0026-895X/95/020238-12$3.00/0

MOLECULAR PHARMACOLOGY, 48:238–249 (1995).

Analysis of Eukaryotic Topoisomerase II Cleavage Sites in the Presence of the Quinolone CP-115,953 Reveals Drug-Dependent and -Independent Recognition Elements

JEFFREY R. SPITZNER,[1] IN KWON CHUNG,[2] THOMAS D. GOOTZ, PAUL R. McGUIRK, and MARK T. MULLER

Ohio State University, Department of Molecular Genetics, Columbus, Ohio 43210 (J.R.S., I.K.C., M.T.M.), and Department of Immunology and Infectious Diseases, Pfizer, Inc., Central Research Division, Groton, Connecticut 06340 (T.D.G., P.R.M.)

Received August 25, 1994; Accepted May 3, 1995

SUMMARY

The quinolone derivative CP-115,953 [6,8-difluoro-7-(4-hydroxyphenyl)-1-cyclopropyl-4-quinolone-3-carboxylic acid] has been shown to induce eukaryotic topoisomerase II-mediated breaks in DNA, producing cleavage patterns that are distinct from those induced by the anticancer drugs amsacrine, etoposide, and teniposide. High levels of the quinolone have been found to inhibit topoisomerase II activity via an interaction with the enzyme and not by DNA unwinding. Topoisomerase II cleavage sites were analyzed on nine DNA fragments, and 85 quinolone-induced sites were sequenced, as well as 86 amsacrine and 134 teniposide sites. A consensus sequence was derived for the quinolone sites that is different from those reported for other drugs; however, because topoisomerase II cleavage sites are double-stranded but not palindromic, different consensus sequences are not easily compared. For this reason, a new, double-stranded, consensus sequence method, the "unique-base analysis," was developed; this was applied to the quinolone sites as well as six other large sets of topoisomerase II sites determined in the absence or presence of drugs. For each of the seven sets of sites, conserved bases were found in the 16-base region spanning positions −6 to +10, relative to the enzyme cleavage site (DNA breakage between −1 and +1). The conserved bases were virtually identical in the regions flanking the cleavage site for all seven data sets. In contrast, the base preferences identified proximal to the cleavage sites were unique to the drug tested. These observations suggest that the selection of cleavage sites by topoisomerase II involves both enzyme-dependent and drug-dependent recognition elements. The single most preferred base in the quinolone sites was a cytosine at −1; the same preference was found with teniposide, and 60 of the 85 quinolone sites co-localized with teniposide sites.

Agents that stabilize the covalent intermediate formed between eukaryotic topo II and DNA *in vivo* are of enormous utility in the treatment of cancers, and a number of chemically unrelated families of compounds have been shown to do this (reviewed in Refs. 1 and 2). These drugs not only stimulate the production of DNA breaks by topo II but also direct these sequence-specific DNA cleavages in patterns that are different from those for the enzyme in the absence of drugs and also different among the various drug classes. The families of topo II inhibitors include both DNA intercalators (such as doxorubicin, AMSA, ellipticine, mitoxantrone, and compounds related to each) and nonintercalators (including epipodophyllotoxins and quinolones), and the clinically valuable compounds demonstrate activities against distinct types of cancers. It is currently not clear whether the altered DNA cleavage specificities of topo II in the presence of these different drugs play a role in their efficacy as anticancer agents, but this is beginning to be investigated (3); for this reason, it is important to analyze the sequences of drug-induced topo II sites and derive consensus sequences for each drug family. In addition, because topo II cleavage sites are generally conserved within drug families (for example, different anthracyclines share cleavage patterns, as do the epipodophyllotoxins VM-26 and etoposide) (see Fig. 3) (4), knowledge of the consensus sequences for each drug family may allow the determination of whether a newly detected topo II inhibitor is novel or belongs to an existing class, without the need to sequence a large number of topo II cleavage sites. Furthermore, research into the sequence specificity of drug-topo II-DNA interactions may help to elucidate the mechanisms through which these drugs act, thus aiding in the design of new anticancer drugs.

Funds for this work were provided by National Institutes of Health Grants RO1-AI28362 and P30-CA16058. J.R.S. was supported by National Institutes of Health Training Grant T32-CA09338, National Institutes of Health National Research Service Award M14335, and the Hereditary Disease Foundation.

[1] Present address: MIT, Department of Biology, Cambridge, MA 02139.

[2] Present address: Yonsei University, Department of Biology, Seoul, 120–749 Korea.

ABBREVIATIONS: topo II, topoisomerase II; CP, CP-115,953; kDNA, kinetoplast DNA; VM-26, teniposide; AMSA, amsacrine; topo I, topoisomerase I.

Consensus sequences for topo II cleavage of DNA have been reported for many different drugs; examples include AMSA (5, 6), VM-26 (6), doxorubicin (7), and ellipticine (8). Many quinolones are in clinical use as antimicrobial agents; these drugs target the bacterial enzyme DNA gyrase and generally have little effect on the eukaryotic homologue (9). A new class of compounds, derived from quinolones and represented by CP, have been reported to stabilize the covalent intermediate formed between eukaryotic topo II and DNA, stimulating the formation of double-stranded breaks in DNA (9). In addition, CP was shown to mediate its cytotoxic effect by targeting topo II, inasmuch as a cell line containing a mutant topo II enzyme was resistant to higher levels of this quinolone than was its wild-type counterpart (10). Therefore, quinolone derivatives like CP may prove useful as a new class of anticancer agents. For this reason, we sought to characterize the DNA sequences cleaved by topo II in the presence and absence of CP.

All of the studies involving sequence analysis of topo II cleavages have had to deal with a complex issue associated with topo II sites. Like the restriction enzyme *Eco*RI, topo II makes double-stranded DNA breaks, with the bottom strand cleavage event being offset 4 bases 5′ to the top strand event (Table 1); however, for topo II, the sequences of the two paired sites are not identical. In fact, the two paired topo II sites often bear no obvious similarities to each other. In deriving the topo II consensus sequences mentioned above, a variety of different methods were used to select one strand from each paired site and then determine a consensus sequence from this simplified data set. A problem with these methods is that they do not follow any common rules, and one consensus sequence (or the matrix of base proportions from which it was derived) is not easily compared with another. We developed a new double-strand consensus sequence method that facilitates comparisons of data sets from different laboratories. In this report, we mapped the sequences of 85 quinolone-induced topo II sites. The data reveal that the quinolone drug induced enzyme cleavage patterns distinct from those of AMSA or VM-26, which were all analyzed on the same sequencing gels. Consensus sequences were derived for the cleavage sites mapped with these drugs and were compared with those of other topo II data sets.

Experimental Procedures

Materials. Human DNA topo II was purified from fresh placenta (5) and yielded a single polypeptide of 170 kDa. Supercoiled plasmid DNA was purified using standard methods. The quinolone CP was synthesized as described (11) and kept as a 15 mM stock solution in 0.1 N NaOH. The drug was dissolved in 10 mM Tris·HCl before use. Restriction enzymes and other DNA-modifying enzymes were provided by commercial sources. Purified calf thymus topo I and the topo II inhibitors AMSA, etoposide, and doxorubicin were provided by TopoGEN (Columbus, OH). VM-26 was obtained from Bristol Myers Squibb. DNA fragments were prepared from several plasmids and end-labeled using polynucleotide kinase, as described previously (5). The construct pJH is pUC19 containing three herpes simplex virus type 1 DR2 repeats cloned into a *Sma*I site. The construct pRYG was described previously (12).

Topoisomerase reactions. Topo II relaxation assays were performed as described (13), and reactions were analyzed on 1% agarose gels lacking or containing 0.5 μg/ml ethidium bromide, as indicated. Decatenation assays were conducted using kDNA, as described previously (14); 1 unit of topo II can decatenate approximately 0.1 μg of kDNA in 15 min at 37°. Cleavage reactions were conducted as described (12), using human p170 topo II and ^{32}P-end-labeled DNA fragments, as specified. Cleavage sites were analyzed on 12% polyacrylamide sequencing gels, along with chemical sequencing markers to unambiguously locate the sites, as described (5, 12). DNA unwinding assays were performed as described (14a), using 0.5 μg of a 6.2-kilobase supercoiled plasmid, drug concentrations as indicated, and an excess of topo I activity, in either topo I buffer (no Mg^{2+}) or topo II decatenation buffer. Products were analyzed by loading one half of the reaction mixture (after termination of the reaction with sodium dodecyl sulfate) onto 0.9% agarose gels containing 0.1% sodium dodecyl sulfate; gels were run at 50 V and then stained with ethidium bromide for visualization of DNA topoisomer distributions.

Analysis of topo II cleavage sites. The sequences of topo II cleavage sites were determined by comparison of the cleaved DNA subfragments with chemical sequencing markers on the gels, as described previously (5), using only sites that were more intense than the markers and were mapped unambiguously. All sequence data analyses were conducted using the Eden-Genesys computer software program (TEAM Associates, Westerville, OH). Each cleavage site was entered as the 34 bases from 15 bases upstream (5′) of the top strand (sequenced) cleavage site (position −15) to 19 bases downstream (3′) of the cleavage site (position +19). Symmetrically, taking into account the 4-base 5′ overhang of topo II sites, the 34-base fragments also include the paired "bottom" or complementary strand sites spanning positions −15 to +19, relative to their cleavage site locations. The different methods used to analyze the data are shown in Table 1, which demonstrates the four consensus sequence methods applied to the quinolone-induced cleavage sites in this report. To illustrate the procedures, only five of the 85 CP sites were selected and only 16 bases are shown (from position −6 to +10, relative to the cleavage site between position −1 and +1) (Table 1). Due to the double-stranded nature of topo II sites, note that each sequence listed, with the cleavage site marked as identified from the sequencing gel, actually represents a pair of cleavage sites, separated by a 4-base 5′ overhang, as shown for the first sequence. Therefore, position +5 on the top strand (a thymine here) is the complement of the base located at the −1 position relative to the bottom strand cleavage site location (Table 1). For simplicity, the methods shown for these sample sites were applied only to positions −6 to −1 for each strand.

