Structural, Steric and Energetic Requirements for Induction of Base Substitutional Mutations by Methylated Guanines and Thymines

Divi Venkateswarlu and R. H. Duncan Lyngdoh *
Department of Chemistry, North-Eastern Hill University, Bijni Complex, Bhagyakul, Shillong 793 003, India

The products of methylation at the N^3-, O^6- and N^7-positions of guanine and at the O^2- and O^4-positions of thymine are subjected to various possibilities for pairing with DNA bases, using calculations at the semiempirical PM3 SCF-MO level. It is predicted that the presence of the Watson–Crick protons in the modified bases would lead to non-mutagenic base-pairing schemes, while their absence facilitates promutagenic pairing schemes, modified guanines behaving like adenine and modified thymines like cytosine. Some degree of competition with non-mutagenic base-pairing schemes is also anticipated. Only the conformers of the *O*-methylated bases with the *O*-methyl group *anti* to the hydrogen bonding side furnish feasible base-mispairing schemes in the double-helical configuration. The *syn* conformers do not pair in the double-helical configuration. Correlation of these results with experimental and theoretically predicted Watson–Crick proton acidities for the nucleoside systems leads to the prediction that N^3- and O^6-methylguanines and O^2- and O^4-methylthymines would be promutagenic bases at biological pH, while N^7-methylguanine would behave without miscoding properties. These predictions are largely confirmed by the reported experimental template properties of these modified DNA bases and are also corroborated by NMR, UV and crystallography studies on some of the modified bases considered here.

Point mutations, including base-pair substitutions, serve as a basis for activation of latent proto-oncogenes to their carcinogenically active form, as documentated for many cases belonging to the *ras* and *neu* oncogene families.[1-3] DNA base alkylation by xenobiotic agents has been linked to point mutation and oncogene activation[4-7] and is believed to exert its genotoxic effect *via* the unusual base-pairing schemes adopted by the methylated base residues, which include the products of alkylation at the N^7-guanine (N^7-G), O^6-guanine (O^6-G), O^4-thymine (O^2-T), O^2-thymine (O^4-T) and N^3-guanine (N^3-G) sites.[8,9]

N^7-Methylguanine does not lead to misincorporation of a non-complementary base when present in templates for nucleic acid polymerases.[10,11] O^6-Methylguanine has been shown to possess the ability to misincorporate thymine residues when present in *in vitro* and *in vivo* templates for DNA and RNA synthesis.[12-15] O^4-Alkylthymines can similarly misincorporate guanine residues during nucleic acid synthesis, as demonstrated for O^4-methyl-, ethyl- and isopropyl-thymines.[16-19] The evidence for O^2-methylthymine is less conclusive,[20] while none exists for N^3-methylguanine. These findings may be summed up by proposing that N^7-methylguanine is non-mutagenic while O^6-methylguanine and O^4-methylthymine are pro-mutagenic, the position of O^2-methylthymine being unclear and that of N^3-methylguanine being unknown.

A Corollary of in vitro *and Theoretical Studies.*—Much *in vitro* and theoretical work has been done to corroborate and add finer detail to the above basic findings, which may be summarized in the following three paragraphs.

X-ray crystallography and NMR studies indicate that substitution of O^6-methylguanine (O^6-MeG) for guanine in synthetic DNA duplexes destabilizes the helical structure.[21] This has the effect of decreasing the T_m of DNA significantly.[22] That these disruptions take on the nature of only local perturbations is indicated by X-ray[23,24] and NMR studies.[25-28] The O^6-MeG–C and O^6-MeG–T pairs are both recognized by the repair enzyme ABC excinuclease *in vitro*,[29] which apparently recognizes the local helical perturbation rather than the lesion itself. Although the conformation adopted by the *O*-methyl group is *syn* to the hydrogen bonding side in the solitary deoxynucleoside[30] and in the single-stranded oligonucleotide, in the double-stranded duplex it goes out of plane, protruding into the major groove for the O^6-MeG–C pair.[31] Proton NMR studies on a right-handed helical dodecamer[25] indicate an *anti* conformation for the methyl group in an O^6-MeG–C pair. The crystal structure of a B-DNA oligonucleotide containing O^6-MeG also revealed an *anti* conformation for the methyl group, where the mismatched O^6-MeG–T pair is basically in Watson–Crick configuration.[32] NMR solution studies (reviewed in ref. 33) have found that the O^6-alkylguanine–T base-pair most likely retains the Watson–Crick alignment, while the O^6-MeG–C pair adopts a wobble configuration. UV melting studies point to a Watson–Crick configuration for the O^6-MeG–C pair provided the modified base is N^1-protonated.[32]

O^4-Alkylthymines are considered to be of greater relevance for carcinogenesis and mutagenesis than O^6-alkylguanines because of their greater resistance to repair.[34-36] A Watson–Crick type of alignment was proposed by Singer[37] for the pair between O^4-methylthymine (O^4-MeT) and thymine, linked by two hydrogen bonds, where the methyl group would not be *syn* to the hydrogen bonding side, but possibly in a conformation intermediate between *syn* and *anti*. This is in contrast to a 2D NMR solution structure obtained[28] which proposed only one hydrogen bond. Both O^4-MeT–G and O^4-MeT–A base pairs were formed by bacterial and fruitfly DNA polymerase fragments acting on a 25-meric oligonucleotide template containing O^4-MeT at a unique site.[38] The efficiency of forming the O^4-MeT–G pair was about ten times greater than that of forming the O^4-MeT–A or T–G pairs and bacterial T4 DNA polymerase allowed for stable incorporation of G opposite O^4-MeT in contrast to incorporation of G opposite T. The solitary nucleoside O^4-MedT has the methyl group *syn* to the hydrogen bonding zone,[39] but when O^4-MedTTP is placed opposite a poly(dA) template, the methyl group reorients to the *anti* conformation which is more favourable for base-pair formation.[40]

