Volume 15 Number 24 1987 Nucleic Acids Research

Mechanisms of DNA sequence selective alkylation of guanine-N7 positions by nitrogen mustards

Kurt W.Kohn, John A.Hartley and William B.Mattes

Laboratory of Molecular Pharmacology, Developmental Therapeutics Program, Division of Cancer Treatment, National Cancer Institute, Bethesda, MD, USA

Received August 5, 1987; Revised and Accepted November 11, 1987

ABSTRACT

Quantitative determinations were carried out of the relative reaction rates of several nitrogen mustards at various guanine-N7 positions in DNA fragments of known sequence. The findings suggest structural hypotheses of the origins of the reaction selectivities. End-labeled DNA fragments were reacted with nitrogen mustards, and the guanine-N7 alkylation sites were analyzed by gel electrophoresis. Relative reaction intensities were determined by computer analysis of digitized densitometer scans. The differences in reaction intensities at different G's were in part attributable to the effects of nearest neighbor base pairs on the molecular electrostatic potential near the reaction site. Uracil and quinacrine mustards have specific sequence preferences for reaction that differ from other mustards. The nature of the specific sequence preferences were determined and hypotheses are proposed to explain their origin.

INTRODUCTION

The bis(2-chloroethyl)amines or nitrogen mustards are still among the most useful drugs in cancer chemotherapy, despite their apparently non-specific chemical reaction mechanisms. Nitrogen mustards may alkylate any accessible nucleophilic atom in nucleic acids, proteins or other biomolecules. The effective nitrogen mustards almost universally have two functional groups, suggesting that the anticancer action involves the formation of crosslinks between macromolecular sites.

DNA is probably the most important target of alkylation by nitrogen mustards, which react predominantly with the N7 positions of guanine residues. Bifunctional products have been obtained in which a nitrogen mustard molecule is bound to two guanine-N7 sites (1), and interstrand crosslinking has been shown by an inhibition of the ability of complementary strands to separate (2). Interstrand crosslinks may occur in the sequence 5'-GC-3' which permits a nitrogen mustard residue to span across the major groove between two guanine-N7 positions on opposite strands. Evidence for intrastrand crosslinking, presumably between adjacent guanines in the same strand, has

been obtained indirectly from the measured excess of diguanyl adducts relative to the number of interstrand crosslinks (3). Interstrand crosslinks and DNA-protein crosslinks have been observed in intact cells (4).

The specific biological effects of nitrogen mustards however may depend on preferential reaction at certain genomic locations. The following considerations suggest this possibility:

1) Nitrogen mustards react predominantly with guanines, and on this basis alone would be expected to react most strongly in G-rich regions. The production of interstrand or intrastrand crosslinks between G's in adjacent base-pairs in GC-rich regions would be especially favored.

2) Regions of unexpectedly high GC-richness are common features in genomes of warm-blooded vertebrates, although their function is unknown (5). The human genome contains an unusually prominent high-GC component (approximately 1.3% of the genome) which includes the c-sis, c-Ha-ras and c-abl oncogenes (6). The 5' flank of the Ha-ras gene, in particular, is very GC-rich.

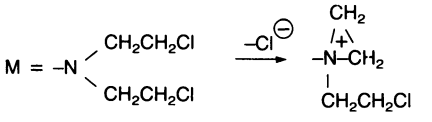
3) The genome of Epstein-Barr virus (EBV) has large regions of extraordinary GC-richness within the 3-kb repeats beginning approximately 3 kb from the replication origin (7; EMBL Nucleotide Sequence Data Library); these repeats have features suggesting an important control function (8). Multiple copies of the EBV genome are present in the nuclei of tumor cells of the endemic African form of Burkitt's lymphoma, and this form of lymphoma is extraordinarily sensitive to chemotherapy (9,10). Even one or two doses of a nitrogen mustard (cyclophosphamide) can produce dramatic regression of this tumor. Such extraordinary sensitivity to treatment with a single agent is not noted in the non-endemic form of the disease, which is rarely associated with EBV.

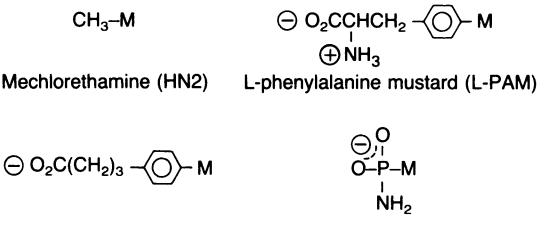
We have previously reported an additional mechanism that may further enhance the selective reaction of nitrogen mustards in G-rich regions (11). We had found that nitrogen mustards react preferentially with G's that are located in the interior of G clusters, possibly due to a strongly negative molecular electrostatic potential in these sequence configuration. The reactive group in the activated form of nitrogen mustards is a positively charged aziridinium group which would be drawn selectively towards the more electronegative regions (12,13). In our experiments, only reactions at guanine-N7, the predominant position of alkylation, are detected. The selective reaction at G's that are adjacent to other G's would favor the formation of intrastrand crosslinks.

We had found that most nitrogen mustards react only weakly at sites of 5′-GC-3′, which are the sites at which interstrand crosslinks are possible. Uracil mustard however deviated from this rule at some, but not all such sites. The reason for this behavior however could not be determined from the previously available data.

Another unusual pattern of reactivity was observed with quinacrine mustard, but again its exact nature and origin could not be determined.

