Local base sequence preferences for DNA cleavage by mammalian topoisomerase II in the presence of amsacrine or teniposide

Yves Pommier*, Giovanni Capranico+, Ann Orr and Kurt W.Kohn
Laboratory of Molecular Pharmacology, Division of Cancer Treatment, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

ABSTRACT

Several classes of antitumor drugs are known to stabilize topoisomerase complexes in which the enzyme is covalently bound to a terminus of a DNA strand break. The DNA cleavage sites generally are different for each class of drugs. We have determined the DNA sequence locations of a large number of drug-stimulated cleavage sites of topoisomerase II, and find that the results provide a clue to the possible structure of the complexes and the origin of the drug-specific differences. Cleavage enhancements by VM-26 and amsacrine (m-AMSA), which are representative of different classes of topoisomerase II inhibitors, have strong dependence on bases directly at the sites of cleavage. The preferred bases were C at the 3′ terminus for VM-26 and A at the 5′ terminus for m-AMSA. Also, a region of dyad symmetry of 12 to 16 base pairs was detected about the enzyme cleavage positions. These results are consistent with those obtained with doxorubicin, although in the case of doxorubicin, cleavage requires the presence of an A at the 3′ terminus of at least one the pair of breaks that constitute a double-strand cleavage (Capranico et al., Nucleic Acids Res., 1990, 18: 6611). These findings suggest that topoisomerase II inhibitors may stack with one or the other base pair flanking the enzyme cleavage sites.

INTRODUCTION

Eukaryotic DNA topoisomerase II catalyses the interconversion of DNA topoisomers by generating a transient DNA double-strand break in a DNA segment through which another segment of DNA duplex passes before the break is resealed (reviewed in 1, 2). Topoisomerase II-mediated double-strand breaks are staggered by 4 base pairs, and each 5′-terminus is covalently bound to a tyrosine residue of the enzyme (3, 4) (Fig. 1). Topoisomerase II-mediated DNA breaks are stabilized by certain anticancer drugs, such as doxorubicin, m-AMSA and VM-26, and each type of drug stabilizes a distinctive set of cleavage sites in a given segment of DNA (5—7). The nuclear matrix associated region (MAR) of the SV40 genome contains many cleavage sites stimulated by anthracyclines, m-AMSA or VM-26, as well as in the absence of drug (7—9), and therefore was selected for investigation of cleavage sites at the DNA sequence level.

Recently we have found that doxorubicin-stabilized sites invariably have an adenine (A) residue at the 3′-terminus of at least one of each pair of strand breaks that would constitute a topoisomerase II double-strand scission (10). This suggested that doxorubicin inhibits the enzyme by binding at one of the two cleavage sites of a topoisomerase II double-strand break. The present study demonstrates analogous but distinctive local DNA sequence requirements for m-AMSA and VM-26. We conclude that the general structure of stabilized ternary complexes of drug, topoisomerase II and DNA may involve the direct interaction of a drug molecule with a base pair immediately adjacent to a cleavage site.

MATERIALS AND METHODS

Drugs, enzymes, and chemicals

m-AMSA (amsacrine: 4′(9-acridinylamino)methanesulfon-m-anisidide) and teniposide (VM-26: 4′-demethylepipodophyllo-toxin-9-(4,6-O-thionylidine-β-D-glucopyranoside) were obtained from the Drug Synthesis and Chemistry Branch, National Cancer Institute, Bethesda, MD. Stock solutions were made in dimethyl-sulfoxide at 10 mM immediately before use. SV40 DNA, restriction endonucleases, T4 polynucleotide kinase, and agarose were purchased from Bethesda Research Laboratories (Gaithersburg, MD). Calf intestine phosphatase was purchased from New England Biolabs (Beverly, MA) and [gamma-^{32}P]-ATP from New England Nuclear Research Products (Boston, MA). DNA topoisomerase II was purified from mouse leukemia L1210 cell nuclei as described previously, and was stored in small aliquots at −70°C in 40% (v/v) glycerol, 0.35 M NaCl, 5 mM

* To whom correspondence should be addressed at Building 37, Room 5C27, National Institutes of Health, Bethesda, MD 20892, USA

+ Present address: Istituto Nationale per lo Studio et la Cura dei Tumori, via G. Venezian 1, 20133 Milan, Italy

$MgCl_2$, 1 mM EGTA, 1 mM KH_2PO_4, 0.2 mM dithiothreitol and 0.1 mM phenylmethanesulfonyl fluoride, pH 6.4. The purified enzyme yielded a single 170 kDa band after silver staining of SDS-polyacrylamide gels (11, 12).

DNA sequencing of topoisomerase II cleavage sites

SV40 DNA fragments were uniquely 5' end-labeled as already described (7, 13) and then purified by electroelution and ethanol precipitation. DNA fragments were reacted with 40−70 ng of topoisomerase II with or without drug in 0.01 M Tris-HCl, pH 7.5, 0.05 M KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM ATP and 15 μg/ml bovine serum albumin for 20 min at 37°C. m-AMSA and VM-26 were used at 10 μM (final concentrations). Reactions were stopped by adding sodium dodecyl sulfate (SDS), EDTA and proteinase K (1%, 10 mM and 250 μg/ml, respectively) and samples incubated for 1 h at 42°C. Samples were ethanol-precipitated and resuspended in 2.5 μl loading buffer (80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue). They were heated at 90°C before loading into DNA sequencing gels (8% polyacrylamide; 29:1, acrylamide:bis) containing 7 M urea in 1×TBE buffer. Most of the cleavage sites used for the analysis are described elsewhere (7).