Pohorille and Loew[41] used a perturbation theory treatment

to study the base-pairing properties of some *O*-methylated bases, all of which were without the relevant Watson–Crick protons. This study predicted the role of conformation of the *O*-methyl groups for favourable base-mismatching in the double-helical configuration, where only the *anti* conformers were conducive to base-pairing of any kind. Ford and Scribner[42,43] examined the possibilities of interstrand proton transfer serving as a basis for mutagenesis following protonation and methylation at the N^7- and O^6-guanine sites using the AM1 SCF-MO method. Quantum chemical calculations[44] have indicated the *syn* conformer of O^6-MeG as being the most stable, while energy-minimization studies using the AMBER methodology show that within a pentameric oligonucleotide, the *anti* conformer is more stable.[44] The conformation and dynamics of oligonucleotides containing modified thymines have been studied using molecular dynamics simulations.[45-47] Parker *et al.*[48] used molecular dynamics simulation studies on oligonucleotides containing O^6-MeG to indicate that the O^6-MeG–T pair is of the Watson–Crick type structure, so that the sequence containing this pair is closer to the normal unmodified sequence than that containing the O^6-MeG–C pair.

Physical Basis for the Induction of Point Mutations.—For the alkylated DNA bases considered here, this physicochemical basis is prepared to include two components, *viz.* (*a*) abstraction of the Watson–Crick proton[49,50] (the N^1-proton for methylated guanines and the N^3-proton for methylated thymines) and (*b*) retention of the alkylating group in a conformation sterically conducive to base-mispairing in the double-helical configuration. The study of these two structural criteria form the objective of this study.

Criterion (*a*) is dealt with by studying each methylated base with and without the proton in question. Criterion (*b*) is incorporated for those cases with exocyclic methyl groups, by including two conformers for study; one with the methyl group *syn* to the hydrogen-bonding region and the other *anti*. The various possibilities arising from joint consideration of these two criteria are examined for their base-pairing properties to see which situations are predicted as feasible.

Theoretical

The equilibrium geometries and energies of the individual bases and base pairs were computed using the PM3 SCF-MO methodology[51] as incorporated in the MOPAC package,[52] full optimization being carried out by the Davidon–Fletcher–Powell algorithm. Two strategies were adopted for optimization of the base pairs; one retaining C_s symmetry for the pair and the other without any symmetry constraints. The enthalpy of pairing E_p, was obtained from the heats of formation of the pair and individual bases. Note was taken of the lengths, r_{HB} of the hydrogen bonds formed, and their number.

The intermolecular configuration of a base-pair was gauged by three markers; the distance r_{NN} between the sugar-bonding nitrogens of the two bases, the angle θ_{NH} between the N–H bonds at these nitrogens and the dihedral angle φ_{NH} between these two N–H bonds. Fig. 1 gives a schematic representation of the configurational markers r_{NN}, θ_{NH} and φ_{NH} used in defining the alignment between base pairs consisting of a purine and a pyrimidine.

Results and Discussion

Apart from the free bases guanine and thymine, five methylated bases were incorporated for study of their base-pairing properties: N^7-methylguanine (N^7-MeG), O^6-methylguanine (O^6-MeG), N^3-methylguanine (N^3-MeG), O^2-methylthymine (O^2-MeT) and O^4-methylthymine (O^4-MeT) whose mutagenic properties were summarized above. Each system was studied in two forms, *i.e.* with and without the Watson–Crick proton, respectively designated by a plus and zero sign, *e.g.* N^7-MeG$^+$ and N^7-MeG^0 for N^7-methylguanine. The *O*-methylated bases have *syn* and *anti* conformers for the methyl group (with respect to the hydrogen-bonding zone) and these are respectively designated by the symbols *s* and *a* in brackets, *e.g.* O^6-MeG$^+$(*a*) and O^6-MeG$^+$(*s*) for cationic O^6-methylguanine. Fig. 2 depicts the five bases in their protonated form with the *O*-methylated bases in their *anti* conformers. Fig. 3 gives the corresponding representations for the deprotonated bases, while Fig. 4 presents the *syn* conformers for the *O*-methylated bases.

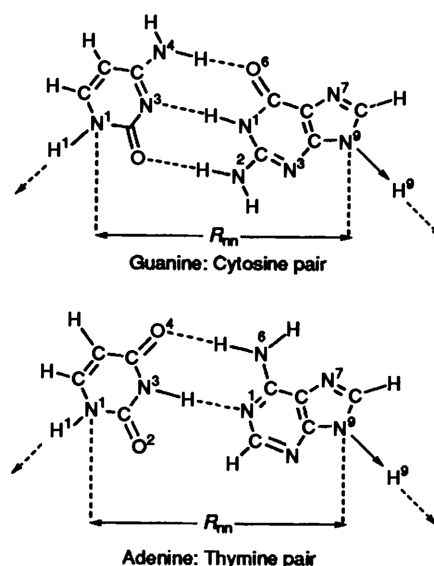

Fig. 1 Configurational marker r_{NN} for the G–C and A–T base pairs. Arrows indicate direction of the N^9–H^9 and N^1–H^1 bonds for purines and pyrimidines respectively, where θ_{NH} is the angle between these bonds for each pair and φ_{NH} is the dihedral angle between these bonds for each pair.