We have now obtained additional reaction intensity data and have analyzed the intensity patterns largely automatically and objectively by means of a computer program. We have determined the nucleotide sequence dependencies of the reactions in some detail and propose structural hypotheses to explain the observed reaction preferences.



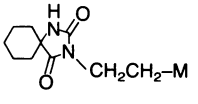


Mechlorethamine (HN2)

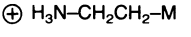
L-phenylalanine mustard (L-PAM)

Chlorambucil

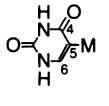
Phosphoramide mustard



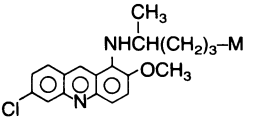
Spiromustine



Mustamine



Uracil mustard



Quinacrine mustard

Figure 1. Chemical structures of the nitrogen mustards used in this study.

Table 1: DNA regions analyzed

Experiment	DNA	Strand	Region analyzed	
031	pBR322	C	4070-4360	
022	"	W	400-600	
055/6	SV40	W	110-240	(72-bp repeat region)
044	hu-H-ras	W	1-170	(GC-rich 5′ flank)

MATERIALS AND METHODS

The nitrogen mustards (Fig. 1) were obtained through the Developmental Therapeutics Program, Division of Cancer Treatment, National Cancer Institute. The DNA restriction fragments listed in Table 1 were labeled with ^{32}P at a 3′ or 5′ end and then purified.

Alkylation reactions and polyacrylamide gel electrophoresis were carried out as previously described (11). In brief, labeled DNA was incubated with alkylating agent in 1 mM EDTA, 25 mM triethanolamine.HCl, pH 7.2 for 60 min at 22-24°C, except that in some experiments NaCl and/or $MgCl_2$ at the indicated concentrations were also included. The reaction was stopped by adding a chilled solution containing sodium acetate, EDTA and tRNA, and the DNA was precipitated with ethanol. The sites of guanine-N7 alkylation were quantitatively converted to strand breaks by treatment with 1 M piperidine at 90°C for 20 min (14). In the current work, alkylation at adenine sites was usually undetectable and always negligible. Polyacrylamide gel electrophoresis was carried out essentially as described by Maxam and Gilbert (15).

Autoradiograms of the electrophoresis gels were prepared using several exposures. Autoradiograms that were judged to be in the linear range of film response were selected for densitometric analysis. Results were checked by comparing different exposures.

Densitometric analysis

The selected autoradiograms were scanned by means of a Beckman model DU8B spectrophotometer equipped with gel scanner. The digitized optical density data were transmitted to a Hewlett-Packard model 9845B desk top computer. A program in BASIC was developed to process the data. The autoradiogram was cut into 4 sections, each less than 20 cm long, this being the limit of the gel scanner. The analysis by section also makes it easier to align each lane so as to scan a homogeneous path along the lane. The corrections for band overlap are done separately for each section. The computer program then merges the data for the different sections into a single data structure.

After a preliminary and conservative manual baseline correction, the program automatically finds and numbers peaks. Spurious peaks are deleted

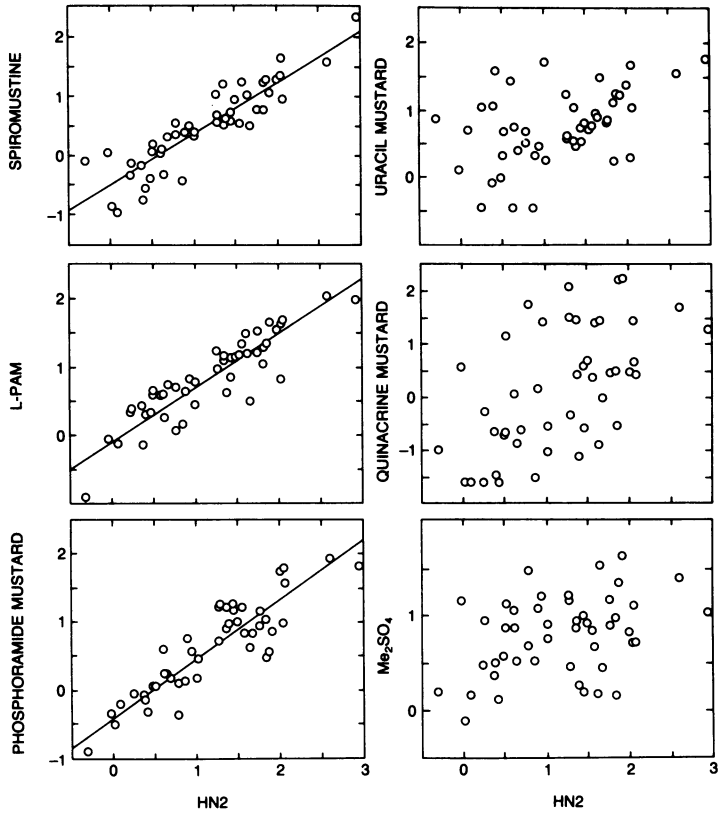


Figure 2. Reaction intensity correlations relative to HN2. Reaction intensities of individual G's with various nitrogen mustards (compound indicated on the vertical axis) are plotted against reaction with HN2. Scale is natural log units.