Statistical tests

The Chi-square one-sample test was used to determine the deviation from the random distribution of bases at each position of the aligned sequences. The expected occurrence of each base averaged over the entire SV40 DNA was A=T=0.296 and C=G=0.204 (10).

Deviation from random distribution for each base at a given position was evaluated by determining the confidence interval for a 0.001 probability of occurrence (P).

Knowing the frequency of occurrence of a given base in the overall SV40 DNA (p=0.296 for A or T, p=0.204 for C or G), and n, the number of sites analyzed, confidence interval intervals were calculated as:

$$p \pm 3.1 \ S.D.$$

where *S.D.* is the standard deviation of the Normal base distribution, calculated as:

$$S.D. = [\ p \ (1-p)/n \ ]^{1/2}$$

Base proportions outside confidence intervals were considered significant (P<0.001). When their value was above the upper limit of the confidence interval they were 'preferred', and when they were below, they were 'excluded'.

Probability calculations

We wish to determine the base preferences attributable to topoisomerase-induced DNA cleavage and estimate the confidence levels for bias in favor or against each possible base at each position relative to the cleavage site. We do this by looking for deviations of the observed from the expected number of occurrences of each possible base. In a set of DNA cleavage sites, we determine, for each position relative to the cleavage site, the number of sites which have a particular base at that position. We then calculate the probability of the chance occurrence of a deviation from the expected number equal to or greater than the observed deviation. This is done for every possible base at every position. In a previous study (10), we calculated the expected number on the basis of the overall base frequencies in

the SV40 genome. In the current data, however, we encountered for the first time a strong preference for a C which, because of the overall deficiency of the CG dinucleotide, resulted in a strong spurious deficiency of G at the following position. We therefore modified our method of calculation to take into account the natural bias in dinucleotide frequencies in particular genomic regions.

We now calculate the probability of occurrence of each possible base taking into account the nearest neighbor bases. Let f(a,x,b) be the probability of finding the base, x, between given upstream and downstream bases, a and b. This set of 64 conditional probabilities was determined for a region of 2000 bases at the 5' or 3' end of the SV40 genome or in the center of the genome, depending upon where the individual sites were located. Thus,

$$f(a,x,b) = \overset{4}{\underset{i=1}{\Sigma}} \ n(a,x,b) \ / \ SUM \ n(a,i,b)$$

where n(a,i,b) is the number of occurrences of the base sequence a,i,b in the region of 2000 bases. (The bases A, C, G, T are represented by the numbers 1, 2, 3, 4.) The expected number of occurrences, e(x), of each possible base at a given position relative to the cleavage site was calculated by summing the conditional probabilities for the individual observed sites: for each possible base, x, the f(a,x,b) were summed over all the sites, using the individual values of a and b for each site.

The probability, p, of a particular base at a particular position was taken as the expected number of occurrences divided by the number of sites (p = e(x)/n). The probabilities of chance deviations between observed and expected numbers of occurrences of each base were then determined as follows (10). The expected number of sites having a given base at any particular location is *pn*. Let *m* be the observed number of sites which have the given base at that location.

If *m* > *pn*, the probability, P, of the chance occurrence of *m* or more instances was calculated as:

$$P = \overset{n}{\underset{i=m}{\Sigma}} \ \frac{n!}{i!(n-i)!} \ p^i(1-p)^{n-i}$$

If *m* < *pn*, then the chance occurrence of m or fewer instances was calculated as

$$P = \overset{m}{\underset{i=0}{\Sigma}} \ \frac{n!}{i!(n-i)!} \ p^i(1-p)^{n-i}$$

Factorials were computed to 10-digit precision as their logarithms using the Lanczos approximation (14). This method of calculating P has the advantage of being valid even for small values of m, not being subject to approximation errors, and providing significance estimates for both base excess and base deficiency.

RESULTS

Nucleotide bias in the vicinity of topoisomerase II cleavage sites stimulated by VM-26 or *m*-AMSA

The base sequences of a total of 618 topoisomerase II DNA cleavage sites in several regions of the SV40 genome were analyzed. The sites included 323 sites stimulated by VM-26, 197 sites stimulated by *m*-AMSA, and 98 sites produced by the enzyme in the absence of drug (7). DNA sequences were aligned at the points of observed phosphodiester bond cleavage in the

5'→ 3' orientation; the bases immediately 5' and 3' to the analyzed break point were numbered −1 and +1, respectively (Fig. 1). At each position relative to the cleavage site, deviations in the base distributions from the global SV40 DNA base frequencies were evaluated by Chi-square analysis.

Figure 2 shows the existence of a region of non-random base composition between positions −3 and +7 for cleavage sites stimulated by VM-26, or by m-AMSA, as well as for cleavage sites induced by the enzyme in the absence of drug. The most significantly biased positions are +1 for m-AMSA, −1 for VM-26 and −1 for enzyme without drug. In the case of VM-26, a significant bias is seen also at position +5, which is the complementary equivalent of position −1 with respect to the dyad symmetry of double-strand cleavage by topoisomerase II (Fig. 1).

The nature of the base preference in the vicinity of the cleavage sites was investigated by a probability analysis similar to that described previously (10), but modified to take into account the identity of the immediate upstream and downstream bases at each position (see Materials and Methods).

VM-26

The strongest base preferences corresponded to the dyadic pair, C at −1 and G at +5 (Fig. 3). Moreover, the combination C(−1) and G(+5) was present in nearly 40% of the most intense sites (intensity 3), but in only 6% of the weakest sites (intensity 1) (Table 1). Conversely, C(−1) and G(+5) were both absent in more than 50% of the intensity 1 sites, but in only 20% of the intensity 3 sites. These differences are statistically very significant. Thus, VM-26-stimulated cleavage is favored by C at the 3' terminus of the observed cleavage site and also by C at the 3' terminus of the potential break position in the opposite strand.

