Reactivity of guanine at m^{5}CpG steps in DNA: evidence for electronic effects transmitted through the base pairs

Arunangshu Das[1], Kit S Tang[2], S Gopalakrishnan[1], Michael J Waring[2] and Maria Tomasz[1]

Background: Mitomycin C (MC), a DNA cross-linking and alkylating agent, targets guanines in the m^{5}CpG sequence with 2–3-fold preference over guanines in unmethylated CpG. Benzo[*a*]pyrenediolepoxide (BPDE) and several other aromatic carcinogens form guanine adducts with an identical selectivity for m^{5}CpG, and in certain cancers G to T transversion mutation 'hotspots' in the p53 tumor suppressor gene are more frequent at this sequence than at guanines in other sequences. MC appears suitable to probe the general mechanism of this selectivity.

Results: A 162-bp DNA fragment containing C, m^{5}C or f^{5}C (5-fluoro cytosine) at all cytosine positions was cross-linked by MC at guanines in CpG steps. The extent of cross-linking increased in the order f^{5}C < C < m^{5}C. Monoalkylation or cross-linking of duplex 12-mer oligonucleotides containing a single CpG, f^{5}CpG or m^{5}CpG step gave yields of adducts that increased in the same order. The rates showed a correlation with the Hammett σ constant of the methyl and fluoro substituents of the cytosine. Only the base-pair cytosine substituent influenced reactivity of guanine.

Conclusions: The 2-amino group of guanine in the m^{5}CpG sequence of DNA has a greater nucleophilic reactivity with mitomycin than CpG. Evidence is presented for a novel mechanism: transmission of the electron-donating effect of the 5-methyl substituent of the cytosine to guanine through H-bonding of the m^{5}C·G base pair. The results explain the enhanced reaction of BPDE at m^{5}CpG in DNA and the origin of G–T mutational hotspots in the p53 gene in cancer.

Addresses: [1]Department of Chemistry, Hunter College, CUNY, New York, NY 10021, USA. [2]Department of Pharmacology, Cambridge University, Tennis Court Road, Cambridge, CB2 1QJ, UK.

Correspondence: Maria Tomasz
E-mail: tomasz@patsy.hunter.cuny.edu

Key words: BPDE adducts, DNA–mitomycin C adducts, guanine nucleophilicity at m^{5}CpG, m^{5}CpG, p53 mutations

Received: **17 February 1999**
Revisions requested: **17 March 1999**
Revisions received: **7 April 1999**
Accepted: **15 April 1999**

Published: **17 June 1999**

Chemistry & Biology July 1999, **6**:461–471
http://biomednet.com/elecref/1074552100600461

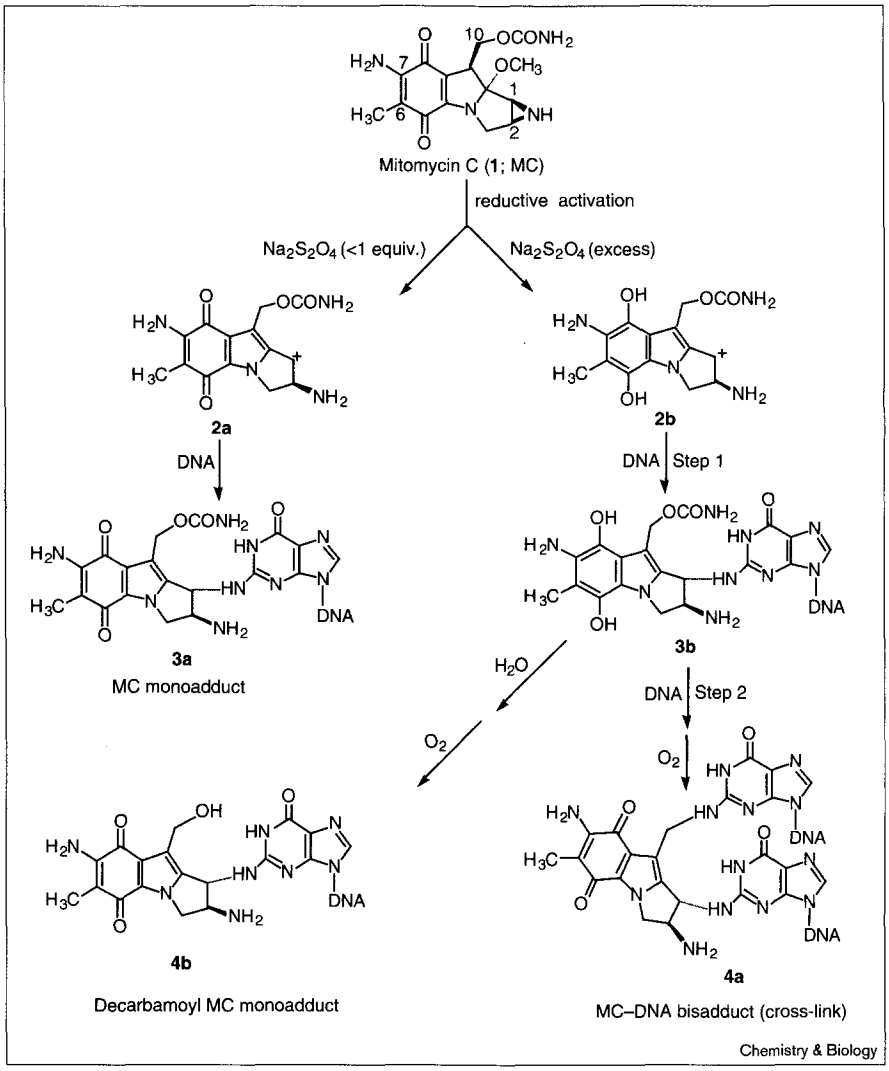
Introduction

The CpG sequence is of special significance in mammalian DNA. It is the exclusive site of cytosine methylation and it is usually methylated; it is estimated that 60–90% of the CpG steps in human DNA contain 5-methyl cytosine (m^{5}C) [1]. The methylation status of CpG steps located in transcriptional regulatory regions of DNA (CpG islands) is a critical determinant of gene expression [2]. The frequency of CpG in mammalian DNA is anomalously low compared with the other 15 dinucleotides: only 1.6%, which corresponds to one fifth of its statistically expected frequency. This is probably because of evolutionary selection for m^{5}CpG to TpG mutations, which are the result of endogenous deamination of m^{5}C to T [3]. The deamination reaction itself proceeds faster with m^{5}C than with C [4]. CpG is frequently mutated in human cancer (reviewed in [5]), consistent with a special sensitivity of m^{5}C to mutagenesis [6].

Methylated CpG was reported recently to affect the recognition for alkylation and cross-linking of DNA by mitomycin C (MC), an agent used in cancer chemotherapy [7,8]. Although CpG·CpG is the exclusive site of MC-induced DNA interstrand cross-links, formed between MC and the exocyclic amino groups of the two guanines [9–11] (Figure 1), cross-link formation is therefore further enhanced (2–4-fold) when the cytosines of CpG·CpG are methylated [7,8]. Guanine monoalkylation by MC (Figure 1) is similarly enhanced at m^{5}CpG·m^{5}CpG compared with CpG·CpG. It was also determined that methylation of the base-paired partner cytosine rather than the 5′-neighbor cytosine of the alkylation-susceptible guanine was responsible for the increased rate of reaction with MC [7]. The mechanism of the effect, however, remains unknown.

In remarkable analogy to the case of MC, Denissenko and coworkers [12] reported that the environmental mutagen and carcinogen benzo[*a*]pyrene diolepoxide (BPDE) forms guanine–N^2 adducts preferentially at m^{5}CpG over CpG or any other NpG sequence. This was shown in the case of p53 tumor suppressor, as well as with DNA of the promoter region of the X-linked *PGK1* gene DNA, each treated with BPDE in cell-free systems. Tang's group [13] examined several other polycyclic aromatic hydrocarbons for m^{5}CpG selectivity using a p53 DNA fragment and found a similar enhancing effect of cytosine methylation

Figure 1



Reductive activation of mitomycin C to a monofunctional or a bifunctional alkylating form (**2a** or **2b**, respectively) and reactions of the reactive species with DNA. Reaction of the monofunctional species **2a** yields MC monoadduct **3a**, whereas the bifunctional form **2b** leads to the MC–DNA bisadduct **4a** and the decarbamoyl MC monoadduct **4b**.

on the reaction of the CpG guanine with the carcinogens. Furthermore, the m^5CpG 'hotspots' of guanine adducts corresponded precisely to mutational hotspots observed frequently in the p53 tumor suppressor gene in human cancers ([12,13], reviewed in [5]).

