucleophile attacks on the *o*-QM faces are not experimentally distinguishable. [b−f] See Table 1 footnotes.

product.[9a] Therefore, having observed above that **P1Me** is the favored product under thermodynamic control, kinetic aspects rather than adduct stability have to be taken into consideration.

Specific solute−solvent interactions, as a result of complexation of an ancillary water molecule, do not introduce any significant change either in relative stability of the alkylation adducts (**P0**−**P2**) or in their stability with respect to reactants.

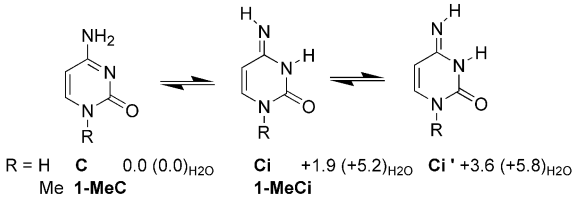
Alkylation of "Naked" Reactants in Gas phase under Kinetic Control. On the PES we have been able to locate six different transition structures (**S0**−**S2**, Figure 1 and Table 2) connecting the reactants to products.

Alkylation at the Oxygen Atom. Alkylation at the oxygen atom of cytosine (O^2) proceeds through **S0** and **S0i** TSs, which both involve H-imino cytosine tautomers **Ci** and **Ci′** (see Scheme 2) less stable, in the gas phase, than the 4-amino derivatives (**C**) by +1.9 and +3.6 kcal mol^{-1}, respectively.

In water solution the instability of **Ci** and **Ci′** becomes even more pronounced than in the gas phase, by +5.2 and +5.8 kcal mol^{-1}.[33,34] Both isomers allow the reactive system to gain a

(33) In good agreement with previous computational investigations at the MP2/6-31G(d,p)//HF/6-31G(d,p)[18d,20c] and B3LYP/6-311+G(2df,2p) levels of theory.[34]

Scheme 2



stabilization by a strong H-bonding (hydrogen bond length H---O of 1.44−1.48 Å and N−H---O angle 174.0°), which involves the QM oxygen atom and the N(3)−H group (see **S0** and **S0′** in Figure 1). Between these two TSs, **S0′** is more stable in the gas phase, showing an activation free energy of +18.9 kcal mol^{-1}. The stabilization of **S0′** relative to **S0** is likely due to a *chelate* H-bonding[35] between reactants in the TS. No TS for an O^2 alkylation pathway involving the most stable **C** tautomer has been located. Methyl substitution at N1 slightly enhances the activation free energy in gas phase of the alkylation process at the oxygen atom (by less than 1 kcal mol^{-1}), in comparison to that of the corresponding unsubstituted counterpart in the gas phase. In fact **S0′Me** lies +19.9 kcal mol^{-1} above free reactants.

Alkylation at NH_2. There are four TSs (**S1**−**S1‴**; see Figure 1) on the pathway from reactants to the alkylation adduct at the NH_2 group, that is **P1** (N^4; for numbering, see Scheme 1). In all the four **S1**−**S1‴** TSs the nucleophilic NH_2 center acts also as hydrogen bond donor. Formally **S1** and **S1′** as well as **S1″** and **S1‴** are pairs of conformers that can be converted into each other by rotation around the C_4–N_4 bond. **S1** and **S1′** exhibit a low dipole moment (3.0 D in gas phase), while those of **S1′** and **S1‴** is considerably higher (8.9 and 9.1 D dipole, respectively, in the gas phase). The remarkable difference in the dipole moments of the **S1**, **S1″** couple as compared to the **S1′**, **S1‴** pairs nicely accounts for the significantly higher stability of the former relative to the latter in the gas phase. Moreover, it suggests that solvent polar effects should play an important role in a selective stabilization of highly polar **S1′** and **S1‴** over the other two less polar TSs. Such a solvent effect has to be taken into account for a realistic depiction of both reaction mechanisms and selectivity of the alkylation processes, particularly in water which has often been used either pure or as cosolvent in numerous experimental investigations.[9−11] Methyl substitution at N1 introduces negligible effects on both TS energy and polarity. Therefore, similar solvent effects can be anticipated also for 1-methylcytosine alkylation through **S1Me** and **S1′Me** TSs (Table 2).