manually and missed peaks are added using cross hairs on the screen to digitize the positions. The program then reads the appropriate base sequence from a file; sequence numbers are assigned to the peaks by a partly manual, partly automatic procedure. The program then sums the optical density areas corresponding to each nucleotide, between boundaries determined by linear interpolation between successive known peaks. A correction is then applied for cutting between each given nucleotide and sites closer to the labelled end. Band overlap corrections are made on the basis isolated bands identified automatically at various positions in the gel. The program collects all of the isolated bands in all of the lanes analyzed on a given film and determines the overlap ratios to the adjacent nucleotide positions as a function of

migration distance. The overlap ratios in the two directions were treated separately, since the bands were generally not symmetrical. Linear regression of the overlap ratios in each section of the film was used to correct the intensities of all the nucleotides by means of an iterative procedure. Finally, the intensities for the G's were automatically corrected relative to background areas for nucleotides in guanine-free regions nearby.

RESULTS

Reaction intensity correlations

The intensity of reaction of a given drug at each guanine was compared with the reaction of mechlorethamine (HN2) as a standard. The reaction

Table 2: Reaction intensity correlations relative to HN2
(LN(intensity) units)

Expt	Drug	N	var	r	b
031	HN2	50	0.71		
022	"	64	0.83		
056	"	36	0.98		
044	"	55	0.84		
031	L-PAM	50	0.70	0.88	0.80
022	"	52	0.77	0.93	0.81
055/6	"	35	0.78	0.92	0.74
044	"	54	0.66	0.87	0.69
031	phosphoramide mustard	49	0.71	0.89	0.88
056	"	36	0.63	0.96	0.62
031	chlorambucil	48	0.59	0.81	0.66
031	spiromustine	47	0.71	0.90	0.86
031	mustamine	48	0.93	0.85	1.14
055/6	"	36	0.87	0.91	0.81
044	"	56	1.13	0.86	1.08
031	dimethylsulfate	48	0.41	0.37	
031	uracil mustard	48	0.55	0.46	
022	"	52	0.62	0.64	
055/6	"	35	0.63	0.52	
044	"	55	0.62	0.58	
031	quinacrine mustard	46	1.15	0.58	
022	"	57	1.24	0.59	
055/6	"	34	1.42	0.69	
044	"	55	1.32	0.46	

N = number of guanine intensity analyzed; var = variance of guanine intensities; r = correlation coefficient; b = slope of linear regression line.

intensities of several nitrogen mustards were closely correlated with each other; this is illustrated in Fig. 2 for spiromustine, L-PAM and phosphoramide mustard relative to HN2. However the reaction intensity patterns of uracil mustard, quinacrine mustard and dimethylsulfate are evidently different.

Table 2 summarizes results obtained in several DNA segments. Five mustards (spiromustine, L-PAM, phosphoramide mustard, chlorambucil and mustamine) correlated with HN2 with correlation coefficients (r) greater than 0.8. The correlation coefficients were substantially lower than this for uracil mustard, quinacrine mustard and dimethylsulfate. Thus 6 mustards show a consistent reaction pattern while three compounds diverge from this pattern.

The dynamic range of the reaction rate is indicated in Table 2 by the variance of the reaction intensities. For most mustards, the variance was typically in the range of 0.70-0.85 natural log units, which is equivalent to approximately a factor of 2. Higher variances were obtained with the intercalator quinacrine mustard and the lowest was seen with dimethylsulfate.

In the 72-base-pair repeat region of SV40 DNA, self-consistent reaction intensity data were obtained in repeats of the same sequence at different migration distances in the gel (Fig. 3).

Dependence of reaction intensities on molecular electrostatic potential (MEP)

The effects of nearest neighbor base pairs on MEP in the vicinity of

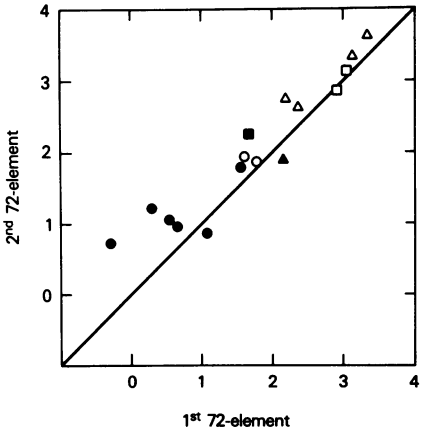


Figure 3. Correlation between the reaction intensities for HN2 at corresponding positions in the 72-bp repeat region of SV40 DNA (expt. 056). Reaction intensities in the first repeat (horizontal axis) are plotted against reaction intensities at the corresponding G in the second repeat (vertical axis). Units are ln(intensity). Symbols: (●) TGC; (△) GGG;(■) TGA; (○) TGG; (▲) AGC; (□) GGA.