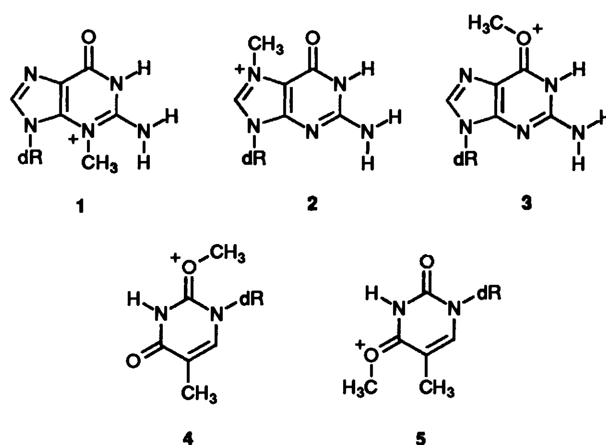

Fig. 2 *anti* Conformers of the five methylated DNA bases in their protonated form

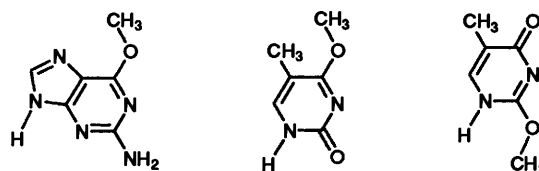

Fig. 3 *anti* Conformers of *O*-methylated DNA bases in their deprotonated form

Table 1 PM3 calculated data for pairing between cytosine and methylguanine systems which retain the N^1-proton

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
Freely optimized geometries						
G–C	−11.79	O^{6}G–H^{4}C	1.804	9.044	−7.00	57.68
		H^{1}G–N^{3}C	1.780			
		H^{2}G–O^{2}C	1.840			
N^3-MeG$^+$–C	−30.67	O^{6}G–H^{4}C	1.832	9.152	2.12	55.00
		H^{1}G–N^{3}C	1.737			
		H^{2}G–O^{2}C	1.740			
N^7-MeG$^+$–C	−23.41	O^{6}G–H^{4}C	1.851	9.081	3.05	52.85
		H^{1}G–N^{3}C	1.733			
		H^{2}G–O^{2}C	1.782			
O^6-MeG$^+$(*a*)–C	−23.33	O^{6}G–H^{4}C	1.929	9.053	−9.91	54.53
		H^{1}G–N^{3}C	1.745			
		H^{2}G–O^{2}C	1.778			
Geometries with C_s symmetry						
G–C	−11.09	O^{6}G–H^{4}C	1.794	9.115	0.00	53.39
		H^{1}G–N^{3}C	1.781			
		H^{2}G–O^{2}C	1.836			
N^3-MeG$^+$–C	−30.68	O^{6}G–H^{4}C	1.828	9.147	0.00	53.71
		H^{1}G–N^{3}C	1.738			
		H^{2}G–O^{2}C	1.742			
N^7-MeG$^+$–C	−23.80	O^{6}G–H^{4}C	1.843	9.076	0.00	51.63
		H^{1}G–N^{3}C	1.730			
		H^{2}G–O^{2}C	1.782			
O^6-MeG$^+$(*a*)–C	−23.54	O^{6}G–H^{4}C	1.924	9.082	0.00	50.34
		H^{1}G–N^{3}C	1.755			
		H^{2}G–O^{2}C	1.762			

Table 2 PM3 calculated data for pairing between cytosine and N^1-deprotonated methylguanine systems

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
Freely optimized geometries						
N^3-MeG^0–C	−3.33	O^{6}G–H^{4}C	1.840	9.250	16.41	54.64
		H^{2}G–O^{2}C	1.848			
N^7-MeG^0–C	−0.17	O^{6}G–H^{4}C	1.839	9.121	10.28	56.30
		H^{2}G–O^{2}C	1.875			
O^6-MeG^0(*a*)–C	−1.76	O^{6}G–H^{4}C	1.945	9.233	−11.95	51.63
		H^{2}G–O^{2}C	1.861			
Geometries with C_s symmetry						
N^3-MeG^0–C	−2.37	O^{6}G–H^{4}C	1.833	9.334	0.00	52.89
		H^{2}G–O^{2}C	1.869			
N^7-MeG^0–C	0.44	O^{6}G–H^{4}C	1.831	9.255	0.00	51.69
		H^{2}G–O^{2}C	1.879			
O^6-MeG^0(*a*)–C	−0.58	O^{6}G–H^{4}C	3.573	9.619	0.00	38.11
		H^{2}G–O^{2}C	1.835			

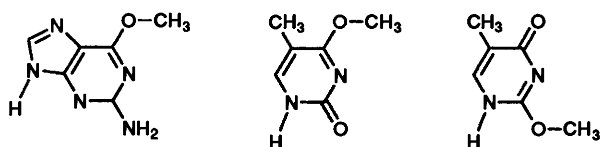

Fig. 4 *syn* Conformers of *O*-methylated DNA bases in their deprotonated form

Base-pairing Properties of Methylated Guanines.—The three methylated guanines in their protonated and deprotonated forms were paired with cytosine and thymine, with O^6-MeG in the *anti* conformation. Table 1 presents results for pairing of the protonated systems with cytosine, giving values of the pairing energy E_p, the three-configurational markers r_{NN}, θ_{NH} and φ_{NH} and the lengths r_{HB} of the hydrogen bonds formed, both optimization options being incorporated, *viz.* free optimization and imposition of C_s symmetry. Table 2 presents the same data for pairing of the protonated systems with thymine. Table 3 presents the data for pairing of the deprotonated guanine systems with thymine.

The data of Table 1 correspond to pH values of the surrounding medium where the base systems exist in a form with the N^1-proton present. From the negative values of the pairing energies E_p, it is immediately obvious that this presence of the N^1-proton allows for the modified guanine systems to pair favourably with cytosine as the free base does. In fact, the net positive charge on the methylated guanines leads to pairing energies even lower than that for the G–C base pair. This may be taken to indicate that, at the appropriately low pH, substitution of these methylated guanines for guanine would stabilize the double-helix, lowering the T_m value for the duplex. These pairing schemes are not mutagenic, not leading to a base substitution. This goes to indicate that the mere steric presence of the methyl group does not in itself constitute a molecular determinant for successful base-mismatching as long as the Watson–Crick proton is still present. The geometries of all the

Table 3 PM3 calculated data for pairing between thymine and N^1-deprotonated methylguanine systems