Alkylation at N3. The most stable TS among those located by us in the gas phase involves the cytosine N3 nucleophilic center (**S2**, in Figure 1). In fact the alkylation reaction at the heterocyclic N3 atom by *o*-QM shows an activation free energy of only +7.30 kcal mol^{-1} in the gas phase. The **S2** TS shows a strong hydrogen bonding between the *o*-QM oxygen atom and a H atom on the exocyclic nitrogen N^4 (H---O distance of 1.62 Å, N−H---O angle of 171.9°), with the forming C--N bond length of 2.08 Å (Figure 1).

Replacement of cytosine with 1-methylcytosine, as alkylated substrate, has little effect on the energetics as well as on solvent effects on TS. In fact, **S2Me** activation free energy is only slightly higher (by less than 0.5 kcal mol^{-1}) than that of the unsubstituted counterpart **S2** (Table 2).

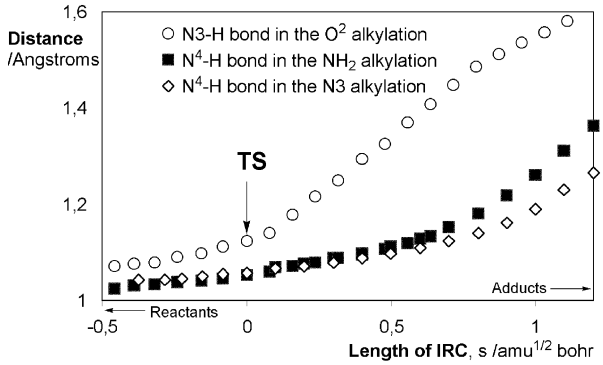


Figure 2. Change of N(3)−H and N^4−H bond lengths along the IRC path, starting from reactants ($s \rightarrow -\infty$) to final alkylation adducts.

Evaluation of the H-Bonding Strength in Gas-Phase TSs. A comparison between geometric features of TSs (**S0**−**S2**) reveals interesting aspects of the intermolecular hydrogen-bonding assistance to the nucleophilic addition. The systematic shortening of the O···HN distance, between the cytosine hydrogen atom being transferred and the *o*-QM oxygen atom, along the TS series **S1** (1.73 Å), **S2** (1.63 Å), and **S0** (1.44 Å) provides geometrical evidence of a parallel increase in H-bonding. Consistent with such bond shortening, the difference between the unscaled stretching frequencies of the N−H bond (involved in the H atom transfer) in the TS and the corresponding bond in the reacting cytosine tautomer (**C** for **S1** and **S2**, **Ci** for **S0** TSs) increases in the same order, namely, $\Delta\nu = 802.6$, 914.9, and 1703.7 cm^{-1} for **S1**, **S2**, and **S0**, respectively. For example, frequency analysis suggests that the weakening of N(3)−H bond passing from **Ci** tautomer (3608.3 cm^{-1}) to **S0** TS (1904.6 cm^{-1}) is noteworthy and clearly demonstrates that such a H atom is involved in a much stronger H-bonding interaction in comparison to the N^4-H in both **S1** and **S2**. IRC calculations [at the B3LYP/6-31G(d) level] confirm that in the oxygen alkylation H-bonding plays a more relevant role than in the nitrogen alkylation pathways. In fact, proton [N(3)−H] transfer to the *o*-QM oxygen occurs earlier in **S0** (similar considerations hold for **S0′**) than that corresponding one involving NH_2 in **S1** and **S2**, as one can easily grasp from inspection of Figure 2. The stretching of the N−H bond along the reaction coordinate from reactants to products is clearly faster for the reaction at the oxygen in comparison to the nitrogen alkylation pathways.

Role of Water on the Alkylation Process: Specific and Bulk Effects. H-bonding interactions between *o*-QM and cytosine in the TSs are particularly effective in the control of TS stabilization. Water molecules may interfere acting as H-bond donors or acceptors in these interactions with resultant important effects on reaction energetics, yet in the gas phase. Moreover, also bulk effects can heavily influence TS relative stability. As a consequence, a thorough investigation in water is mandatory, since biologically relevant modifications of DNA take place in such a medium. We have recently shown that the chemoselectivity of NH_3, H_2O, and H_2S alkylation reactions by *o*-QM is strongly affected by both specific and bulk solvent effects.[7] In fact, modeling the reactivity of *o*-QM only in the

(34) Russo, N.; Toscano, M.; Grand, A. *J. Am. Chem. Soc.* **2001**, *123*, 10272.
(35) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997; Chapter 2, pp 11−32.