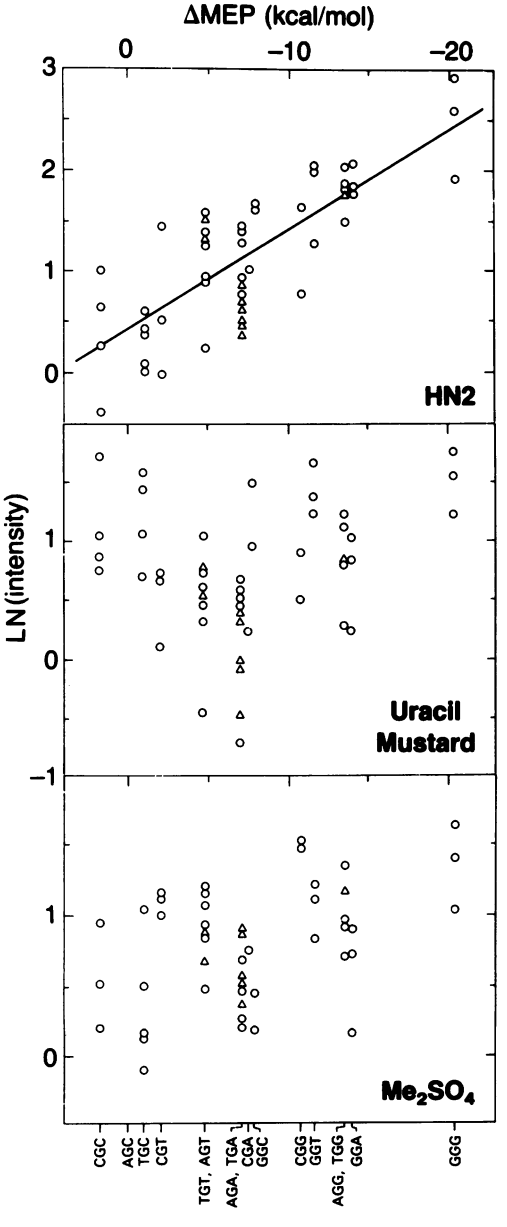


Figure 4. Dependence of reaction intensity on molecular electrostatic potential (MEP). The reaction intensities at various G-N7 positions in SV40 DNA are plotted against the effect of nearest neighbor base pairs on MEP as calculated by Pullman and Pullman (16). The corresponding base sequences, with the reacted G in the center, are marked on the bottom of the graph. Panels: top, HN2; middle, uracil mustard (UM); bottom, dimethylsulfate (DMS).

Table 3: Correlations of reaction intensities with effects of nearest neighbor base pairs on MEP (LN(intensity) units)

Expt	Drug	N	var	r	b
031	HN2	50	0.71	0.82	0.099
022	"	64	0.83	0.79	0.104
056	"	36	0.97	0.82	0.100
044	"	56	0.90	0.73	0.093
031	L-PAM	50	0.68	0.75	0.086
022	"	53	0.76	0.75	0.087
055	"	37	0.75	0.94	0.091
044	"	54	0.66	0.76	0.071
031	phosphoramide mustard	49	0.71	0.63	0.076
056	"	39	0.61	0.90	0.071
031	chlorambucil	50	0.58	0.77	0.077
031	spiromustine	50	0.73	0.77	0.095
031	mustamine	49	0.93	0.62	0.099
056	"	39	0.85	0.67	0.073
044	"	56	1.13	0.53	0.085
031	dimethylsulfate	50	0.53	0.31	...
031	uracil mustard	49	0.58	0.21	...
022	"	53	0.61	0.25	...
055	"	37	0.61	0.25	...
044	"	55	0.62	0.12	...
031	quinacrine mustard	49	1.17	0.54	...
022	"	59	1.22	0.58	...
055	"	36	1.43	0.89	0.163
044	"	55	1.32	0.65	0.121

N = number of guanine intensities analyzed; var = variance of guanine intensities; r = correlation coefficient; b = slope of linear regression line for reaction intensities versus MEP (b is in units of mol/kcal).

various positions in B-DNA have been calculated by Pullman and Pullman (16). We asked whether the values for the guanine-N7 position would correlate with the reaction intensities. The sum of the calculated effects of the base-pairs on the 3′ and 5′ sides of the reactive G was used to make the comparison.

Figure 4 shows examples of plots of reaction intensity against MEP. In the reaction with HN2, the most negative sequence, GGG, is seen to react most strongly, whereas the most positive sequences, which have a C on the 3′ side of the reacting G, are seen to react most weakly. This was true also for L-PAM, spiromustine, phosphoramide mustard, chlorambucil and mustamine, and in several different DNA segments. In the reaction with uracil mustard, however, the sequence 5′-GC-3′ often reacted as intensely as GGG.

Table 4: Effects of 2mM Mg and 100mM Na on reaction intensities
(Relative to TEA buffer. LN(intensity) units.)

Drug	Cation	N	var	r	b
HN2	TEA	36	0.98	...	...
"	+Mg	36	0.56	0.92	0.53
"	+Na	36	0.72	0.95	0.71
"	+Mg+Na	36	0.66	0.91	0.62
L-PAM	TEA	37	0.77	...	...
"	+Mg	37	0.49	0.83	0.52
"	+Na	35	0.57	0.95	0.72
"	+Mg+Na	36	0.53	0.94	0.64
phosphoramide mustard	TEA	38	0.63	...	...
"	+Mg	38	0.58	0.98	0.92
"	+Na	38	0.53	0.97	0.85
mustamine	TEA	39	0.87	...	...
"	+Mg	39	0.41	0.91	0.44
"	+Na	39	0.44	0.86	0.45
uracil mustard	TEA	37	0.63	...	...
"	+Mg	37	0.41	0.63	0.42
"	+Na	37	0.47	0.94	0.70
"	+Mg+Na	37	0.42	0.79	0.52
quinacrine mustard	TEA	36	1.40	...	...
"	+Mg	36	1.36	0.97	0.95
"	+Na	36	1.53	0.98	1.07
"	+Mg+Na	36	1.48	0.98	1.03

SV40 DNA, 72-bp repeat region, W-strand. TEA = 25 mM triethanolamine.HCl (pH 7.2); N = number of guanine intensities analyzed; var = variance of guanine intensities; r = correlation coefficient for reaction intensity in presence of added cation versus TEA only; b = slope of the linear regression of reaction in the presence versus the absence of cations.

