Theor Chem Account (2009) 122:179–188
DOI 10.1007/s00214-008-0497-5

REGULAR ARTICLE

Density functional study of the influence of C5 cytosine substitution in base pairs with guanine

Adam Moser · Rebecca Guza · Natalia Tretyakova · Darrin M. York

Received: 15 August 2008 / Accepted: 2 December 2008 / Published online: 24 December 2008

Abstract The present study employs density-functional electronic structure methods to investigate the effect of chemical modification at the C5 position of cytosine. A series of experimentally motivated chemical modifications are considered, including alkyl, halogen, aromatic, fused ring, and strong σ and π withdrawing functional groups. The effect of these modifications on cytosine geometry, electronic structure, proton affinities, gas phase basicities, cytosine–guanine base pair hydrogen bond network and corresponding nucleophilicity at guanine are examined. Ultimately, these results play a part in dissecting the effect of endogenous cytosine methylation on the reactivity of neighboring guanine toward carcinogens and DNA alkylating agents.

1 Introduction

Methylation at the C5 position of cytosine (^{Me}C) is an important endogenous nucleobase modification in mammalian genomic DNA [1–3] found in approximately 1% of total bases in mammalian genome [4]. This epigenetic modification, which is a conserved change to DNA [5], influences chromatin structure [6–8] and mediates gene expression [9]. Many tumors exhibit altered methylation patterns, leading to activation of proto-oncogenes and silencing of tumor suppressor genes [10]. Endogenously methylated CG within the coding region of the *p53* tumor suppressor gene [11] are known lung cancer mutational hotspots [12, 13]. The presence of mutational hotspots at methylated CG dinucleotides can be caused by an increased chemical reactivity toward carcinogens.

It has been shown that ^{Me}C can modulate the reactivity of neighboring guanine toward carcinogens and DNA alkylating drugs [14–23] and its influence depends on the attacking species. Guanine reactivity toward the tobacco carcinogen benzo[a]pyrene [24], which is metabolized into the reactive diolepoxide (+)-*anti*-7*r*,8*t*-dihyrdoxy-*c*9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene (BPDE), is doubled upon cytosine methylation [25]. Another prominent tobacco carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), exhibits reduced reactivity at methylated CG dinucleotides [23, 26].

While the ability of ^{Me}C to influence the reactivity of neighboring guanine is well established, the question of how such a relatively small chemical modification on cytosine can influence specific adduct yields at the neighboring guanine bases remains unanswered. As illustrated in Fig. 1, C5 methylation can influence the susceptibility of neighboring guanine bases toward carcinogen attack by modifying DNA structure, pre-covalent binding in the DNA grooves, intercalation, and local structure and electronics at the ^{Me}CG step. Further, methylation is known to affect the conformations of the resulting DNA adducts, which may play a role in repair [27, 28]. To paint a complete mechanistic picture that will provide insight into carcinogenesis, all these steps must be considered.

One strategy uniquely suited to elucidate this multi-step process is to compare C5 cytosine analogs to ^{Me}C. In a study using mitomycin C, fluorine was used in place of the methyl substituent [18]. This electron withdrawing group

A. Moser · D. M. York (✉)
Department of Chemistry, University of Minnesota,
207 Pleasant St. SE, Minneapolis, MN 55455-0431, USA
e-mail: york@umn.edu

R. Guza · N. Tretyakova
Department of Medicinal Chemistry and the Cancer Center,
University of Minnesota, Minneapolis, MN 55455, USA

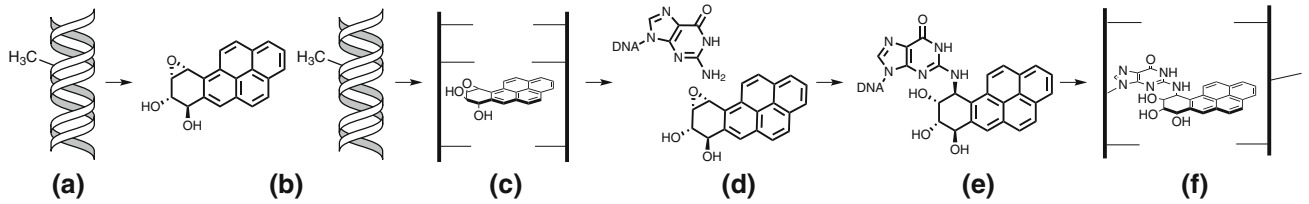

Fig. 1 General steps involved in BPDE reaction with DNA containing 5-methyl cytosine. **a** DNA duplex containing C5-methyl group in the major groove. **b** BPDE pre-covalently binds at the DNA grooves. **c** BPDE intercalates into the DNA duplex. **d** BPDE reaches a reactive conformation with the exocyclic amine of guanine. **e** BPDE reacts at N^2 of guanine forming a covalent adduct. **f** The adduct reaches a stable final conformation

gave insight into the importance of electronic effects on reactivity. Another study showed that C5 halogen substituted cytosine and uracil can mimic methylation and affect ^{Me}C signals observed in tumors [29]. C5 cytosine analogs have also been studied computationally [30, 31]. However, one cannot unambiguously interpret the effects of a single, or even a few, chemical modifications. In fact, the complexity of the general mechanism of DNA adduct formation, as depicted in Fig. 1 with BPDE, requires a very broad range of analogs that exhibit distinct, systematic chemical trends to make any definitive statements about ^{Me}C role in individual reaction steps.

Theoretical methods provide a powerful tool to aid in the interpretation of experimental data and to describe the effects of chemical modifications, including their effect on local electronic structure properties of individual C5 substituted bases [30], base pairs [31–35], and their more global effect on the helical base stack [36–39]. These factors will influence the biologically relevant chemical reactivity. In the present work, we undertake the study of a systematic series of cytosine analogs shown in Fig. 2, with the goal of providing insight into the electronic structure properties of the C5 modified cytosine bases and their base pairs with protonated and unprotonated guanine. Modifications include alkyl, alkenyl and alkynyl chains, halogens, aromatic, fused ring, and strong σ and π withdrawing and electron donating functional groups. This information will be useful, ultimately, in arriving at a consensus view of methylated cytosine's role in mediating DNA reactivity with carcinogens and drugs.

