Volume 13 Number 4 1985 Nucleic Acids Research

Drosophila topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site

Miriam Sander and Tao-shih Hsieh

Duke University Medical Center, Department of Biochemistry, Durham, NC 27710, USA

Received 27 December 1984; Accepted 29 January 1985

ABSTRACT

In order to study the sequence specificity of double-strand DNA cleavage by Drosophila topoisomerase II, we have mapped and sequenced 16 strong and 47 weak cleavage sites in the recombinant plasmid pπ25.1. Analysis of the nucleotide and dinucleotide frequencies in the region near the site of phosphodiester bond breakage revealed a nonrandom distribution. The nucleotide frequencies observed would occur by chance with a probability less than 0.05. The consensus sequence we derived is 5'GT.A/TAY↓ATT.AT..G 3', where a dot means no preferred nucleotide, Y is for pyrimidine, and the arrow shows the point of bond cleavage. On average, strong sites match the consensus better than weak sites.

INTRODUCTION

DNA topoisomerases catalyze the interconversion of topological isomers of DNA. These enzymes are ubiquitous in nature. In procaryotic organisms they are involved in most chromosome functions, including transcription, replication, daughter chromosome segregation, and recombination (for reviews see ref. 1-3). The best characterized topoisomerases are ω protein, and DNA gyrase from Escherichia coli, which are prototypes of the two classes of topoisomerase, type I and type II, respectively. In eucaryotes including yeast, Xenopus, Drosophila, rat and human, type I and type II topoisomerases are known; the nature of the involvement of these topoisomerases in eucaryotic chromosome function still remains to be established.

Topoisomerases catalyze knotting, unknotting, catenation, decatenation, and supercoiling and/or relaxation of DNA molecules. In these reactions, strand breakage and joining events are tightly coupled. For all topoisomerases, it has been possible to uncouple the breakage/joining events and to observe the DNA cleavage half-reaction. Topoisomerases cleave DNA and form a covalent DNA/protein complex as the cleavage is executed (4-9). This complex is thought to represent an

intermediate in the catalytic cycle of topoisomer interconversion.

The sites of topoisomerase mediated DNA cleavage are nonrandomly distributed in DNA fragments, suggesting that the reaction is sequence dependent. Whether the sequence dependence of this reaction reflects that of the catalytic reaction is not known; however, cleavage sites must be at least a subset of the sites with which the enzyme interacts. This view has support in that, for DNA gyrase, not all binding sites are cleavage sites (10).

Previous studies indicate that a low level of sequence homology exists between the cleavage sites of a particular topoisomerase (11-14). For ω protein, the nucleotide 4 bases 5' to the phosphodiester bond which is cleaved is a conserved cytidine residue (12). In other studies degeneracy is observed at all positions near the cleavage site, although nucleotide frequency distributions are nonrandom (13,14).

In this work the nucleotide sequence specificity of the Drosophila type II topoisomerase cleavage reaction is considered. 16 strong and 47 weak cleavage sites in the insert DNA of the recombinant plasmid pπ25.1 (15, 16) were mapped and sequenced. Analysis of nucleotide and dinucleotide frequencies among the cleavage sequences allows the derivation of an 11-base consensus sequence for the region surrounding the cleavage site. Most strong cleavage sites match the consensus in 7 or more positions; most weak sites match the consensus in 4, 5 or 6 positions. This suggests that the level of homology to the consensus sequence may determine the frequency of cleavage at a given site.

MATERIALS AND METHODS

Enzymes and DNA

T4 polynucleotide kinase and T4 DNA polymerase were purchased from Bethesda Research Laboratories. Bacterial alkaline phosphatase was from Worthington. Restriction endonuclease EcoRI was a generous gift of Dr. P. Modrich. Other restriction enzymes were purchased from Bethesda Research Laboratories or from New England Biolabs. Drosophila topoisomerase II was purified according to a previously published procedure (17). Plasmid DNA was purified by the procedure of Katz et al (18).

Labeling of DNA

5' ends of DNA fragments were labeled with T4 polynucleotide kinase and γ-^{32}P-ATP (19). 3' ends were labeled by T4 DNA polymerase and α-^{32}P-dTTP (20). Either restriction digested plasmid DNA, or gel purified DNA

fragments were substrates for the labeling reactions. After purification of labeled DNA by isopropanol precipitation, and subsequent restriction digestion to generate fragments with a unique end-label, the DNA fragments were separated by electrophoresis in agarose or polyacrylamide gels. Gels were autoradiographed to identify the fragments to be isolated. DNA from the appropriate gel slices were then electroeluted onto DEAE-cellulose paper (Bio-rad), and eluted from the paper with buffer containing 1 M NaCl (fragments <1 kb) or 3 M sodium acetate (fragments >1 kb). DNA was concentrated by ethanol precipitation.