The observed correlations between reaction intensity (expressed in natural log units) and MEP are summarized in Table 3. For most nitrogen mustards, the correlation coefficient (r) was usually in the range, 0.75-0.90. The slope of the linear regression line (b) was usually in the range, 0.085-0.100 mol/kcal. However, phosphoramide mustard and chlorambucil, which bear negatively charged groups, gave somewhat lower slopes. The special behaviors of uracil and quinacrine mustards will be considered in detail later.

Effects of cations

The experiments described above were carried out at low ionic strength in triethanolamine buffer. In order to distinguish between electrostatic factors and other factors that may influence reaction selectivity, the effects of 2 mM $MgCl_2$ and/or 100 mM NaCl in the solvent were examined. Increased ionic

Table 5: Band intensity correlations for uracil mustard relative to HN2

	All G's			5'-GC-3' excluded			
expt	N	var	r	N	var	r	slope
031	48	0.55	0.46	40	0.58	0.76	0.68
022	52	0.62	0.64	36	0.67	0.90	0.78
056	35	0.63	0.52	22	0.56	0.93	0.72
044	55	0.62	0.58	34	0.43	0.87	0.54

(Analysis similar to Table 1, showing the effect of excluding G's that are flanked on the 3' side by C. r = correlation coefficient.)

strength produced the expected general reduction in reaction rates, so that higher drug concentrations were needed to produce the experimentally optimum extent of reaction; our focus of attention however is on changes in the reaction selectivity among different G's.

Mg and Na did not substantially alter the reactivity ranking of G's; this is shown in Table 4 by the generally high values of r, the correlation coefficient for the presence versus the absence of cations. (The special properties of uracil and quinacrine mustards will be considered separately.)

Although the rank order of reactivities of the G's was preserved, the cations did alter the degree of selectivity. The degree of selectivity is measured by the slope (b) of the linear regression for reaction in the presence versus the absence of cations (Table 4). For HN2 and L-PAM, 2 mM Mg reduced the selectivity (i.e. slope, b) by approximately 50%. Na (100 mM) reduced the selectivity by only 30%, and the combination of Mg and Na had an intermediate effect. Phosphoramide mustard, which bears a negatively charged group, showed no reduction in selectivity by either Mg or Na. Mustamine, which bears a positive ammonium group, showed greater reductions in selectivity.

Sequence preferences specific for uracil mustard

The special reactivity pattern of uracil mustard was attributable to G's that have a C on the 3' side. When the instances of 5'-GC-3' were excluded from consideration, the remaining reaction intensities correlated well with HN2 (Table 5). Thus sequences other than 5'-GC-3' behave similarly to the case of other mustards.

However not all of the 5'-GC-3' sequences exhibited enhanced reactivity with uracil mustard. This is shown in a plot of reaction intensity with uracil mustard versus HN2 (Figure 5). The open symbols represent sequences other than 5'-GC-3'; the linear regression line for this set of points is shown and has a slope similar to that of other mustards. The solid triangles represent 5'-YGC-3' sequences, and the solid circles represent 5'-RGC-3'

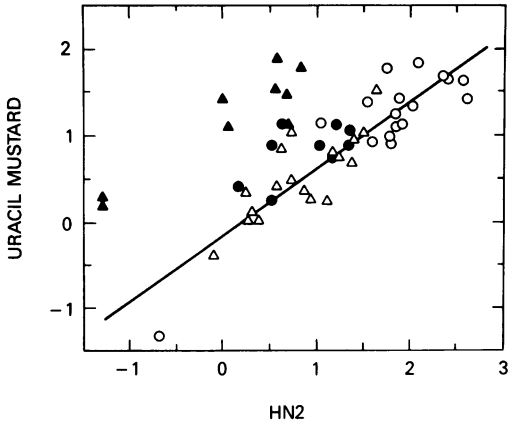


Figure 5. Strong reaction intensities specific for uracil mustard. The reaction intensities (in natural log units) for uracil mustard are plotted versus HN2 for a region of pBR322 DNA (expt. 022). Symbols: (○) RGD; (△) YGD; (●) RGC; (▲) YGC. (R = purine; Y = pyrimidine; D = not C)

sequences (Y=pyrimidine; R=purine). The 5′-YGC-3′ sequences show enhanced reaction with uracil mustard (relative to the regression line), whereas the 5′-RGC-3′ sequences show little or no enhancement.

A similar analysis was carried out for each of the DNA segments, and the combined results are shown in Table 6. It is evident that the peculiarity of uracil mustard is due almost entirely to enhanced reactivity with 5′-YGC-3′ sequences.

The reaction of uracil mustard with 5′-YGC-3′ sequences also has the characteristic that it is less suppressed by 2 mM Mg than are reactions at other sequence sites (Figure 6).

Table 6: Sequence-specific enhancement of reaction with uracil mustard
(Combined data for experiments 031, 022, 055/6 and 044.)