2 Methods

All calculations were performed in accord with the standardized protocol, previously detailed [40], used to construct the *QCRNA* database, a recently developed online database of quantum calculations for RNA catalysis [41]. In brief, all structures were optimized in the gas phase with with B3LYP/6-31++G** as implemented in the Gaussian03 suite of programs [42]. Frequency calculations at the optimized geometries were performed to establish the nature of all stationary points and used to calculate thermodynamic quantities. Electronic energies and other properties of the density were further refined via single point calculations at the optimized geometries using the 6-311++G(3df,2p) basis set. Solvation effects were treated by single-point calculations based on the gas phase optimized structures using the polarizable continuum model (PCM) [43–45] and a variation of the conductor-like screening model (COSMO) [46] with the 6-311++G (3df,2p) basis set.

The molecular enthalpies and free energies are based on the refined gas phase single point energy calculations and along with the frequency analysis are used to obtain the proton affinity (PA) and gas phase basicity (GPB). As discussed in previous work [47, 48], pK_a values are calculated using a standard thermodynamic cycle (e.g. Scheme 1(a) in 49) to give an equation for the pK_a.

$$\mathrm{p}K_a = \frac{\log(e)}{RT}\left[\Delta G^{\circ}_{gas} - \Delta G^{\circ}_{solv}(HA) + \Delta G^{\circ}_{solv}(H^+) + \Delta G^{\circ}_{solv}(A^-)\right] \quad (1)$$

The solvation free energy of the proton, $\Delta G_{solv}{}^{\circ}$ (H^+), is −265.87 kcal/mol as determined by Tissandier et al. [50, 51]. Relative pK_a values (ΔpK_a) are always given in the following form

$$\Delta\mathrm{p}K_a = \mathrm{p}K_a(\text{analog}) - \mathrm{p}K_a(\text{native}) \quad (2)$$

No counterpoise corrections were calculated to correct for basis set superposition errors in the case of the hydrogen bonded base pairs. At the basis set levels used in the present work, these corrections have been shown to be fairly small [31, 52], with differences in relative values typically on the order of 0.1 kcal/mol.

For single base calculations, the N^1 proton was substituted with a methyl group to provide a more realistic model of the base connected to a deoxyribose for the single base cytosine calculations [31]. Substituted Cytosine:Guanine base pairs were calculated in their Watson–Crick hydrogen bonding scheme (Fig. 3). The same base pairs were

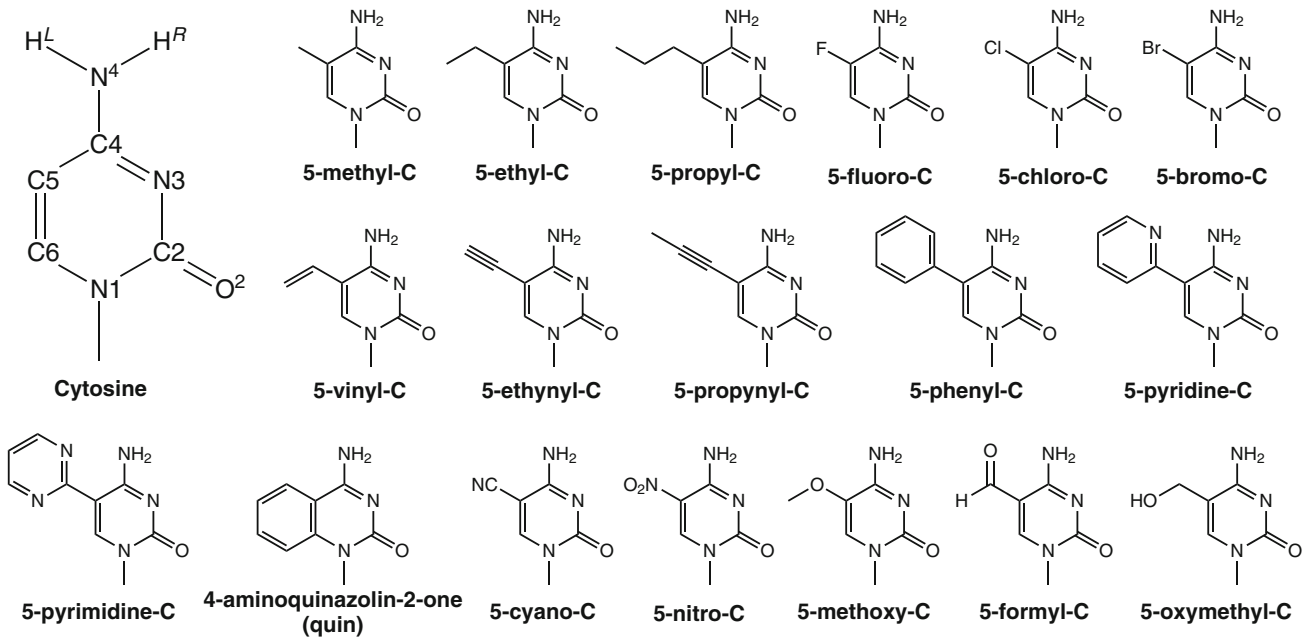

Fig. 2 Numbered cytosine (*top left*) along with cytosine analogs with substitution at the fifth position

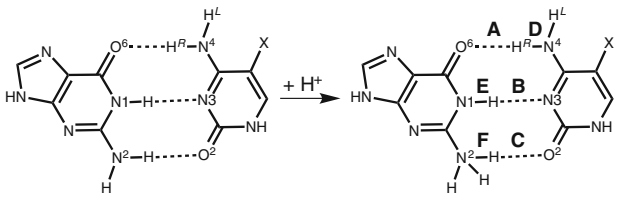

Fig. 3 Watson–Crick GC base pair and protonated N^2 CG basepair with hydrogen bond lengths labeled

calculated with the N^2 of guanine protonated, resulting in a quaternary amino group. All base pair structures retain a coplanar purine and pyridine without constraint.