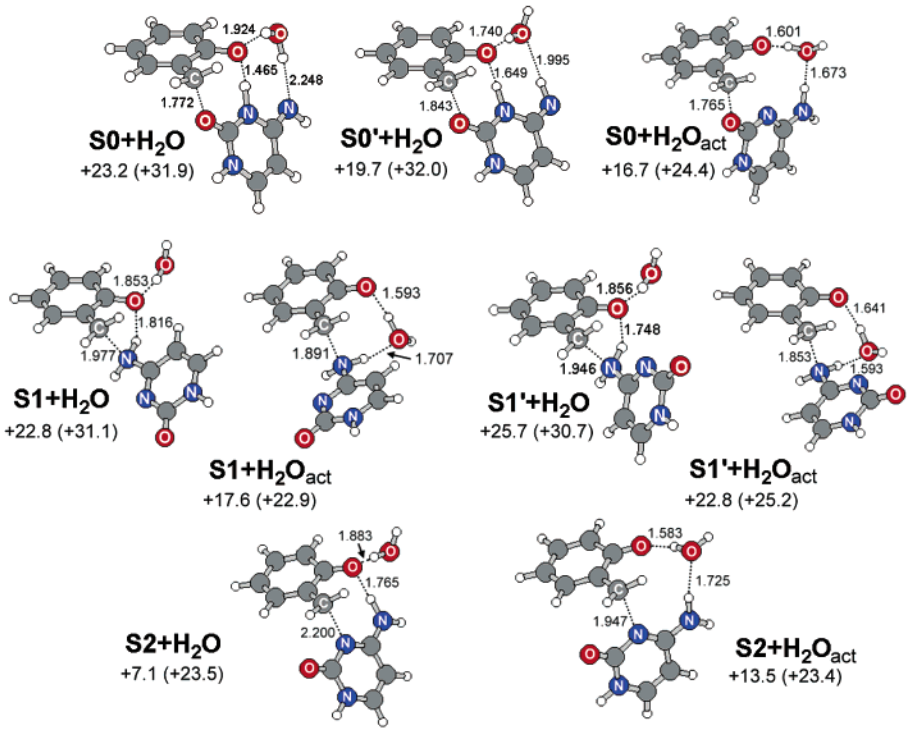


Figure 3. Optimized B3LYP/6-31G(d) TSs of the *o*-QM cytosine alkylation reaction, with an explicit water molecule. In the **S0**+$\mathbf{H_2O_{act}}$, **S1**+$\mathbf{H_2O_{act}}$, **S1′**+$\mathbf{H_2O_{act}}$, and **S2**+$\mathbf{H_2O_{act}}$ TSs, the water molecule is directly involved in an *active* process. In the **S0**+$\mathbf{H_2O}$, **S0′**+$\mathbf{H_2O}$, **S1**+$\mathbf{H_2O}$, **S1′**+$\mathbf{H_2O}$, and **S2**+$\mathbf{H_2O}$ TSs, the water molecule is complexed to the *o*-QM oxygen atom without taking part in the proton-transfer process (*passive* process). Bond lengths are in Å, and activation Gibbs free energies (at the B3LYP/6-311+G(d,p)//B3LYP/6-31G(d) level) in the gas phase ($\Delta G'_{gas}$) and water solution ($\Delta G'_{sol}$, in parentheses) are in kcal/mol.

gas phase resulted in the wrong chemoselectivity scale in comparison to solution experimental data.[6]

Reaction of Water-Complexed *o*-QM. It has recently been well documented that contemporary continuum models can adequately describe reactions in solution.[36] Nevertheless, such an approach cannot properly describe the solvent effect on reactivity, if one or more solvent molecules are directly involved in the reaction mechanism.[7,36−38]

Thus, an appropriate depiction of QM reactivity with nucleobases in water must take care of both specific and bulk effects of the polar protic solvent. Thus, we decided first to examine *the effect of an explicit water molecule in the alkylation reaction of* **C** (the smallest pyrimidine base) by *o*-QM, to elucidate *the water H-bonding effect* on reaction mechanism and on the reactivity of *o*-QM as benzylating agent. Such a mechanism will be called from now on "*water-assisted*". Second we addressed the *effect of the solvent bulk on both water-assisted and -unassisted mechanisms*. We achieved the first goal by locating several TSs on the PES governing the alkylation reaction of cytosine in the presence of a specific water molecule complexed to the *o*-QM oxygen atom (TSs in Figure 3).