In addition to alkylation at guanines, methylated CpG in p53 DNA was shown to endow selectivity for another type of mutagenic reaction with DNA: sunlight-induced pyrimidine dimer formation [14,15]. Again, chemical hotspots and mutational hotspots in the p53 gene in skin cancer dramatically coincided. These findings led to the hypothesis that the well-established propensity of m^5C for mutation [6] may be due to its increased reactivity not only with water but also with environmental or endogenous mutagens [6,14].

Enhanced hydrolytic deamination of cytosine in m^5CpG has been attributed to the higher pK_a value of m^5C than that of cytosine [4,16], although another mechanism involving the action of DNA (cytosine-5) methyltransferase has been considered [16–18]. To explain the enhancement of sunlight-induced pyrimidine dimerization by m^5CpG it was proposed that the 5-methyl group has a direct, electronic effect on cytosine photoreactivity [14]. It is less obvious, however, how alkylation of guanine can be affected by the same 5-methyl group of cytosine. In view of the relevance of this question to molecular mechanisms of mutagenesis we investigated the possibility that guanine had intrinsically higher reactivity in m^5CpG. We hypothesized that the electron-donating effect of the methyl group of m^5C could be transmitted to the partner guanine through H-bonding, resulting in increased nucleophilic reactivity of the guanine 2-amino group. If this were true then 5-fluorocytosine would have the opposite effect: decreased reactivity of guanine due to the electron-withdrawing effect of the 5-fluoro substituent of the cytosine. To test the validity of

this hypothesis, three variants of a 162 base pair (bp) DNA restriction fragment were synthesized, containing either C, m^{5}C or f^{5}C at all cytosine positions, using a polymerase chain reaction (PCR) method developed previously for preparation of base analog-substituted DNA [19]. A comparison of the extent of cross-linking of the three DNAs by MC was used to assess the relative reactivities of guanine at CpG, m^{5}CpG and f^{5}CpG steps. The MC cross-linking reaction is ideally suited to this purpose because of its specificity for guanines in the CpG sequence [9–11]. In addition to DNA, synthetic 12-mer duplex oligonucleotides containing the three cytosine variants at a single CpG step were probed by MC for their susceptibility to both monoalkylation and cross-linking. Remarkably, the relative reactivities were found to correlate with the Hammett σ constants [20] of the methyl and fluorine substituents on the cytosine, suggesting that electronic substituent effects are indeed transmitted to the guanine partner through H-bonding. In support of this interpretation, an ^{1}H-nuclear magnetic resonance (NMR) experiment showed that in DNA–MC adducts the guanine-N^2-H chemical shift is dependent on the 5-substituent of the partner cytosine. On the basis of these findings, a mechanism is proposed for the enhanced alkylation of guanine by MC at the m^{5}CpG sequence, which is likely to apply generally to analogous guanine alkylation reactions by other agents including BPDE [21]. The mechanism therefore suggests a general basis for the increased G→T mutation frequency of guanine at m^{5}CpG, observed in the p53 tumor suppressor gene in human cancer.

Results

Cross-linking of *tyrT* DNA by MC: effects of substituting m^{5}C and f^{5}C for C on the efficiency of cross-linking

The three radioactively labeled PCR-developed DNAs, designated as natural, m^{5}C- and f^{5}C-DNA, were subjected to cross-linking by reductively activated MC, then after heat denaturation were subjected to denaturing polyacrylamide gel electrophoresis (PAGE). The observed band patterns (Figure 2a) indicated two groups of DNA molecules in each case: the faster moving group (marked SS DNA) consisted of single-stranded 162 bp DNA as the major band, together with several distinct slower bands

Figure 2

(a) Cross-linking of *tyrT* DNA containing cytosine, 5-methyl cytosine or 5-fluoro cytosine, assayed using denaturing PAGE at various concentrations of MC. The control lanes are free from MC or $Na_2S_2O_4$. The remaining lanes show DNA treated with 0–100 μM MC in the presence of 2.8 mM $Na_2S_2O_4$. XL DNA, single-stranded DNA containing MC intrastrand cross-links and monoadducts. SS DNA, single-stranded DNA. **(b)** Plots of the overall extent of cross-linking as a function of MC concentration.

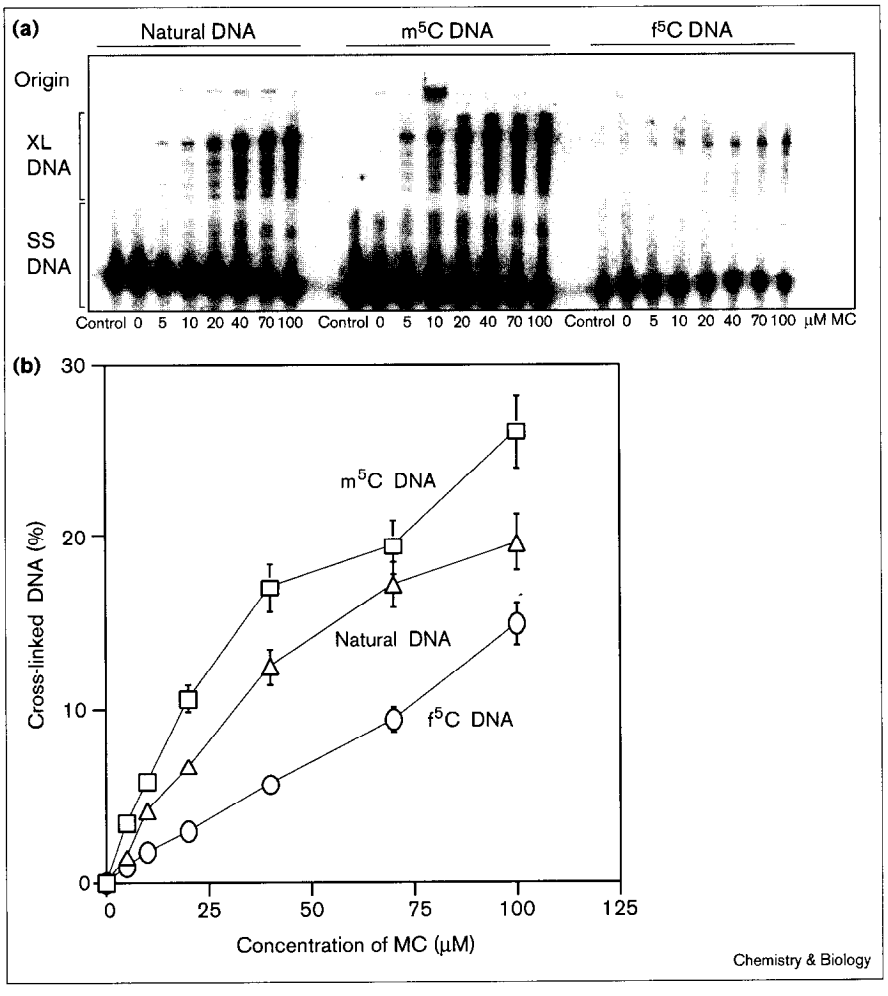
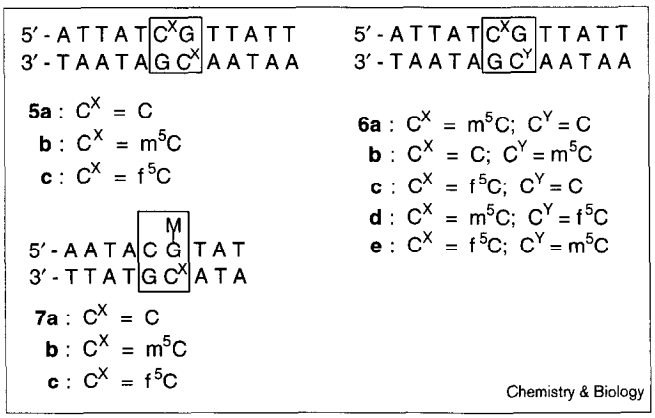
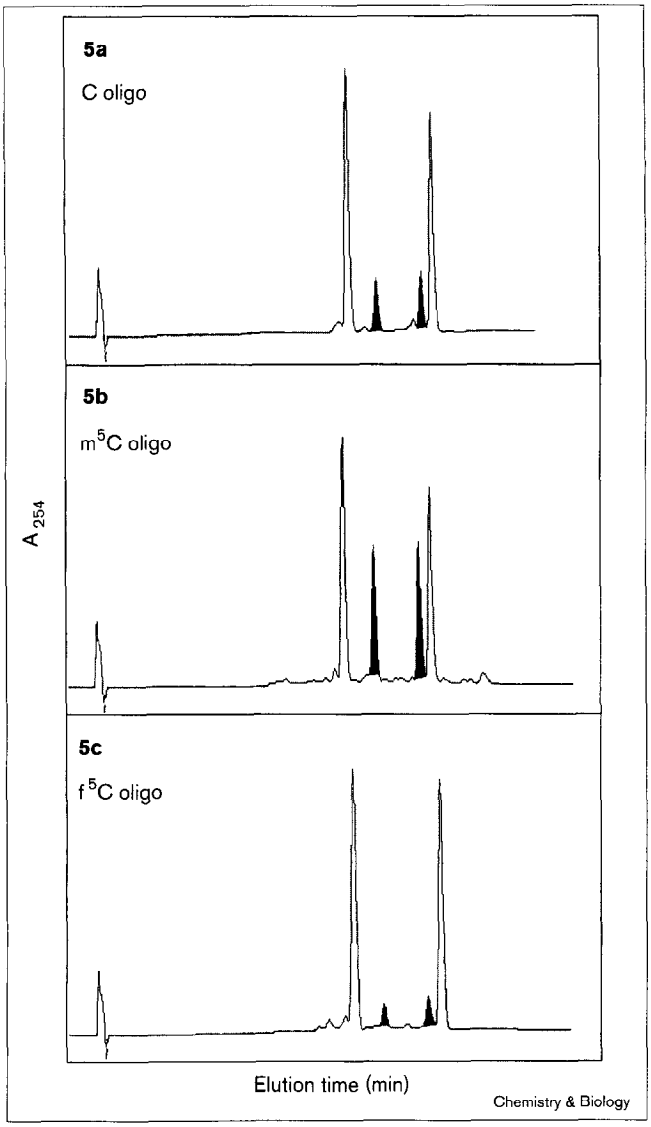