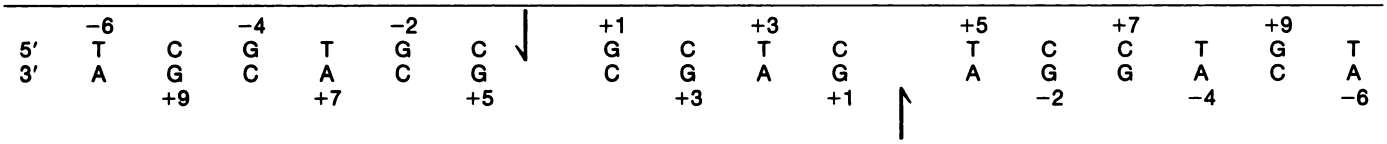
Method A is the conventional procedure for derivation of consensus sequences. The sequenced DNA fragments are aligned with one another (for topo II sites this is always fixed at the cleavage sites), and the base proportions are calculated, independently, for each position.

In method B, the five double-stranded cleavages, which actually represent a total of 10 breakages of DNA, are treated as 10 independent sites. Each site (sites 1–5) is shown as the top strand site and also as the bottom strand site, which is read the same way (5′ to 3′ on the bottom strand) as the top strand site, with numbering and alignment relative to its own cleavage site. Thus, if the opposite 5′ end of these DNA fragments had been radiolabeled instead, it would have been these bottom strand cleavage events that were sequenced, with exactly the same top plus bottom sites listed for the double-stranded model as listed here. In method B, the matrix of base proportions is computed by summing the number of occurrences of each base, at each position, and then dividing by the total number of sites (in this example, 10 sites). Thus, at the 10 positions that existed 6 bases 5′ of the cleavage site (the −6 position for each strand), 6 of these bases were adenines (a proportion of 0.6); therefore, adenine is the preferred base and is assigned as a consensus base.

In method C, all five pairs of sites (10 total sites, as shown above) are considered, but from each pair only one (either the top strand or the bottom strand site) is selected, remaining aligned by its corresponding cleavage site position. Previous topo II consensus sequence reports have all used this method to generate consensus sequences; they have varied only in how the "best" strand is selected for each

TABLE 1
Comparative methods for analysis of consensus sequences

Configuration of double-strand cleavage of sequence 1



Data set of five sequenced topo II sites

1. TCGTGC^GCTCTCCTGT
2. GCGTAA^TCATGGTCAT
3. ATATGT^ATATGTATGT
4. GAATGC^ATATATATGT
5. AAGTTC^TGCTATGTGG

		Base	Proportion					
			−6[a]	−5	−4	−3	−2	−1
A. As sequenced	1. TCGTGC	A	0.4	0.4	0.4	0.0	0.2	0.2
	2. GCGTAA	C	0.0	0.4	0.0	0.0	0.0	0.6
	3. ATATGT	G	0.4	0.0	0.6	0.0	0.6	0.0
	4. GAATGC	T	0.2	0.2	0.0	1.0	0.2	0.2
	5. AAGTTC	Consensus	R	N	G	T	G	C
B. Double-stranded model	1. TCGTGC ACAGGA	A	0.6	0.2	0.6	0.1	0.4	0.2
	2. GCGTAA ATGACC	C	0.1	0.6	0.0	0.1	0.1	0.5
	3. ATATGT ACATAC	G	0.2	0.0	0.4	0.1	0.4	0.0
	4. GAATGC ACATAT	T	0.1	0.2	0.0	0.7	0.1	0.3
	5. AAGTTC CCACAT	Consensus	A	C	A	T	R	C
C. Best single-strand model	1. TCGTGC	A	0.6	0.2	0.4	0.0	0.6	0.2
	2. GCGTAA	C	0.0	0.8	0.0	0.0	0.0	0.6
	3. ACATAC	G	0.2	0.0	0.6	0.0	0.2	0.0
	4. ACTAT	T	0.2	0.0	0.0	1.0	0.2	0.2
	5. AAGTTC	Consensus	A	C	G	T	A	C
D. Unique-base consensus model	1. tCgTGC AcAgga = ACATGC	A	1.0	0.4	0.8	0.2	0.8	0.4
	2. gCGTAa AtgacC = ACGTAC	C	0.2	1.0	0.0	0.2	0.2	1.0
	3. AtATgt aCatAC = ACATAC	G	0.4	0.0	0.6	0.2	0.6	0.0
	4. gaATgC ACatAt = ACATAC	T	0.2	0.4	0.0	1.0	0.2	0.6
	5. AagTtC cCAcAt = ACATAC	Consensus	A	C	A	T	A	C

[a] Base position.

pair. The object is to select the one strand from each pair with the most homology to all of the other selected best strands from the data set. The strategies obviously require some variety of iterative methods, whether done by eye or by a computer program. In this report, we selected the best single-strand strands by maximizing the total information content within the 16-base window from position −6 to +10; the details are described in Results.

Method D, the unique-base model, is a novel method that includes contributions from cleavage sites in both strands but does not require any arbitrary strand selections. The first step of method D is essentially the independent derivation of a consensus sequence for each base position. In method C, one complete strand is selected from each paired site and the other strand is ignored. With method D, both strands are analyzed, position by position (both −6 positions, both −5 positions, etc.), to determine how many of the sites have at least one of each base occurrence. In Table 1, example D, five paired sites (a total of 10 cleavage events) are considered. The sequences can be viewed as being aligned in method B (double-stranded model showing both paired cleavage sites, reading 5′ to 3′) but, rather than simply summing the total occurrences of each base, in method D the number of paired sites (remaining linked as pairs) that contain at least one adenine at position −6, cytosine at −6, . . . adenine at −5, etc., is determined. Thus, at the −6 position here, all five paired sites have at least one adenine at −6, only one site has a cytosine at −6, two sites have a guanine at −6, and one site has a thymine at −6. The total number of these "unique" occurrences is divided by the number of paired sites (in this example, five not 10) to determine the unique-base proportions, as indicated in the base proportions matrix in Table 1, example D. At position −5 relative to the cleavage site, two sites have at least one adenine, all five have at least one cytosine, no sites have a guanine, and two sites have at least one thymine, yielding the unique-base proportions indicated. Method D differs

from method B in that a paired site with a cytosine at position −5 on each strand (as in site 1 here) scores only one cytosine. In addition, unless all of the paired sites analyzed have the same base identities on each pair, the proportions for each position sum to greater than unity. These unique-base proportions are determined for each position (in this example, only positions −6 to −1), and then a unique-base consensus sequence is determined by comparing these base proportion values with a threshold equal to twice the expected (background) base proportions. A consensus sequence for these five sites is shown in Table 1, example D; the consensus bases are indicated in the unique-base sequence listing as capital letters to indicate that, even if a paired topo II site contains two −3 thymines, only one contributes to consensus sequence derivation. The unique-base consensus sequence removes the "strandedness" of topo II sites, retaining only conserved base information from either strand. The first sequence in Table 1, for example, contains a purine-rich sequence on the bottom strand site that is not homologous to other cleavage sites; the unique-base analysis identifies from this sequence only the bases that match the alternating purine-pyrimidine characteristics of the other topo II sites.

Statistical methods. Information content is a measure of the nonrandomness of base frequencies at each aligned base position for a set of DNA fragments (5, 15, 16). Adjusted information content values were calculated as described (5, 15, 16), as implemented by the Eden-Genesys sequence analysis computer program. The conventional χ^2 goodness of fit test was used for each base position, as described (16), using the sequence background base proportions as the expected values. The consensus sequence shown in Fig. 6 was derived by comparison of χ^2 values with the critical value, with 3 degrees of freedom at a confidence level of 99.5% ($p < 0.005$).

Results

Comparison of etoposide and quinolone effects on relaxation and cleavage of DNA by topo II. Reactions containing supercoiled plasmid DNA, human topo II, and etoposide or the quinolone CP were incubated and then terminated as described in Experimental Procedures. The samples, containing drug concentrations as indicated in Fig. 1, were analyzed on agarose gels containing (Fig. 1A) or lacking (Fig. 1B) ethidium bromide. The results indicated that CP induced topo II to make more double-stranded breaks in the DNA substrate (form III DNA) at 4 or 20 μg/ml CP than did etoposide even at 500 μg/ml (Fig. 1, compare *lanes 10* and *11* with *lanes 3*), whereas the two drugs were roughly comparable in inducing nicked circle (form II) DNA. An additional difference in the activities of the two drugs was that, at the higher concentrations of CP (Fig. 1, *lanes 8* and *9*), both the relaxation and cleavage activities of topo II were completely inhibited, whereas etoposide did not display this effect (at the concentrations tested).

In an experiment analogous to that of Fig. 1, cleavage reactions were performed on a linear, end-labeled DNA substrate, with topo II and varying concentrations of etoposide or CP; the results were analyzed on a sequencing gel, as shown in Fig. 2A. Consistent with the results shown in Fig. 1, CP exhibited its maximal stimulation of DNA cleavage by topo II at a final concentration of 20 μg/ml, and high CP concentrations (100 or 500 μg/ml) inhibited DNA cleavage by topo II (note that 500 μg/ml etoposide is an approximately 15-fold molar excess of drug, compared with 20 μg/ml CP). Etoposide induced topo II cleavage of DNA at all concentrations tested, and, on this DNA fragment, many more cleavage sites were observed than with CP.

Analysis of topo II inhibition by high levels of quinolone. Additional experiments were carried out to investigate the cause of the inhibition of topo II activities (both catalytic and cleavage) by higher concentrations of the quinolone. Two possibilities existed; either the drug interacted with the DNA substrate, precluding topo II activity (as did DNA-binding drugs like doxorubicin), or the drug interacted with the enzyme at a step before formation of a covalent cleavable complex between topo II and DNA. A test was designed to determine which of the two reagents needed to be added in excess to overcome the inhibition (either DNA or enzyme). Fig. 2, B and C, shows the results of decatenation assays containing kDNA and topo II, with or without the quinolone.