3 Results

The series of cytosine substitutions were chosen as follows. First, consecutively longer alkyl chains were chosen (methyl, ethyl, and propyl) to investigate the possibility that hydrophobic groups may increase carcinogen binding at analog sites, thereby increasing the possibility of reaction at that position. Conversely, there is a possibility larger substituents may begin to impede carcinogens from reaching the reaction site. Secondly, we consider a series of halogens (fluoro, chloro, and bromo). These σ-electron withdrawing substituents can help probe the electronic effects with fairly minimal, systematic steric effects that are complementary to the σ-electron donating alkyl groups. Thirdly, vinyl, ethynyl, and propynyl added to the list as possible intercalation enhancers. Originally, phenyl was chosen as an aromatic intercalation enhancer, but preliminary calculations indicated the minimum energy conformation of 5-phenyl-C does not lie in the plane of the base pair due to a steric clash between the H_L amino hydrogen and a nearby hydrogen on the phenyl ring (the in-plane conformer is $\sim$9 kcal/mol higher in free energy). Therefore, fourthly, in addition to phenyl, we considered the set of 2-pyrimidine, 2-pyridine, and 4-aminoquinazolin-2-one (quin) to determine how the relative torsion of these larger aromatic substituents may effect pre-covalent binding and intercalation. Fifthly, we consider strong electron withdrawing and donating groups, including a strong σ-withdrawing cyano, strong π-withdrawing nitro and formyl, and electron donating methoxy and oxymethyl group to further probe electronic effects. All analogs are shown in Fig. 2.

The first step in unraveling the effects of these analogs is to consider in detail their effects on individual bases and base pairs. Since cytosine is not covalently bonded to guanine, any electronic or structural effect that influences carcinogen reactivity is likely to occur through the hydrogen bond network, which is known to be cooperative [53]. The results that follow focus on the atoms involved in GC base pair hydrogen bonding. Results are separated into three subsection: cytosine base, cytosine:guanine base pair, and cytosine:guanine base pair with protonated N^2 guanine.

All calculation have been deposited in the *QCRNA* database [41] and can be downloaded, viewed, and manipulated on-line.

3.1 Cytosine base

Initially, the effect on cytosine base geometry was considered to determine if functionalization at C5 makes any significant changes to base structure. Bond length and angle changes upon substitution were fairly negligible, less that 0.01 Å and 3°, respectively. The results agree with previously reported structural changes by Dannenberg and Tomasz [32]. The pyridine, pyrimidine, and phenyl analogs were chosen in particular for future work to investigate steric hindrance during intercalation. The absolute minimum dihedral angle is defined by C4–C5–C_a–X, where C_a is the substituent carbon attached to the C-5 of cytosine and X is bonded to C_a and gives the lowest absolute dihedral value. The dihedrals for the absolute minimum energy conformers are 0.0°, 21.9°, and 57.3° for pyrimidine, pyridine, and phenyl, respectively.

Table 1 compares atomic charges, total charge of the substituent, dipole moment, and polarizability for each analog. The total substituent charge is defined as the sum of the ChelpG atomic charges [54] of all the atoms that make up the substituent. Overall, the substituents introduce mild changes to the atomic charges compared to the native cytosine. These changes reduce in magnitude the further the atom is from the substituent.

Substituents at the fifth position of cytosine project out from major groove and depending on their length, charge, and polarizability, may influence pre-covalent binding of carcinogen metabolites. The fluoro, propynyl, quin, and nitro analogs all have negative total charges of at least 0.1 electrons, while ethyl, propyl, phenyl, pyridine, and pyrimidine have positive total charges of at least 0.1 electrons. All analogs show a decrease in total charge compared to the unsubstituted cytosine's hydrogen (0.174 *e*). All analogs show an increase in total polarizability, and for analogs with aromatic substituents, there is a marked increase in polarizability.

In unsubstituted cytosine, the molecular dipole points along the C2–O^2 bond, which is generally true for the analogs though it can be slightly modified. In the CG base pair this dipole would be directed at the N^2 of guanine, so changes in dipole may play a key role in understanding reactivity changes at this position. The magnitude of the dipole moment is significantly reduced for nitro (72%) and cyano (67%) substituents and amplified considerably for methoxy (18%), pyridine (22%), and pyrimidine (22%) substituents. The total isotropic polarizability is increased for all analogs compared to native cytosine, except for fluorine.

Table 2 provides the PA and GPB for the four protonation/deprotonation sites: protonation at the O^2 and the N3 and deprotonation of the left (H_L) and right (H_R) N^4 protons (see Fig. 2). Analogs show a broad range of effect on the PA and GPB values relative to the native cytosine ($\sim$20 kcal/mol). Trends between analogs are generally conserved between the PA and GPB.

Table 1 Partial atomic charges of selected atoms (*e*), molecular dipole (Debye), and isotropic polarizability (Å^3) for cytosine analogs

Substituent	H_L	H_R	N^4	C4	N3	O^2	C2	Sub.[a]	Dipole	Pol.
Cytosine	0.363	0.385	−0.872	0.848	−0.778	−0.614	0.858	0.174	6.2079	13.00
Methyl	0.347	0.371	−0.804	0.689	−0.736	−0.620	0.835	0.058	6.6209	14.88
Ethyl	0.335	0.371	−0.812	0.743	−0.755	−0.622	0.850	0.160	6.8450	16.61
Propyl	0.342	0.374	−0.822	0.746	−0.751	−0.620	0.835	0.142	6.9513	18.46
Fluoro	0.384	0.398	−0.848	0.689	−0.720	−0.605	0.824	−0.192	4.8087	12.97
Chloro	0.378	0.415	−0.900	0.827	−0.774	−0.607	0.866	−0.094	4.6273	14.80
Bromo	0.361	0.415	−0.889	0.861	−0.798	−0.609	0.861	−0.060	4.7239	16.00
Vinyl	0.313	0.362	−0.746	0.640	−0.729	−0.613	0.851	−0.015	5.7886	17.45
Ethynyl	0.370	0.411	−0.882	0.755	−0.753	−0.604	0.846	−0.069	5.2418	17.04
Propynyl	0.358	0.400	−0.840	0.680	−0.728	−0.604	0.821	−0.104	6.5634	19.49
Phenyl	0.280	0.356	−0.723	0.717	−0.765	−0.616	0.860	0.113	6.4332	23.97
Pyridine	0.378	0.390	−0.881	0.853	−0.795	−0.614	0.862	0.151	7.5697	24.10
Pyrimidine	0.410	0.422	−0.979	0.953	−0.829	−0.608	0.880	0.168	7.5434	23.51
Quin	0.313	0.364	−0.738	0.622	−0.730	−0.609	0.904	−0.045	6.1922	20.01
Cyano	0.414	0.429	−0.943	0.846	−0.764	−0.587	0.865	−0.061	2.0792	15.68
Nitro	0.450	0.429	−0.947	0.841	−0.766	−0.574	0.850	−0.149	1.7085	16.25
Formyl	0.435	0.411	−0.925	0.874	−0.774	−0.576	0.814	−0.025	4.2652	15.83
Methoxy	0.349	0.402	−0.842	0.706	−0.751	−0.623	0.858	−0.078	7.3528	15.59
Oxymethyl	0.354	0.364	−0.812	0.779	−0.770	−0.611	0.850	0.073	7.2132	15.80