α-^{32}P-dTTP and phosphate were purchased from New England Nuclear. γ-^{32}P-ATP was synthesized according to a modification of the procedure of Johnson and Walseth (21).

Gel electrophoresis

Agarose gel electrophoresis was carried out in TPE buffer (36mM Tris base, 30 mM NaH_2PO_4, 1 mM Na_3EDTA, pH 7.7) at a voltage gradient of 2 V/cm with recirculation of buffer. Nondenaturing acrylamide gels were prepared in TBE buffer (100 mM Tris-borate, 2 mM Na_3EDTA, pH 8.3) and were preelectrophoresed prior to use. Sequencing gels and markers were prepared according to the method of Maxam and Gilbert (22).

Topoisomerase cleavage reactions

Conditions for topoisomerase cleavage reactions were as follows: 1 µl *Drosophila* topoisomerase II (0.25 - 0.50 mg/ml) was added to a 20 µl reaction mixture containing 10 mM Tris-HCl pH 7.9, 50 mM KCl, 0.1 mM Na_3EDTA, 10 mM $MgCl_2$, 25-50 µg/ml bovine serum albumin (Pentex), 1.25 mM ATP, and 0.05 - 0.10 µg ^{32}P-labeled DNA substrate (10^4-10^5 cpm). After 3-5 min incubation at 30°C, the reaction was stopped with 1 µl of 10% SDS. Control reactions were carried out with no enzyme added, and by adding 1 µl 0.2 M Na_3EDTA to stop the reaction instead of SDS. The DNA was deproteinized by proteinase K (50 µg/ml, 45 min at 45°C). For nondenaturing acrylamide gels, samples were electrophoresed after the addition of marker dyes. For sequencing gel samples, DNA was precipitated by ethanol in the presence of 2 M ammonium acetate and 20 µg tRNA, and resuspended in sequencing gel loading buffer.

Other methods

DNA sequence analysis utilized software purchased from International Biotechnologies, Inc., and a program designed by us to calculate nucleotide and dinucleotide frequencies.

Densitometry was performed using a Zeineh soft laser scanning densitometer.

RESULTS

Mapping and sequencing topoisomerase cleavage sites

Topoisomerase cleavage sites were mapped on the insert DNA of the recombinant plasmid pπ25.1 (15, 16). This plasmid contains the Drosophila transposable element, P element, and flanking sequence from locus 17C, cloned into the BamH1 site of pBR322 (see map, Fig. 1C). Regions of the DNA containing strong topoisomerase cleavage sites were identified using the three uniquely end-labeled fragments, A-C, which are shown in Fig. 1A. Topoisomerase cleavage reactions were carried out on these fragments (see methods) and the products of the reaction were resolved by electrophoresis on 3.5% nondenaturing polyacrylamide gels. The autoradiogram of the dried gels is shown in Fig. 2. Densitometry on the

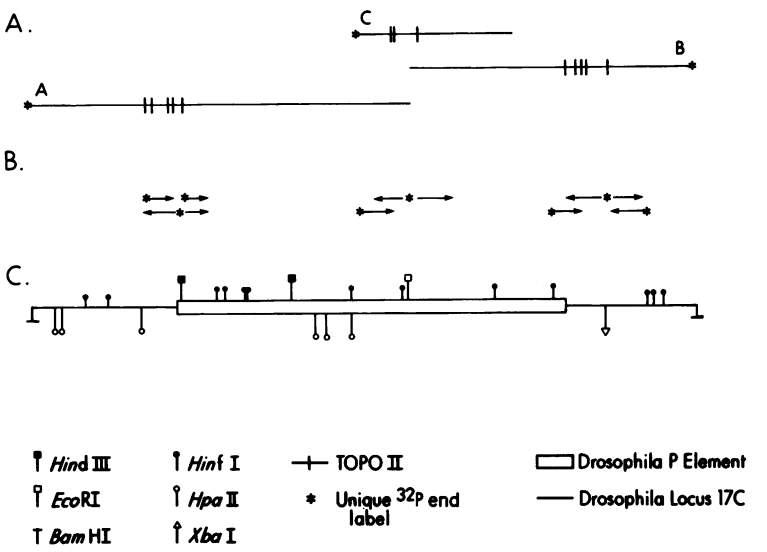

Fig. 1- Restriction map and strategy used for cleavage site mapping- A) Fragments A-C, with unique ^{32}P-end label, were purified from agarose gel, and used to locate strong topoisomerase cleavage sites with respect to the map of pπ25.1. B) Sites of ^{32}P-end label (asterisk) and the direction and extent of DNA sequencing for topoisomerase cleavage sequence determination. C) Restriction map of pπ25.1 insert DNA: 4.8 kb of Drosophila DNA is shown. This fragment is cloned into the BamH1 site of pBR322.