Reaction sites (G)	<+σ	>+σ <+2σ	>+2σ <+3σ	>+3σ	Totals
5′- Y G C -3′	2	3	5	29	39
5′- R G C -3′	9	6	7	1	23
5′- N G D -3′	106	24	2	0	132
Totals	117	33	14	30	194

σ = variance of deviations from the linear regression of LN(intensity) values for uracil mustard versus HN2, calculated for each experiment, excluding G's that are followed by 3′C. The table lists the number of reaction sites of each type falling within the indicated limits of σ. (R = purine; Y = pyrimidine; D = not C; N = any base.)

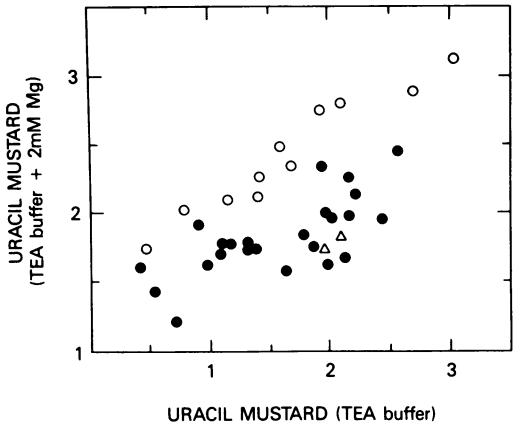


Figure 6. Effect of 2mM Mg on the reaction intensities of uracil mustard with various sites in SV40 DNA (72-bp repeat region, expt. 055). Units are ln(intensity). Symbols: (○) TGC; (△) AGC; (●) NGD. (D = not C, N = any base)

6-Methyluracil mustard behaved like most other mustards and did not exhibit the specific reaction preferences of uracil mustard (data not shown).

Sequence preferences specific for quinacrine mustard

The reaction preferences of quinacrine mustard were distinctive and unaffected by Mg or Na (Table 4). The sequence preferences of quinacrine mustard were discerned by sorting the sequences from highest to lowest reactivity. The sequence selectivity was very consistent and was found to depend on the 2 next bases in the 3′ direction from the reacting G (Table 7). The identity of the base on the 5′ side has no discernable effect. The reaction preference depends on the 2 bases on the 3′ side of the reacting G as

Table 7: Quinacrine mustard reaction intensities at various sequence configurations in pBR322 DNA. (Experiments 022 and 031.)

	expt. 022			expt. 031		
	n	LN(intensity)	delta-MEP	n	LN(intensity)	delta-MEP
GGR	11	1.99 ± 0.28	-14.9 ± 4.2	6	1.80 ± 0.33	-13.9 ± 4.0
GGY	15	0.55 ± 0.35	-12.8 ± 3.5	4	0.75 ± 0.33	-14.2 ± 2.9
GTR	6	1.65 ± 0.44	-7.4 ± 4.5	7	1.37 ± 0.42	-4.0 ± 2.9
GTY	2	0.34 ± 0.41	-7.2 ± 3.4	6	0.46 ± 0.17	-5.6 ± 3.7
GAR	1	0.51	-12.9	5	-0.57 ± 0.62	-8.9 ± 3.3
GAY	4	-0.39 ± 1.01	-8.0 ± 2.8	6	-0.67 ± 0.23	-6.1 ± 0.0
GCR	10	-0.90 ± 0.65	-4.1 ± 3.3	2	-0.11 ± 0.16	1.4 ± 1.4
GCY	10	-0.67 ± 0.83	-1.6 ± 4.3	5	-0.87 ± 0.48	0.3 ± 3.7

The reaction is at the underlined G; the 2 bases following in the 3′ direction are indicated (R=purine, Y=pyrimidine, n=number of sites).

follows:

GR = TR > GY = TY > AN = CN

(R = purine; Y = pyrimidine; N = any base).

DISCUSSION

Nitrogen mustards in aqueous solution decompose spontaneously with the loss of chloride to form a highly reactive aziridinium intermediate (12,13). The positively charged aziridinium group is the species that reacts with DNA. In the case of aromatic nitrogen mustards such as L-PAM, chlorambucil and uracil mustard, the aziridinium group may be less stable or may be distorted, compared with aliphatic mustards such as HN2, and this may affect the reaction mechanism (17). However the distinction between aliphatic and aromatic nitrogen mustards has little effect on the DNA sequence selectivity of the reactions at guanine-N7, since we observed no substantial differences in sequence selectivity between HN2 and L-PAM or chlorambucil.

One would expect the positively charged aziridinium to react preferentially with the most negative G's. The influence of the nearest neighbor base pairs on the molecular electrostatic potential (MEP) in the vicinity of guanine-N7 positions in B-DNA, as calculated by Pullman and Pullman (16), was found to correlate with the reaction intensities of six nitrogen mustards (including all of those studied, except uracil and quinacrine mustards, whose special behaviors will be discussed separately); as expected, the reaction intensities tended to increase with the degree of negativity of the MEP.

The nearest neighbor base pairs however do not fully account for the variations in reaction intensities, since there was often considerable variation among different G's having given base pair neighbors. The longer range sequence dependence remains to be determined. Such an analysis may be useful for mapping the electrostatic environment of DNA.