Such a water molecule may be directly involved in transferring a proton from cytosine to the *o*-QM oxygen atom in a cascade process, or it may be complexed to the *o*-QM oxygen atom without directly taking part in the proton-transfer process (such as in **S0**+$\mathbf{H_2O}$, **S0′**+$\mathbf{H_2O}$, **S1**+$\mathbf{H_2O}$, **S1′**+$\mathbf{H_2O}$, and **S2**+$\mathbf{H_2O}$ TSs). The first type of TS (labeled by the subscript "act", such as **S0**+$\mathbf{H_2O_{act}}$, **S1**+$\mathbf{H_2O_{act}}$, **S1′**+$\mathbf{H_2O_{act}}$, and **S2**+$\mathbf{H_2O_{act}}$) characterizes, according to Williams's definition,[39] the "*active*" water-assisted nucleophilic addition, where the ancillary water molecule acts as both donor and acceptor of hydrogen bonds in a cyclic array. The second type of TS defines the "*passive*" water-assisted process[39] because the water molecule works monofunctionally as hydrogen bond donor.

Change in both geometry and activation free energy (in gas phase)[40] for the passive water-assisted mechanism for all the alkylation processes (through **S0**+$\mathbf{H_2O}$, **S0′**+$\mathbf{H_2O}$, **S1**+$\mathbf{H_2O}$, and **S2**+$\mathbf{H_2O}$ TSs), in comparison to their unassisted counterparts (**S0**, **S0′**, **S1**, and **S2**), is negligible. Data in Table 3 demonstrate that the favorable enthalpic effects introduced by the additional water molecule is largely, for **S1**+$\mathbf{H_2O}$ and **S2**+$\mathbf{H_2O}$, counterbalanced by entropic factors (see data in Tables 2 and 3).

In the alkylation of the NH_2 group, the active processes always feature, as expected, remarkable changes in TSs geometries, which in some cases are accompanied by large changes in activation free energy. The presence of a discrete water molecule, directly involved in the proton-transfer process in the active water assisted mechanism, allows the reactive system to reach a more perpendicular approach of the N^4 nucleophilic center to the *o*-QM exocyclic methylene group, in

(36) Arnaud, R.; Adamo, C.; Cossi, M.; Millet, A.; Vallée, Y.; Barone, V. *J. Am. Chem. Soc.* **2000**, *122*, 324 and references cited therein.
(37) Pardo, L.; Osman, R.; Weinstein, H.; Rabinowitz, J. R. *J. Am. Chem. Soc.* **1993**, *115*, 8263.
(38) (a) Yamabe, S.; Ishikawa, T. *J. Org. Chem.* **1997**, *62*, 7049. (b) Yamabe, S.; Ishikawa, T. *J. Org. Chem.* **1999**, *64*, 4519. (c) Okumoto, S.; Fujita N.; Yamabe, S. *J. Phys. Chem.* **1998**, *102*, 3991.
(39) Williams, I. H. *J. Am. Chem. Soc.* **1987**, *109*, 6299.
(40) Activation free energy for the water-assisted alkylation reactions has been referred to free reactants (i.e., *o*-QM, cytosine, and water). The choice has been suggested by the fact that water-complexed cytosine is strongly destabilized in water bulk ($\Delta G'_{sol} = +7.3$ kcal mol^{-1}), in comparison to free cytosine and free water in water solution. In the gas phase the water-complexed cytosine is slightly more stable than free cytosine and water ($\Delta G'_{gas} = +1.0$ kcal mol^{-1}).

Table 3. B3LYP/6-31G(d) and B3LYP/6-311G+(d,p)//B3LYP/6-31G(d) Electronic Activation Energy ($\Delta E^{\ddagger}$, in kcal mol^{-1}), Enthalpy ($\Delta H^{\ddagger}$), Entropy ($\Delta S^{\ddagger}$), Free Energy in the Gas Phase ($\Delta G^{\ddagger}_{gas}$), Solvation Free Energy (δG_{sol}), Solvent Effect on Free Energy ($\Delta\Delta G_{sol}$), and Free Activation Energy in Solution ($\Delta G^{\ddagger}_{-ol}$) for the Water-Assisted Alkylation Reaction of **C** by *o*-QM[a]