[a] Sum of ChelpG atomic charges of all substituent atoms not including the C5 of cytosine

Table 2 Proton affinity, gas phase basicity, and relative pK_a at various positions of cytosine

Substituent	Proton affinity				Gas phase basicity				ΔpK_a^{PCM}		ΔpK_a^{COSMO}	
	$O^2(+)$	N3(+)	$N^4{}_L(-)$	$N^4{}_R(-)$	$O^2(+)$	N3(+)	$N^4{}_L(-)$	$N^4{}_R(-)$	$O^2(+)$	N3(+)	$O^2(+)$	N3(+)
Cytosine	231.8	232.4	350.8	355.6	223.6	224.5	342.9	347.5	0.0	0.0	0.0	0.0
Methyl	235.1	234.7	349.7	355.4	227.1	227.0	341.7	347.2	0.9	0.7	1.0	0.7
Ethyl	236.3	235.9	349.5	355.7	228.2	228.1	341.3	346.9	1.1	0.8	1.2	0.9
Propyl	236.8	236.4	349.5	355.7	228.7	228.7	341.3	347.1	1.1	0.8	1.2	0.9
Fluoro	228.6	226.5	347.0	349.5	219.7	217.8	340.1	342.4	−1.5	−2.9	−1.3	−2.9
Chloro	228.5	227.4	346.4	349.4	220.0	219.2	338.9	341.6	−1.7	−2.7	−1.5	−2.5
Bromo	228.8	227.8	345.9	349.0	220.6	219.9	338.0	340.9	−1.4	−2.6	−1.1	−2.5
Vinyl	233.2	233.2	346.8	352.8	225.4	225.8	338.6	345.0	0.0	−0.1	0.2	0.0
Ethynyl	229.4	229.8	348.6	351.1	221.2	221.7	339.8	342.8	−2.4	−2.3	−2.0	0.2
Propynyl	233.7	233.9	351.0	353.6	225.9	226.4	342.9	345.3	−1.0	−1.1	−0.8	−1.0
Phenyl	235.7	236.2	347.6	353.2	227.9	228.8	339.5	345.5	−0.1	−0.1	0.1	0.0
Pyridine	236.6	239.1	346.7	358.2	228.8	231.6	338.2	349.4	−0.5	0.3	−0.4	0.2
Pyrimidine	235.6	238.1	357.4	357.6	227.5	230.2	349.3	348.5	−1.2	−0.4	−1.2	−0.6
Quin	234.9	234.2	342.9	349.5	227.0	226.5	335.6	342.1	0.2	0.2	0.4	0.3
Cyano	219.4	219.8	340.1	343.3	211.3	211.9	332.2	335.2	−4.8	−4.8	−4.6	−4.6
Nitro	217.6	218.9	345.1	343.9	209.7	211.2	336.6	335.5	−5.9	−5.6	−5.5	−5.3
Formyl	223.1	225.9	351.8	350.5	215.2	218.1	342.8	342.1	−4.1	−3.0	−4.1	−3.0
Methoxy	237.0	234.5	352.9	355.0	228.4	226.2	345.9	347.7	0.8	−0.9	0.9	−0.9
Oxymethyl	234.5	235.5	341.0	356.8	226.7	227.9	333.2	347.7	0.4	0.5	0.0	0.2

Proton affinities and gas phase basicities are reported in kcal/mol. Relative pK_a values (ΔpK_a) are relative to the native cytosine, see Eq. 2

The O^2 PA and GPB increase for alkyl, vinyl, propynyl, aromatic compounds, and methoxy. Of specific interest is the large decreases for both cyano and nitro. The same trend is seen for N3 protonation and given the cooperativity of the hydrogen bonds might also be important to carcinogen reactivity. Similarly, changes to the $N^4{}_R$ PA and GPB may affect hydrogen bonding to guanine due to its proximity to the substituent. Deprotonation of $N^4{}_L$ is also shown.

Relative pK_a values are summarized in Table 2 for the protonation O^2 and N3, calculated using Eq. 1. Neither the PCM nor COSMO model could give reasonable solvation energies for the anionic nucleobases, so pK_a values for the deprotonation at N^4 are not shown. The PCM and COSMO models give very similar results and only slightly different trends for the protonation of O^2 and N3. As a guide to the absolute accuracy of these numbers, Zhang and Mathews [55] cites the cytosine pK_a for the N3 native and methylated analog protonation as 4.45 and 4.6, respectively. Our calculations for the same analogs give 6.29 and 6.98, respectively. Sowers [56] references a 1.8 pK_a drop between 5-fluoro-C and cytosine, while we show a 2.9 pK_a decrease. While the absolute pK_a difference is $\sim$2 pK_a units, the relative values that we are relying on are within $\sim$1.0 pK_a units.