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
Freely optimized geometries						
N^3-MeG^0–T	−4.92	N^{1}G–H^{3}T	1.820	9.069	−38.60	60.30
		H^{2}G–O^{2}T	1.805			
N^7-MeG^0–T	−5.84	N^{1}G–H^{3}T	1.782	9.026	−41.25	59.40
		H^{2}G–O^{2}T	1.821			
O^6-MeG^0(*a*)–T	−5.68	N^{1}G–H^{3}T	1.830	9.610	−20.41	54.38
		H^{2}G–O^{2}T	1.816			
Geometries with C_s symmetry						
N^3-MeG^0–T	−2.63	N^{1}G–H^{3}T	1.857	9.195	0.00	53.83
		H^{2}G–O^{2}T	1.785			
N^7-MeG^0–T	−4.51	N^{1}G–H^{3}T	1.802	9.118	0.00	51.40
		H^{2}G–O^{2}T	1.792			
O^6-MeG^0(*a*)–T	−3.87	N^{1}G–H^{3}T	1.837	9.152	0.00	49.11
		H^{2}G–O^{2}T	1.793			

base-pair systems more or less conform to the double-helical configuration typified by the G–C pair. This is in line with the experimental evidence[32] for the Watson–Crick configuration of the O^6-MeG$^+$–C pair, where the O^6-methylguanine is N^1-protonated. The imposition of the C_s symmetry constraint does not lead to results or conclusions markedly differing from those obtained without any constraints.

Table 2 furnishes the data for the pairing between cytosine and the guanine systems minus their N^1-protons, which would in each case correspond to a pH value for the medium which is higher than those corresponding to the systems considered in Table 1. Since in each case, the base pairing with these guanine systems is cytosine, none of these pairing schemes may be described as mutagenic. The first observation is that the pairing energies in each case are appreciably less negative than those of Table 1, indicating lesser favourability of base-pairing. This is because of the loss of the N^1-proton resulting in the disappearance of one hydrogen bond. These neutral methylated systems are characterized by low pairing energies and only two hydrogen bonds. As such, these pairs would be expected to destabilize the DNA duplex at the appropriate pH, as indicated by the destabilizing effect noted experimentally[21,22] by substitution of O^6-MeG for guanine. The two types of optimization strategies lead to different pairing configurations, particularly for the O^6-MeG^0–C pair. If the fairly small degree of torsional twist φ_{NH} in the freely optimized geometries be regarded as accommodatable in the double-stranded situation, the values of the r_{NN} and φ_{HH} markers show some departure from the Watson–Crick alignment typified here by the G–C pair of Table 1. If the freely optimized geometries be taken as more representative of the real situation than the symmetry-constrained geometries, these results may be taken as predictive of a wobble type of pairing configuration for these neutral alkyl guanines with cytosine, which has been indicated experimentally[33] by NMR solution studies for the O^6-MeG^0–C pair

Table 3 gives the data for pairing between thymine and the deprotonated guanine systems, which pairings (like the pairings of Table 2) may be expected to occur at pH values higher than those for the systems in Table 1. Pairing between a guanine system and thymine would be mutagenic if successful, leading to a base transition (G ⟶ A base substitution). Here again, pairing occurs *via* two hydrogen bonds and the pairing energies are smaller for each case in comparison with the pairing between the protonated guanines and cytosine. But the pairing in each case is predicted to be favourable, as attested to by the negative values of E_p. The results obtained from free optimization and from the C_s symmetry constraint differ to some extent here. While the geometries derived from C_s symmetry imposition could be possibly accommodated into the Watson–Crick configuration, there is a large departure from this configuration for the freely optimized geometries. This raises the question of whether the latter could fit into the normal Watson–Crick double helix at all. If not, then one would have to depend upon the results obtained from C_s symmetry imposition to get a picture of what might be the actual situation prevailing in the normal double-helix. In this case, the small variations from Watson–Crick alignment as typified by the G–C pair (Table 1) may be interpreted in terms of the minor local perturbations in helical structure which have been indicated by experimental studies pertaining to this situation.[23–28] That the configuration for the O^6-MeG^0(*a*)–T pair is close to the Watson–Crick alignment is in line with the results of X-ray structure elucidation.[32] It is also in good accord with the predictions of molecular dynamics simulation studies[49] which indicate that the oligonucleotide sequence containing the O^6-MeG^0–T pair is closer to the normal one than that containing the O^6-MeG^0–C pair.

From the data of Tables 2 and 3, it is predicted that O^6-MeG^0 can pair with both cytosine and thymine, although the latter is definitely preferred energetically, being closer to the normal Watson–Crick configuration. This points to the competition observed between incorporation of thymine and incorporation of cytosine when both are present for incorporation into a DNA strand by action of a bacterial DNA polymerase.[53] The mutation efficiency has also been noted *in vivo* to be *ca.* 0.75,[15] indicating a definite preference for the mutagenic O^6-MeG^0–T pair over the non-mutagenic O^6-MeG^0–C pair.

Base-pairing Properties of Methylated Thymines.—The two methylated thymines in their protonated and deprotonated forms were paired with guanine and adenine, the methylated bases each in their *anti* conformation. Table 4 presents results for pairing of both the protonated and deprotonated systems with adenine, giving values of the pairing energy E_p, the three configurational markers r_{NN}, θ_{NH} and φ_{NH} and the lengths r_{HB} of the hydrogen bonds formed. Table 5 presents the same data for pairing of the deprotonated systems with guanine.