In Fig. 2B, *lane 1*, is a marker of catenated kDNA that remained in the wells. In Fig. 2B, *lanes 2–4*, 100, 200, and 400 ng of kDNA, respectively, were incubated with 0.7 unit of topo II in the absence of CP; decatenation was complete in all cases. Reactions in Fig. 2B, *lanes 5–7*, were identical to those in Fig. 2B, *lanes 2–4*, except that the drug was present at 40 μg/ml (a level selected to give a detectable amount of decatenated kDNA product). In the presence of CP, reactions containing 200 or 400 ng of kDNA yielded no increase in the amount of product decatenated by topo II (only the amount of catenated kDNA in the wells increased); thus, additional quantities of DNA did not increase the level of activity detected. Fig. 2B, *lane 8*, is identical to Fig. 2B, *lane 5*, except that 1.5 units of topo II were reacted with kDNA in the presence of 40 μg/ml CP. Activity was nearly complete in this reaction, suggesting that additional enzyme overcame the inhibition of enzyme activity by the quinolone; this conclu-

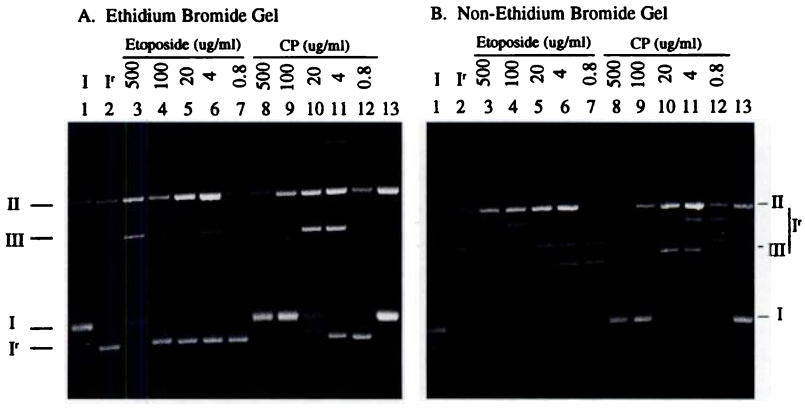


Fig. 1. Drug titration of topo II relaxation. Topo II was incubated with supercoiled plasmid DNA and either etoposide (*lanes 3–7*) or quinolone (CP) (*lanes 8–12*), at the final drug concentrations indicated, and the relaxation assays were analyzed as described in Experimental Procedures. *Lanes 1*, untreated supercoiled DNA; *lanes 2*, relaxed DNA marker (no drugs); *lanes 13*, an older preparation of the same DNA, to mark the position of nicked open circular DNA. A and B, Identical samples, but the gel in A was run in the presence of 0.5 μg/ml ethidium bromide. The gel mobilities for the different DNA forms are indicated for each gel, where form I is supercoiled, form Ir is relaxed, covalently closed, circular DNA, form II is nicked circles, and form III is linearized plasmid.

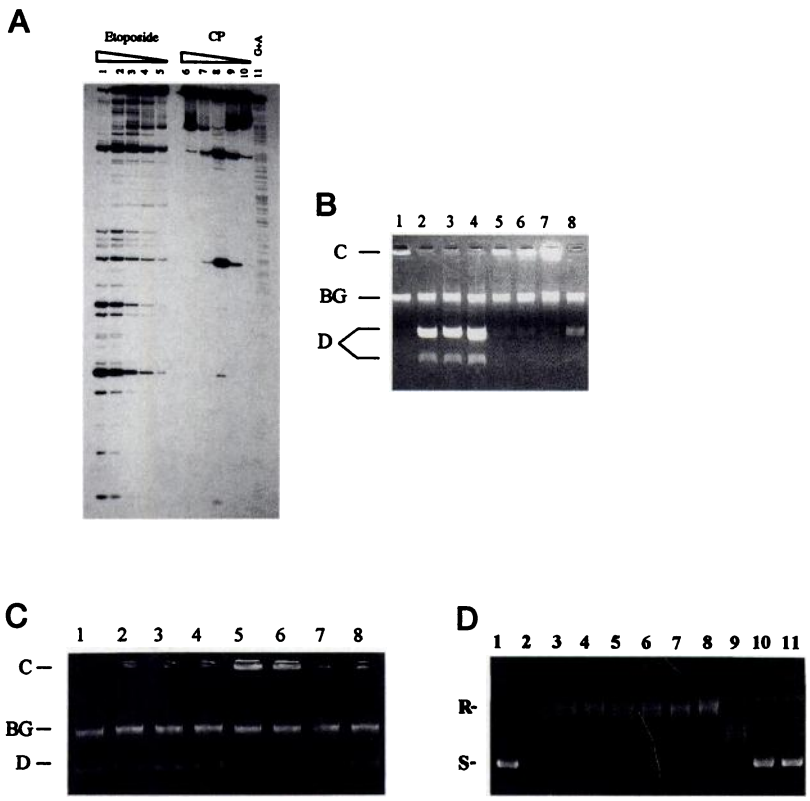


Fig. 2. Drug titration of topo II DNA cleavage. A, Topo II cleavage assays were performed on a *HindIII/BglI* end-labeled DNA fragment from pJH in reactions containing etoposide (*lanes 1–5*) or quinolone (CP) (*lanes 6–10*), as described in Experimental Procedures. Final concentrations of the indicated drugs were 500 μg/ml (*lanes 1* and *6*), 100 μg/ml (*lanes 2* and *7*), 20 μg/ml (*lanes 3* and *8*), 4 μg/ml (*lanes 4* and *9*), and 0.8 μg/ml (*lanes 5* and *10*). *Lane 11*, purine (guanine plus adenine) chemical sequencing reaction of the DNA fragment. Samples were prepared and run on a 12% polyacrylamide gel, as described by Spitzner *et al.* (12). B and C, Topo II decatenation reactions were analyzed on agarose gels, as described in Experimental Procedures. The positions of catenated kDNA (*C*) and decatenated products (*D*) are indicated, as well as an irrelevant background nucleic acid band (*BG*). B, *Lane 1*, kDNA marker (100 ng); *lanes 2–7*, decatenation reactions (incubated for 30 min at 37°) containing 0.7 unit of topo II and either no drug (*lanes 2–4*) or 40 μg/ml CP (*lanes 5–7*), with kDNA amounts of 100 ng (*lanes 2* and *5*), 200 ng (*lanes 3* and *6*), or 400 ng (*lanes 4* and *7*); *lane 8*, 1.5 units of topo II, 40 μg/ml CP, and 100 ng of kDNA. C, *Lanes 1–8*, 100 ng of kDNA; *lanes 1–4*, no drug; *lanes 5–8*, 40 μg/ml CP. Reactions (incubated for 30 min at 37°) contained topo II amounts of 0.25 unit (*lanes 1* and *5*), 0.5 unit (*lanes 2* and *6*), 1.0 unit (*lanes 3* and *7*), or 2.0 units (*lanes 4* and *8*). D, Results of a DNA unwinding experiment using supercoiled plasmid DNA, topo I, and topo I buffer (see Experimental Procedures), in the presence of drugs as specified, are shown. *Lane 2*, no drug (relaxed DNA marker) (*R*); *lanes 3–7*, with CP at final concentrations of 2, 8, 20, 40, and 100 μg/ml, respectively; *lanes 8–11*, with doxorubicin at 1, 2, 4, and 8 μM, respectively; *lane 1*, no enzyme, showing the gel mobility of the supercoiled plasmid DNA (*S*).

sion is confirmed in Fig. 2C (in which all reactions contained 100 ng of kDNA). This experiment reveals the effects of enzyme titration, in the absence (Fig. 2C, *lanes 1–4*) or presence (Fig. 2C, *lanes 5–8*) of 40 μg/ml CP, on decatenation of kDNA by topo II. In the absence of the quinolone the amount of decatenated product rose proportionally with the topo II levels; with 0.25 unit of enzyme virtually no decatenation product was visible (Fig. 2C, *lane 5*), whereas with 2 units (Fig. 2C, *lane 8*) the decatenation of kDNA neared completion (compare with Fig. 2C, *lane 4*, lacking drug). These results show that the inhibition of topo II activity by high concentrations of CP is due to a titration of the amount of enzyme available and that this interaction of CP and topo II takes place before formation of a covalent cleavable complex between topo II and DNA (because both cleavage and catalytic activities were inhibited).

To ensure that the inhibition was not due to DNA intercalation by CP, DNA unwinding experiments were carried out as described (14a). The results, shown in Fig. 2D, reveal that even at the highest quinolone concentration tested (100 μg/ml, or 280 μM) Fig. 2D, *lane 7*), which was greater than the amount required for topo II inhibition (see above), no DNA unwinding was observed. In contrast, doxorubicin displayed extensive DNA unwinding at 4 μM (Fig. 2D, *lane 10*). A complete lack of DNA unwinding was also found when the same reactions with CP were performed in the topo II decatenation buffer, which contained 10 mM $MgCl_2$ (data not shown). Note that these experiments do not preclude either nonintercalative binding of CP to DNA or DNA intercalation at a higher concentration of CP (above that required for topo II inhibition).