Even though there were no additional strong preferences for individual bases in the vicinity of the cleavage sites, there was a significant symmetry of the most biased bases (Fig. 2B, middle panel). A dyad symmetry centered between positions +2 and +3 is observed for 6 out of 8 base comparisons in the region −6 to +10 (p < 0.01), the most positively biased bases being

```
 −6            −1 +1      +4 +5              +10
  A  C  G  A  A  C  A . C  C  G  T  T  C  G  T
```

This symmetry could be a true dyad symmetry for individual DNA molecules; alternatively, the sequence recognition may
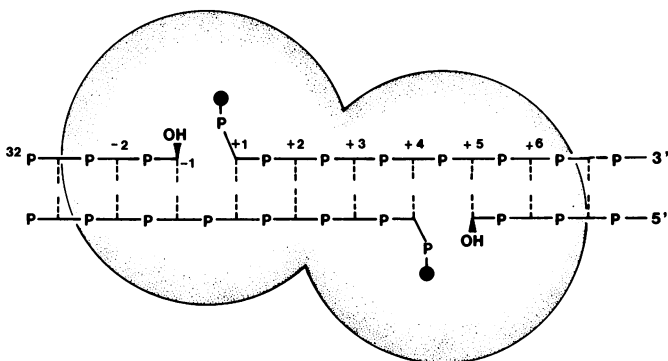
occur at one or the other dyadic site but not necessarily at both within the same DNA molecule. In order to distinguish these two possibilities, subsets of the cleavage sites were analyzed as shown in Fig. 4. The absence of dyad symmetry in the cleavage site subsets and the absence of consistent base preferences among the subsets argue against a requirement for dyad symmetry within individual molecules. However, among 62 sites having a C at −1, 28 also had a G at +5 (Fig. 4B, upper two panels). This indicates that, although these bases do not have to be present simultaneously within the same DNA molecule, their simultaneous presence does enhance the probability of cleavage.

POSITION FROM CLEAVAGE SITE
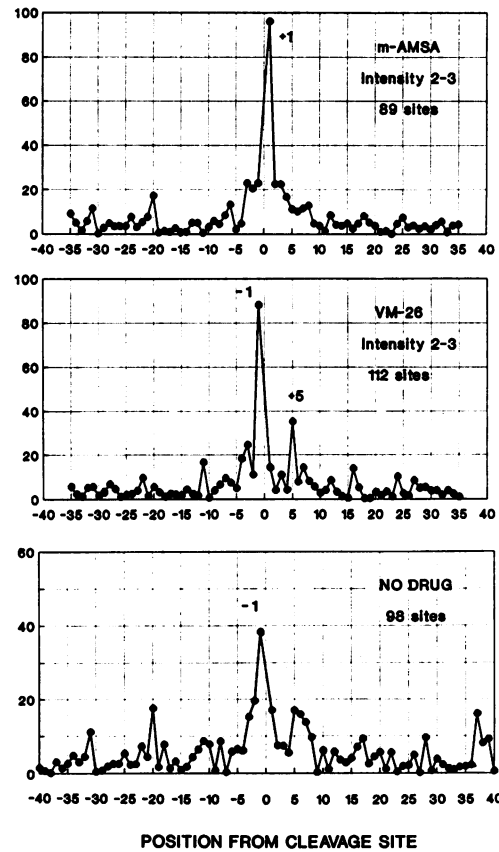
Figure 2. Chi-square values of the nucleotide distribution at each position of the cleavage site. In the case of m-AMSA and VM-26, only the strong and moderately strong sites are included. The chi-square values for p = 0.05 and p = 0.01 are 5.99 and 11.34, respectively (3 degree of freedom).

Table 1. Preference for C(−1) and G(+5) at the VM-26-induced sites and relationship to cleavage intensity.

			Number of sites		
Intensity	C(−1) G(+5)	C(−1) not-G(+5)	not-C(−1) G(+5)	not-C(−1) not-G(+5)	Total
3	10 *(38)*	8 *(31)*	3 *(12)*	5 *(19)*	26 *(100)*
2	18 *(21)*	26 *(30)*	15 *(17)*	27 *(31)*	86 *(100)*
1	13 *(6)*	49 *(23)*	40 *(19)*	109 *(52)*	211 *(100)*
Total	41 *(13)*	83 *(26)*	58 *(18)*	141 *(44)*	323 *(100)*

Numbers in italic and parentheses are percentages of total sites of that intensity. Chi-square analysis shows a very significant association between presence of C(−1) and/or G(+5) and cleavage intensity (chi-square = 37.1 for 6 degree of freedom; p < 0.00001 [two-tailed test]).

Figure 1. Schematic representation of a topoisomerase II-DNA complex. Base positions are numbered from the cleavage site on the upper strand (^{32}P label indicated at the 5'-DNA terminus). Black dots represent enzyme tyrosine residues that are covalently linked to the 5'-DNA termini of the breaks.