For HN2, L-PAM and mustamine, the degree of preference for the most negative G's was greatest at low ionic strength. The addition of Na or Mg cations reduced the degree of preference without any substantial change in the order of preference. Mg (2 mM) was more effective than Na (100 mM) in this regard. Interestingly, the addition of both cations together produced an intermediate rather than an additive effect. The effect of Mg may be due to specific ion binding to the DNA phosphodiester groups, whereas the effect of Na may be predominantly in the solvent environment of the DNA. When Na and Mg are present together, the Na may reduce the specific ion binding of Mg. The specific ion binding of doubly charged Mg cations could have an especially

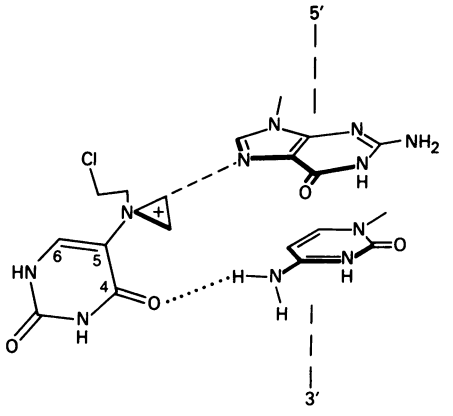


Figure 7. Possible configuration of uracil mustard (aziridinium form) at the initiation of the reaction with guanine-N7, showing the proposed interaction of the uracil-O4 atom with the N-H of a 3'-cytosine. The attacking aziridinium carbon is modelled at a van der Waals distance from the guanine-N7 atom and in the plane of the guanine ring.

strong effect on the molecular electrostatic potential, and this may explain the observed intermediate result with the combination of the two cations.

Several of the mustards used have positively or negatively charged groups which may affect the degree of the MEP-dependent sequence selectivity. Phosphoramide mustard and chlorambucil have negatively charged groups, and therefore their aziridinium intermediates would have no net charge. The negative groups however are some distance away from the aziridinium, and hence may not completely abolish the electrostatic interactions. Thus the magnitude of the MEP-dependent sequence selectivity of phosphoramide mustard and chlorambucil was only 20-30% less than that of HN2 and L-PAM (Table 3). The effect of Na or Mg however was markedly reduced (Table 4). In the case of mustamine, which has a positively charged amino group, and whose aziridinium intermediate therefore has two positive charges, the magnitude of the MEP-dependent selectivity at low ionic strength was not enhanced detectably (Table 3); however the suppression by Na or Mg was greater than in the case of HN2 or L-PAM (Table 4). Suppression by Mg therefore may be a good indicator of whether long-range electrostatics plays a major role in sequence selectivity.

Uracil mustard

The results with uracil mustard demonstrate that the substituent attached to the nitrogen mustard group can impose a distinct DNA sequence preference for reaction. The sequence preference specific for uracil mustard was found

to be 5′-YGC-3′ (Y = pyrimidine). The mechanism for the specific preference differs from the mechanism of reaction at other sites in being less dependent upon long-range electrostatic interactions, as shown by a lesser degree of suppression of the reaction by Mg ions.

A hypothesis to explain this specific sequence preference can be proposed. The reason that a 3′-C tends to suppress the G-N7 reaction of most mustards may be that the amino group of the 3′-C nearly overlies the G-N7, and the dipole of the amino N-H exerts a positive electrostatic field in the vicinity of the G-N7. Thus the neighboring base having the greatest positive influence on the MEP near the G-N7 is a 3′-C (16). Molecular modelling suggest that, as the aziridinium of uracil mustard approaches the G-N7, the uracil-O4 atom could interact with the N-H of the 3′-C, thereby countering the positive electrostatic influence of the latter (Figure 7).

Two observations remain to be expained: (1) the specific reaction preference requires a 5′-pyrimidine, and (2) the specific reaction preference is abolished by adding a methyl group to the 6 position of the uracil moiety.

Computer modelling shows that the geometry of the proposed interaction between the uracil-O4 and the cytosine N-H would be significantly improved by a displacement of the reacting guanine towards its sugar-phosphate backbone. When the guanine is situated between two pyrimidines, a displacement of the guanine in this direction would be anticipated on the basis of the sliding displacement of the central base pair predicted by the Calladine-Dickerson rules (18). The hypothesized displacement would arise because of a steric clash between the central guanine and the two propeller-twisted purines on the opposite strand in the adjacent base pairs. Thus the interaction geometry between the uracil-O4 and the cytosine N-H may only be adequate for enhanced reaction if the reacting guanine has pyrimidines on both sides.

Computer modelling also suggested an explanation for the suppressive effect of a methyl group on the uracil 6 position. The interaction between the uracil-O4 and the cytosine N-H would be favored by a rotation about the C5-N5 bond of the uracil mustard, and a methyl group on the uracil 6 position could block this rotation because of steric clash with the chloroethyl group.

If this hypothesis is principally correct, it should be possible to design new nitrogen mustards that have either an enhanced or an altered sequence preference.