Sequencing of quinolone-induced topo II cleavage sites and comparison with sites induced by other drugs. Cleavage reactions were performed on a number of end-labeled DNA fragments with topo II in the presence of CP, as well as with no drug, AMSA, VM-26, or etoposide. The results obtained with three of these DNA fragments are shown in Fig. 3. The sequencing gels in Fig. 3 (and others not shown) revealed that etoposide and VM-26 displayed very similar DNA cleavage patterns and that these drugs induced more sites of DNA cleavage by topo II than did AMSA or CP (which induced about equal numbers of sites) (see also Table 3). Many of the topo II sites observed with CP were also observed with etoposide (and VM-26), and many CP sites not common to that set were instead observed with AMSA. In other words, CP induced few novel topo II sites; in fact, only 16 of the 85 CP sites did not co-localize with either VM-26 or AMSA sites (a comparison of drug cross-reactivities is discussed below). Fig. 3C shows the results of topo II cleavage assays with a DNA fragment containing a 54-base alternating purine-pyrimidine tract (Fig. 3C, *RY*) from the 5′ region of the human β-globin gene. This sequence was used in a previous study to show that topo II displays a remarkable cleavage activity at these repeats (12); clearly, quinolone-induced topo II sites also exhibit a marked preference for alternating purine-pyrimidine repeats. Within the RY repeat in Fig. 3C, topo II, alone or with any of the four drugs, cleaved the DNA at sites relatively conserved for all five cases, consistent with the findings reported previously (12).

Analysis of quinolone-induced topo II sites by methods A and B. A total of 85 CP topo II sites were sequenced from nine different DNA fragments containing a variety of background base proportions. These sites were entered into the Eden-Genesys sequence analysis computer program, as described in Experimental Procedures, and 34 base subfragments spanning the bases −15 to +19 (relative to cleavage site positions on either strand) were subjected to the data analyses indicated below. Method A (Table 1) is the conven-

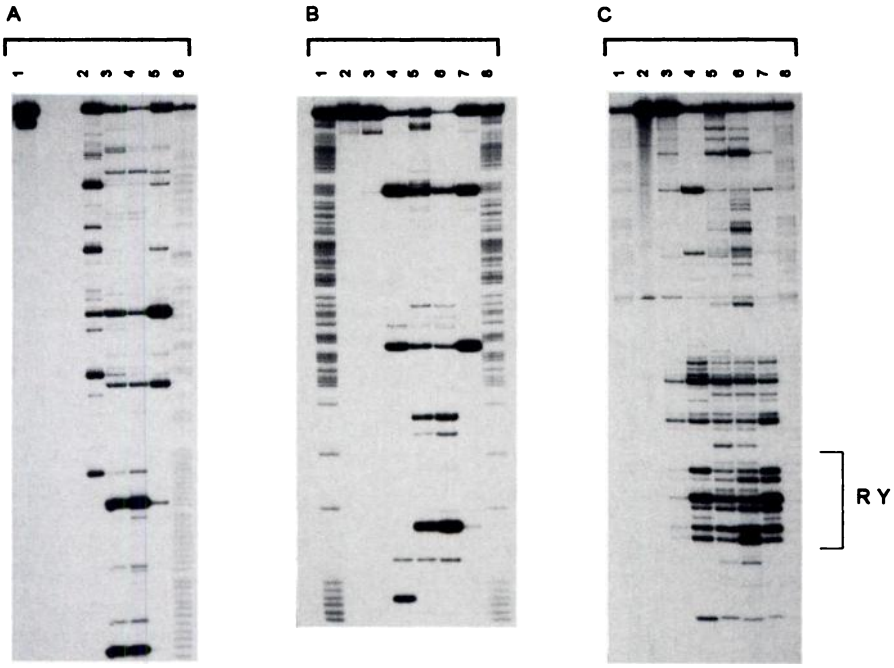


Fig. 3. Sequencing of drug-induced topo II cleavage sites on DNA. Topo II sites were mapped in the presence of the different drugs on nine end-labeled DNA fragments; the sequencing gels for three of these DNA fragments are shown. A, The DNA substrate is an *Eco*RI/*Pvu*II fragment of pJH; B, the gel contains a *Hind*III/*Bgl*I fragment of pJH (see Fig. 2, legend); C, the gel contains a *Hind*III/*Nde*I fragment of plasmid pRYG (12). Cleavage reactions were performed and analyzed as in Fig. 2. Final drug concentrations were 20 μg/ml quinolone (CP), 500 μg/ml etoposide, 500 μg/ml VM-26, and 50 μg/ml AMSA. A, *Lane 1*, DNA alone; *lane 2*, topo II plus AMSA; *lane 3*, topo II plus VM-26; *lane 4*, topo II plus etoposide; *lane 5*, topo II plus quinolone; *lane 6*, guanine plus adenine sequencing markers. B, *Lanes 1* and *8*, guanine plus adenine sequencing markers; *lane 2*, DNA alone; *lane 3*, topo II; *lane 4*, topo II plus AMSA; *lane 5*, topo II plus VM-26; *lane 6*, topo II plus etoposide; *lane 7*, topo II plus quinolone. C, *Lanes 1–8*, as in B. *RY*, the 54-base alternating purine-pyrimidine sequence contained in this DNA fragment.

tional manner for consensus sequence determination and uses only the bases of the DNA strand that was actually sequenced (the top strand). Base frequencies were tabulated for each of the 34 top strand base positions. Using these base frequencies, the information content of the data set was calculated at each position; this is shown graphically in Fig. 4A. Information content is a measure of the nonrandomness of base occurrences (5, 15, 17, 18), which, unlike χ^2 analysis, is independent of sample size. Fig. 4A shows that the base frequencies of the aligned CP data set are essentially random. Only positions −1 and +5 (which are symmetric; position +5 is immediately 5′ to the cleavage site on the bottom strand) provide some information, and there is clearly not sufficient information from this data treatment to derive a useful consensus sequence for CP topo II sites.

Topo II makes double-stranded breaks in DNA. Thus, each of the 85 CP-induced topo II sites contains one top strand cleavage site (which was identified on the sequencing gels) and one bottom strand site located on the complementary strand, offset 4 bases from the top strand site (as with the restriction enzyme *Eco*RI). In method B (Table 1), the top and bottom strand sites were pooled and treated as 170 separate sites, each aligned relative to the cleavage site on its own strand. The information content of these 170 sites is plotted in Fig. 4B; naturally, the plot is symmetric (the axis of dyad symmetry lies between positions +2 and +3). Within the 170 pooled sites, the only base identities conserved were at positions −2 and −1 (+6 and +5 are the corresponding bottom strand positions), which contain 43% adenines and 43% cytosines, respectively. This does not, however, mean that the observed base frequencies are random. In fact, as revealed by the χ^2 plot in Fig. 4C, these pooled topo II sites contain base proportions that are substantially different from background proportions. Most of the positions within the 16 bases spanning position −6 to position +10 have χ^2 values above the 99.5% significance value of 12.84 ($p < 0.005$) for 3 degrees of freedom. Thus, the CP data set contains considerable total base information, which is simply not best reduced to a consensus sequence by method B.

Analysis of best single strand from topo II sites by method C. An alternative to the objective analyses of topo II sites described above requires the selection of one strand from each of the 85 paired CP sites; the strand selected should have the most in common with each of the other 84 selected strands. This is the method used for other topo II DNA consensus sequence derivations, with the individual strands being chosen either by human selection (19) or by iterative computer-assisted optimization (5). An example of a strand-selected consensus sequence is shown as method C in Table 1. To determine whether an informative consensus sequence could be derived in this manner, the Eden-Genesys computer program was used to select cleavage sites in random orders and, with each subsequent selection, to determine which strand of each site pair yielded a larger information content total when added to the previous result. After one strand from each pair had been selected for all 85 sites, the total information content was calculated for this new aligned data set. The complete alignment process was repeated many times, using random orders each time, until no order produced a better final alignment. It should be noted that these alignments are different from typical alignments, in that there is no changing of the alignment point (no sliding 5′ or 3′) but only the selection of one strand or the other at its cleavage site.

An optimal strand-aligned set of the 85 CP sites was analyzed for information content, by position; the results are shown graphically in Fig. 5A. A plot of χ^2 analysis of this same aligned set (versus background base proportions) is shown in Fig. 5B. Both results reveal that the "best-strand" data set contains base frequencies that are significantly nonrandom in nature, particularly in a region spanning positions −6 to +9, relative to the topo II cleavage site. The base proportions for this set of 85 CP sites are shown in Fig. 5C, along with a consensus sequence determined by χ^2 calcula-