A

B

Base proportion (percent)
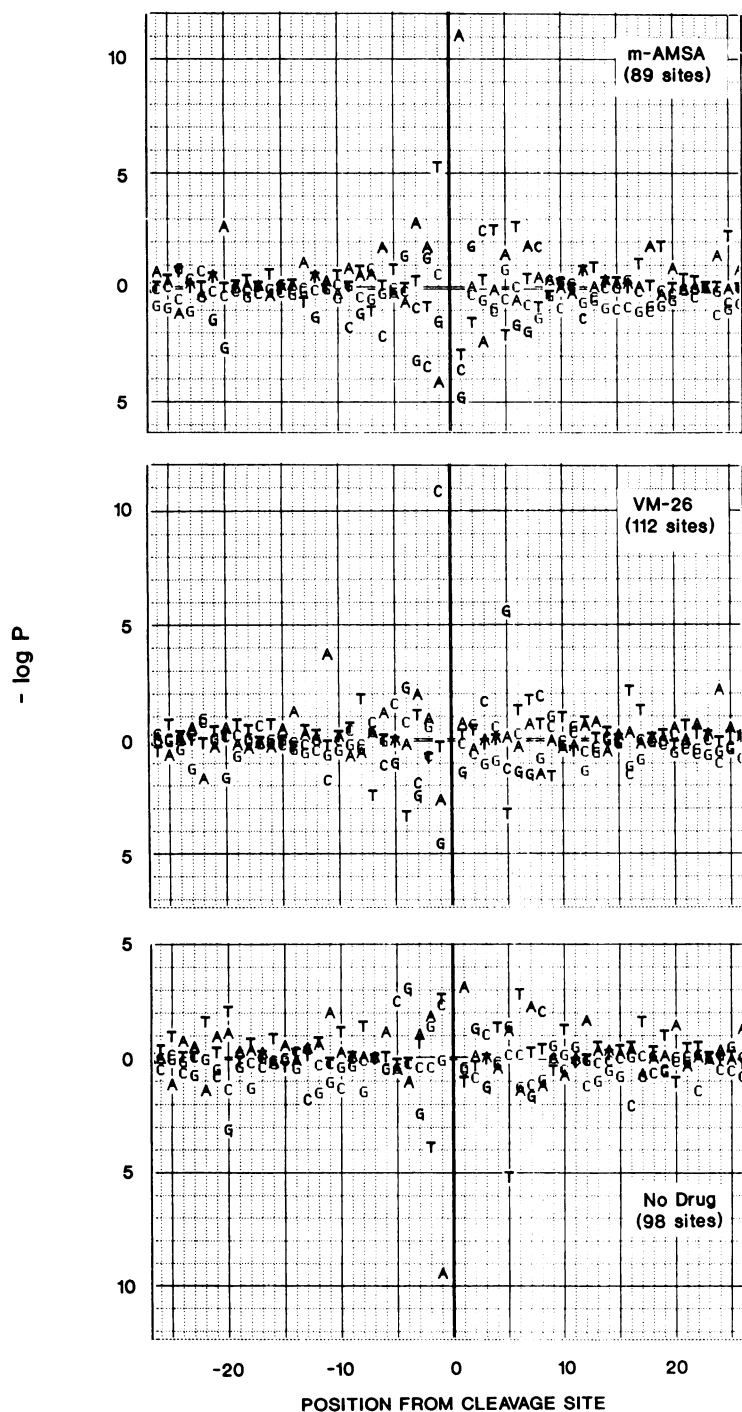
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12
											m-AMSA (89 sites)											
A	27	38	36	37	**45**	27	24	**48**	39	13	**76**	33	**15**	30	38	25	42	35	37	31	25	36
C	22	11	17	28	10	17	20	11	**6**	30	**3**	15	**38**	12	15	24	16	**33**	24	12	24	12
G	15	20	13	15	15	20	29	**7**	**33**	11	**3**	**38**	15	10	29	10	9	12	13	24	22	13
T	36	30	34	20	30	36	27	34	22	**45**	17	**15**	33	**47**	18	42	34	20	26	33	29	38
											VM-26 (112 sites)											
A	31	23	26	32	41	31	29	**46**	34	**16**	39	26	29	36	31	25	34	20	28	27	26	37
C	20	25	19	27	15	28	29	10	16	**55**	19	16	**32**	21	13	26	13	28	25	19	25	18
G	18	18	15	24	18	16	29	9	30	**5**	**7**	27	13	13	**41**	13	12	18	26	18	25	11
T	31	34	40	**17**	26	25	**13**	36	20	23	35	31	26	30	**14**	37	**42**	35	21	37	24	35
											No drug (98 sites)											
A	34	30	31	30	40	27	22	42	39	**5**	**48**	33	30	29	36	21	42	21	30	24	29	40
C	12	22	19	22	17	31	21	14	14	**38**	15	14	30	16	23	27	14	31	21	14	19	14
G	15	17	10	19	14	16	30	8	**33**	17	11	30	12	15	30	9	9	14	21	21	24	17
T	39	21	40	29	29	27	27	36	**14**	40	26	23	29	40	**11**	43	35	34	28	40	28	29

POSITION FROM CLEAVAGE SITE

Position from cleavage site

Figure 3. Base preference at topoisomerase II cleavage sites induced in the presence of *m*-AMSA (upper panels), VM-26 (middle panels), or no drug (lower panels) in SV40 DNA. Position 0 corresponds to the cleavage site. **A.** Percentage of base occurrence at each position. Standard deviations of the expected base frequency (A=T=29.6%; G=C=20.4% in SV40) were calculated, and limits of confidence (p <0.001) determined as 3.1 standard deviations. Bold and underlined numbers represent base frequencies significantly greater than expected; underlined numbers, base frequencies significantly lower than expected. Confidence intervals for A or T were [20%−40%] (*m*-AMSA), [17%−43%] (VM-26), and [15%− 45%] (No Drug). Confidence intervals for C or G were [15% −33%] (*m*-AMSA), [8%−32%] (VM-26), and [> 33%] (No Drug). **B.** Probabilities of the observed base frequency deviations from expectation. In the y-axis, P is the probability of observing that deviation or more, either as excess (above base line) or deficiency (below base line) relative to the expected frequency of each individual base.