experimental lanes for fragments A and B (labeled + topo II), shows a 4-fold variation in band intensity. A similar experiment was performed using a 4 kb fragment of pBR322, which was 5'-^{32}P-labeled at the BamHI site, coordinate 375, and extended to the EcoRI site. Estimates of the cleavage efficiency on fragments A, B, and the pBR322 fragment were obtained by performing densitometry on autoradiographs and calculating the fraction of the starting material which was cleaved by topoisomerase. The Drosophila

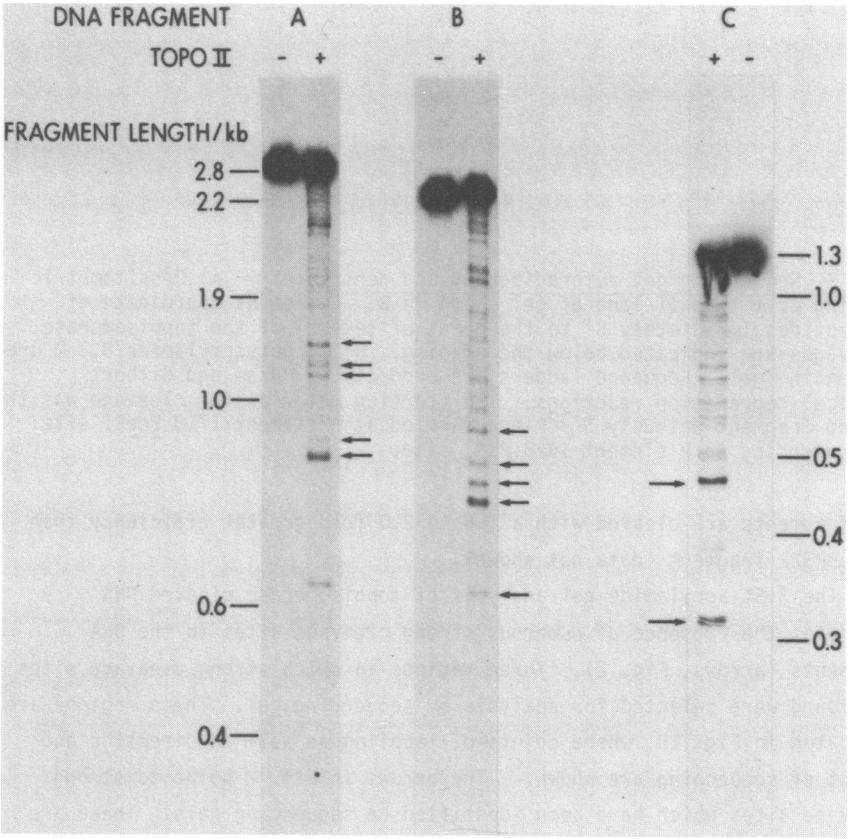

Fig. 2- Topoisomerase cleavage site mapping- DNA substrates for topoisomerase cleavage reactions are the fragments A, B and C of pπ25.1 insert DNA, as shown in Fig. 1A. Topoisomerase cleavage reaction was carried out as described (see methods). Products of the cleavage reaction for fragments A and B were resolved on one 3.5% nondenaturing polyacrylamide gel; fragment C cleavage products were resolved on a separate gel. The migration of size markers are shown. Arrows indicate strong topoisomerase cleavage sites for which the DNA sequence was determined.