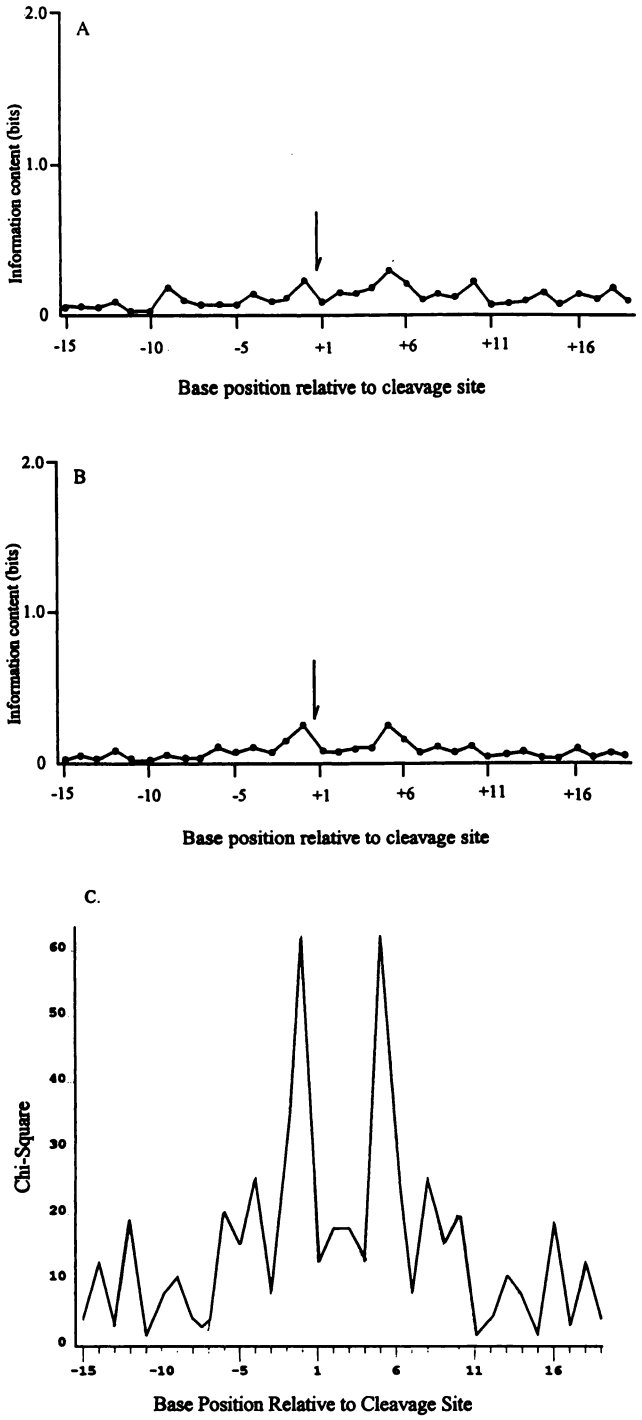


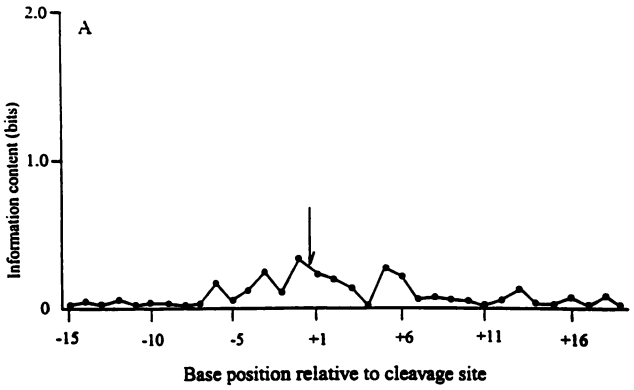
Fig. 4. Information content plots of quinolone-induced topo II data set (85 CP sites). A, Sites aligned by method A (top or sequenced strand). B, Sites aligned by method B (double-stranded analysis). The analysis methods for the quinolone data set are described in the text. The graphs show the information content values calculated for each position, relative to the cleavage site (*arrow*), for the entire data set, plotted for positions −15 to +19. An information content value near 0 bits indicates essentially random (nonconserved) base frequencies at that position; if, for example, every one of the 85 sites had a cytosine at −1, the information content for that position would be the maximum of 2.0 bits. C, Same data from method B, plotted using the calculated χ^2 values at each position, relative to the topo II cleavage sites (with total background base proportions used as the expected frequencies). The 99.5% significance level for these χ^2 values is 12.84, indicating that a number of positions, most notably −1 and +5, contain base frequencies that are not random; little consensus sequence information can be deduced from these results, however.

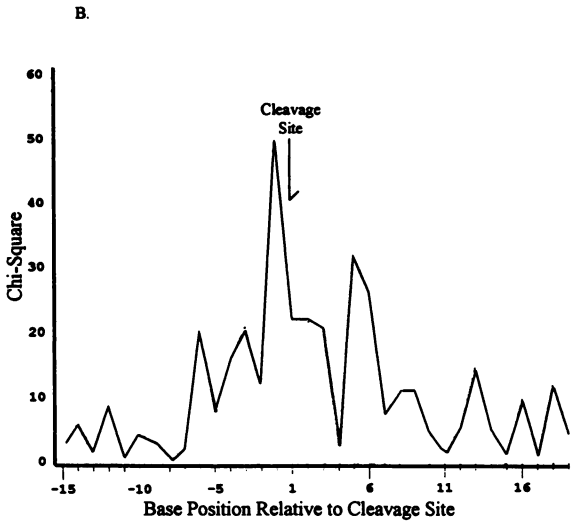
tions of single or double bases (at 99.5% significance). These results further reveal that the CP sites contain base information that is clearly nonrandom.

Unique-base consensus model for topo II sites. The best-strand results do not appear to be an ideal solution to the analysis of topo II sites. The consensus sequence is highly degenerate, and the strand selection process is somewhat arbitrary. The consensus matrix itself may be a better representation of topo II sites than is a consensus sequence (5, 18), but it is not easy to compare the matrix of one drug with that of another. We previously showed that the unique model for topo II sites was useful for site prediction (18); in that report we used a consensus sequence to examine whether a test site had a match to each consensus base on one strand or the other (but with no additional score for matching on both strands). Here, we extend this work and describe the unique-base proportion analysis (Table 1, method D), which requires no strand selections but reveals the contributions of both strands of topo II sites without the dilution of information observed with a normal double-stranded analysis (Fig. 4B). In this method, for each base position (relative to the cleavage site) the proportion of sequences that have each nucleotide at that position (on one strand or the other) is determined. The unique-base proportions for the 85 CP-induced topo II sites are shown in Fig. 6. Thus, 63.5% of the 85 sites had a cytosine on at least one of the strands at the −1 position (immediately 5′ of the cleavage event), whereas only 14.1% had a guanine at one of the −1 positions (either a −1 guanine or a +5 cytosine). Note that this method provides information that is different from the conventional single- or double-stranded analyses.

Are the results from the unique-base analysis meaningful? A consensus sequence for CP sites was determined by comparing these base proportions with thresholds equal to twice the expected (background) base proportions from the sites (this is analogous to a cutoff of 50% when background proportions are 25% in a normal analysis); the unique consensus sequence is shown in Fig. 6. In contrast to the best-strand consensus sequence (Fig. 5C), this CP-induced topo II site consensus sequence is less degenerate. The unique-base model clearly identifies the bases that were conserved in the consensus sequence in Fig. 5C but, in addition, detects preferences that were not easily identified by the other methods. Thus, the unique-base model may be a more sensitive measure for identifying conserved bases from double-stranded topo II sites.

Comparison of drug-induced topo II sites using unique consensus sequences. To test the usefulness of the unique-base model for characterizing sets of topo II cleavage sites, unique-base proportions were determined for the seven different data sets of topo II sites described in Table 2. Unique consensus sequences were derived for each data set (exactly as for the CP sites); these are shown in Table 2. The unique consensus sequences are strikingly similar for all seven topo II data sets at positions −6 to −2 (and symmetrically at positions +6 to +10, which represent positions −6 to −2 on the bottom strand). In contrast, the sets differ somewhat at position −1 (+5), varying among preferences for cytosine, thymine, or any pyrimidine, and are quite different at positions +1 and +2 (and symmetrically at positions +3 and +4). The base conservations appear independent of drug identities or of any drug inclusion, suggesting that these





C.

Position	Raw Frequencies A	C	G	T	Base Proportions A	C	G	T	% Redundancy	Consensus sequence Based on 99.5% chi-square
-6	41	17	14	12	0.488	0.202	0.167	0.143	9.94	A
-5	13	23	17	31	0.155	0.274	0.202	0.369	3.81	N
-4	29	15	30	10	0.345	0.179	0.357	0.119	6.56	G/A
-3	35	8	9	32	0.417	0.095	0.107	0.381	13.78	A/T
-2	31	10	27	16	0.369	0.119	0.321	0.190	6.13	A/G
-1	17	45	3	20	0.200	0.529	0.035	0.235	19.46	C
1	40	11	8	26	0.471	0.129	0.094	0.306	13.19	A
2	14	7	28	36	0.165	0.082	0.329	0.424	11.19	T/G
3	26	16	34	9	0.306	0.188	0.400	0.106	7.70	G/A
4	19	24	16	26	0.224	0.282	0.188	0.306	1.33	N
5	40	11	28	6	0.471	0.129	0.329	0.071	15.53	A/G
6	7	17	19	42	0.082	0.200	0.224	0.494	12.73	T
7	29	23	21	12	0.341	0.271	0.247	0.141	3.20	N
8	16	26	11	32	0.188	0.306	0.129	0.376	5.63	T/C
9	24	11	31	19	0.282	0.129	0.365	0.224	4.52	N
10	17	22	16	30	0.200	0.259	0.188	0.353	2.39	N

Fig. 5. Analysis of 85 quinolone-induced topo II sites aligned by method C (best single strand). A, Plot of information content, by position, for the CP cleavage sites (compare with plots of the same data set aligned by methods A and B in Fig. 4). B, Graph of χ^2 values computed for each aligned position, from −15 to +19. *Arrow*, cleavage site. The critical value for a χ^2 test with 3 degrees of freedom and a 99.5% confidence level (=0.005) is 12.84. C, Sequence details of the 85 quinolone-induced topo II sites, as aligned by method C, for the region containing the greatest amount of sequence information, which spans positions −6 to +10, relative to the cleavage site location. The raw frequencies are the numbers of each base occurrence at each position for the 85 CP sites. The base proportions represent the proportions of each base at each position in the data set, as determined by dividing the raw frequencies by the number of sites (i.e., 85). Percentage redundancy is a measure of information content in which the value (in bits) at each base position (see A) is divided by the maximal possible information content value for the set and then multiplied by 100%. The consensus sequence for the 85 quinolone-induced topo II sites aligned by method C was determined by computing χ^2 confidence intervals independently at each position, either for single bases (adenine, cytosine, guanine, and thymine) or, if no single base lay outside this interval, for pairs of bases (adenine plus cytosine, adenine plus guanine, adenine plus thymine, and their complements). If a position did not contain base frequencies with significant χ^2 values (at 99.5%) for single- or double-base occurrences, then an N (any base) was assigned to the consensussequence.