The 'C(−1) and G(+5)' subset of VM-26 cleavage sites shows no other clear base preference (Fig. 4B, upper panel). The 3 other subsets do suggest other possible base preferences, but these are difficult to define because of the small sample sizes. This suggests that the presence of C(−1) at the 3' termini of both cleavage sites may be a strong determinant of VM-26 action, while the presence of only one C(−1) may necessitate the presence of certain bases at other positions.

An interesting subset is the 'not-C(−1) and not-G(+5)' subset because it consists of a relatively large number of sites (32 sites,

A

B Base proportion (percent)

	C (-1) and G (+5) (28 sites)											
A	32	29	29	50	-	18	39	14	18	25	21	39
C	29	29	32	4	-	32	21	25	39	18	39	25
G	0	21	7	14	G	14	7	14	21	21	29	11
T	39	21	32	32	-	36	32	46	21	36	11	25

A	32	29	29	32	50	36	29	43	21	-
C	25	32	25	29	4	25	25	11	14	C
G	18	21	18	25	14	29	32	11	36	-
T	25	18	29	14	32	11	14	36	29	-

	C (-1) and not-G (+5) (34 sites)											
A	38	24	24	24	59	35	35	18	24	35	26	38
C	15	12	47	41	29	24	18	18	26	12	24	24
G	0	21	18	6	-	9	9	29	38	12	26	12
T	47	44	12	29	12	32	38	35	12	41	24	26

A	26	29	21	21	24	26	35	59	35	-
C	18	26	21	24	26	38	41	12	15	C
G	21	12	9	26	29	15	15	3	38	-
T	35	32	50	29	21	21	9	26	12	-

	not-C (-1) and G (+5) (18 sites)											
A	56	22	33	39	-	22	44	22	39	17	28	39
C	0	11	22	0	-	22	0	28	11	33	22	0
G	22	44	6	17	G	6	11	22	17	17	11	17
T	22	22	39	44	-	50	44	28	33	33	39	44

A	33	11	28	39	44	22	17	44	39	28
C	22	33	17	28	17	17	28	0	6	-
G	11	17	17	22	11	17	39	11	28	11
T	33	39	39	11	28	44	17	44	28	61

	not-C (-1) and not-G (+5) (32 sites)											
A	38	28	31	34	47	22	22	25	34	25	28	31
C	25	13	22	25	16	25	6	41	19	19	16	16
G	13	28	19	19	-	19	19	6	22	22	28	6
T	25	31	38	22	38	34	53	28	25	34	28	47

A	34	19	28	41	50	38	28	34	41	41
C	16	13	13	28	13	25	19	13	25	-
G	19	22	19	22	13	6	38	13	19	13
T	31	47	41	9	25	31	16	41	16	47

-20   -10   0   10   20
POSITION FROM CLEAVAGE SITE

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1   1 2 3 4 5 6 7 8 9 10 11 12
Position from cleavage site

Figure 4. Base preference at the 4 subsets of VM-26-induced cleavage sites. From top to bottom: C(-1) and G(+5), C(-1) and not-G(+5), not-C(-1) and G(+5), not-C(-1) and not-G(+5). A. Percentage of base occurrence at each position. Standard deviations of the expected base frequency and limits of confidence (p <0.001) were calculated (see Fig. 3A). Confidence intervals for A or T were [4% - 56%], [6% - 54%], [> 63%], and [5% - 55%] for each subset (top to bottom). Confidence intervals for G or C were [> 44%], [> 41%], [> 49%], and [>42%] for each subset (top to bottom). B. Probabilities of the observed base frequency deviations from expectation (see Fig. 3B).

approximately 30% of the most intense sites) (Fig. 4, lower panels). Here, the most significant base preferences are T(+7) and C(+8).

m-AMSA

In contrast to VM-26 and doxorubicin, for which the major preference is at -1, m-AMSA shows a major preference at +1,

which corresponds to the nucleotide at the 5' rather than the 3' terminus of the cleavage site (Fig. 3, upper panels; for data on doxorubicin, see ref 10). The major preferred base for both m-AMSA and doxorubicin is an A, which however is located on opposite sides of the cleavage site for the two drugs. 76% of the m-AMSA sites had an A at the +1 position (Fig 3A, top panel); the probability of this occurring by chance is less than

A

B
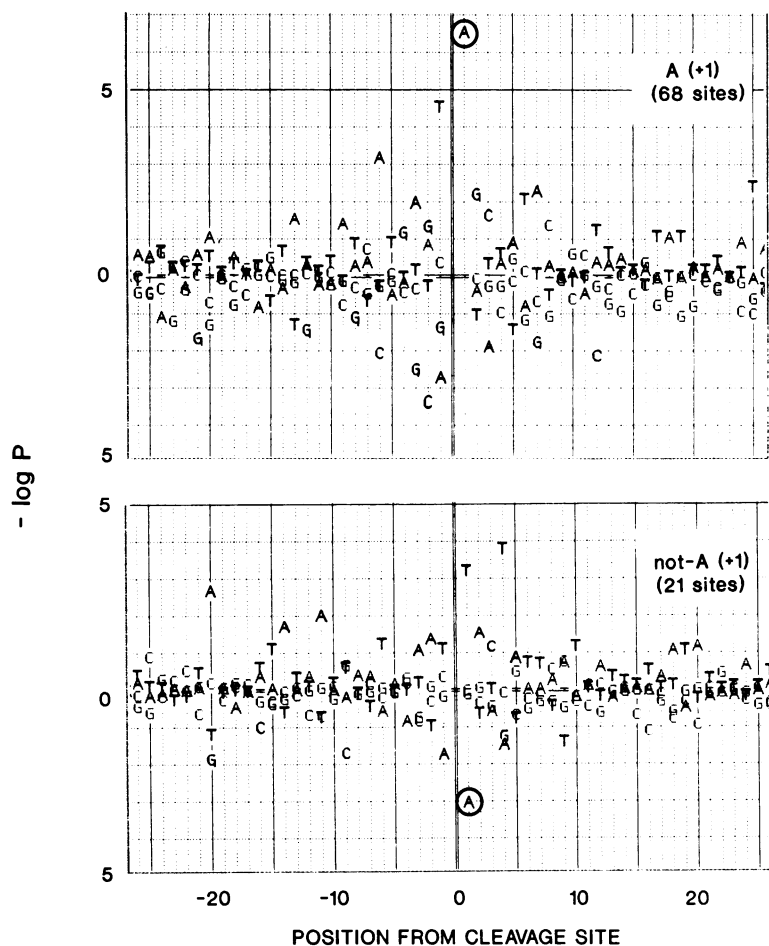
Base proportion (percent)