The data of Table 4 for protonated *O*-methylthymines pertains to pH values corresponding to retention of the N^3-T proton, where pairing proceeds *via* two hydrogen bonds. It is predicted that these protonated bases, while pairing favourably with adenine, adopt pairing configurations that deviate only somewhat from the double-helical one as typified by the A–T pair. None of these pairs are mutagenic, since the modified thymines pair with adenine, indicating that retention of the N^3-proton would allow for non-mutagenic pairing schemes. The

Table 4 PM3 calculated data for pairing between adenine and N^3-protonated thymine systems

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
Freely optimized geometries						
A–T	−6.10	O^{4}T–H^{6}A	1.823	9.125	6.64	42.46
		H^{3}T–N^{1}A	1.778			
O^4-MeT$^+$(*a*)–A	−12.75	O^{4}T–H^{6}A	2.523	8.843	−4.90	51.23
		H^{3}T–N^{1}A	1.721			
O^2-MeT$^+$(*a*)–A	−14.69	O^{4}T–H^{6}A	1.864	9.045	6.61	51.23
		H^{3}T–N^{1}A	1.683			
Geometries with C_s symmetry						
A–T	−5.61	O^{4}T–H^{6}A	1.811	9.134	0.00	52.60
		H^{3}T–N^{1}A	1.766			
O^4-MeT$^+$(*a*)–A	−10.93	O^{4}T–H^{6}A	2.432	8.888	0.00	50.63
		H^{3}T–N^{1}A	1.725			
O^2-MeT$^+$(*a*)–A	−14.90	O^{4}T–H^{6}A	1.860	9.044	0.00	45.53
		H^{3}T–N^{1}A	1.684			

Table 5 PM3 calculated data for pairing between adenine and N^3-deprotonated thymine systems

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
Freely optimized geometries						
O^2-MeT^0(*a*)–A	0.28	O^{4}T–H^{6}A	1.851	9.086	39.09	62.09
O^4-MeT^0(*a*)–A	1.85	O^{4}T–H^{6}A	2.848	9.186	11.63	45.99
Geometries with C_s symmetry						
O^2-MeT^0(*a*)–A	5.28	O^{4}T–H^{6}A	1.851	10.768	0.00	76.86
O^4-MeT^0(*a*)–A	3.11	O^{4}T–H^{6}A	3.117	9.936	0.00	47.39

Table 6 PM3 calculated data for pairing between guanine and N^3-deprotonated thymine systems

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
Freely optimized geometries						
O^2-MeT^0(*a*)–G	−1.62	N^{3}T–N^{1}G	1.973	9.297	21.16	50.82
		O^{2}T–N^{2}G	1.887			
O^4-MeT^0(*a*)–G	−6.13	N^{3}T–N^{1}G	1.868	9.014	−14.44	52.42
		O^{2}T–N^{2}G	1.833			
Geometries with C_s symmetry						
O^2-MeT^0(*a*)–G	−0.60	N^{3}T–N^{1}G	1.968	9.302	0.00	48.94
		O^{2}T–N^{2}G	1.906			
O^4-MeT^0(*a*)–G	−4.47	N^{3}T–N^{1}G	1.834	9.122	0.00	48.99
		O^{2}T–N^{2}G	1.858			

values of E_p are larger in magnitude for modified thymines than for normal thymine, leading to the prediction that, at the appropriately acidic pH, substitution of O^2- or O^4-methylthymine for thymine would result in stabilization of the double-helix. Imposition of C_s symmetry and free optimization both lead to essentially the same conclusions.

The data of Table 5 for the N^3-deprotonated bases pertain to higher pH values where the proton is abstracted and pairing with adenine has to occur through just one hydrogen bond. None of these pairs are predicted as energetically favoured, with C_s symmetry imposition and free optimization leading to the same result. The pairing configurations for these pairs appear to deviate even more from double-helical than those of Table 4. This all leads to the inference that O^2- and O^4-methylthymines in their N^3-deprotonated form would not form stable pairs with adenine in the double-helical configuration. The substitution of thymine by these *O*-methylthymines in a DNA duplex would thus be expected to destabilize markedly the helix locally.

The data of Table 6 constitute the possibilities for pairing between deprotonated *O*-methylthymines and guanine, both optimization strategies being used. These data pertain to a pH corresponding to abstraction of the N^3-protons and indicate potentially mutagenic base-pairing situations. The results do not depend much upon the optimization strategy employed. Both O^2- and O^4-methylthymines are seen to base-pair favourably with guanine, although the pairing energy for the former is appreciably lower than that for the latter, indicating that O^4-methylthymine might be a more efficient mutation-inducing agent than O^2-methylthymine. The base-pairing configurations for these systems are not far from double-helical, especially for the O^4-MeT^0(*a*)–G pair, indicating that these pairs could be accommodated into this configuration. These findings are basically in accord with the mutagenic base-pair between O^4-MeT and G as proposed by Singer [37] and do not agree with the single hydrogen-bonded base-pair proposed by Kalnik *et al.*[27] The *anti* conformation for the *O*-methyl group is predicted to be favourable to these mutagenic pairing schemes, as indicated by experimental studies.[41] That the O^4-MeT–G pair is clearly more favourable than the O^4-MeT–A pair (*cf.* Tables 5 and 6) is also predicted here, being in line with the relative efficiencies of formation of these pairs by action of polymerase fragments upon a 25-meric oligonucleotide template containing O^4-MeT.[39]

Table 7 PM3 calculated data for pairing between normal DNA bases and the *syn* conformers of *O*-methylated bases (C_s symmetry imposed in all cases)

System	E_p/kcal mol^{-1}	Hydrogen bond	r_{HB}/Å	r_{NN}/Å	φ_{NH} (°)	θ_{NH} (°)
O^6-MeG^0(*s*)–T	−4.25	N^{1}G–H^{3}T	2.840	9.610	0.00	35.36
		H^{2}G–O^{2}T	1.836			
O^2-MeT^0(*s*)–G	−6.40	N^{1}G–H^{3}T	2.428	9.958	0.00	60.91
		H^{2}G–O^{2}T	3.705			
O^4-MeT^0(*s*)–G	−4.63	N^{1}G–H^{3}T	2.693	9.555	0.00	36.56
		H^{2}G–O^{2}T	1.813			

Conformational Role of O-*Methyl Groups*.—The exocyclic *O*-methyl group of O^6-methylguanine, O^2- and O^4-methylthymines may exist in the *syn* or *anti* conformation, defined with respect to the hydrogen-bonding zone. The conformation adopted is of import for the feasibility of base-pairing in the double-helical configuration, as is shown by the data of Table 7 on the pairs formed between the *syn* conformers and normal DNA bases. Data on the *anti* conformers is represented by the preceding Tables.