struct	$\Delta E^{\ddagger}$	$\Delta H^{\ddagger}$	$\Delta S^{\ddagger}$	$\Delta G^{\ddagger}_{gas}$	δG_{sol}[b]	$\delta\Delta G_{sol}$[c]	$\Delta G^{\ddagger}_{sol}$[d]	μ_{gas}	μ_{sol}
				O Water-Assisted Alkylation TSs					
S0+$\mathbf{H_2O_{act}}$	−13.7[e]	−9.8	−62.2	8.7	−17.4	+8.1	16.8	9.7	12.5
	−5.8[f]			16.7			24.4		
S0+$\mathbf{H_2O}$	−5.4[e]	−2.0	−60.8	16.1	−16.8	+8.7	24.8	11.2	14.6
	+1.7[f]			23.2			31.9		
S0′+$\mathbf{H_2O}$	−4.8[e]	−0.8	−58.8	16.8	−13.2	+12.3	29.0	7.0	10.1
	−1.8[f]			19.7			32.0		
				NH_2 (N^4) Water-Assisted Alkylation TSs					
S1+$\mathbf{H_2O_{act}}$	−12.2[e]	−7.7	−60.5	10.3	−20.2	+5.3	15.6	4.5	4.7
	−4.9[f]			17.6			22.9		
S1+$\mathbf{H_2O}$	−5.3[e]	−0.7	−55.9	15.9	−17.1	+8.3	24.3	4.3	4.3
	+1.5[f]			22.8			31.1		
S1′+$\mathbf{H_2O_{act}}$	−8.6[e]	−4.4	−60.0	13.5	−23.0	+2.4	15.9	9.8	10.4
	+0.1[f]			22.8			25.2		
S1+$\mathbf{H_2O}$	−1.8[e]	+2.4	−55.5	18.9	−20.5	+4.9	23.9	10.0	13.7
	+4.9[f]			25.7			30.7		
				N3 Water-Assisted Alkylation TSs					
S2+$\mathbf{H_2O_{act}}$	−16.0[e]	−11.9	−61.3	6.4	−15.6	+9.9	16.3	8.4	8.6
	−8.8[f]			13.6			23.4		
S2+$\mathbf{H_2O}$	−21.2[e]	−17.0	−58.6	0.5	−9.1	+16.4	16.9	4.5	6.1
	−14.6[f]			7.1			23.5		

[a] See footnote *a* in Table 1, considering also that, for a reaction with A + B + C = D stoichiometry (such as the water-assisted alkylation mechanism), the correction for $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, and $\Delta G^{\ddagger}$ are $2RT$, $2(R + R \ln R^{\ddagger}T)$, and $-2RT \ln R^{\ddagger}T$. At 298 K the corrections amount to +1.18 and −3.79 kcal mol^{-1} for $\Delta H^{\ddagger}$ and $\Delta G^{\ddagger}$ and +16.68 eu for $\Delta S^{\ddagger}$. [b−f] See Table 1 footnotes.

comparison to the passive mechanism (compare **S1**+$\mathbf{H_2O_{act}}$ to in **S1**+$\mathbf{H_2O}$). Such a favorable approach results in a generalized stabilization of the active water-assisted mechanism over the passive one (both referred to H_2O + cytosine + *o*-QM). Nevertheless, active assistance does not significantly change TS activation energies involved in the NH_2-alkylation process in comparison to the unassisted mechanism. In fact, from the energetic point of view, TS stabilization induced by water assistance in the gas phase is moderate for the NH_2 (N^4) alkylation. For example, **S1**+$\mathbf{H_2O_{cat}}$ and **S1′**+$\mathbf{H_2O_{act}}$ (Figure 3) activation free energy values are lower than those of the related unassisted counterparts (**S1** and **S1′**, Figure 1), by −4.0 and −2.0 kcal mol^{-1} (in gas phase), respectively.

Concerning N(3) alkylation reaction, the water-assisted process in the gas phase (through **S2**+$\mathbf{H_2O_{act}}$ TS) is strongly disfavored over its passive counterpart (**S2**+$\mathbf{H_2O}$) by more than +6 kcal mol^{-1}, probably because the added water molecule disrupts the strong NH---O hydrogen bonding which is operative in **S2**+$\mathbf{H_2O}$ TS, pushing also the cytosine oxygen atom closer to *o*-QM (compare **S2**+$\mathbf{H_2O}$ vs **S2**+$\mathbf{H_2O_{act}}$ geometries, in Figure 3).