													A (+1) (68 sites)									
A	26	43	35	37	53	25	26	47	37	16		A	28	15	37	37	22	46	34	34	32	22	34
C	24	15	18	29	9	18	18	13	4	31		-	16	37	12	18	26	16	32	19	13	28	9
G	15	16	12	13	13	21	29	7	34	10		-	43	16	13	28	10	7	12	16	28	21	15
T	35	26	35	21	25	37	26	32	25	43		-	13	32	38	18	41	31	22	31	26	29	43

													not-A (+1) (21 sites)									
A	29	24	38	38	19	33	14	52	48	5		-	48	14	10	43	33	29	38	48	29	33	43
C	19	0	14	24	14	14	29	5	10	29		14	10	43	14	5	14	14	33	38	10	10	24
G	14	33	19	19	19	19	29	5	29	14		14	24	10	0	33	10	14	14	5	10	29	10
T	38	43	29	19	48	33	29	38	14	52		71	19	33	76	19	43	43	14	10	52	29	24

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1   1 2 3 4 5 6 7 8 9 10 11 12

Figure 5. Base preference at the 2 subsets of m-AMSA-induced cleavage sites. Upper panels: A(+1); lower panels: not-A(+1). **A.** Percentage of base occurrence at each position. Standard deviations of the expected base frequency and limits of confidence (p <0.001) were calculated (see Fig. 3A). Confidence intervals for A or T were [13%− 47%] (upper panel) and [0%−61%] (lower panel). Confidence intervals for C or G were [5% −35%] (upper panel) and [0%−47%] (lower panel). **B.** Probabilities of the observed base frequency deviations from expectation (see Fig. 3B).

1 in 10^{10} (fig. 3B, top panel). Moreover, the presence of A(+1) correlates positively with cleavage intensity: A(+1) occurs in 88%, 72% and 59% of the intensity 3, 2 and 1 sites, respectively (Table 1).

In contrast to VM-26 and doxorubicin, probability analysis for m-AMSA revealed relatively little dyadic symmetry (Fig. 3B, upper panel). Thus the dyadic base corresponding to A(+1) would be T(+4), which shows only borderline preference. The most preferred bases in the vicinity of the cleavage site were

```
   -4        -1 +1      +4        +5 +8
   G  A    T  A  G    C  T    A  T  A  C
```

This sequence shows dyadic correspondence for 5 out of 6 comparisons (p <0.01).

Cleavage sites that lack A(+1) however did show a distinct preference for T(+4) (76% of the not-A[+1] sites) (Fig. 5A), indicating that the dyadic positions can influence cleavage at some sites.

Absence of both A(+1) and T(+4) was seen in only 11% of all m-AMSA cleavage sites (21 sites of 197; Table 2). Among these sites, 76% (16 sites) were weak (intensity 1), 24% (5 sites) were intensity 2, and none were strong (intensity 3). Thus, lack

Table 2. Preference for A(+1) and T(+4) at the m-AMSA-induced sites and relationship to cleavage intensity.

			Number of sites		
Intensity	A(+1) T(+4)	A(+1) not-T(+4)	not-A(+1) T(+4)	not-A(+1) not-T(+4)	Total
3	8 (32)	14 (56)	3 (12)	0 (0)	25 (100)
2	18 (28)	28 (44)	13 (20)	5 (8)	64 (100)
1	26 (24)	38 (35)	28 (26)	16 (15)	108 (100)
Total	52 (26)	80 (41)	44 (22)	21 (11)	197 (100)

Numbers in italic and parentheses are percentages of total sites of that intensity. Grouping of the intensity 2 and 3 sites together, shows a significant association between the presence of A(+1) and T(+4) (p < 0.003, two-tailed test).

of A(+1) and T(+4) is seen primarily at weak sites and all of the most intense sites have either A(+1), T(+4) or both (Table 2).

Cleavage sites stimulated by m-AMSA in the SV40 MAR were determined in both DNA strands (Fig. 6). Most of the sites corresponded to double-strand cleavage pairs of similar intensities. However some sites, usually of weak intensity, were solitary. In a few cases, the cleavage intensities of a double-strand pair were unequal. Single-strand cleavage by m-AMSA therefore
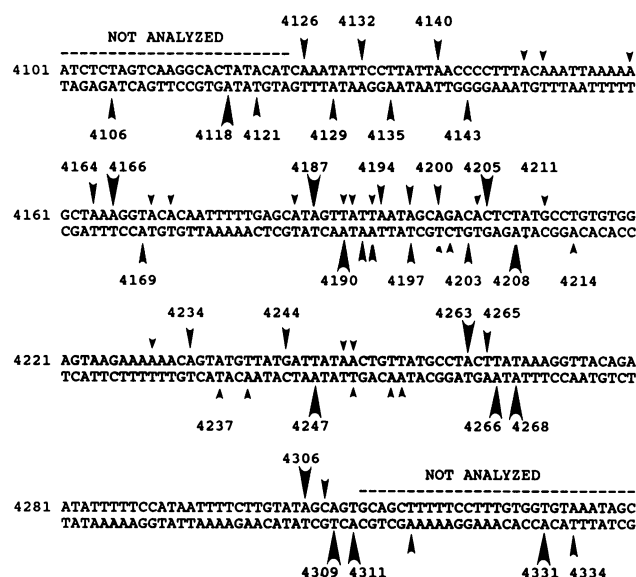
Figure 6. Topoisomerase II cleavage sites stimulated by m-AMSA in the SV40 nuclear matrix associated region (MAR). Arrowheads point to the covalently linked base (+1 position), and size of the arrowhead represents cleavage intensity.