Position	Raw Frequencies A	C	G	T	Base Proportions A	C	G	T	Consensus sequence Based on 99.5% chi-square
-6	55	32	32	29	0.647	0.376	0.376	0.341	A
-5	27	46	25	43	0.318	0.541	0.294	0.506	C
-4	47	22	44	24	0.553	0.259	0.518	0.282	G/A
-3	42	26	28	47	0.494	0.306	0.329	0.553	T
-2	53	25	40	22	0.624	0.294	0.471	0.259	A/G
-1	20	54	12	43	0.235	0.635	0.141	0.506	C
1	45	22	28	41	0.529	0.259	0.329	0.482	N (A)
2	21	38	36	48	0.247	0.447	0.424	0.565	T
3	48	36	38	21	0.565	0.424	0.447	0.247	A
4	41	28	22	45	0.482	0.329	0.259	0.529	N (T)
5	43	12	54	20	0.506	0.141	0.635	0.235	G
6	22	40	25	53	0.259	0.471	0.294	0.624	T/C
7	47	28	26	42	0.553	0.329	0.306	0.494	A
8	24	44	22	47	0.282	0.518	0.259	0.553	C/T
9	43	25	46	27	0.506	0.294	0.541	0.318	G
10	29	32	32	55	0.341	0.376	0.376	0.647	T

Fig. 6. Analysis of 85 quinolone-induced topo II sites aligned by method D (unique-base model). The unique-base analysis (see text and Table 1) of the quinolone cleavage sites yielded the unique raw frequencies of bases shown here for positions −6 to +10. Division of these unique-base frequencies by the total number of CP cleavage sites (85) resulted in the unique-base proportion values shown. The threshold for consensus sequence assignment was set at twice the expected base proportions, which was calculated for the DNA fragments used in these experiments to be adenine = thymine = 0.2745 and cytosine = guanine = 0.2255; therefore, any unique-base proportions above adenine or thymine > 0.549 or cytosine or guanine > 0.451 were assigned as consensus sequence bases. Positions +1 and +4 had no unique-base proportions above these thresholds, so they were assigned as N (any base), even though the values for adenine and thymine, respectively, were greater than 0.5 (0.529). Note that the expected background unique-base proportions would be cytosine = guanine = 0.4001 [calculated as [0.2255 + 0.2255 − (0.2255 × 0.2255)]] and adenine = thymine = 0.4736.

positions are truly preferences of the enzyme that were not affected by the presence of drugs.

Examination of all seven consensus sequences shows that, independently of drug effects, topo II prefers sequences with an adenine at a −6 position (equal to a thymine at +10), cytosine at −5, any purine at −4, and thymine at −3. All seven data sets contain a purine at position −2 on one of the strands, but the preference for adenine, guanine, or either appears somewhat dependent upon the drug used (or the DNA sequences analyzed), although the differences in base proportions were not large. Somewhat greater differences were observed at the −1 position. All consensus sequences have pyrimidines at position −1, but only the quinolone (CP) and VM-26 exhibited preferences for cytosine over thymine; this is consistent with the examination of VM-26-induced topo II sites by others (6). The greatest differences between the data sets were found at position +1 (immediately 3′ of the top strand cleavage event). In the presence of AMSA, as with the enzyme alone, there was a strong preference for adenine at +1 (equal to a thymine at +4). In contrast, ellipticine-induced topo II cleavages showed a cytosine conserved at position +1, whereas the CP and VM-26 data sets displayed only weak base preferences at +1 (for adenine and guanine,

TABLE 2

Comparison of unique-base topo II consensus sequences

Seven different sets of DNA sites cleaved by topo II were each independently subjected to the unique-base method for analysis of topo II sites. The rules for assignment as consensus sequence bases were the same for all data sets; for each set, the unique-base proportions calculated were compared with threshold values of twice the background base proportions observed within that data set. For each of the seven sets of topo II sites, the resulting consensus sequences are shown; in each set, the single most conserved base occurrence is indicated in bold type, for each strand. The one or two bases with unique-base proportions above the calculated thresholds are shown at the appropriate base positions, relative to the cleavage site (between −1 and +1). Base identities shown in parentheses [such as (A) at +1 for the CP set] indicate that the unique-base proportion for this base was greater than 0.5 but less than the threshold value (in other words, close but not great enough for true assignment). Bases shown in brackets [only [G] at +1 and [C] at +4 for VM-26] indicate that unique-base proportions were above the threshold value but were less than 0.5. The seven cleavage site data sets are as follows: CP, the 85 quinolone-induced topo II sites described in this report; AMSA, the 86 AMSA-induced topo II sites collected in this study from the same DNA fragments and sequencing gels as the quinolone sites; VM-26, the 134 VM-26-induced topo II sites collected in this study from the same DNA fragments and sequencing gels as the quinolone sites and the AMSA sites described above; 71 AMSA, the 71 AMSA-induced chicken topo II sites reported previously (5, 18); no drug (65), the 65 chicken topo II sites identified in the absence of any drugs (18); *Drosophila*, the 77 *Drosophila* topo II sites sequenced in the absence of any drugs (19, 27) and additional data kindly provided by T. Hsieh; ellipticine, the 25 calf thymus topo II sites induced by an ellipticine derivative (8).

Data set (no. of sites sequenced)	Base (cleavage between −1 and +1)															
	−6	−5	−4	−3	−2	−1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
CP (85)	A	C	G/A	T	A/G	**C**/(T)	N/(A)	T	A	N/(T)	**G**/(A)	T/C	A	C/T	G	T
AMSA (86)	A	C/(T)	A/G	T	A/G	T/C	**A**	C	G	**T**	A/G	T/C	A	T/C	G/(A)	T
VM-26 (134)	A	C	G/(A)	T	G/(A)	**C**	[G]	C	G	[C]	**G**	C/(T)	A	C/(T)	G	T
71 AMSA (71)	A/T	C	A/G	A/T	G/A	T/C	**A**	G/C	C/G	**T**	A/G	C/T	T/A	T/C	G	T/A
No drug (65)	T	C	G/A	T	G	T	**A**	G	C	**T**	A	C	A	C/T	G	A
Drosophila (77)	A	C/(T)	G/A	T	A	T/C	**A**	T	A	**T**	A/G	T	A	C/T	G/(A)	T
Ellipticine (25)	A	C	G	T	A/G	**T**	C	C	G	G	**A**	T/C	A	C	G	T

respectively). There were also marked differences at the +2 position (symmetrical with +3), although these base preferences were not as conserved within the individual data sets as were those at the −1 or +1 positions (data not shown). An interesting characteristic of all of the consensus sequences is a strong tendency toward being of alternating purine-pyrimidine nature, consistent with the observation that topo II is extremely reactive towards such alternating sequences (12).

The best way to compare sequence specificities of drug-induced topo II sites is to gather the data from the same DNA fragments and same sequencing gels, as was done in this study for the quinolone (CP), AMSA, and VM-26 (Table 2). This method eliminates any differences among sequence backgrounds from which the data were collected and allows empirical determination of which sites were cleaved by topo II in the presence of different drugs. A chart of drug overlap results is shown in Table 3. It can be seen that VM-26 induced considerably more enzyme cleavage sites than did AMSA or the quinolone; this is consistent with previous findings (6). It seems that VM-26 sites are somewhat less well specified than the other drug data, and the total value for the information content of the VM-26 data set is lower than that for AMSA or CP (data not shown). The results in Table 3 suggest that the recognition signals for quinolone (CP)-induced topo II cleavage are more similar to those for VM-26 than to those for AMSA, because 60 of the 85 CP sites were also observed with VM-26, whereas only 41 co-localized with AMSA sites (note that these are empirical data and not predictions). About one half of the AMSA sites were identical to either of the other drug sites, and 28 sites were observed that were cleaved by topo II in the presence of all three drugs (data not shown). A comparison of these three consensus sequences in Table 2 suggests that the factor common to VM-26 and CP sites is the strong preference for cytosine at the −1 position, whereas these drugs display different preferences at the less conserved +1 to +4 positions.

TABLE 3

Sequence co-localization of topo II sites induced by different drugs

Drug-induced topo II cleavage sites were sequenced from nine different DNA fragments in this study, and quinolone (CP), VM-26, and AMSA sites were collected simultaneously (see Fig. 3). This allowed empirical determination of which cleavage sites identified with one drug were also (to the base pair) observed with a different drug. This table shows the number of the 85 CP sites also induced by VM-26 or AMSA, and so forth. Thus, 60 of the 85 quinolone sites were also VM-26-induced topo II sites, whereas only 45 of the 134 VM-26 sites were also cleaved in the presence of AMSA.

	Co-localized sites		
	CP	VM-26	AMSA
85 CP sites	85/85	60/85	41/85
134 VM-26 sites	60/134	134/134	45/134
86 AMSA sites	41/86	45/86	86/86

To address the similarities and differences between the CP sites and the AMSA or VM-26 sites in greater detail, the 85 CP sites were divided into two subsets for each drug comparison. As suggested in Table 3, one subset consisted of the 60 CP sites that were also cleaved by topo II in the presence of VM-26 and the other contained the 25 CP sites not in the VM-26 data set; likewise, the second pair of subsets was broken down into the 41 CP sites also observed with AMSA and the 44 CP sites that were not. Each of the four subsets was analyzed individually by the unique-base method described above; the resulting consensus sequences are shown in Table 4. Comparison of CP sites also cleaved by VM-26 (Table 4, CP and VM-26) with those not co-localized with VM-26 sites (Table 4, CP not VM-26) reveals differences between the data sets at a number of base positions. The CP and VM-26 sites have a strongly conserved −1 cytosine; notably, the predominant −1 base in the CP not VM-26 data is a thymine (in the data sets in Table 2, it can be seen that a −1 cytosine is a stronger preference for VM-26 than even for CP). Other differences include high proportions in the CP not VM-26 sites of −3 guanine, +1 thymine, +2 guanine, and +3 cytosine, none of which were preferred bases for the whole VM-26 data set. Thus, it may be that particular base occurrences (or combinations) that are acceptable for CP-mediated topo II cleavage preclude the interaction with VM-26; due to the small number of sites in the CP not VM-26 data set, however, strong conclusions should not be drawn, except to

TABLE 4

Analysis of conserved and nonconserved CP sites

The CP topo II sites were broken into subsets that either were (and) or were not (not) also co-localized with VM-26 or AMSA sites, as described in the text. Unique-base consensus sequences were derived from each of these subsets exactly as described for Table 2.