The alkyl, vinyl, propynyl, aromatic, and methoxy analogs all produce minimal pK_a changes for both N3 and O^2 ($\leq$1.2 pK_a units). Of the remaining substituents, the pK_a values decrease for both N3 and O^2 by as much as 5.9 pK_a units. These pK_a decreases indicate that the halogen, ethynyl, cyano, and nitro analogs are decreased hydrogen bond acceptors at N3 and O^2 compared to unmodified cytosine.

3.2 CG base pair

Hydrogen bond lengths and angles for the modified CG base pairs were determined. Similar to the single base calculations, no significant structural deviations are found between base pairs containing C5-cytosine analogs and cytosine. Cyano and nitro analogs show a slight lengthening in the cytosine O^2–guanine N^2 hydrogen bond along with a 1°–2° change in the angle.

A summary of the binding energy for each analog with and without solvation is shown in Table 3. The table breaks down the contribution into electronic energy, enthalpy, entropic contribution, and free energy in the gas phase, and in aqueous solution using both PCM and COSMO solvation models. The binding enthalpy for the native CG pair in the present study (−23.1 kcal/mol) is reasonably close to both the experimental value of Yanson an co-workers [57] of −21.0 kcal/mol as well as various theoretical calculations [58–60] including the value of −23.8 kcal/mol reported by Meng et al. [31].

The gas phase binding enthalpy values span a range of 3.5 kcal/mol. The trends are similar, but smaller in

Table 3 GC base pair binding electronic energy, enthalpy, entropic contribution, free energy, and free energy with solvation corrections

Substituent	ΔE	ΔH	$-T\Delta S$	ΔG	ΔG_{aq}^{PCM}	ΔG_{aq}^{COSMO}
Cytosine	−24.8	−23.1	12.0	−11.2	10.8	11.0
Methyl	−25.3	−23.6	12.1	−11.5	10.5	10.4
Ethyl	−25.3	−23.7	11.8	−11.9	10.5	10.5
Propyl	−25.4	−23.7	11.8	−12.0	10.5	10.6
Fluoro	−24.5	−22.8	12.2	−10.6	11.3	11.4
Chloro	−24.2	−22.5	12.0	−10.4	11.3	11.4
Bromo	−24.1	−22.4	12.3	−10.1	11.6	11.8
Vinyl	−24.5	−22.9	11.7	−11.2	10.5	10.7
Ethynyl	−23.8	−22.2	12.0	−10.2	11.4	11.6
Propynyl	−24.3	−22.7	11.3	−11.4	10.6	10.7
Phenyl	−24.7	−23.0	11.9	−11.1	11.0	11.2
Pyridine	−23.4	−21.7	11.8	−9.9	11.2	11.3
Pyrimidine	−23.5	−21.8	11.9	−9.9	11.4	11.5
Quin	−25.1	−23.6	11.7	−11.9	10.2	10.5
Cyano	−22.8	−21.2	11.8	−9.4	11.7	11.7
Nitro	−21.8	−20.3	11.4	−8.9	11.7	11.7
Formyl	−22.0	−20.3	11.5	−8.9	11.5	11.5
Methoxy	−25.1	−23.4	12.2	−11.4	10.9	10.5
Oxymethyl	−23.8	−22.3	11.6	−10.7	10.5	10.2

All values are give in kcal/mol

magnitude for the gas phase binding free energy. Solvation, which disfavors the nucleobase association, decreases this range to 1.6 kcal/mol. In the gas phase, the alkanes, propynyl, quin, and methoxy analogs all show increased binding for the GC base pair, while the halogens, ethynyl, pyridine, pyrimidine, formyl, oxymethyl, cyano, and nitro show decreased binding. This trend of electron withdrawing substituents destabilizing binding energy is consistent with work by Kawahara et al. [52]. The phenyl and vinyl analogs show negligible change to free energy. Due to the overall small changes in binding free energy, it is likely that the observed changes in melting temperature of DNA helices with these analogs [61] are only mildly affected by changes in base pair hydrogen bonding, and that other contributions, such as base stacking and solvation, are more significant.

3.3 Protonated CG base pair

Table 4 gives the CG base pair geometry with the guanine N^2 protonated (Fig. 3). As discussed below, this protonation is a model reaction for electrophilic carcinogen attack. These structures are planar, first order transition states as discussed by Dannenberg and Tomasz [32] with an imaginary frequency attributed to inversion about the

Table 4 GC base pair hydrogen bonding geometry with N^2 protonated guanine

Substituent	Bond length						Angle		
	A	D	B	E	C	F	A/D	B/E	C/F
Cytosine	1.751	1.043	1.875	1.027	1.033	1.676	178.8	178.7	170.5
Methyl	1.750	1.042	1.879	1.028	1.029	1.689	179.4	178.6	170.5
Ethyl	1.753	1.042	1.881	1.028	1.028	1.692	179.7	178.6	170.5
Propyl	1.754	1.041	1.882	1.028	1.027	1.695	179.6	178.5	170.4
Fluoro	1.856	1.031	1.773	1.045	1.394	1.154	172.2	179.7	171.8
Chloro	1.849	1.031	1.780	1.046	1.387	1.157	173.8	179.6	171.8
Bromo	1.846	1.030	1.779	1.046	1.383	1.159	174.1	179.7	171.8
Vinyl	–	–	–	–	–	–	–	–	–
Ethynyl	1.862	1.030	1.773	1.047	1.375	1.164	173.1	180.0	171.6
Propynyl	1.762	1.041	1.876	1.028	1.029	1.688	178.1	178.8	170.6
Phenyl	1.764	1.040	1.883	1.028	1.028	1.692	178.9	178.7	170.3
Pyridine	–	–	–	–	–	–	–	–	–
Pyrimidine	1.794	1.037	1.893	1.028	1.030	1.681	179.2	179.1	170.3
Quin	1.733	1.044	1.882	1.029	1.026	1.696	179.1	178.5	169.8
Cyano	1.845	1.031	1.783	1.046	1.458	1.123	173.5	179.2	172.5
Nitro	1.847	1.031	1.804	1.044	1.466	1.119	175.0	178.9	172.8
Formyl	1.878	1.029	1.782	1.047	1.444	1.129	172.9	179.6	172.3
Methoxy	1.758	1.041	1.874	1.028	1.026	1.705	177.1	178.7	170.6
Oxymethyl	–	–	–	–	–	–	–	–	–

Bond lengths are given in Å. Angles are given in degrees

Bonds and angles correspond to Fig. 3

exocyclic, protonated N^2 amino group. Fully optimized structures of the protonated base pair are not planar in the gas phase making it difficult to compare to the planar bases within a DNA helix, therefore the analysis was restricted to the first order transition, planar structures. Values for vinyl and pyridine analog are not reported as the planar transition state structures were not found.