BASE PREFERENCE

DOXORUBICIN	A (-1)
m-AMSA	A (+1)
VM-26/VP-16	C (-1)

Figure 7. Schematic representation of the proposed ternary complex of the drug (black rectangle), topoisomerase II, and the DNA, showing stacking of the drug with the base pairs flanking the cleavage site. Base preference for each drug is indicated.

Table 3. Preference for lack of A(−1) and T(+5) at topoisomerase II cleavage sites in the absence of drug.

Intensity	not-A(−1) not-T(+5)	A(−1) not-T(+5)	not-A(−1) T(+5)	A(−1) T(+5)	Total
3	14 *(93)*	0 *(0)*	1 *(7)*	0 *(0)*	15 *(100)*
2	21 *(81)*	2 *(8)*	3 *(12)*	0 *(0)*	26 *(100)*
1	47 *(84)*	3 *(5)*	7 *(11)*	0 *(0)*	57 *(100)*
Total	82 *(84)*	5 *(5)*	11 *(11)*	0 *(0)*	98 *(100)*

Numbers in italic and parentheses are percentages of total sites of that intensity.

is not uncommon, in agreement with the relatively high single-strand break signal observed by alkaline elution assays of *m*-AMSA-treated cells (as compared to doxorubicin [15] or VM-26 [16]) and with the findings of Muller et al. using erythrocyte topoisomerase II (17).

Topoisomerase II without drug

Cleavage sites in the absence of drug had been analyzed previously (10). Analysis using the new method of probability calculation and with the inclusion of some additional sites was in agreement with our previous conclusions (10). The bias is mainly for absence of A at −1 and for absence of T at +5 (Fig. 3B, lower panel). Among the 98 enzyme alone sites, more than 80% have neither A(−1) nor T(+5), and none have A(−1) and T(+5) simultaneously (Table 3). Therefore, an A must be absent from one or both 3' termini of the dyadic cleavage sites.

DISCUSSION

Most previous attempts to account for topoisomerase II-induced cleavage sites have focussed on long-range DNA sequence patterns which could influence DNA conformation or protein interactions (18, 19). There have been no previous attempts to explain the differences in site locations between different drugs.

In a previous study of sites stimulated by doxorubicin, we noted a base requirement immediately at the sites of cleavage (10); the requirement was for an A at the 3' terminus of at least one of two break positions. The current work shows that other drugs, in particular VM-26 and *m*-AMSA, have unique base preferences immediately flanking the cleavage sites; thus, the strongest local base preference was always at the 3' or at the 5' terminus of a break.

The current work utilized a new refinement of the previous probability analysis (10), taking into account the bias for or against certain base pair doublets in the region of sequence analyzed. This was important because, for the first time, a C was observed as a preferred base; when analyzed by the previous method, this caused a spurious bias against G in the following position because of the bias against 5'-GC-3' doublets. The current work presents a general method of compensating for this king of bias.

For VM-26, as in the case of doxorubicin, the strongest base preference was at the 3' terminus of the cleavage site. However, the preferred base in this position was C for VM-26, in contrast to the strong preference for A in the case of doxorubicin. In the case of doxorubicin, sites lacking an A at the 3' terminus of the cleavage site invariably had a T at position +5 which corresponds to an A at the 3' terminus of the potential topoisomerase II break position on the opposite strand (10). In the case of VM-26, sites lacking a C at the 3' terminus of the cleavage site usually had a G at position +5, especially for the stronger cleavage sites. In contrast to doxorubicin, VM-26 did not show an absolute requirement for an essential base to be present at least at one or the other cleavage site. Hence, the absence of an essential base may be compensated by other long range base recognition for VM-26 but not for doxorubicin.

Contrary to the previous cases, the major base preference for *m*-AMSA was at the 5' terminus of the break. As with doxorubicin, the major preferred base for *m*-AMSA was an A, but located on the other side of the break. Therefore the stabilization of cleavable complexes may involve an interaction with the base pair on either side of the break, depending on the drug.

m-AMSA differed from the other drugs studied, as well as from the pattern produced by the enzyme alone, in that there was only a weak base preference at the dyadic site on the strand opposite to the observed cleavage. This suggests that *m*-AMSA usually

stabilizes complexes in a state in which only one strand is cleaved, whereas the other drugs more frequently stabilize double-strand cleavage complexes. This possibility is consistent with alkaline elution measurements in intact cells and isolated nuclei which show a higher ratio of single to double-strand breaks with *m*-AMSA than with most other topoisomerase II inhibitors, including VM-26 and doxorubicin (15, 16, 20). Although *m*-AMSA sometimes showed cleavage at dyadic sites on both strands (Fig. 5), it is possible that in most individual DNA molecules only one site or the other of a dyadic pair was cleaved at any one time. Intensity differences between some of the dyadic pairs indicate that single-stand cleavage complexes occur at least in these cases.