Figure 3



Structures of duplex oligodeoxyribonucleotides used in this work. In the **7a–c** series the top strand guanine is substituted by MC, as indicated by the letter M linked to the letter G. The substitution corresponds to monoadduct **3a** (Figure 1).

Figure 4



Monoalkylation of the duplex oligonucleotides **5a–c** by MC, assayed using HPLC. Shaded peaks, alkylated strands; unshaded peaks, unmodified strands.

that were assigned as single-stranded DNA containing MC intrastrand cross-links and monoadducts because such modifications are known to impart slightly retarded mobility compared with control DNA on denaturing PAGE [22]. The bands in the second group (marked XL DNA) correspond to cross-linked DNA molecules containing 324 nucleotides that migrate much slower than the single-stranded control (marked SS DNA). The ability of denaturing PAGE to resolve several distinct cross-linked species derives from the varying position of the cross-link within the DNA. As the level of cross-linking was less than 37% at all drug concentrations employed (Figure 2b), 'single hit' cross-linking, that is, not more than one cross-link per DNA molecule, may be assumed. *Tyr*T DNA contains 13 CpG steps (i.e. 13 sites cross-linkable by MC) [22]. The position of the cross-link affects the mobility such that centrally cross-linked molecules migrate most slowly [23]. The six major bands observed in the cross-linked groups seen in Figure 2a have been assigned in previous work to specific cross-linked CpG positions [22]. The mobility and distribution of the cross-linked bands from the three DNAs were identical, as expected. The yields of total cross-linked DNA varied considerably, however (Figure 1b): m^5C-DNA gave higher yields, whereas f^5C-DNA gave lower yields, than natural DNA, at all concentrations of MC tested. The MC concentrations required to cross-link 15% of the DNA (ISC_{15}) are 34, 60 and 100 μM for m^5C-, C- and f^5C-substituted *tyr*T DNA, respectively, estimated from the data of Figure 2b.

Monoalkylation and cross-linking of C-, m^5C- and f^5C-containing oligonucleotides

The yields of monoalkylation of oligonucleotides **5a–c** (Figure 3) were determined by separating the reaction mixtures into alkylated and nonalkylated oligonucleotide strand components using high-performance liquid chromatography (HPLC; a total of four components) and measuring their relative amounts (Figure 4). The yields decreased in the order $m^5C > C > f^5C$ (Table 1, Figure 5). The yields in the cross-linking reaction were determined independently by separating and measuring the A_{260} of the cross-linked and non-cross-linked fractions using Sephadex G-50 chromatography (Figure 6). The yields again decreased in the order $m^5C > C > f^5C$ (Table 1, Figure 5).

Selective influence of the cytosine-5 substituent on the alkylation of the base-paired guanine

To investigate this selectivity we varied the 5-substituent attached to the cytosines of $C^XpG{\cdot}C^YpG$ in the top and

Table 1

Monoalkylation and cross-linking of guanine at CpG, m^{5}CpG and f^{5}CpG sequences by MC: yields, relative rate constants, Hammett σ constants of the cytosine-5 substituent and NMR δ value of the N^2-H of the alkylated guanine.

	Oligonucleotides				*tyr* T DNA		
Cytosine-5 substitution	% yield of monoalkylation*	% yield of cross-linking†	$(k/k_H)_{monoalkylation}$‡ (log)	$(k/k_H)_{cross\text{-}link}$‡ (log)	$(k/k_H)_{cross\text{-}link}$‡,§ (log)	σ_p# (Hammett constant)	δ (ppm)¶
H	18.05 ± 0.45	24.1	1.00 (0)	1.00 (0)	1.00 (0)	0	9.56
CH_3	38.65 ± 0.95	58.4	1.48 (0.17)	3.17 (0.5)	1.67 (0.22)	−0.14	9.76
F	8.46 ± 1.2	14.7	0.27 (−0.57)	0.58 (−0.24)	0.33 (−0.48)	+0.15	9.08

*Based on % guanine residues monoalkylated. †Based on % duplex **5** cross-linked.‡k is the pseudo-first-order rate constant of the monoalkylation or crosslinking reaction of the DNA substrate with MC; k_H is that of a substrate which contains unsubstituted cytosine. Ratios of k/k_H were calculated from the yields observed under standard conditions using the formula: $k/k_H = \log(100 - \%\text{ yield of reaction k})/\log(100 - \%\text{ yield of reaction } k_H)$. §Calculated from the % yield of cross-linked *tyr*T DNA using 20 μM MC (Figure 2b). #[20]. ¶Data from Figure 8.

bottom strands of the oligonucleotides independently. The various duplex substrates (**6**; Figure 3) were monoalkylated by MC at their guanines in the usual way. The yields of alkylated guanine in the top and bottom strands were determined using HPLC to separate the alkylated and nonalkylated oligonucleotide species (a total of four components) and measuring their proportions. The HPLC separation is shown in Figure 4 and the results are summarized in Figure 7a,b. They demonstrate that 5-methylation of a cytosine residue increases the yield of alkylation of guanine in the opposite strand more than twofold, whereas no increase occurs in the same strand (Figure 7a). Similarly, 5-fluoro-substitution of cytosine leads to significantly decreased alkylation of guanine in the opposite strand but the yield in the same strand is not affected (Figure 7b). The selective effect of the cytosine 5-substituent on the nucleotide in the opposite strand is especially striking in the case of **6e**, in which the yield of alkylation of the top-strand guanine is threefold higher than that of the bottom-strand guanine. The results indicate that both the stimulatory effect of 5-methyl cytosine and the inhibitory effect of 5-fluoro cytosine are selective to the guanine that is paired with the modified pyrimidine. Adjacent base pairs appear to be affected by their respective cytosine substituents independently.