Distinct from the N alkylation processes, active water-assisted alkylation at the cytosine oxygen (**S0**+$\mathbf{H_2O_{act}}$) induces a pronounced change not only in TS geometry but also in activation energy. At variance with the passive process that involves both the less stable cytosine tautomers **Ci** and **Ci′** (through **S0**+$\mathbf{H_2O}$ and **S0′**+$\mathbf{H_2O}$, in Figure 3), *o*-QM reacts with the most stable cytosine tautomer **C** in the water-assisted reaction (**S0**+$\mathbf{H_2O_{act}}$). Moreover, the water molecule in **S0**+$\mathbf{H_2O_{act}}$ allows a proton transfer from the NH_2 group to the QM oxygen atom, while an almost ideal alignment of an oxygen lone pair with the π-system at the QM carbon is attained without significant angle strain. The more favorable geometry of the nucleophile attack (from a stereoelectronic standpoint) on the most stable isomer and its higher nucleophilicity[41] in comparison to the tautomers (**Ci** and **Ci′**) are important factors in lowering the activation free energy of the alkylation process at the cytosine oxygen atom from +24.9 to +16.7 kcal mol^{-1} (in gas phase). Although the stabilization induced by the water complexation on O-alkylation (**S0**+$\mathbf{H_2O_{act}}$) is remarkable, it is still not sufficient to divert the primary target from cytosine N3 in the gas phase (see Table 1). In fact, the free activation energy for oxygen alkylation ($\Delta G'_{gas}$ = +16.7 kcal mol^{-1}) is much higher than that for N3 alkylation ($\Delta G'_{gas}$ = +7.1 kcal mol^{-1}) in the gas phase.

In summary, all of *the three alkylation pathways (at O^2, N^4, and N3) follow a water-assisted mechanism in the gas phase, which is the "active" one for both NH_2 and oxygen benzylation but is the "passive" one for the process at the heterocyclic nitrogen atom (N3).*

Role of the Electrostatic Effect of the Solvent Bulk. We have dealt so far with the effect of specific hydrogen-bonding interaction of water on *o*-QM reactivity with cytosine. Nevertheless, since the TSs and products located by us span a wide range of dipole moments (from 3.0 to 9.8 D), we have to examine the electrostatic effects of the bulk on the energetics of both water-assisted (passive and active) and uncatalyzed alkylation of cytosine, to reliably describe solvent effect on the activation free energy of the alkylation reactions by *o*-QM. We achieved that goal by computing the free energy of solvation on reactants, products, and TSs (δG in Tables 1−3) by single point calculations on the frozen gas-phase structures with the C-PCM model. The validity of this approach is supported by Barone's results, which show that for the hydrogen cyanide and methanimine reactive system[36] important geometrical changes

(41) Nucleophilicity has been judged by a higher CHelpG charge on the oxygen atom of **C** in comparison to **Ci**, 0.60 vs 0.55 electrons, and by the analysis of the electrostatic potential surface in the region of the oxygen lone pairs for **C** and **Ci**. For a comprehensive review of the use of molecular electrostatic potentials for the comparison and description of stable molecules, see ref 42.

(42) Tasi, G.; Palinko, I. *Top. Curr. Chem.* **1995**, *174*, 45.
(43) Benson, S. *Thermochemical Kinetics*; Wiley: New York, 1968; p 8.
(44) Rastelli, A.; Bagatti, M.; Gandolfi, R. *J. Am. Chem. Soc.* **1995**, *117*, 4965.

have little influence on the solvation energy. We have also confirmed very recently, for the nucleophilic attack by NH_3 on *o*-QM,[7] that fully optimized stationary points in water display free solvation energies similar to those obtained from single point calculations on the optimized gas-phase structures. This approach allowed us to calculate the solvent effect on both the reaction free energy and the activation free energy ($\delta\Delta G_{sol}$) as $\delta G_{sol} - \delta G_{reactants}$ (with $\delta\Delta G_{sol}$ referred to the adduct and TS, respectively), which are quantitative measurements of adducts and TSs solvation in comparisons to free reactants. They can be added to ΔG_{gas} and $\Delta G^{\ddagger}_{gas}$ to give ΔG_{sol} and $\Delta G^{\ddagger}_{sol}$, respectively (see Tables 1−3).