Quinacrine mustard

Quinacrine mustard is capable of intercalating in DNA prior to covalent reaction with a guanine-N7. A rapid initial non-covalent binding is indicated

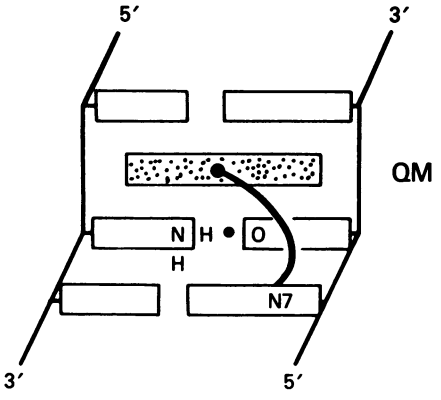


Figure 8. Schematic view into the major groove, showing a quinacrine mustard (QM) molecule intercalated between the first and second base pair 3′ to a guanine-N7 reaction site. The hydrocarbon side chain which tethers the mustard group to the quinacrine ring system is shown curving over an intervening thymine residue.

by the unusually low concentrations required for reaction. Of all the mustards studied, quinacrine mustard showed the largest discrimination among different guanines. The degree of discrimination among guanines was not reduced by Mg or Na cations, suggesting that the sequence discrimination is not due to long-range electrostatic interactions.

The reaction intensity was found to depend upon the two bases in the 3′ direction, and not on the identity of the 5′ base. This can be rationalized on the basis of the right-handed twist of B-DNA. Models show that simultaneous intercalation and guanine-N7 alkylation by quinacrine mustard can only occur if the intercalation site is 3′ to the alkylation site (Figure 8).

The strongest determinant of intense reaction was found to be the presence of a G or a T immediately 3′ to the reacting G, the two being equally effective. A possible reason for this emerges from the examination of a CPK space-filling model. With the mustard group linked to a guanine-N7 and the quinacrine ring intercalated between the first and second base pairs in the 3′

5′. . .TT┌CCATT┌CCA . . . 3′
3′. . .AA└GGTAA└GGT . . . 5′

Figure 9. Regularly spaced strong quinacrine mustard sites in a repeated DNA family (DYZ1) prominent in human Y chromosomes.

direction, the hydrocarbon side chain connecting these two parts of the molecule is closely apposed to the edge of the intervening 3′ base in the major groove (Figure 8). If this base is A or C then the amino group of the base would be situated behind the hydrocarbon chain. This may be a relatively unstable configuration compared to the case of G or T where, instead of an amino group, there would be an oxygen atom.

Given G or T as the first base on the 3′ side, then the reaction is more intense if the second base is a purine than if it is a pyrimidine. If the first base is an A or a C then the reaction is weak, and the nature of the second base has no apparent effect. The second base may affect the stability of intercalation of the quinacrine ring system.

Although it remains to be determined to what degree the sequence preferences are preserved in chromatin, this possibility is suggested in the case of quinacrine mustard by the insensitivity of the reaction preferences to solvent cations.

In the use of quinacrine mustard as a stain for chromosome banding, it has been noted that one of the most intensely stained regions is a heterochromatic area in the human Y chromosome (19,20) which has been found to contain 3-10x10^5 tandem repeats of the pentanucleotide TTCCA (21). This arrangement could accommodate regularly spaced strong quinacrine mustard sites at close to maximum density near saturation in the staining reaction (Figure 9) and thus could explain the intense staining.

REFERENCES

1. Brookes,P. and Lawley,P.D. (1961) Biochem. J. 80, 496-503.
2. Kohn,K.W., Spears,C.L. and Doty,P. (1966) J. Mol. Biol. 19, 266-288.
3. Chun,E.H.L., Gonzales,L., Lewis,F.S., Jones,J. and Rutman,R.J. (1969) Cancer Res. 29, 1184-1194.
4. Ewig,R.A. and Kohn,K.W. (1977) Cancer Res. 37, 2114-2122.
5. Bernardi,G., Olofsson,B., Filipski,J., Zerial,M., Salinas,J., Cuny,G., Meunier-Rotival,M. and Rodier,F. (1985) Science 228,953-958.
6. Zerial,M., Salinas,J., Filipski,J. and Bernardi,G. (1986) Eur. J. Biochem. 160, 479-485.
7. Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Suguin,C. et al. (1984) Nature 310, 207-211.
8. Karlin,S. (1986) Proc. Natl. Acad. Sci. USA 83, 6915-6919.
9. Burkitt,D. (1967) Cancer 20, 756-759.
10. Ziegler,J.L. (1981) N. Engl. J. Med. 305, 735-745.
11. Mattes,W.B., Hartley,J.A. and Kohn,K.W. (1986) Nucleic Acids Res. 14, 2972-2987.
12. Levins,P.L. and Papanastassiou,Z.B. (1965) J. Amer. Chem. Soc. 87, 826-831.
13. Williamson,C.E. and Witten,B. (1967) Cancer Res. 27, 33-38.

14. Mattes,W.B., Hartley,J.A. and Kohn,K.W. (1986) Biochim. Biophys. Acta 868, 71-76.
15. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
16. Pullman,A. and Pullman,B. (1981) Q. Rev. Biophys. 14, 289-380.
17. Bardos,T.J., Datta-Gupta,N., Hebron,P., and Triggle,D.J. (1965) J. Med. Chem. 8, 167-174.
18. Dickerson,R.E. (1983) J. Mol. Biol. 25, 419-441.
19. Caspersson,T., Zech,L. and Johansson,C. (1970) Exp. Cell Res. 60, 315-319.
20. George,K.P. (1971) Stain Technol. 46,34-36.
21. Nakahori,Y., Mitani,K., Yamada,M., and Nakagome,Y (1986) Nucleic Acids Res. 14, 7569-7580.