Compared to the CG base pairs, the protonated CG base pairs have much larger deviations from the native geometric values. The most interesting change observed is for the cytosine O^2 acceptor and guanine N^2 donor. For the alkyl, propynyl, aromatic, and methoxy analogs, the proton has completely transferred from the guanine amino group to the cytosine carbonyl. This proton transfer is very important as it may mimic proton movement during N^2-guanine alkylation by carcinogens at the guanine N^2.

PA, GPB, and relative pK_a values for the protonation of GC base pair at the guanine N^2 position are summarized in Table 5. Similar to the single cytosine results, the addition of solvation makes minimal changes to the analog trends and decreases the overall range of the deviation from the native GC base pair. The PA and GPB correlate well with the protonation of single cytosine O^2, see Table 2. Halogen, ethynyl, cyano, and nitro substituents lead to decreased PA and GPB, while alkanes, propynyl, aromatics, and methoxy increase PA and GPB. It should be noted that given the correlation between the single cytosine and base pair calculations, it is likely that vinyl and pyridine increase PA and GPB. Dannenberg and Tomasz [32] reported the GPB for fluoro and methyl as −2.7 and 2.6 kcal/mol, respectively, which compare well to our −3.6 and 2.5 kcal/mol.

Table 5 Proton affinity (ΔH), gas phase basicity (ΔG) and relative pK_a (ΔpK_a) of N^2 guanine in the GC base pair

Substituent	(ΔH)	(ΔG)	ΔpK_a^{PCM}	ΔpK_a^{COSMO}
Cytosine	224.4	213.9	0.0	0.0
Methyl	226.7	216.4	0.4	0.3
Ethyl	227.6	217.2	0.7	0.7
Propyl	228.1	217.7	0.8	0.9
Fluoro	220.9	210.3	−1.6	−1.6
Chloro	221.1	210.4	−1.9	−1.9
Bromo	221.4	211.1	−1.7	−1.6
Vinyl	–	–	–	–
Ethynyl	222.1	211.5	−2.1	−2.0
Propynyl	226.2	216.7	−0.1	0.1
Phenyl	227.5	217.2	−0.1	0.1
Pyridine	–	–	–	–
Pyrimidine	228.1	217.8	−0.6	−0.6
Quin	227.2	216.7	−0.1	0.1
Cyano	215.8	205.2	−3.2	−3.2
Nitro	215.1	204.3	−3.7	−3.5
Formyl	218.8	208.1	−3.0	−3.0
Methoxy	228.3	217.8	0.9	0.5
Oxymethyl	–	–	–	–

Proton affinity and gas phase basicity are given in kcal/mol. Relative pK_a values are calculated as the analog pK_a minus the cytosine pK_a, see Eq. 2

4 Discussion

To determine how cytosine methylation affects reactivity of neighboring guanine, we consider likely ways the extra methyl group might influence each reaction step. This is particularly difficult given that various carcinogens may have significantly different mechanisms. In this study, we have analyzed the geometric and electronic properties of both the single cytosine and CG base pair with a variety of C5 substituents and protonation states. We can apply this knowledge to the possible reaction steps shown in Fig. 1.

Geometric properties of cytosine and CG base pair analogs show minimal change compared to the native structures, likely indicating little impact on the local helical structure. Further, the C5 substituents protrude from the major groove, which is up to 12 Å wide for B DNA [62], and therefore are unlikely to sterically distort the helix. On the other hand, the polarizability was increased by as much as a 100% upon C5 functionalization. While some polarizability of cytosine lies outside the base stacking region, this enhancement likely influences base stacking observed experimentally [63].

Pre-covalent binding of reactive species to DNA likely precedes the chemical reaction, so substituent effects on the binding interaction will be a key factor in understanding the overall reaction. Cytosine analogs show a range of 13.0–24.1 $Å^3$ in polarizability and −0.192 to 0.168 *e* for total substituent charge (see Table 1). The range of the total substituent charges goes from electronically negative to electronically positive with a range of polarizabilities signifying each analogs creates a unique local electrostatic environment that may alter solvation and binding affinity. Further, some of the substituents protrude as much as 4.3 Å farther out of the major groove than the native cytosine hydrogen, indicating this electronic environment is well presented to a possible reactive species.

For some carcinogens, such as BPDE, intercalation is known to occur prior to chemical reaction [64]. The ability for a carcinogen to intercalate depends on the stacking ability of the bases and the ease of entry for the intercalator into the helix. If stacking ability is proportional to polarizability, as suggested previously [65], then all of the analogs considered here (except for fluoro) should show increased base stacking. While some C5 substituents are

quite flexible (e.g. propyl) others are conformationally constrained (e.g. propynyl). The flexibility may thus influence the steric barrier to intercalation. The pyrimidine, pyridine, and phenyl analogs were chosen specifically to explore the intercalation enhancement along with changing steric barrier. If enhanced polarizability correlates with increased pre-covalent retention times and reactivity and steric barrier to intercalation plays no role, then the pyridine, pyrimidine, and phenyl, quin analogs will show significant reactivity increases. If instead the relative reactivity is shown to follow quin > pyrimidine > pyridine > phenyl, then steric obstruction to intercalation is likely occurring.