Bulk Effects on Unassisted Mechanism. The free energy of solvation (δG) is important and stabilizes both prereaction intermediates and TSs less than free reactants, as shown quantitatively by a positive $\delta\Delta G_{sol}$, with the exception of **S1′** and **S1‴** (Table 2). As a result, the bulk effect of the solvent induces dissociation of the complex **IC** (which becomes less stable than free reactants by +9.2 kcal mol^{-1}) into reactants and increases the activation free energies of all uncatalyzed alkylation reactions (by 2.1−6.6 kcal mol^{-1}), with the exception of cytosine alkylation at NH_2 (N^4) through **S1′** and **S1‴** TSs, which $\delta\Delta G_{sol}$ is reduced by −4.2 and −5.3 kcal mol^{-1}, respectively (Table 2). This exception finds a reasonable explanation on the polar nature of both **S1′** and **S1‴**. In fact, the latter TSs exhibit the highest dipole moments (8.9 and 9.1 D in the gas phase, respectively, and 9.3 and 13.8 D in bulk, respectively) among the TSs unassisted by water.

Nonelectrostatic terms have a negligible effect, since they are generally much smaller in absolute value than the electrostatic contribution and overall they are very similar among unassisted TSs.

Bulk Effects on the Water-Assisted Mechanism. Passive vs Active Mechanisms in Water. The water bulk has an important effect on water-assisted alkylation reactions. In more details, all water-assisted TSs are considerably destabilized in comparison to both free reactants and to the related unassisted mechanism. An important factor in determining such a selective destabilization is the nonelectrostatic term of the solvent bulk (due mainly to repulsion between solvent and reactants), which is always higher for water-assisted TSs in comparison to that of the unassisted counterpart (by roughly +3.0 kcal mol^{-1}). Furthermore, the passive mechanism is generally disfavored in comparison to the active one. For clarity it is useful comparing free solvation energy values for the following pair of active vs passive TSs: **S0**+$\mathbf{H_2O_{act}}$ (−17.4) vs **S0**+$\mathbf{H_2O}$ (−16.8); **S1**+$\mathbf{H_2O_{act}}$ (−20.2) vs **S1**+$\mathbf{H_2O}$ (−17.1); **S2**+$\mathbf{H_2O_{act}}$ (−15.6) vs **S2**+$\mathbf{H_2O}$ (−9.1 kcal mol^{-1}). Therefore, among the water-assisted mechanisms, the passive one, which was slightly favored in gas phase only for the alkylation at the heterocyclic nitrogen atom (N3), is always and clearly ruled out in water bulk.

Summary of Cytosine and 1-Methylcytosine Reactivity and Selectivity in Water. Gas-phase calculations on "naked" cytosine and *o*-QM are capable of anticipating qualitatively, assuming kinetic control, the correct experimentally observed product (**P2Me**). However, it is important to become aware of solvation effects, since not only they dramatically influence the relative energetics of the three alkylation pathways (reducing the nucleophilicity of O^2 and N-3 centers in comparison to NH_2 group) but even more importantly they determine the choice between the reaction mechanisms from uncatalyzed to water-catalyzed processes. In more detail, nitrogen alkylations of cytosine at both N3 and NH_2 (N^4) positions are processes that are better described by an uncatalyzed model through **S2** and **S1′** TSs, respectively (see Figure 1). On the other hand, oxygen alkylation of cytosine in water (similarly to the nucleophilic addition of water to *o*-QM previously studied by us)[9] follows an active water-catalyzed mechanism. In fact, the activation free energy of the process through the **S0**+$\mathbf{H_2O_{act}}$ TS is slightly lower than that through the **S0** and **S0′** TS (respectively by −2.6 and −1.1 kcal mol^{-1}). *Therefore, a quantitative evaluation of the selectivity in the cytosine alkylation by o-QM, under kinetic control, can be obtained from direct comparison of the free activation energies for* **S2** (+13.7 kcal mol^{-1}), **S1′** (+20.7 kcal mol^{-1}), and **S0**+$\mathbf{H_2O_{act}}$ (+24.4 kcal mol^{-1}) TSs (see Figures 1 and 2). These data demonstrate that *o-QM is a highly selective alkylating agent of cytosine in water under kinetic control, with the highly dominant attack at the N3 nucleophilic center to give* **P2** *product*. Very similar conclusions can be drawn for 1-methycytosine alkylation reaction by *o*-QM. In fact, the gap between free activation energy for the **S2Me** TS (+14.2 kcal mol^{-1}) and that of its most stable competitor **S1′Me** TS (+20.7 kcal mol^{-1}) in water is 6.5 kcal mol^{-1}.