When the *syn* conformers are considered, the pairs between O^6-methylguanine and thymine, O^2-methylthymine and guanine and O^4-methylthymine and guanine are predicted as energetically favourable, as seen from the negative values of E_p. The resultant geometries, however, indicate configurations far from double-helical. These results compare well with those of Pohorille and Loew[41] who predicted through perturbational calculations that the *syn* conformers of *O*-methylated bases would furnish strongly repulsive interactions with normal DNA bases in the double-helical configuration. This is not the case for the *anti* conformers, as the preceding Tables show. It is interesting to observe the general agreement between our results and previously reported structural studies,[25,31,32] including crystallographic, proton NMR and NMR solution studies, which all indicate that the methyl group of O^6-methylguanine adopts the *anti* conformation when faced with a pairing situation.

There are conflicting reports on the possible orientation of the methyl group in the base-pair between O^4-methylthymine and guanine. Singer and co-workers[38] postulated a base-pairing scheme where the O^4-methyl group lies intermediate between *syn* and *anti* positions. The 2D NMR studies of Kalnik *et al.*[28] proposed a different scheme in which the O^4-methyl group adopts the *syn* orientation. Our studies suggest that though the O^4-methyl group may adopt both orientations when pairing, the base-pair with the *syn* conformer clearly deviates from the double helical conformation (Table 7), while the one with the *anti* conformer closely approximates to the Watson–Crick alignment (Table 6).

The inference is that the *syn* conformers of these three *O*-methylated bases would not undergo mutagenic base pairing with the appropriate base under normal conditions (the usual Watson–Crick double-helical configuration). This means that they would not be of consequence for induction of base-substitutional mutations. A corollary of this is that the *syn* to *anti* rotational barrier would be a factor of relevance for the feasibility of an *O*-methylated base to induce point mutations. These concepts are currently being worked upon (Lyngdoh and Haorah, unpublished results) to explain the mutagenic and carcinogenic inactivity of all alkylating agents and *N*-nitroso compounds containing *tert*-alkyl groups.

Summary of Favoured Pairing Schemes.—In the data of Tables 1–6, the feasibility of the base-pairs predicted to be favourable in the double-helical configuration is decided on energetic grounds, eliminating those pairs which occur out of this configuration, so that the favoured pairs include the following (*a*)–(*d*).

(*a*) For protonated alkylguanines, pairing with cytosine [*viz.* the N^3-MeG$^+$–C, N^7-MeG$^+$–C, O^6-MeG$^+$(*a*)–C pairs] is favoured, all of which would be innocuous for mutation since they simulate the scheme present in the normal non-mutagenic G–C pair.

(*b*) For deprotonated alkylguanines, mutagenic pairing with thymine [*viz.* the N^3-MeG^0–T, N^7-MeG^0–T and O^6-MeG(*a*)–T pairs] is energetically favoured over non-mutagenic pairing with cytosine [*viz.* the N^3-MeG^0–C, N^7-MeG^0–C and O^6-MeG^0(*a*)–C pairs]. That both are nevertheless energetically allowed points to the possibility of competition between mutagenic and non-mutagenic pairing possibilities, the former being preferred.[53]

(*c*) For protonated alkylthymines, the *anti* conformers only being considered, pairing with adenine [*viz.* the O^2-MeT$^+$(*a*)–A and O^4-MeT$^+$(*a*)–A pairs] are favoured, both of which are of no consequence for point mutations since they mimic the scheme present in the normal non-mutagenic T–A pair.

(*d*) For deprotonated alkylthymines (only *anti* conformers) pairing with guanine is energetically favoured over adenine. Though the pairing of O^4-MeT^0 with guanine is energetically more favourable than O^2-MeT^0, both of them are significant for inducing point mutations.

Situation at Biological pH.—Out of the various favoured pairs mentioned in the last section, the question of which actually occur at biological pH depends upon whether the Watson–Crick proton in question (the N^1-proton for alkylguanines and the N^3-proton for alkylthymines) would be retained at this pH. The acidities of these protons have been linked to the pK_a of the base or nucleoside in question. This aspect has been treated in earlier studies (unpublished results) and the outcome of theoretical predictions coupled with experimental pK_a values is that at biological pH (close to neutral), the methylated base systems would exist as their deoxynucleosides in the forms summarized as follows.

N^3-Methylguanine, N^1-deprotonated (neutral), N^3-MeG^0.
O^6-Methylguanine, N^1-deprotonated (neutral), O^6-MeG^0.
N^7-Methylguanine, N^1-protonated (cationic), N^7-MeG$^+$.
O^2-Methylthymine, N^3-deprotonated (neutral), O^2-MeT^0.
O^4-Methylthymine, N^3-deprotonated (neutral), O^4-MeT^0.

The pairs that they would form at biological pH in the double-helical configuration thus include the following.

N^3-Methylguanine, N^3-MeG^0–C (non-mutagenic) and N^3-MeG^0–T (mutagenic).
O^6-Methylguanine, O^6-MeG^0(*a*)–C (non-mutagenic) and O^6-MeG^0(*a*)–T (mutagenic).
N^7-Methylguanine, N^7-MeG$^+$–C (non-mutagenic).
O^2-Methylthymine, O^2-MeT^0–G (mutagenic).
O^4-Methylthymine, O^4-MeT^0–G (mutagenic).