	Base															
	−6	−5	−4	−3	−2	−1	+1	2	3	4	5	6	7	8	9	10
CP and VM-26	A	C/T	A/G	T/A	A/G	C	A	T/[C][a]	A/[G]	T	G	T/C	A/T	T/C	A/G	T
CP not VM-26	(A)/(T)	[C]/[G]	N	[G]	A	T/[C]	(A)/(T)	[G]/(T)	[C]/(A)	(A)/(T)	A/[G]	T	[C]	N	[C]/[G]	(T)/(A)
CP and AMSA	A	C/T	A/G	T	A	T/C	A	[T]/[C]	[A]/[G]	T	A/G	T	A	C/T	A/G	T
CP not AMSA	A	[C]	[G]	A/[T]	G	C	T	T	A	A	G	C	T/[A]	[C]	[G]	T

[a] Brackets, value greater than 0.5 but not greater than twice the expected base proportions; parentheses, value greater than twice the expected base proportions but not greater than 0.5.

note that these sites contain sequences different from those associated with the VM-26 sites. The comparison of sites cleaved by both CP and AMSA with those sites not cleaved in the presence of AMSA yields readily interpretable information. At the +1 position in the CP and AMSA set, an adenine is preferred; this +1 adenine is the single most conserved base for AMSA sites (see both AMSA sets in Table 2). The CP not AMSA sites, however, lack this +1 adenine, instead showing a preference for thymine at +1. Several other distinct differences are also found between these two subsets over the region spanning position −2 to +6, relative to the cleavage sites. These co-localization analyses provide evidence that the unique-base model yields information that is useful for detecting similarities and differences between the sites of interaction of different drug classes.

Another way to test whether a consensus sequence model is useful is to ask whether it actually resembles the raw data from which it is derived and matches better to sites than to nonsites. The simplest way to convert the CP unique-base consensus sequence into a matrix representation is to assign a value of unity to consensus base matches and a value of 0 to all nonconsensus bases. In a matrix for the CP sites, position −6 would contain values of adenine = 1.0, cytosine = 0, guanine = 0, and thymine = 0, and a degenerate position such as −4 would contain adenine = 1.0, cytosine = 0, guanine = 1.0, and thymine = 0. If matrix match scores are then calculated for all sites (85 sites) and nonsites (>2200 verified nonsites) as in the report by DiDonato *et al.* (16), we find that the average matrix match score for sites is 0.47, whereas that for nonsites is 0.32. If this same CP matrix is applied to the 134 VM-26 sites, the average score is 0.43, whereas the match to the 86 AMSA sites scores 0.36 (the nonsites for each have the same average score as the CP nonsites); these values are consistent with the similarities between these sets, as reported above. If this CP matrix is applied to an unrelated set of DNA-binding fragments (herpes simplex virus type 1 ICP4 binding sites) (16), both sites and nonsites score 0.31. One way to lower the number of consensus matrix matches due to chance is to assign the degenerate bases values of 0.5, rather than 1.0. Then the values for position −6 (adenine) would be the same as described above, but those for −4 (any purine) would be adenine = 0.5, cytosine = 0, guanine = 0.5, and thymine = 0. After analysis of the sites and nonsites as described above but using this matrix, the true sites have an average score of 0.37, whereas the true nonsites have an average score of 0.25. The highest site score identified was 0.71, which was observed in only two instances, both of which were sequenced on perfect alternating purine-pyrimidine repeats; most sites scored closer to the average. If the same type of matrix is derived from the unique-base consensus sequence for the 134 VM-26 sites and is applied to these VM-26 sites, the average score is 0.34, whereas the nonsites score is 0.25. When the AMSA consensus matrix is applied to the 86 AMSA sites, true sites score 0.38 and nonsites average 0.25. These results show that the unique-base consensus sequences derived from these sets do resemble the topo II sites from which they were derived and can distinguish, on average, between sequences that were or were not cleaved by the enzyme; these findings are consistent with our previous matrix analyses of DNA binding sites (12, 16, 18).

Discussion

In previous studies, topo II sites were gathered from a variety of DNA fragments of differing background base proportions, and consensus sequences were determined using different methodologies (5–8, 20–23). Although these studies yielded useful information about topo II-DNA interactions (with or without drug), the results are not easily compared with one another. Our strategy was to characterize a consensus recognition element for the interaction of a novel quinolone (CP) with eukaryotic topo II and DNA and then to compare this compound with other topo II inhibitors. The first step was to determine the optimal drug concentration to use for analyses. Reactions were carried out using purified topo II, plasmid DNA, and either the quinolone CP or etoposide. CP generated much greater conversion of DNA substrate to its linear form (indicative of double-stranded breaks) than did etoposide at any concentration tested (Fig. 1). Also, CP exhibited maximal stimulation of enzyme-induced, double-stranded DNA breaks at 4–20 μg/ml final concentration. A similar experiment was conducted using a linear, uniquely end-labeled DNA fragment, and a concentration of 20 μg/ml was determined to be optimal for inducing DNA cleavage by topo II (Fig. 2).

An interesting observation made during the drug titration experiments (Figs. 1 and 2A) was that high concentrations of the quinolone completely inhibited DNA relaxation by topo II, as well as the induction of single- or double-stranded DNA breaks by the enzyme. We performed decatenation reactions in which the levels of either the DNA substrate or the enzyme were titrated in the presence of a constant amount of CP (Fig. 2, B and C). Only the addition of excess topo II was able to overcome the drug inhibition; excess DNA had no effect. This result is consistent with the observation that the inhibitory levels of CP were the same for the kDNA decatenation reactions containing up to 0.4 μg of DNA as for the DNA cleavage

reactions containing only a few nanograms of DNA. Furthermore, DNA unwinding experiments showed no evidence of DNA intercalation by CP at the concentrations tested (up to 100 μg/ml, 280 μM), unlike the doxorubicin control (Fig. 2D; other data not shown). We conclude that the inhibition effect is therefore due to an interaction between CP and topo II before the formation of a covalent cleavable complex between enzyme and DNA. This topo II inhibition by the quinolone is different from that by other agents that are DNA-binding compounds (nonspecific DNA intercalators such as ethidium bromide or topo II poisons such as doxorubicin). Thus, unlike other topo II poisons, CP has two modes of interaction with topo II, one that stabilizes the covalent enzyme-DNA complex and one that apparently prevents the initiation of topo II activity. This suggests that CP derivatives may be synthesized that optimize one mode of inhibition or the other.

To characterize topo II sites induced by CP, cleavage reactions were performed with nine different DNA fragments, using an optimal concentration of the quinolone. For comparison, topo II sites were also mapped on these DNA fragments in the presence of no drug, AMSA, VM-26, or etoposide. Qualitative and quantitative analyses of the data were described above, and conventional sequence analysis methods were shown not to be of great utility. In an attempt to derive a more useful characterization of topo II sites, a new method was devised, which we call the unique-base analysis. This method is analogous to the unique analysis described before (18), in which a consensus sequence was applied simultaneously to both strands to predict the relative strengths of cleavage sites. The difference here is that this new model essentially derives, independently, a consensus sequence for each base position, relative to the cleavage site. The procedure is described as method D in Table 1; the data for 85 CP sites are shown in Fig. 6. The result of applying this method to the CP data is the symmetric consensus sequence ACRTRC NTANGYAYGT (cleavage occurs at the carat and R is any purine, Y is any pyrimidine, and N is any base). This consensus sequence is similar to that of the best-strand method (Fig. 5C) but is less degenerate and required no strand selections in the analysis. This unique-base analysis was applied to six other topo II data sets, and the resulting consensus sequences are shown in Table 2. The surprising finding is that all seven data sets showed conserved base identities at positions −6 to −2 (ACRTR), and symmetrically at positions +6 to +10, but differed at positions −1 to +5 (note that positions −6 and +10 may also each be labeled "adenine or thymine"). These results suggest that the sequence recognition elements in the regions flanking the cleavage sites on either strand are dictated by preferences of topo II alone and that the different drugs alter these preferences only immediately around the cleavage site (note that position +5 is the −1 position for the bottom strand paired cleavage site). These observations imply that topo II inhibitors induce enzymatic cleavages at sites that in the flanking regions contain sequences sufficient for the binding of topo II and immediately around the cleavage site contain sequences at which the drugs can interact with the local DNA bases to form a drug-enzyme-DNA ternary complex. An interesting feature of all these consensus sequences is the degree to which they resemble alternating purine-pyrimidine repeats (even though relatively few of the cleavage sites analyzed contained such repeats). This is consistent with the observations that these repeat sequences are hotspots for topo II cleavage (12) (Fig. 3C) and suggests that this affinity is primarily directed by interactions of the enzyme with DNA, independently of drug effects. These results also show that different topo II enzymes bind similar DNA sites; Table 2 shows results for enzymes from four distinct species.

The unique-base analysis of sets of topo II cleavage sites permits rapid characterization of the sequences without any arbitrary or complicated strand selection procedures and yields a double-stranded representation of the topo II sites that can be directly compared with the results obtained from other data sets. Naturally, this methodology cannot generate base information that is not present in the topo II data sets; indeed, application of the unique-base analysis to random sequences yields no consensus sequences (data not shown). Rather, this procedure allows the identification of bases that are conserved among topo II sites but escape detection by more conventional means because of the degenerate nature of topo II sites. It is clear that base information is present in these data sets, because certain base occurrences are greatly pronounced (such as a −1 cytosine in the CP set) and are identified using any of the sequence analysis methods. Furthermore, χ^2 analysis of the CP sites (single or double stranded) reveals that the base frequencies at many positions flanking the cleavage sites are quite distinct from background proportions; in fact, numerous base positions display base frequencies that differ from the expected values even when examined using a significance level of 99.5% ($p < 0.005$).