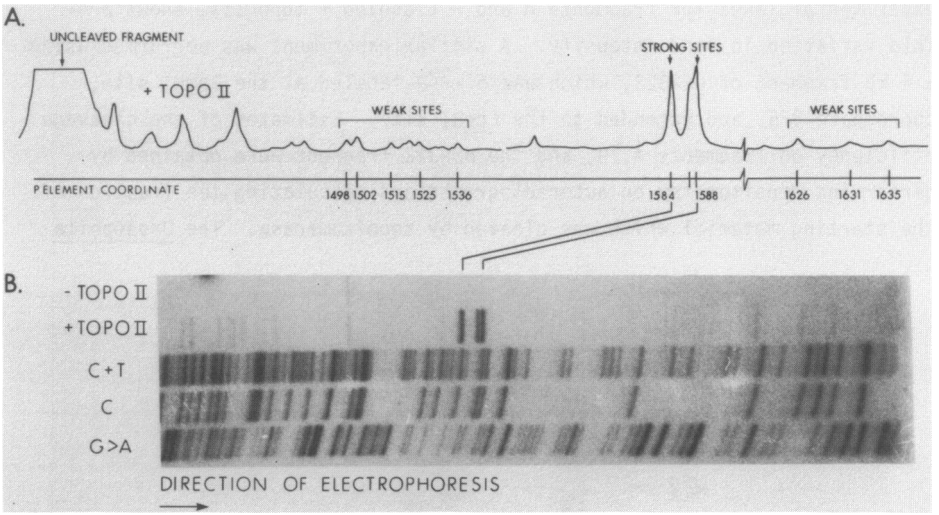

Fig. 3- Sequencing gel autoradiograph and densitometry- A) Densitometric tracing of + topo II lane of gel shown in B. P element coordinate of nucleotides immediately 5' to the point of several of the topoisomerase cleavages are indicated below the tracing. B) 8% polyacrylamide/8.3 M urea sequencing gel. Sequence ladders are products of Maxam and Gilbert chemical degradation reactions. The starting material for cleavage was the 468 bp fragment uniquely 5'^{32}P-end labeled at P element 1710 EcoRI site, and extending to P element 1428 PstI site.

DNA fragments are cleaved with a 1.5 to 2.0 fold greater efficiency than the pBR322 fragment (data not shown).

The 3.5% acrylamide gel analysis of topoisomerase cleaved DNA indicates the presence of numerous strong cleavage sites in the DNA fragments (arrows, Fig. 2). Three regions in which strong cleavage sites are found were selected for analysis by sequencing gel. These regions are indicated in Fig. 1B, where points of labeling as well as direction and extent of sequencing are shown. The arrows in Fig. 2 point to strong cleavage sites which have been identified on sequencing gels. These are also indicated, aligned with the restriction map of Fig. 1C, by the vertical slash marks in Fig. 1A.

The autoradiograph of a sequencing gel showing prominent topoisomerase cleavages, and densitometric tracing of the experimental lane are shown in Figs. 3A and 3B. Cleavage sites were designated strong or weak, based on band intensity on the sequencing gels. The designation of "strong site" was also made dependent on the detectability of a corresponding band in the 3.5% acrylamide gel system.