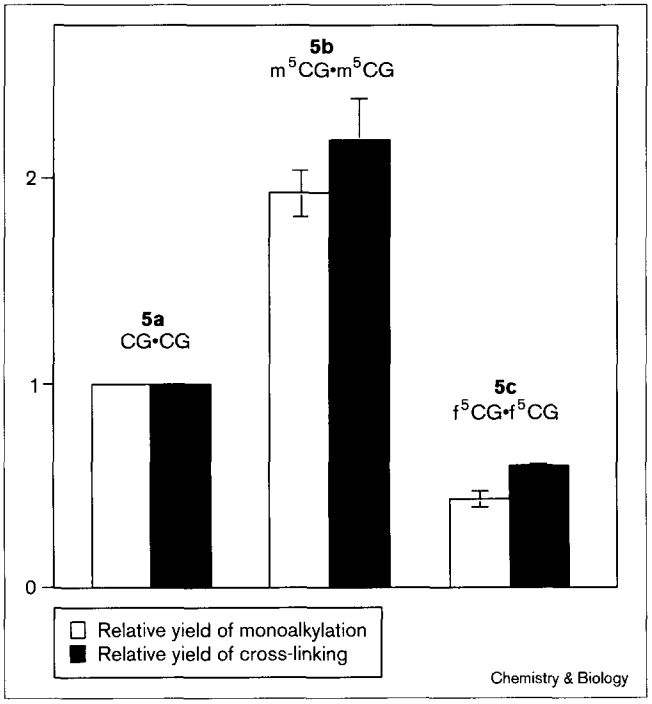
Variation of the NMR chemical shift of the N^2-proton of guanine opposite C, m^{5}C or f^{5}C in MC–oligonucleotide complexes 7a–c

The guanine-2-amino proton of the MC monoadduct **3a** incorporated in duplex oligonucleotides is detectable in H_2O using ^{1}H-NMR [24]. In order to probe directly the influence of the opposite-strand cytosine analog on the electron density of this proton the ^{1}H-NMR spectra of the duplexes **7a–c** were determined. As revealed in Figure 8, the chemical shift of the top-strand guanine-N^2 proton varied with the substitution in the bottom-strand cytosine: 9.56 ppm for C, 9.76 for m^{5}C and 9.08 for f^{5}C. These values are also shown in Table 1.

Correlation of the reaction rates with the Hammett sigma constants and with ^{1}H-NMR chemical shifts

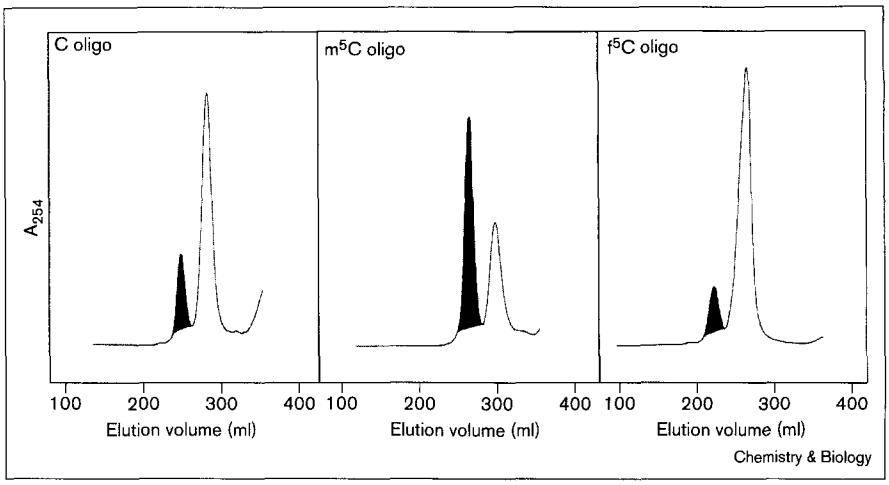
Absolute rates of reaction of activated MC with DNA cannot be measured as activated MC decays very fast in solution because of several competing reactions—its tautomerization to 2,7-diaminomitosene, reactions with water, phosphate, bisulfite ions and DNA bases—all proceed too fast to measure. Because the reaction with DNA is likely to follow pseudo first-order kinetics with

Figure 5



Relative yields of monoalkylation and cross-linking from the duplex oligonucleotides **5a–c** by MC, calculated from the observed percentage yields of monoalkylation and cross-linking reactions listed in Table 1.

Figure 6



Cross-linking of the duplex oligonucleotides **5a–c** by MC, assayed using Sephadex G-50 gel chromatography. Shaded peak, cross-linked duplex; unshaded peak, uncross-linked strands.

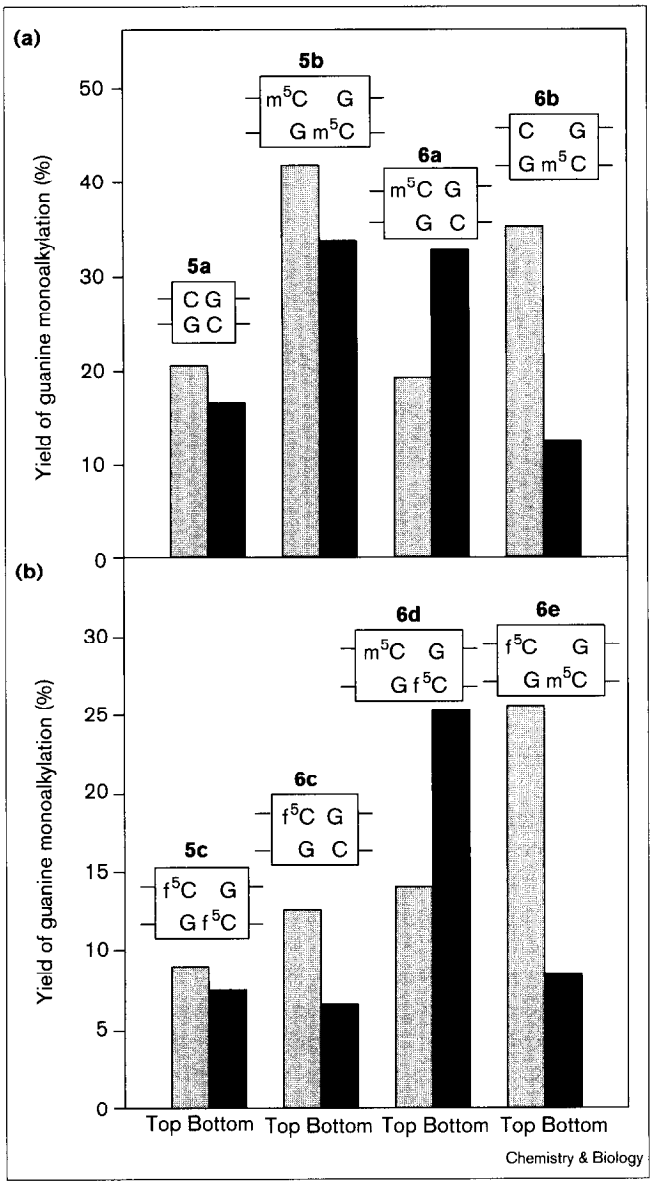
respect to DNA until abrupt self-destruction of the active MC species [9,11,25], however, the relative yields of alkylation of the various DNA substrates under identical conditions may be used to calculate relative rate constants [11,25]. The calculated k/k_H values for reactions of m^5C- and f^5C-oligonucleotides compared with C-oligonucleotides are shown in Table 1. Corresponding values for *tyr*T DNA cross-linking calculated from the yields obtained with 20 μM MC (Figure 2b) are also listed.