Experimental results for the reactivity changes of mytomycin C using the fluoro and methyl analogs have been attributed to be a transmitted electronic effect from the substituent through the hydrogen bonds to the guanine N^2 amino group [18]. In essence, the substituent donates or withdraws its nucleophilicity through the cytosine O^2 to the guanine N^2 resulting in an increase or decrease in reactivity toward electrophiles. Previous calculations [32] used protonation of the guanine N^2 in the CG base pair as an indicator of the strength of this effect. These results showed methyl favoring protonation by 2.6 kcal/mol and fluoro disfavoring the reaction by 2.7 kcal/mol compared to the native cytosine free energy. We see similar results, 2.5 and 3.6 kcal/mol for methyl and fluoro, respectively.

If this electronic effect plays the dominant role in the reaction, then by inference, the present results would suggest the following order of analog reactivity toward electrophiles:

$$\begin{aligned} &methoxy, pyridine, pyrimidine, propyl, phenyl, ethyl > \\ &quin, propynyl, oxymethyl, methyl, vinyl > \\ &native > \\ &ethynyl, bromo > chloro, fluoro > formyl > cyano, nitro \end{aligned}$$

(see Table 5). We have grouped substituents with similar values based on the errors known for the QCRNA protocol (the standard deviation for relative GPB values is ~1 kcal/mol) [47, 48]. The placement of vinyl, oxymethyl, and pyridine was estimated from the single cytosine gas phase basicities, which generally mirrors the base pair GPB. This trend is also seen in the work by Dannenberg and Tomasz [32].

The electronic effect is also consistent with the changes in hydrogen bond distances for the protonated GC. Figure 4 plots the base pair GPB versus the hydrogen bond distance between the cytosine O^2 and the proton within the hydrogen bond to the guanine N^2. Results indicate that substituents that donate nucleophilicity enhance this proton transfer in the protonated base pair, while electron withdrawing substituents will disfavor proton transfer. In

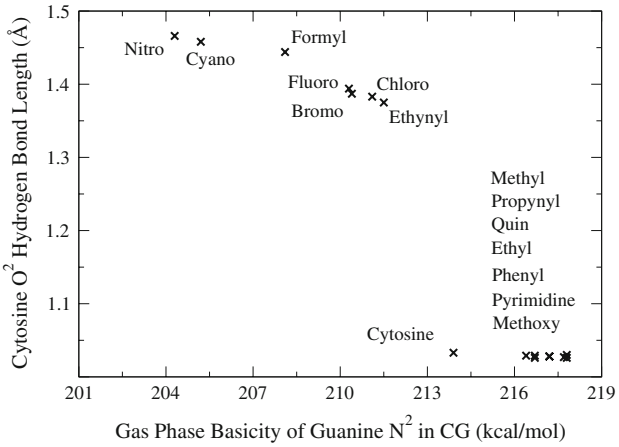

Fig. 4 Gas phase basicity and hydrogen bond distance between cytosine O^2 and guanine N^2 after protonation

summary, if the electronic effect is a significant factor in the chemical steps of the reaction, analogs that can donate nucleophilicity through the hydrogen bonds should increase the GPB, facilitate proton transfer to the cytosine carbonyl, and increase reactivity at the guanine amine.

In order to fully probe chemical mechanism with theoretical methods, for even a particular carcinogen, a multifaceted approach is required that includes molecular simulation, quantum chemical calculations, and correlation with experimental data. Nonetheless, a first step toward providing insight into reactivity of methylated CG base pairs as dictated by experimental chemical modification of cytosine, is to characterize substitution effects on the electronic structure of the individual base and base pair. The data provided herein establishes a baseline characterization across a broad range of chemical modifications that may help to guide and interpret the results of future experiments.

5 Conclusion

Cytosine methylation at C5 is an important endogenous base modification that influences gene expression [9] and is correlated to known lung cancer mutational hot spots in the *p53* tumor suppressor gene [11]. While it is widely known that this relatively small chemical modification significantly changes reactivity of the base paired guanine toward electrophiles, the actual mechanism is both unknown and likely composed of many steps. We have proposed a variety of C5 cytosine substitutions chosen specifically to elucidate the relative importance of likely reaction step.

In this work, we used density functional theory to calculate various cytosine analog properties of both the single cytosine bases and in the CG base pairs. These properties include hydrogen bonding structure, atomic charges,

polarizability, PA, GPB, and pK_a. Our results show that the cytosine analogs present a distinct electronic environment to an incoming carcinogen, which will likely influence the pre-covalent major groove binding. Further, previous work suggested that reactivity changes are based on an electronic effect that is transferred from the substituent through the cytosine carbonyl to the guanine amino group. We have quantified this effect for our analogs such that future experimental work can conclude the importance of this electronic effect. As whole, these results provide a guide to future interpretation into the relative importance of the reaction steps required for carcinogen and drug reactivity at methylated CG steps.

Acknowledgments D. Y. is grateful for financial support provided by the National Institutes of Health (Grant GM62248) and the University of Minnesota Biomedical Informatics and Computational Biology program, and NT is grateful for financial support provided by National Cancer Institute (CA-100670). AM is grateful for support through a Department of Energy NDSEG fellowship, and RG thanks the National Institutes of Health for support through the Chemistry–Biology Interface Training Grant (T32-GM08700). Computational resources were provided by the Minnesota Supercomputing Institute.