The inference from the above predictions is that N^3- and O^6-methylguanines could pair in mutagenic fashion (although competing with non-mutagenic pairing), while N^7-methyl-

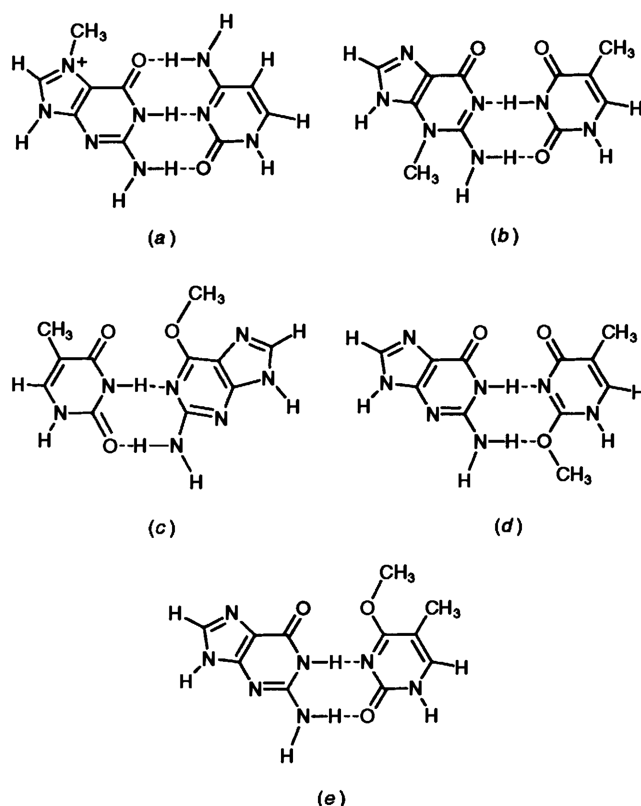

Fig. 5 Base-pairs expected to predominate at biological pH: (*a*) N^7-methylguanine (cationic) with cytosine; (*b*) N^3-methylguanine (neutral) with thymine; (*c*) O^6-methylguanine (neutral) with thymine; (*d*) O^2-methylthymine (neutral) with guanine; (*e*) O^4-methylthymine (neutral) with guanine

guanine would be non-mutagenic. The pairing properties of cationic N^7-methylguanine with thymine have been examined here using the same methodology, and although the pairing energy is negative (-9.29 kcal mol^{-1}),† the configurational parameters are such as to indicate that the potentially mutagenic pairing here would not be possible in the double-helical configuration chiefly owing to the presence of the N^1-proton. The values of the configurational markers are: r_{HB}, N^1-G–N^3-T = 2.874; r_{NN}, 9.28 Å; r_{HB}, N^2-G–O^2-T = 1.805; θ_{NH}, 48.53°.

O^2- and O^4-methylthymines also possess the ability to induce mutagenic pairing with guanine, while the possibility of non-mutagenic pairing with adenine is energetically unfavoured. The former is predicted to give a rather small pairing energy with guanine, while the latter pairs quite favourably with guanine, so that we may conclude that O^4-methylthymine would be more efficient in inducing base-substitutions than O^2-methylthymine. These two methylated bases differ from N^3- and O^6-methylguanines in that here the chances for mutagenic pairing competing with non-mutagenic pairing weighs heavily in favour of the former, so that the alkylthymines (especially O^4-methylthymine) may be expected to be much more specific in their inducement of base-substitutions than the alkylguanines.

Fig. 5 portrays the base-pairs which would be expected to predominate for each species at biological pH.

Conclusions

The criteria proposed for successful induction of base-substitutional mutations, *viz.* absence of the Watson–Crick proton and appropriate conformation of the methyl group in *O*-methylated bases, are seen to be borne out by the results of these base-pairing calculations. Application of theoretical and experimentally-based considerations on the acidities of the Watson–Crick protons in question lead to a correct assignment of mutagenic potential for N^7- and O^6-alkylguanines and the *O*-methylated thymines, furnishing also a prediction that N^3-methylguanine would be mutagenic.

† 1 cal = 4.184 J.

Acknowledgements

D. V. thanks the University Grants Commission, Government of India, for the award of a Senior Research Fellowship. R. H. D. L. is grateful for financial assistance from the Board of Research in Nuclear Sciences, Department of Atomic Energy, Government of India, through the research project No. 4/4/90-G. Thanks are due to the Computer Centre of this University, all calculations having been made on the VAX 3100-40 system.