The data were tested to fit the Hammett structure–reactivity relationship $\log k/k_H = \sigma\rho$ [20], where, in our case, k_H is the rate constant for reactions at unsubstituted C·G and k is the rate constant for that at m^5C·G or f^5C·G. In the original Hammett relationship σ is a constant that sums up the total electronic effects (resonance plus field) of substituent X on the rate of a reaction at Y in the system $X\text{-}C_6H_4\text{-}Y$. A positive value of σ indicates an electron-withdrawing group and a negative value an electron-donating group. The effects of X are transmitted to Y through the π electrons of the benzene ring, thereby influencing the reaction rate. The value of ρ is constant for a given type and conditions of reaction [20]. In the present application the C·G base pair takes the place of the benzene ring; X is the cytosine-5 substituent and Y is the 2-amino group of the guanine. The fit of the data is close to linear (Figure 9a), analogously to $X\text{-}C_6H_4\text{-}Y$ systems from which σ was derived. The correlation of the electronic effect of the X substituent, characterized by its σ_p value, that includes both field and resonance effects [20] with the rate of reaction at Y is therefore apparent from the plots. The similar correlation of the ^{1}H-NMR chemical shift of the guanine-N^2H of the MC–guanine monoadduct in duplexes **7a–c** with $\log k/k_H$ (Figure 9b) further substantiates the hypothesis that the electronic properties of the cytosine substituent influence the electron density of the guanine-N^2 proton in the opposite strand and, in turn, this influences the rate of the reaction at the purine 2-amino group.

Discussion

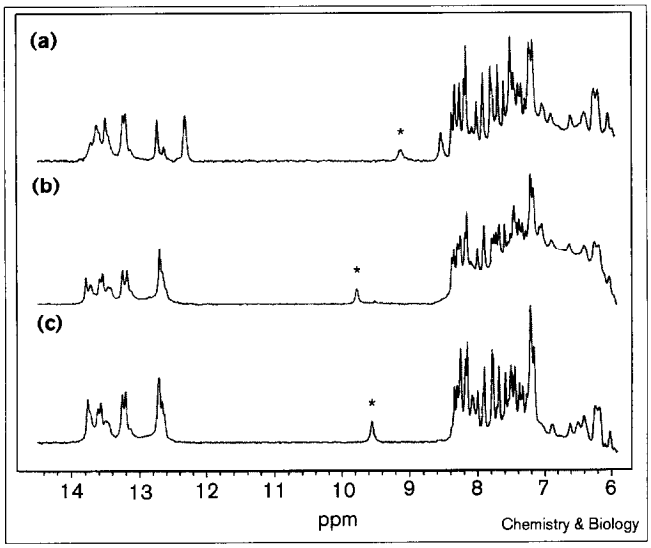
Enhanced alkylation of guanine at methylated CpG·CpG sites in DNA has been clearly demonstrated by at least two biologically important agents, mitomycin C and BPDE, both of which alkylate the exocyclic amino group of guanine in the minor groove of the double helix. The magnitude of the enhancement observed in cell-free systems is similar: 2–3-fold for MC and 2–4-fold for BPDE [12,13]. Assay of BPDE adducts formed in intact cells indicated similar results [12]; enhanced formation of MC adducts in intact cells has not yet been tested. Given the considerable differences in structure between MC and BPDE a common origin of the effect seemed more likely to reside in an electronic modulation of guanine reactivity than in modulation through steric changes in DNA. The cytosine 5-methyl group is located in the major groove, whereas the site of alkylation is in the minor groove; furthermore, methylation of CpG cytosine does not cause any appreciable change in DNA conformation under physiological conditions, as concluded from crystal [26,27] and solution structures [28] of oligonucleotides and from studies using electrophoretic mobility assays of structural changes upon methylation of DNA restriction fragments [29]. A slight localized change in the twist of the m^5C·G base pair compared with that of C·G in the same sequence context has nevertheless been noted [26,27] and therefore a steric origin of the increased efficiency of alkylation could not be entirely excluded *a priori*. The present results provide a compelling case for an electronic origin of this phenomenon, however. First, the 5-methyl- and 5-fluoro-substituted cytosines exert opposite effects on the reactivity of guanine. Second, the Hammett σ_p constant of the methyl, fluoro and H substituents, which is a relative measure of their electronic effect on the nucleophilic reactivity of a

Figure 7



Selective effect of the cytosine 5-substituent on the base-paired guanine: percentage yields of guanine monoalkylation in the top and bottom strands of duplex oligonucleotides in the **5** and **6** series. **(a)** Effect of m^5C. **(b)** Effect of f^5C.

Figure 8

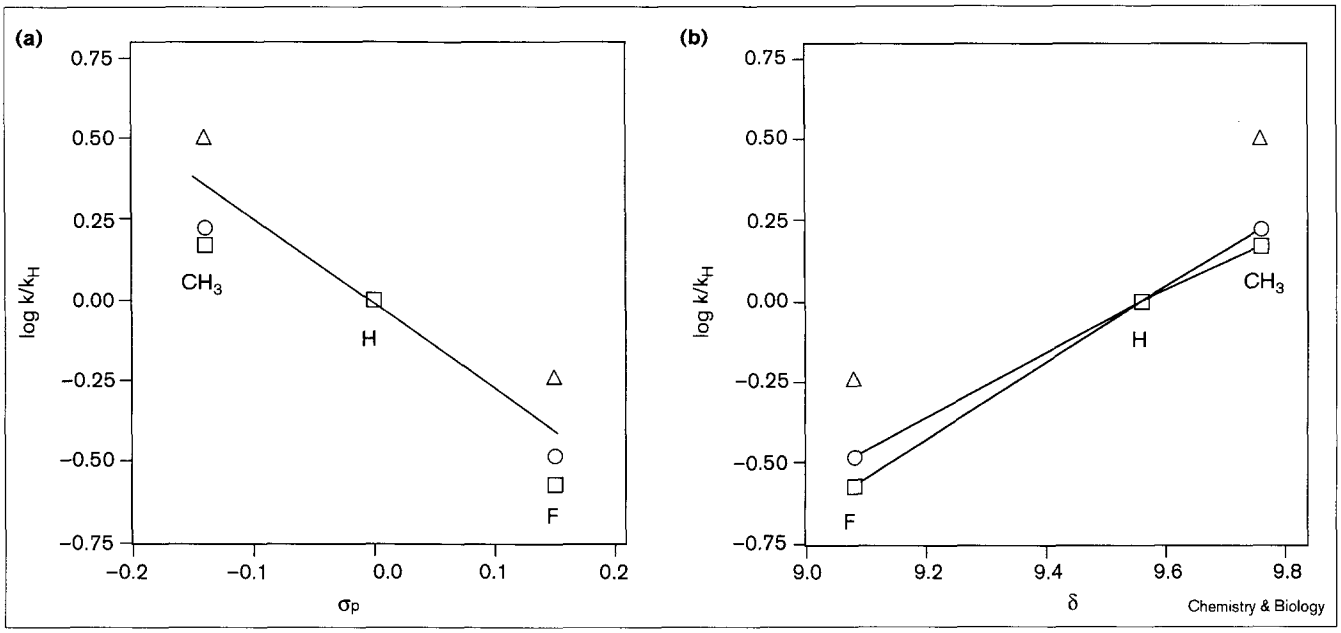


Expanded one-dimensional NMR plots of MC monoadduct duplexes d(AATAC[M]GTAT)·d(ATAC^{X}GTATT; **7**). **(a)** C^X, f^5C; **(b)** C^X, m^5C; **(c)** C^X, C. Spectra were obtained at 10°C using jump and return pulse sequence. *Denotes the [M]G-N^2H resonance in each duplex. Pairing the f^5c opposite [M]G shifts the [M]G-N^2H resonance upfield by 0.48 ppm; a similar effect is observed for [M]G-N^1H resonance also. Pairing m^5c with [M]G shifts the [M]G-N^2H resonance downfield by 0.20 ppm. The opposite shifts are consistent with the opposite inductive effects of the methyl and fluoro substituents.