A number of observations suggest that these unique-base consensus sequences represent useful characterizations of topo II sites. First, these consensus sequences specify the bases for more positions than do the other analysis methods (methods A, B, and C) but do contain all the information that was discernible from those analyses (compare with the method C consensus sequence). Second, the unique consensus sequences derived for all seven data sets shown in Table 2 show differences in base frequencies immediately around the cleavage site (where the drug-specific differences are expected to occur) but reveal highly conserved base identities across all sets at the flanking positions (−6 to −2 and +6 to +10); this consistency would be very unlikely to be due to chance. Third, the unique consensus sequences were shown to display quantitative differences between their similarities to true sites versus verified nonsites. Sites were shown to match the derived consensus matrices with scores approximately 50% greater than those for nonsites for all topo II data sets tested; this value is higher than for matrices derived from any of the other models tested (data not shown) and is also higher than differences reported for other predictive analyses of topo II sites (12, 18). Fourth, the unique-base consensus sequences derived for the various topo II data sets that were collected simultaneously using the same DNA substrates (quinolone, VM-26, and AMSA; see Tables 2–4) demonstrate relationships that are in accord with the empirical results. The quinolone sites co-localized with more VM-26 sites than AMSA sites, and the unique-base analyses revealed that CP and VM-26 share the feature of having a position −1 cytosine as the single most conserved base among the cleavage sites; in contrast, the most conserved base for AMSA was a position +1 adenine. The analyses of quinolone sites that did or did not co-localize with VM-26 or

AMSA sites (Table 3) show subset consensus sequences consistent with the similarities and differences found in the complete drug sets.

The relatively low frequencies of conserved bases in topo II cleavage sites (compared with many other, more specific, DNA-binding proteins) have made the study of drug-dependent effects a rather difficult issue upon which to draw definitive conclusions. However, it is clear that this issue is of importance, because the anticancer drugs that function by stabilizing the topo II-DNA covalent intermediate all have preferred DNA binding sites for interaction with the enzyme (for some drug classes, such as anthracyclines, by virtue of direct strong binding of DNA and for others by strong binding to DNA only as part of a complex with topo II). Therefore, to understand how topo II poisons function, it is necessary to investigate the mechanisms underlying the trapping of the covalent enzyme-DNA complex, including the aspect of the DNA sequences at which this takes place. The DNA sequence-specific interactions of the drugs may be directly related to their modes of action; for example, our findings that the majority of CP sites were also observed with VM-26 and that the most significant base preference for each is a −1 cytosine are consistent with the report by Robinson *et al.* (10) that a cell line containing a mutant topo II enzyme, rendering it resistant to VM-26, is also partially cross-resistant to quinolones. The observation that mitoxantrone induced topo II cleavage of DNA at many sites in common with VM-26 (25) and different from those of the anthracycline Adriamycin (7) suggests that knowledge about sequence specificities of topo II poisons reveals aspects of the activities of these drugs that would not be predicted based upon chemical structures. In particular, Capranico *et al.* (26) showed that two unrelated drugs, AMSA and bisantrene, each have a +1 adenine preference and that, based upon molecular modeling studies, they may share certain structural features when bound to DNA. It may also be significant that certain drug classes exhibit the strongest base preferences at the +1 position (e.g., AMSA), whereas other drug classes display the greatest base specificity at the −1 position (as does CP) (Table 2). Thus, the study of sequence preference classes for topo II poisons may be valuable both for the characterization of existing inhibitors and for the design of novel inhibitors; the sequence analysis method we present here is a simple and useful tool for these studies.

Acknowledgments

We thank Christine Sommer-Furbee for critical reading of the paper.

References

1. Liu, L. F. DNA topisomerase poisons as antitumor drugs. *Annu. Rev. Biochem.* **58:**351–375 (1989).
2. Kreuzer, K. N. DNA topoisomerases as potential targets of antiviral action. *Pharmacol. Ther.* **43:**377–395 (1989).
3. Gewirtz, D. A. Does bulk damage to DNA explain the cytostatic and cytotoxic effects of topoisomerase II inhibitors? *Biochem. Pharmacol.* **42:** 2253–2258 (1991).
4. Capranico, G., F. Zunino, K. W. Kohn, and Y. Pommier. Sequence-selective topoisomerase II inhibition by anthracycline derivatives in SV40 DNA: relationship with DNA binding affinity and cytotoxicity. *Biochemistry* **29:**562–569 (1990).
5. Spitzner, J. R., and M. T. Muller. A consensus sequence for cleavage by vertebrate DNA topoisomerase II. *Nucleic Acids Res.* **16:**5533–5556 (1988).
6. Pommier, Y., G. Capranico, A. Orr, and K. W. Kohn. Local base sequence preferences for DNA cleavage by mammalian topoisomerase II in the presence of amsacrine or teniposide. *Nucleic Acids Res.* **19:**5973–5980 (1991).
7. Capranico, G., K. W. Kohn, and Y. Pommier. Local sequence requirements for DNA cleavage by mammalian topoisomerase II in the presence of doxorubicin. *Nucleic Acids Res.* **18:**6611–6619 (1990).
8. Fosse, P., R. Brigitte, M. Le Bret, C. Paoletti, and J.-M. Saucier. Sequence requirements for mammalian topoisomerase II mediated DNA cleavage stimulated by an ellipticine derivative. *Nucleic Acids Res.* **19:**2861–2868 (1991).
9. Gootz, T. D., and N. Osheroff. Quinolones and eukaryotic topoisomerases, in *Quinolone Antimicrobial Agents* (D. C. Hooper and J. S. Wolfson, eds.). American Society for Microbiology, Washington, D. C., 139–160 (1993).
10. Robinson, M. J., B. A. Martin, T. D. Gootz, P. R. McGuirk, M. Moynihan, J. A. Sutcliffe, and N. Osheroff. Effects of quinolone derivatives on eukaryotic topoisomerase II: a novel mechanism for enhancement of enzyme-mediated DNA cleavage. *J. Biol. Chem.* **266:**14585–14592 (1991).
11. Gilligan, P. J., P. R. McGuirk, and M. J. Witty. Antibiotic derivatives of 7-phenyl-1,4-dihydro-4-oxo-quinoline carboxylic acids. United States Patent 4,623,650 (1986).
12. Spitzner, J. R., I. K. Chung, and M. T. Muller. Eukaryotic topoisomerase II preferentially cleaves alternating purine-pyrimidine repeats. *Nucleic Acids Res.* **18:**1–11 (1990).
13. Muller, M. T., J. R. Spitzner, J. A. DiDonato, V. B. Mehta, and K. Tsutsui. Single-strand DNA cleavages by eukaryotic topoisomerase II. *Biochemistry* **27:**8369–8379 (1988).
14. Chung, I. K., and M. T. Muller. Aggregates of oligo(dG) bind and inhibit topoisomerase II activity and induce formation of large networks. *J. Biol. Chem.* **266:**9508–9514 (1991).

14a. Pommier, Y., J. M. Covey, D. Kerrigan, J. Markovits, and R. Pham. DNA unwinding and inhibition of mouse leukemia L1210 DNA topoisomerase I by intercalators. *Nucleic Acids Res.* **15:**6713–6731 (1987).

15. Stormo, G. Computer methods for analyzing sequence recognition of nucleic acids. *Annu. Rev. Biophys. Biophys. Chem.* **17:**241–263 (1988).
16. DiDonato, J. D., J. R. Spitzner, and M. T. Muller. A predictive model for DNA recognition by the herpes simplex virus protein ICP4. *J. Mol. Biol.* **219:**451–470 (1991).
17. Stormo, G. D. Probing information content of DNA-binding sites. *Methods Enzymol.* **208:**458–468 (1991).
18. Spitzner, J. R., and M. T. Muller. Application of a degenerate consensus sequence to quantify recognition sites by vertebrate DNA topoisomerase II. *J. Mol. Recognit.* **2:**63–74 (1989).
19. Sander, M., and T. S. Hsieh. *Drosophila* topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. *Nucleic Acids Res.* **13:**1057–1072 (1985).
20. Pommier, Y., A. Orr, K. W. Kohn, and J.-F. Riou. Differential effects of amsacrine and epipodophyllotoxins on topoisomerase II cleavage in the human c-*myc* protooncogene. *Cancer Res.* **52:**3125–3130 (1992).
21. Pommier, Y., G. Capranico, A. Orr, and K. W. Kohn. Distribution of topoisomerase II cleavage sites in simian virus 40 DNA and the effects of drugs. *J. Mol. Biol.* **222:**909–924 (1991).
22. Capranico, G., S. Tinelli, F. Zunino, K. W. Kohn, and Y. Pommier. Effects of base mutations on topoisomerase II DNA cleavage stimulated by *m*AMSA in short DNA oligomers. *Biochemistry* **32:**145–152 (1993).
23. Shen, L. L., J. Baranowski, J. Fostel, D. A. Montgomery, and P. A. Lartey. DNA topoisomerases from pathogenic fungi: targets for the discovery of antifungal drugs. *Antimicrob. Agents Chemother.* **36:**2778–2784 (1992).
24. Chen, G. L., L. Yang, T. C. Rowe, B. D. Halligan, K. M. Tewey, and L. F. Liu. Nonintercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian DNA topoisomerase II. *J. Biol. Chem.* **259:**13560–13566 (1984).
25. Capranico, G., P. De Isabella, S. Tinelli, M. Bigioni, and F. Zunino. Similar sequence specificity of mitoxantrone and VM-26 stimulation of *in vitro* DNA cleavage by mammalian DNA topoisomerase II. *Biochemistry* **32:** 3038–3046 (1993).
26. Capranico, G., M. Palumbo, S. Tinelli, M. Mabilia, A. Pozzan, and F. Zunino. Conformational drug determinants of the sequence specificity of drug-stimulated topoisomerase II DNA cleavage. *J. Mol. Biol.* **235:**1218–1230 (1994).
27. Lee, M. P., M. Sander, and T. Hsieh. Nuclease protection by *Drosophila* DNA topoisomerase II: enzyme/DNA contacts at the strong topoisomerase II cleavage sites. *J. Biol. Chem.* **264:**21779–21787 (1989).

Send reprint requests to: Mark T. Muller, Ohio State University, Department of Molecular Genetics, 484 W. 12th Ave., Columbus, OH 43210.