References

1 H. Varmus, *Ann. Rev. Genet.*, 1984, **18**, 553.
2 C. I. Bargmann, M. C. Hung and R. A. Weinberg, *Cell*, 1986, **45**, 649.
3 C. I. Bargmann and R. A. Weinberg, *EMBO J.*, 1988, **7**, 2043.
4 P. D. Lawley, in *Chemical Carcinogens*, 2nd edn., ACS Monograph 173, Am. Chem. Soc., Washington DC, 1976, p. 83.
5 P. D. Lawley, in *Chemical Carcinogens*, 2nd edn., revised and expanded, ACS Monograph 182, Am. Chem. Soc., Washington DC, 1984, p. 485.
6 S. Hirani-Hojatti, J. R. Milligan and M. C. Archer, in *The Relevance of N-Nitroso Compounds to Human Cancer: Exposure and Mechanisms*, ed. H. Bartsch, I. K. O'Neill and R. Schulte-Hermann, IARC, Lyon, France, 1987, p. 26.
7 S. Sukumar, V. Notario, D. Martin-Zanca and M. Barbacid, *Nature*, 1983, **306**, 658.
8 B. Singer, W. J. Bodell, J. E. Cleaver, G. H. Thomas, M. F. Rajewsky and W. Thon, *Nature*, 1978, **276**, 85.
9 B. Singer, *Nature*, 1976, **264**, 333.
10 R. Schoental, *Biochem. J.*, 1969, **114**, 55.
11 D. B. Ludlum, *J. Biol. Chem.*, 1970, **245**, 477.
12 L. L. Gerchman and D. B. Ludlum, *Biochim. Biophys. Acta*, 1973, **308**, 310.
13 P. J. Abbott and R. Saffhill, *Biochim. Biophys. Acta*, 1979, **562**, 51.
14 J. R. Mehta and D. B. Ludlum, *Biochim. Biophys. Acta*, 1978, **521**, 770.
15 O. S. Bhanot and A. Ray, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 7348.
16 B. Singer, J. Sagi and J. T. Kusmierek, *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 4884.
17 B. Singer, S. J. Spengler, H. Fraenkel-Conrat and J. T. Kusmierek, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 28.
18 B. D. Preston, B. Singer and L. A. Loeb, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 8501.
19 B. D. Preston, B. Singer and L. A. Loeb, *J. Biol. Chem.*, 1987, **262**, 13 821.
20 R. Saffhill and P. J. Abbott, *Nucleic Acid Res.*, 1978, **5**, 1971.
21 J. M. Voigt and M. D. Topal, *Biochemistry*, 1990, **29**, 5012.
22 B. L. Gaffney, L. A. Marky and R. A. Jones, *Biochemistry*, 1989, **28**, 5881.
23 T. Brown, W. N. Hunter, O. Kennard, G. Kneale and D. Rabinovich, *Nature*, 1985, **315**, 604.
24 W. N. Hunter, T. Brown, N. N. Anand and O. Kennard, *Nature*, 1986, **320**, 552.
25 D. J. Patel, L. Shapiro, S. A. Kozlowski, B. L. Gaffney and R. A. Jones, *Biochemistry*, 1986, **25**, 1027.
26 D. J. Patel, L. Shapiro, S. A. Kozlowski, B. L. Gaffney and R. A. Jones, *Biochemistry*, 1986, **25**, 1036.
27 M. W. Kalnik, M. Kouchadjian, B. F. L. Li, P. F. Swann and D. J. Patel, *Biochemistry*, 1988, **27**, 100.
28 M. W. Kalnik, M. Kouchadjian, B. F. L. Li, P. F. Swann and D. J. Patel, *Biochemistry*, 1988, **27**, 108.
29 J. M. Voigt, B. Van Houten, A. Sancar and M. D. Topal, *J. Biol. Chem.*, 1989, **264**, 5172.
30 R. Parthasarathy and S. M. Fridey, *Carcinogenesis*, 1986, **7**, 221.
31 S. L. Ginell, S. Kuzmich, R. A. Jones and H. M. Berman, *Biochemistry*, 1990, **29**, 10 461.

32 G. A. Leonard, J. Thomson, W. P. Watson and T. Brown, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 9573.
33 B. F. Li, P. F. Swann, M. W. Kalnik, M. Kouchadjian and D. J. Patel, in *NMR Spectroscopy in Drug Research*, ed. J. W. Jaroszewski, K. Schaumberg and H. Koford, Munskgaard, Copenhagen, 1988.
34 J. A. Swenberg, M. C. Dyroff, M. A. Bedell, J. A. Popp, N. Huh, U. Kirstein and M. F. Rajewsky, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 1692.
35 F. C. Richardson, M. C. Dyroff, J. A. Boucheron and J. A. Swenberg, *Carcinogenesis*, 1985, **6**, 625.
36 T. P. Brent, M. E. Dolan, H. Fraenkel-Conrat, J. Hall, P. Karran, F. Laval, G. P. Margison, R. Montesano, A. E. Pegg, P. M. Potter, B. Singer, J. A. Swenberg and D. B. Yarosh, *Proc. Natl. Acad. Sci. USA*, 1988, **885**, 1759.
37 B. Singer, in *Carcinogenesis: Fundamental Mechanisms and Environmental Effects*, ed. B. Pullman, P. O. P. T'so and H. Gelboin, D. Reidel, Dordrecht, Holland, 1980, p. 91.
38 M. K. Dosanjh, J. M. Essigman, M. F. Goodman and B. Singer, *Biochemistry*, 1990, **29**, 4698.
39 B. D. Allore, A. Queen, W. J. Blonski and F. E. Hruska, *Can. J. Chem.*, 1983, **61**, 2397.
40 J. D. Engel and P. H. von Hippel, *J. Biol. Chem.*, 1987, **253**, 927.
41 A. Pohorille and G. H. Loew, *Biophys. Chem.*, 1985, **22**, 37.
42 G. P. Ford and B. Wang, *Int. J. Quantum Chem.*, 1992, **44**, 587.
43 G. P. Ford and B. Wang, *Int. J. Quantum Chem.*, 1992, **44**, 604.
44 L. G. Pederson, T. A. Darden, D. W. Deerfield II, H. M. W. Anderson and D. G. Hoel, *Carcinogenesis*, 1988, **9**, 1553.
45 *Molecular Dynamics: Applications in Molecular Biology*, ed. J. M. Goodfellow, Macmillan, London, 1991.
46 J. M. Goodfellow and M. A. Williams, *Curr. Opin. Struct. Biol.*, 1992, **2**, 211.
47 L. Cruzeiro-Hannson, P. F. Swann, L. Pearl and J. M. Goodfellow, *Carcinogenesis*, 1992, **13**, 2067.
48 K. Parker, L. Cruzeiro-Hannson and J. M. Goodfellow, *J. Chem. Soc., Faraday Trans.*, 1993, 2637.
49 R. H. Duncan and G. S. Davies, *J. Theor. Biol.*, 1989, **140**, 345.
50 R. H. D. Lyngdoh, *Proc. Indian Acad. Sci.* (*Chem. Sci.*), 1993, **105**, 253.
51 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 209; 2215.
52 J. J. P. Stewart, *QCPE Bull.*, 1983, **3**, 43.
53 P. J. Abbott and R. Saffhill, *Biochim. Biophys. Acta*, 1979, **562**, 51.

Paper 4/04291C
Received 14*th July* 1994
Accepted 24*th November* 1994