functional group [20], correlates well with the observed rates of the DNA–mitomycin reactions, including both oligonucleotides and DNA restriction fragments as substrates. Third, the substituent effect is restricted to the single C·G basepair in which the cytosine is substituted; the reactivity of the adjacent C·G base pair is not affected. Fourth, the 5-substituent of cytosine has a direct influence on the ^1H chemical shift of the N^2–H–proton of the partner guanine in the opposite strand in accordance with its Hammett σ constant and the chemical shifts correlate linearly with the observed reaction rates (Figure 9b). This was established by preparing an oligonucleotide in which the guanine residue was monoalkylated at its 2-amino group by MC; this construct was then hybridized with its complement, containing either C, m^5C or f^5C opposite the site of the guanine adduct (**7a**, **b** or **c**, respectively). In MC monoadduct duplexes the single guanine-N^2 proton, H-bonded to cytosine, exhibits a sharp signal between 9 and 10 ppm due to its reduced rate of exchange with H_2O [24]. Thus, **7a–c** provide a system for observing the effect of the cytosine-5 substituent on the chemical shift of the N^2-proton of the partner guanine. Relative to unsubstituted cytosine, an upfield shift in the case of methyl cytosine and downfield shift in the case of fluoro cytosine were observed. This indicates a greater electron density at the guanine-N^2-atom in the m^5C·G basepair than in the C·G base pair, with the electron density of G to N^2 in the f^5C·G base pair the lowest of all three. The modulating signal must originate from the substituent variation in the other strand. Furthermore, as in duplex DNA the only contacts between C and G in a C·G base pair are the H-bonds, the observed electron-donating effect of the methyl group (Hammett $\sigma = -0.14$) and the electron-withdrawing effect of the fluoro group ($\sigma = +0.15$) [20] must be transmitted from C to G via the H-bonds.

These conclusions readily explain how the electronic properties of the cytosine 5-substituent modulate the rate

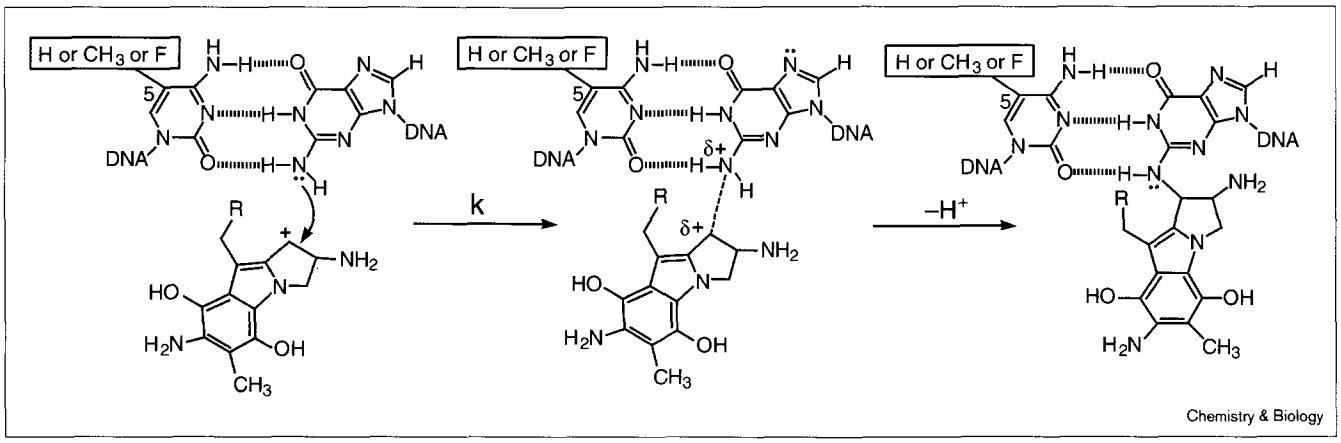
Figure 9



(a) Plots of relative rates of different reactions as a function of Hammett sigma from data in Table 1. **(b)** Same, as a function of the ^{1}H-NMR chemical shift of [M]G-N2-H of 7a–c from data in Table 1. Squares, monoalkylation reactions; triangles, cross-linking of oligonucleotides; circles, cross-linking of *tyr*T DNA.

of covalent reaction with guanine in a C·G base pair (Figure 10): nucleophilic attack by the guanine-2-amino group on the MC delocalized 1-carbonium ion creates a partial positive charge on the N^2-nitrogen in the transition state. Transmission of negative charge to N^2 from m^5C lowers the energy of the transition state, thereby increasing the rate; conversely, electron deficiency transmitted from f^5C increases the transition-state energy, resulting in a lower rate. This applies to both the monoalkylation and cross-linking processes, because the rate-determining step is the same in both ('Step 1', Figure 1). The effects on the rates of the reactions are relatively small: the maximal threefold increase observed in the case of methylated cytosine [7] corresponds to a decrease of only 0.65 kcal of activation energy at 25°C. This suggests that the transmission of electronic effects across H-bonded DNA base pairs is less efficient than that via aromatic systems on which the Hammett σ–rho relationship is based [20]. Further study is needed for a greater understanding of this phenomenon. For example, it would be of interest to examine

Figure 10



Nucleophilic attack on the MC 1-carbonium ion by the 2-amino group of guanine base-paired with 5-substituted cytosine.

whether other guanine-N^2-specific alkylating drugs respond in the same way as MC to methyl- and fluorocytosine substitution and whether all nucleophilic reactive sites of guanine (N^2, O^6 and N7) are affected similarly by an appropriate electrophile. This is probably not the case because the latter two positions are not H-bond donors to C in the G·C basepair, like N^2. In addition, N7 is a much stronger nucleophile than N^2 and O^6 [30].

Significance

Guanine in the m^{5}CpG sequence of DNA exhibits enhanced covalent reactivity with benzo[*a*]pyrenediolepoxide (BPDE) [12,31] and various other aromatic carcinogens and mutagens [13], as well as with the antitumor agent, mitomycin C (MC) [7,8]. This study identified the mechanistic origin of this phenomenon in the case of MC, a guanine-N2-specific alkylator. The mechanism is based on a novel property of duplex DNA: transmission of the electronic effect of a cytosine substituent to the base-paired guanine through H-bonding. The similar magnitude of the enhancing effect of cytosine methylation on the MC and BPDE adduct yields argues strongly for a common origin; that is, enhanced nucleophilicity of the guanine-2-amino group, the principal target of both agents [11,12].

Enhanced formation of BPDE adducts at m^{5}CpG in the p53 gene was found to coincide with an increased average frequency of mutation at the same sites in human cancer. This coincidence was interpreted as a signature for an environmental carcinogen playing a causative role in certain tumors ([12,13]; reviewed in [5]). As a compelling example, it was noted [13] that a G to T transversion, generally indicative of guanine damage has been observed as a mutational hotspot at an m^{5}CpG step (codon 157) in smoking-associated lung cancer and linked to the presence of benzpyrene in cigarette smoke [32]. The present study suggests that the ultimate origin of such G→T mutational hotspots is the intrinsically higher nucleophilic reactivity of guanines toward environmental or endogenous mutagens when base-paired with methylated cytosine.

Materials and methods

Materials

Mitomycin C was obtained from Dr. D.M. Vyas, Bristol-Myers Squibb Co., Wallingford, CT. Oligonucleotides were synthesized on an automated DNA synthesizer (model 380B, Applied Biosystems, Inc.), using the β-cyanoethyl phosphoramidite method. For incorporation of f^{5}C into oligonucleotides 'convertible F-dC' (TMP-F-dU-CE phosphoramidite from Glen Research, Sterling, VA) was used.

The 162-bp *Escherichia coli* tyrT promoter-containing *tyr*T (A93) DNA used as a PCR template was isolated from pAT plasmid by digestion with restriction enzymes *Aval* (Promega) and *Eco*RI (Boehringer) and purified by 6% nondenaturing PAGE (19:1 acrylamide: bisacrylamide). The DNA, located under UV light, was eluted out of the crushed gel slice using a unidirectional electroelutor (IBI Model UEA) at 120 V for 2 h. The DNA was further purified with a QIAquick PCR purification kit (QIAGEN) according to the suggested standard protocol. Primers were synthesized and quality-checked using mass spectrometry at the Protein and Nucleic Acid Chemistry Facility of Cambridge University Biochemistry Department, UK. The Crick primers were 5′ end-labeled with [γ-^{32}P]-ATP (DuPont/NEN) of 6000 Ci/mmol specific activity with T4 polynucleotide kinase (Promega). Unincorporated [γ-^{32}P]-ATP was separated from the primer by using a QIAquick nucleotide removal kit (QIAGEN) according to the manufacturer's protocol. Nucleoside triphosphates dATP, dGTP, dCTP and dm^{5}CTP were purchased from Pharmacia Biotech. df^{5}CTP was obtained from Sierra Bioresearch, Tucson, AZ.