References

1. Jost J-P, Hofsteenge J (1992) Proc Natl Acad Sci USA 89:9699
2. Holman MR, Ito T, Rokita SE (2007) J Am Chem Soc 129:6
3. Ehrlich M, Gama-Sosa MA, Huang L-H, Midgett RM, Kuo KC, McCune RA, Gehrke C (1982) Nucleic Acids Res 10:2709
4. Gal-Yam EN, Saito Y, Egger G, Jones PA (2008) Annu Rev Med 59:267
5. Bird A (2007) Nature 447:396
6. Romanov GA, Zhavoronkova EN, Savel'ev SV, Vanyushin BF (1983) Neurosci Behav Physiol 16:285
7. Yu D, Berlin JA, Penning TM, Field J (2002) Chem Res Toxical 15:832
8. Hausheer FH, Rao SN, Gamcsik MP, Kollman PA, Colvin OM, Saxe JD, Nelkin BD, McLennan IJ, Barnett G, Baylin SB (1989) Carcinogenesis 10:1131
9. Bird AP (1986) Nature 321:209
10. Riggs AD (1983) Adv Cancer Res 40:1
11. Denissenko MF, Chen JX, Tang M-S, Pfeifer GP (1997) Proc Natl Acad Sci USA 94:3893
12. Hoffmann D, Hoffmann I (1997) J Toxicol Environ Health 50:307
13. Hecht SS (1999) J Natl Cancer Inst 91:1194
14. Sendowski K, Rajewsky MF (1991) Mutat Res 250:153
15. Mathison BH, Said B, Shank RC (1993) Carcinogenesis 14:323
16. Chen JX, Zheng Y, West M, shong Tang M (1998) Cancer Res 58:2070
17. Weisenberger DJ, Romano LJ (1999) J Biol Chem 274:23948
18. Das A, Tang KS, Gopalakrishnan S, Waring MJ, Tomasz M (1999) Chem Biol 6:461
19. Ross MK, Mathison BH, Said B, Shank RC (1999) Biochem Biophys Res Commun 254:114
20. Li V-S, Reed M, Zheng Y, Kohn H, Tang M-S (2000) Biochemistry 39:2612
21. Pfeifer GP, Tang M, Denissenko MF (2000) Curr Top Micro Biol Immunol 249:1
22. Burdzy A, Noyes KT, Valinluck V, Sowers LC (2002) Nucleic Acids Res 30:4068
23. Rajesh M, Wang G, Jones R, Tretyakova N (2005) Biochemistry 44:2197
24. Hecht SS (2000) J Natl Cancer Inst 92:782
25. Tretyakova N, Matter B, Jones R, Shallop A (2002) Biochemistry 41:9535
26. Ziegel R, Shallop A, Upadhyaya P, Jones R, Tretyakova N (2004) Biochemistry 43:540
27. Zhang N, Lin C, Huang X, Kolbanovskiy A, Hingerty BE, Amin S, Broyde S, Geacintov NE, Patei DJ (2005) J Mol Biol 346:951
28. Rodríguez FA, Cai Y, Lin C, Tang Y, Kolbanovskiy A, Amin S, Patel DJ, Broyde S, Geacintov NE (2007) Nucleic Acids Res 35:1555
29. Valinluck V, Liu P, Kang JI Jr, Burdzy A, Sowers LC (2005) Nucleic Acids Res 33:3057
30. Jang YH, Sowers LC, Çağin T, Goddard WA, III (2001) J Phys Chem A 105:274
31. Meng F, Liu C, Xu W (2003) Chem Phys Lett 373:72
32. Dannenberg JJ, Tomasz M (2000) J Am Chem Soc122:2062
33. Kryachko ES, Nguyen MT (2002) J Phys Chem A 106:9319
34. Sherer EC, York DM, Cramer CJ (2003) J Comput Chem 24:57
35. Giese TJ, Sherer EC, Cramer CJ, York DM (2005) J Chem Theory Comput 1:1275
36. Šponer J, Leszczynski J, Hobza P (2001) J Mol Struct (Theochem) 573:43
37. Řeha D, Kabeláč M, Ryjáček F, Šponer J, Šponer JE, Elstner M, Suhai S, Hobza P (2002) J Am Chem Soc 124:3366
38. Kumar A, Elstner M, Suhai S (2003) Int J Quantum Chem 95:44
39. Dabkowska I, Jurečka P, Hobza P (2005) J Chem Phys 122:204322
40. Giese TJ, Gregersen BA, Liu Y, Nam K, Mayaan E, Moser A, Range K, Nieto Faza O, Silva Lopez C, Rodriguez de Lera A, Schaftenaar G, Lopez X, Lee T, Karypis G, York DM (2006) J Mol Graph Model 25:423
41. QCRNA, http://theory.chem.umn.edu/Database/QCRNA
42. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, revision C.02. Gaussian, Inc., Wallingford
43. Tomasi J, Persico M (1994) Chem Rev 94:2027
44. Cossi M, Barone V, Cammi R, Tomasi J (1996) Chem Phys Lett 255:327
45. Mineva T, Russo N, Sicilia E (1998) J Comput Chem 19:290
46. Barone V, Cossi M (1998) J Phys Chem A 102:1995
47. Range K, Riccardi D, Cui Q, Elstner M, York DM (2005) Phys Chem Chem Phys 7:3070
48. Range K, López CS, Moser A, York DM (2006) J Phys Chem A 110:791
49. Li G, Cui Q (2003) J Phys Chem B 107:14521
50. Tissandier MD, Cowen KA, Feng WY, Gundlach E, Cohen MH, Earhart AD, Coe JV, Tuttle TR Jr (1998) J Phys Chem A 102:7787
51. Camaioni DM, Schwerdtfeger CA (2005) J Phys Chem A 109:10795

52. Kawahara S, Kobori A, Sekine M, Taira K, Uchimaru T (2001) J Phys Chem A 105:10596
53. Asensio A, Kobko N, Dannenberg JJ (2003) J Phys Chem A 107:6441
54. Breneman CM, Wiberg KB (1990) J Comput Chem 11:361
55. Zhang X, Mathews CK (1994) J Biol Chem 26:7066
56. Sowers LC (2000) J Biomol Struct Dyn 17:713
57. Yanson IK, Teplitsky AB, Sukhodub LF (1979) Biopolymers 18:1149
58. Šponer J, Leszczynski J, Hobza P (2002) Biopolymers 61:3
59. Šponer J, Jurečka P, Hobza P (2004) J Am Chem Soc 126:10142
60. Mo Y (2006) J Mol Model 12:665
61. Valinluck V, Wu W, Liu P, Neidigh JW, Sowers LC (2006) Chem Res Toxicol 19:556
62. Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry. W. H. Freeman and Co., New York
63. Norberg J, Vihinen M (2001) J Mol Struct Theochem 546:51
64. Geacintov NE, Yoshia H, Ibanez V, Jacobs SA, Harvey RG (1984) Biochem Biophys Res Commun 122:33
65. Sowers LC, Shaw BR, Sedwick WD (1987) Biochem Biophys Res Commun 148:790