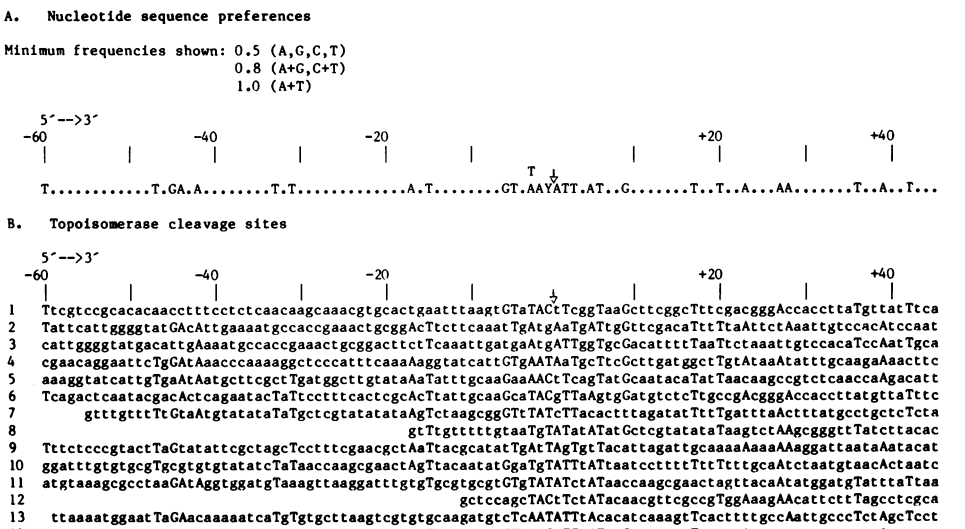

Fig.4- A) Nucleotide sequence preferences for set of topoisomerase cleavage sequences shown in B. Minimum frequencies are indicated above. Dot indicates no preferred nucleotide; Y is for pyrimidine; arrow indicates site of topoisomerase cleavage. B) Topoisomerase cleavage sequences aligned at point of phosphodiester bond cleavage. Position -1 and +1 are immediately 5' and 3' to the bond cleaved, respectively. For sequence 1 to 16, the -1 position has the following coordinate in P element or flanking DNA: 1, P46; 2, P1579; 3, P1583; 4, P1763; 5, P1805; 6, P2856; 7, F-197; 8, F-235; 9, F2982; 10, F3019; 11, F3055; 12, F+nd; 13, F+nd; 14, F+nd; 15, F-20; 16, F-55. (Coordinate given is for the noncoding strand of P element (P), or for the strand contiguous with it in flanking DNA (F). Flanking DNA is numbered increasingly negative in the 5' direction away from P element.) nd- not determined

Two methods were used to confirm that the cleavage sites identified were due to the cleavage activity of topoisomerase and not to that of any contaminating nuclease which might be present in the enzyme preparation. First, the dependence of the cleavage on the addition of a protein denaturant was verified (data not shown). The requirement for addition of denaturant has been reported in previous studies on DNA cleavage by Drosophila topoisomerase II (9), and is unlikely to be characteristic of any nuclease activity. Second, it was also confirmed for a portion of the sites mapped, that the cleavage occurred on complimentary strands with a 4 bp stagger (also a known characteristic of cleavage by Drosophila topoisomerase II).

Strong site analysis

Shown in Fig. 4B are the sequences of 16 strong topoisomerase cleavage sites which were identified by sequencing gel analysis. These are aligned at the point of phosphodiester bond cleavage, which is shown by the arrowhead marked in sequence 1. The nucleotides immediately 5' and 3' to the point of cleavage are numbered -1 and +1 respectively, and the span from positions -60 to +45 is shown (sequence information in the 5' region is lacking for several sites). For each sequence, the P element or flanking DNA coordinate of the -1 position nucleotide is given in the figure legend.

Each topoisomerase DNA cleavage event involves breakage of both DNA strands to generate 4 base long 5'-protruding ends. In the cleavage site sequences, no dyad symmetry axis is found that is coincident with the pseudosymmetry axis which relates the points of phophodiester bond breakage. Thus, the choice of which strand to use in aligning the sixteen strong cleavage site sequences will have an effect on the apparent level of homology between the sequences. The alignment shown in Fig. 4B was chosen by generating a preliminary consensus sequence and then maximizing the observed homology by choosing the strand which best fit the preliminary consensus sequence.

Inspection of the aligned sequences reveals that no unique sequence pattern is found in common between them. However, nucleotide preferences are evident near the site of phosphodiester bond cleavage. A tabulation of the preferred nucleotides is shown in Fig. 4A; included are all single nucleotides present at a frequency greater than 0.5, and all pyrimidine (Y) or purine (P) pairs present at a frequency greater than 0.8 (if no single nucleotide is above a frequency of 0.5 at that position). The single occurrence of A + T at a frequency of 1.0 is also shown. A dot marks positions where no nucleotide preference is observed. Sequences in Fig. 4B are printed in upper and lower case type; upper case indicates a match with the preferred nucleotide(s) for that position.

When a random alignment of 16 sequence segments from pπ25.1 insert DNA is made, it is found that nucleotide frequencies of 0.5 or greater occur at a frequency of 2 per 10 nucleotides. This is also approximately the occurrence of preferred nucleotides in the tabulation of Fig. 4A in the regions +10 to +45 and -10 to -60. However, in the 20 nucleotide block surrounding the topoisomerase cleavage site, 11 preferred nucleotides are tabulated; in this local region the frequency of nucleotides satisfying

A.

Nucleotide frequency x 100

Nucleotide	-10					-5					-1	+1				+5						+10		
(34)A	25	25	44	38	44	19	13	38	**56**	**63**	0	**56**	19	6	19	**50**	19	44	19	31	25	31	38
(20)G	25	25	0	31	13	**50**	25	31	0	13	13	6	13	0	38*	19	6	6	13	**56**	0	0	19
(17)C	6	31	25	6	6	6	13	19	0	6	**44***	6	0	25	13	13	6	6	44*	6	31	31	19
(29)T	44	19	31	25	38	25	**50**	13	**44**	19	**44**	31	**69**	**69**	31	19	**69**	44	25	6	44	38	25
5'-->3' Consensus	.	.	.	.	.	G	T	.	A/T		A	Y	A	T	T	.	A	T	.	.	G	.	.	.