Cytosine alkylation at the oxygen atom is not achievable using *o*-QM in water, not only because the pathway is kinetically disfavored but mainly because the resulting adducts are strongly unstable (+13.3 kcal mol^{-1}) in comparisons to free reactants. The same conclusions hold for N-Me cytosine demonstrating that our results are useful to predict reactivity of deoxycytidine in water solution.

It is interesting to note that there does not exist a good parallelism between nucleophilicity in Table 4 (evaluated by activation free energy values for the *o*-QM alkylation reactions) and basicity (Table 5) measured by relative free energy (both in the gas phase and water solution) for the three protonated forms of cytosine (Scheme 3). In more detail in the gas phase

Table 4. Nucleophilicity Scale of **C** and **1-MeC** Nucleophilic Centers in the Gas Phase and in Water Solution, As Judged by Comparison of the Computed Activation Free Energy for the Alkylation Processes with *o*-QM

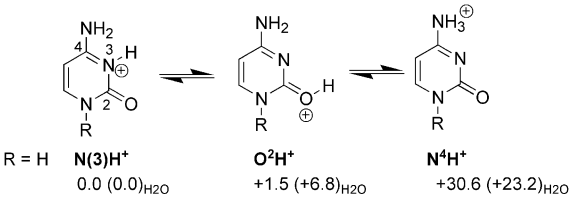
nucleophile	gas phase[a]	water soln
C and **1-MeC**	N3[b] ≫ O2 [c] ≥ NH_2(N^4)[c]	N3[d] ≫ NH_2(N^4)[d] ≫ O^2 [c]

[a] In the presence of an ancillary water molecule. [b] Passive water-assisted process. [c] Active water-assisted process. [d] Unassisted process.

Table 5. Basicity Scale of the Nucleophilic Centers of Tautomer **C**, in the Gas Phase and in Water Solution, As Judged by Comparison of the Computed Free Energy, in the Gas Phase and in Solution, of Cytosine Cations Resulting from **C** Protonation at NH_2 (**N^4H^+**), N-3 (**N(3)H$^+$**), and O^2 (**O^2H^+**)

nucleophile	gas phase	water soln
C	N3 ≥ O^2 ≫ NH_2(N^4)	N3 > O^2 ≫ NH_2(N^4)

Scheme 3



cytosine N-3 is by far the most nucleophilic site although its basicity is only slightly higher in comparison to that of oxygen atom. In water solution N3 still represents both the most nucleophilic and basic site, but the basicity and nucleophilicity orders for NH_2 and oxygen atom are reverted.

Conclusion

Calculations at the C-PCM-B3LYP/6-31G(d)//B3LYP/6-31G(d) level of theory predict that (i) the alkylation reactions in water at the N nucleophilic centers of cytosine (NH_2 and N3) are better modeled by an uncatalyzed process, where the solvent acts mainly through its bulk properties. In contrast, oxygen alkylation (at O^2), which is much more susceptible to H-bonding assistance, is slightly better described by a water-catalyzed process, where the water molecule plays an *active* role, transferring the proton from the NH_2 group to the *o*-QM oxygen atom (Table 4).

Our data also suggest that (ii) nucleophilic addition of the cytosine oxygen atom to *o*-QM, although kinetically competitive in gas phase, is not an experimentally observable process in water due to both kinetic and thermodynamic reasons. As far as intramolecular chemoselectivity, among all the possible mechanisms, our calculations unambiguously suggest that the most nucleophilic site both in gas phase and in solution is the heterocyclic nitrogen atom (N3) for both cytosine and 1-methylcytosine. Our computational results suggest that both uncatalyzed and active water-catalyzed reactive models (for N- and O-centered nucleophiles, respectively) could be useful in a thorough evaluation (under neutral conditions) of DNA base alkylation by soft and polarizable electrophiles such as QM-like structures.

Acknowledgment. Financial support from Pavia University ("Fondo Giovani Ricercatori 1999") is gratefully acknowledged. We also thank the CICAIA (Modena University) and CINECA (Bologna University) for computer facilities and Prof. Remo Gandolfi for useful discussions and advice.

Supporting Information Available: Electronic energies and Cartesian coordinates of all stationary points, in the gas phase, optimized at the B3LYP/6-31G(d) level of theory (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA028732+