Methods

Quantities of oligonucleotides were measured by UV spectrometry using the molar extinction coefficient 10,000 at 260 nm (average value per mononucleotide residue; 10 A_{260} units of an oligonucleotide correspond to 1 μmol mononucleotide). The quantity of MC was measured using the molar extinction coefficient 21,840 at 360 nm. Quantities of nucleosides and nucleoside–MC adducts were measured using the following molar extinction coefficients at 254 nm: dA, 13,300; dG 13,000; dT, 6,600; dC, 6,300; dm^{5}C, 4,725; df^{5}C, 5,500; adduct **3a**, 24,000; adduct **4a**, 30,000.

HPLC

A Beckman instrument equipped with System Gold programmable solvent module 125 and diode array detector 168 controlled by System Gold Chromatography Software was used.

Monofunctional alkylation of oligonucleotides by MC: substoichiometric anaerobic $Na_2S_2O_4$ as activating agent [25]

Duplex oligonucleotides (self-complementary or complementary strands in a 1:1 molar strand ratio, 1 mM mononucleotide residue; usually 2–10 μmol total scale) were mixed with MC (4 mM) in 0.1 M sodium phosphate, pH 7 buffer at 4°C, in an ice bath under stirring in air. Fresh anaerobically prepared $Na_2S_2O_4$ solution (40 mM in the same buffer) was added, to give 2 mM final concentration. After 1 h the mixture was chromatographed on a 2.5 × 56 cm Sephadex G-25 column with 0.02 M NH_4HCO_3 as eluant. The first UV absorbing peak fraction containing the alkylated and unalkylated oligonucleotide(s) was then collected and lyophilized. Alkylated oligonucleotides were isolated using HPLC.

Cross-linking of oligonucleotides: excess anaerobic $Na_2S_2O_4$ as activating agent [11]

Duplex oligodeoxynucleotides as above (1 mM mononucleotide residue) were exposed to MC (2 mM) in 0.1 M Tris, pH 7.4 buffer at 4°C with stirring. The solution was deaerated and kept anaerobic by purging with argon. Fresh anaerobically prepared $Na_2S_2O_4$ stock solution (0.16 M in same buffer) was added to 3 mM final concentration in 5 portions at 10 min intervals. After 1 h the reaction mixture was chromatographed on a Sephadex G-50 (2.5 × 56 cm) column at 40°C. The pure cross-linked oligonucleotide was isolated from the first-eluted fraction.

Characterization of oligonucleotide–MC adducts formed by monoalkylation and bifunctional alkylation reactions

The monoalkylation reaction product, isolated by Sephadex G-25 column chromatography, was digested by SVD and AP using a standard protocol [25]. The digest was separated into its nucleoside and MC–nucleoside adduct components using HPLC on a C-18 reverse phase column; these components were identified by coelution with authentic standards. The bifunctional alkylation reaction product, isolated from the Sephadex G-50 column, was analyzed in identical fashion [11].

PCR for the synthesis of natural, m^{5}C- and f^{5}C-substituted tyrT *DNAs*

PCR mixtures (100 μl each, 50 mM KCl, 10 mM Tris–HCl, 0.1% Triton) each containing 1 mM Watson 1 primer, 20 μl of the labeled Crick 1 primer, 20 mg of *tyr*T template, and 0.25 mM of each appropriate dNTP were heated at 94°C for 1 min. Two units of *Taq* Polymerase (Promega) were then added to each reaction. Finally 20 μl of mineral oil

was added to prevent evaporation. Amplification cycles each cycle consisting of the following segments: (a) denaturation at 94°C for 1 min, template–primer annealing at 37°C for 2 min, and polymerization at 72°C for 10 min, were performed 30 times. Following the last cycle the reaction mixtures were slowly cooled (5 min at 60°C and 5 min at 37°C) to facilitate complete annealing of complementary DNA strands. The PCR products were then purified and recovered by 6% nondenaturating PAGE, electroelution, QIAquick PCR purification and lyophilization.

Bifunctional alkylation (cross-linking) of tyrT *DNA: excess $Na_2S_2O_4$ as activating agent*

A stock solution of 1 mM MC was prepared in 30% methanol and stored at 20°C in the dark. It was diluted to working concentrations with appropriate volumes of 30% methanol. The reaction mixture (25°µl) of TE buffer, pH 7.0, containing 10 µg of DNA and an appropriate amount of MC, was deaerated by bubbling with nitrogen for 2 min and then capped tightly. Water (500 µl) was deaerated for 5 min, which was followed by addition of sodium dithionite ($Na_2S_2O_4$) to make up a concentration of 20 mM. The solution was further deaerated for 5 min. The reaction mixture was supplemented with 4 µl of the 20 mM sodium dithionite solution, followed by deaeration for 30 s. It was capped tightly and kept in an ice bath in the dark for 1 h. The cross-linking reaction was stopped by adding 2 µg of calf thymus DNA (Sigma) in 2 µl of pH 7.4 TE buffer, 7.5 µg of tRNA (Sigma) in 25 µl of 0.3 M sodium acetate and 130 µl cold ethanol. It was kept on dry ice for 10 min and then microcentrifuged at 13,000 rpm for 10 min. The supernatant was discarded. The pellet was dried in air, then dissolved in 10 µl of strand separation buffer (30% dimethyl sulfoxide, 50% formamide, 1 mM EDTA, and 10% Sigma gel loading solution).

Gel electrophoresis and quantification by phosphorimaging

Cross-linked products were resolved using 8% denaturating PAGE (19:1 acrylamide:bisacrylamide, 50% w/v urea, 40 cm × 30 cm × 0.3 mm) run with 1 X TBE at 1000 V for 45 min. Such gels were prepared by mixing 20 ml of 40% acrylamide and bisacrylamide solution with 80 ml of 1 X TBE and caused to polymerize by adding 0.5 ml of 10% w/v ammonium persulfate and 50 µl of TEMED. Before loading the samples were denatured at 94°C for 4 min followed immediately by chilling in an ice bath. The gels were pre-run at 1000 V until the temperature reached 60°C. After electrophoresis they were soaked in 10% acetic acid for 10 min, transferred to Whatman 3 MM paper, and dried on a gel-drier (Bio-Rad) at 80°C for 45 min. They were then exposed to storage phosphor screens overnight at room temperature. A Molecular Dynamics 425 E PhosphorImager was used to collect the data from the storage screens. Baseline corrected scans were analyzed by integrating all band densities between two selected boundaries using ImageQuant version 3.3 software.

Synthesis of a 9-mer oligonucleotide–MC monoadduct complex for use in ^{1}H–NMR studies

The partially self-complementary oligonucleotide d(AATACGTAT) was monoalkylated by MC using the general monoalkylation protocol, described above. The desired adducted strand d(AATAC[M]GTAT) was purified by HPLC using a C-18 preparative reversed phase column and 10% to 14% acetonitrile in 60 min gradient in 0.1 M TEA, pH 7.0 buffer. Its identity was verified by quantitative analysis of its deoxynucleoside and adduct composition using methods described previously [25]. This adducted strand was used in the NMR experiments as follows.

NMR samples

Three NMR samples were made using the d(AATAC[M]GTAT) complex and hybridizing it with an equal amount (in A_{260} units) of complementary d(ATACGTATT), or d(ATAm^{5}CGTATT), or d(ATAf^{5}CGTATT) (**7a–c**). A typical sample was prepared as follows: The purified adduct strand d(AATAC[M]GTAT) was dissolved in milliQ water and exactly 50 A_{260} unit was taken in a polypropylene tube. The purified complementary strand (50 A_{260} unit) dissolved in milliQ water was added to the tube. NMR buffer (0.1 M NaCl, 0.0 M sodium phosphate, pH 7.4; 0.5 ml) was also added and the solution was lyophilized to dryness. The dry sample was dissolved in 0.5 ml 10% D_2O and transferred to an NMR tube.