Both VM-26 and *m*-AMSA showed a detectable dyadic symmetry of the most preferred bases centered between the expected topoisomerase II cleavage sites. This symmetry extended for 8 or 6 base pair in each direction from the symmetry axis for VM-26 and *m*-AMSA, respectively. The size of the DNA region covered by topoisomerase II has been estimated by DNase I footprinting as approximately 25 base pairs (21). The slightly larger size of the DNase I footprint is consistent with the extent of the observed base preference symmetry. The base preference symmetry may reflect a recognition pattern for topoisomerase II cleavage. The absence of this symmetry in the subsets of the cleavage sites analyzed in Figs. 4 & 5 suggests that the recognition occurs at one or the other but not necessarily at both break positions on opposite strands.

The strong base preferences for doxorubicin had suggested a model for the enzyme complex in which the drug interacts by stacking with the base pair at the 3'terminus of the break (10). The current results with VM-26 and *m*-AMSA further support this model in two respects (Fig. 7). First, the finding that alteration of the ring system changes the base pair that is recognized agrees with the predicted stacking of drug and base pair. Second, the finding that *m*-AMSA has a major preference for the base pair at the 5'terminus agrees with the assumed location of the drug ring system between the base pairs at the 3' and 5'termini. Moreover, the fact that any given drug favors either one or the other but not both base pairs suggests that the complex does not involve a true intercalation but rather a stacking with one or the other base pair which may be separated in space, perhaps in a configuration resembling an intermediate in the strand passage process. The well known differences in site selection among different classes of topoisomerase II blockers (5, 7, 13) can now be attributed, at least in part, to base pair recognition immediately at the cleavage sites (Fig. 7).

Although the question of whether or not etoposides, such as VM-26, bind to DNA is controversial (22), the molecules have a planar ring system which may stack with a base pair within a topoisomerase II complex. Similarly, the topoisomerase I blocker, camptothecin, does not bind extensively to DNA but tends to induce cleavage sites having a G at the 5'terminus, consistent with an analogous stacking model for complexes with this enzyme (23).

ACKNOWLEDGEMENTS

REFERENCES

1. Wang, J.C. (1985) *Annu. Rev. Biochem.* **54**, 665–697.
2. Wang, J.C. (1987) *Biochim. Biophys. Acta* **909**, 1–9.
3. Liu, L.F., Rowe, T.C., Yang, L., Tewey, K.M., and Chen, G.L. (1983) *J. Biol. Chem.* **258**, 15365–15370.
4. Sander, M., and Hsieh, T. (1983) *J. Biol. Chem.* **258**, 8421–8428.
5. Tewey, K.M., Rowe, T.C., Yang, L., Halligan, B.D., and Liu, L.F. (1984) *Science* **226**, 466–468.
6. Pommier, Y., and Kohn, K.W., (1989) In Glazer, R.I. (ed.), Developments in Cancer Chemotherapy. CRC Press, Inc., Boca Raton, FA, vol 2, pp. 175–196.
7. Pommier, Y., Capranico, G., Orr, A., and Kohn, K.W. (1991) *J. Mol. Biol.*, in press.
8. Pommier, Y., Cockerill, P.N., Kohn, K.W., and Garrard, W. (1990) *J. Virol.* **64**, 419–423.
9. Capranico, G., Zunino, F., Kohn, K.W., and Pommier, Y. (1990) *Biochemistry* **29**, 562–569.
10. Capranico, G., Kohn, K.W., and Pommier, Y. (1990) *Nucleic. Acids. Res.* **18**, 6611–6619.
11. Minford, J., Pommier, Y., Filipski, J., Kohn, K.W., Kerrigan, D., Mattern, M., Michaels, S., Schwartz, R., and Zwelling, L.A. (1986) *Biochemistry* **25**, 9–16.
12. Pommier, Y., Kerrigan, D., Schwartz, R.E., Swack, J.A., and McCurdy, A. (1986) *Cancer Res.* **46**, 3075–3081.
13. Capranico, G., Jaxel, C., Roberge, M., Kohn, K.W., and Pommier, Y. (1990) *Nucleic. Acids. Res.* **18**, 4553–4559.
14. Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., (1986) In The art of scientific computing. Cambridge University Press, pp. 156–160.
15. Zwelling, L.A., Michaels, S., Erickson, L.C., Ungerleider, R.S., Nichols, M., and Kohn, K.W. (1981) *Biochemistry* **20**, 6553–6563.
16. Long, B.H., Musial, S.T., and Brattain, M.G. (1985) *Cancer Res.* **45**, 3106–3112.
17. Muller, M.T., Spitzner, J.R., DiDonato, J.A., Mehta, V.B., Tsutsui, K., and Tsutsui, K. (1988) *Biochemistry* **27**, 8369–8379.
18. Sander, M., and Hsieh, T.S. (1985) *Nucleic. Acids. Res.* **13**, 1057–1072.
19. Spitzner, J.R., and Muller, M.T. (1988) *Nucleic. Acids. Res.* **16**, 5533–5556.
20. Pommier, Y., Kerrigan, D., Schwartz, R., and Zwelling, L.A. (1982) *Biochem. Biophys. Res. Commun.* **107**, 576–583.
21. Lee, M.P., Sander, M., and Hsieh, T.-S. (1989) *J. Biol. Chem.* **264**, 21779–21787.
22. Chow, K.C., Macdonald, T.L., and Ross, W.E. (1988) *Mol. Pharmacol.* **34**, 467–473.
23. Jaxel, C., Capranico, G., Kerrigan, D., Kohn, K.W., and Pommier, Y. (1991) *J. Biol. Chem.*, in press.