NMR experiments

All NMR experiments were performed using the 400 MHz VARIAN Unityplus or JEOL NMR spectrometers. One-dimensional spectra in water were obtained using the jump and return pulse sequence [33]. The carrier frequency was centered on HDO resonance. A spectral width of 22 ppm, jump-return delay of 70 microseconds, recycle delay of 1 s, and 128 transients were used for obtaining the spectrum. Temperature was maintained at 10°C. A set of one-dimensional NOE experiments were also performed. For this purpose, the jump-return pulse sequence was modified to include a presaturation period before the pulse. One presaturation frequency was set outside the spectral range to serve as a control while the other frequency was set on the [M]G–N^{2}H resonance. The power level for presaturation was set low enough to saturate only the desired resonance. The presaturation period was set at 150 ms to avoid spin diffusion.

Acknowledgements

This research was supported by a grant (CA28681) from the National Cancer Institute and a Research Center in Minority Institution award (RR003037) from the Division of Research Resources, NIH (both to M.T.), and by grants from the Cancer Research Campaign, the Association for International Cancer Research, and the European Union (to M.J.W.). K.S.T. thanks the British Council, the Hong Kong British Trade Commission, the British Foreign and Commonwealth Office and the Cambridge Commonwealth Trust for awarding a Chevening Scholarship.

References

1. Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213.
2. Tate, P. & Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **3**, 226-231.
3. Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499-1504.
4. Zhang, X. & Mathews, C.K. (1994). Effect of DNA cytosine methylation upon deamination-induced mutagenesis in a natural target sequence in duplex DNA. *J. Biol. Chem.* **26**, 7066-7069.
5. Perwez Hussain, S. & Harris, C.C. (1998). Molecular epidemiology of human cancer: contribution of mutation spectra studies of tumor suppressor genes. *Cancer Res.* **58**, 4023-4037.
6. Jones, P., Rideout, W., Shen, J-C., Spruck, C. & Tsai, Y. (1992). Methylation, mutation and cancer. *Bioessays* **14**, 33-36.
7. Johnson, W.S., He, Q-Y. & Tomasz, M. (1995) Selective recognition of the m^{5}CpG dinucleotide sequence in DNA by mitomycin C for alkylation and cross-linking. *Bioorg. Med. Chem.* **3**, 851-860.
8. Millard, J.T. & Beachy, T.M. (1993). Cytosine methylation enhances mitomycin C cross-linking. *Biochemistry* **32**, 12850-12856.
9. Teng, S.P., Woodson, S.A. & Crothers, D.M. (1989). DNA sequence specificity of mitomycin crosslinking. *Biochemistry* **28**, 3901-3907.
10. Millard, J.T., Weidner, M.F., Raucher, S. & Hopkins, P.B. (1990). Determination of the DNA crosslinking sequence specificity of reductively activated mitomycin C at single nucleotide resolution: deoxyguanosine residues at CpG are crosslinked preferentially. *J. Am. Chem. Soc.* **112**, 3637-3641.
11. Borowy-Borowski, H., Lipman, R. & Tomasz, M. (1990). Recognition between mitomycin C and specific DNA sequences for cross-link formation. *Biochemistry* **29**, 2999-3004.
12. Denissenko, M.F., Chen, J.X., Tang, M.-S. & Pfeifer, G.P. (1997). Cytosine methylation determines hotspots of DNA damage in the human p53 gene. *Proc. Natl Acad. Sci. USA* **94**, 3893-3898.
13. Chen, J.X., Zheng, Y., West, M. & Tang, M.S. (1998). Carcinogens preferentially bind at methylated CpG in the p53 mutational hotspots. *Cancer Res.* **58**, 2070-2075.
14. Tonnassi, S., Denissenko, M.F., & Pfeifer, G.P. (1997). Sunlight induces pyrimidine dimers preferentially at 5-methyl cytosine bases. *Cancer Res.* **57**, 4727-4730.
15. Drouin, R. & Therrien, J. (1997). UVB-Induced cyclobutane pyrimidine dimer frequency correlates with skin cancer mutational hotspots in p53. *Photochemistry and Photobiology* **66**, 719-726.

16. Perakyla, M. (1998). A model study of the enzyme-catalyzed cytosine methylation using *ab initio* quantum mechanical and density functional theory calculations: pK_a of the cytosine N3 in the intermediates and transition states of the reaction. *J. Am. Chem. Soc.* **120**, 12895-12902.
17. Yebra, M.J. & Bhagwat, A.S. (1995). A cytosine methyltransferase converts 5-methylcytosine in DNA to thymine. *Biochemistry* **34**, 14752-14757.
18. Smith, S.S., Kaplan, B.E., Sowers, L.C. & Newman, E.M. (1992). Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation. *Proc. Natl Acad. Sci. USA* **89**, 4744-4748.
19. Bailly, C., Payet, D., Travers, A.A. & Waring, M.J. (1996). PCR-based development of DNA substrates containing modified bases: An efficient system for investigating the role of the exocyclic groups in chemical and structural recognition by minor groove binding drugs and proteins. *Proc. Natl Acad. Sci. USA* **93**, 13623-13628.
20. March, J. (1992). *Advanced Organic Chemistry.* (4th edn) pp.278-286, Academic Press, New York.
21. Harvey, R.G. (1981). Activated metabolites of carcinogenic hydrocarbons. *Accounts Chem. Res.* **14**, 218-226.
22. Tomasz, M., *et al.*, & Waring, M.J. (1998). The purine 2-amino group as the critical recognition element for sequence-specific alkylation and cross-linking of DNA by mitomycin C. *J. Am. Chem. Soc.* **120**, 11581-11593.
23. Millard, J.T., Spencer, R.J. & Hopkins, P.B. (1998). Effect of nucleosome structure on DNA interstrand cross-linking reactions. *Biochemistry* **37**, 5211-5219.
24. Sastry, M., Fiala, R., Lipman, R., Tomasz, M. & Patel, D.J. (1995). Solution structure of the monoalkylated mitomycin C-DNA complex. *J. Mol. Biol.* **247**, 338-359.
25. Kumar, S., Lipman, R., & Tomasz, M. (1992). Recognition of specific DNA sequences by mitomycin C for alkylation. *Biochemistry* **31**, 1399-1407.
26. Heineman, U. & Hahn, M. (1992). C-C-A-G-G-C m^{5}C-T-G-G helical fine structure, hydration and comparison with C-C-A-G-G-C-C-T-G-G. *J. Biol. Chem.* **267**, 7332-7341.
27. Mayer-Jung, C., Moras, D. & Timsit, Y. (1997). Effect of cytosine methylation on DNA–DNA recognition at CpG steps. *J. Mol. Biol.* **270**, 328-335.
28. Delepierre, M., Langlois, D., Estantoi, Igolen, J. & Roques, B.P. (1986). Conformational studies of d(m^{5}CpGpm^{5}CpG) and d(CpGpCpG) by ^{1}H and ^{31}P–NMR. *Eur. J. Biochem.* **161**, 571-577.
29. Hodges-Garcia, Y. & Hagerman, P. (1995). Investigation of the influence of cytosine methylation on DNA flexibility. *J. Biol. Chem.* **270**, 197-201.
30. Dipple, A. (1995). Commentary: DNA adducts of chemical carcinogens. *Carcinogenesis* **16**, 437-441.
31. Dennissenko, F.M., Pao, A., Tang, M.-s. & Pfeifer, G.D. (1996). Preferential formation of benzo *(a)* pyrene adducts at lung cancer mutational hotspots in p53. *Science* **274**, 430-432.
32. Hernandez, T.M. & Hainaut, P.A. (1998). Specific spectrum of p53 mutations in lung cancer from smokers review of mutations compiled in the IARC p53 database. *Environ. Health. Perspect.* **106**, 385-391.
33. Plateau, P. & Gueron, M. (1982). Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.* **104**, 7310-7311.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.