B.

Dinucleotide frequency x 100

Dinucleotide	-10					-5					-1	+1				+5						+10	
(13)AA	13	19	25	6	0	13	13	31	0	0	13	0	0	6	0	0	13	6	6	6
(6)AG	0	6	0	19	13	0	0	0	0	0	13	0	0	6	6	0	0	6	0	0
(5)AC	0	0	0	0	6	0	0	6	**38**	0	0	0	0	6	6	6	**19**	6	0	13
(9)AT	13	19	13	19	0	0	25	19	25	0	**31**	19	6	0	**38**	13	13	0	25	6
(7)GA	0	0	6	0	6	13	**25**	0	0	13	0	0	0	13	0	0	0	0	19	0
(4)GG	0	0	6	6	6	0	0	0	0	0	0	0	0	13	0	0	0	0	19	0
(4)GC	**25**	0	0	0	6	13	0	0	0	0	0	0	0	6	0	0	0	0	**25**	0
(5)GT	0	0	**19**	6	**31**	0	6	0	13	0	6	13	0	6	**19**	6	6	0	13	0
(6)CA	**25**	13	0	6	0	6	13	0	0	6	0	0	6	0	6	6	0	**25**	0	13
(4)CG	0	13	0	0	0	0	0	0	0	6	0	0	6	0	0	0	0	**19**	0	0
(3)CC	0	0	6	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	6	0
(4)CT	6	0	0	0	6	6	6	0	0	**31**	6	0	13	13	6	0	6	0	0	**19**
(7)TA	6	6	13	6	6	6	6	**31**	0	**38**	6	6	13	**31**	13	**38**	6	0	0	13
(6)TG	0	13	6	**25**	6	**31**	0	13	13	0	0	0	**31**	0	0	6	13	19	0	0
(5)TC	0	6	0	6	0	6	0	0	0	6	0	**25**	13	0	0	0	25	0	0	**19**
(10)TT	13	6	6	0	13	6	6	0	6	0	25	**38**	13	0	6	25	0	6	6	13

Fig. 5- Nucleotide and dinucleotide frequencies for the 16 topoisomerase cleavage sequences shown in Fig. 4B, from position -11 to +12 (nucleotide), or -10 to +11 (dinucleotide). Calculated frequencies for all nucleotide and dinucleotides, using the known sequence information for P element and flanking DNA, are shown in parentheses at left of figure. A) Nucleotide frequencies: Topoisomerase consensus sequence is shown below the frequency table. Symbols are as in Fig. 4A. Boxes mark frequencies of the nucleotides in the consensus sequence. B) Dinucleotide frequencies: Boxes mark dinucleotide frequencies more than 3.5x greater than the calculated dinucleotide frequency.

the above-stated cutoff levels is 2.75-fold greater than observed in the random alignment of DNA sequences. It is probable that the increased density of preferred nucleotides, observed near the site of topoisomerase catalyzed DNA cleavage, reflects the existence of sequence determinants which are recognized by Drosophila topoisomerase II. Thus, we propose that the sequence 5' GT.A/TAY↓ATT.AT..G 3' can be considered a putative consensus sequence for Drosophila topoisomerase II cleavage sites.

Nucleotide and dinucleotide frequency tables of the 22 base region near the cleavage site are shown in Fig. 5. In these tables the sequence coordinate runs horizontally (from -11 to +12). Each row lists the

frequency of a nucleotide or dinucleotide in the set of 16 strong cleavage sites. Calculated frequencies of the nucleotides and dinucleotides for all of P element and all the known flanking sequence are shown in parentheses at the left of the figure. In Fig. 5A the boxes indicate the frequencies for the nucleotides present in the consensus sequence.

The following features of the nucleotide frequency distribution are of interest. At position -3, only A or T is allowed; G and C are excluded bases. Nucleotides are excluded at three other positions: -1 position, A excluded; +2 position, C excluded; and +3 position, G excluded. Exclusion of these nucleotides is not absolute, however, since they occur in weakly cleaved sites, although at very low frequencies (see below for further discussion). In the -1 position, the pyrimidine frequency is 0.88 with no bias between C and T. At position +3, the pyrimidine frequency is 0.94, with T strongly preferred, occurring at a frequency of 0.69. Indicated with asterisks are guanine and cytidine frequencies, which are below the 0.5 cutoff level for the consensus sequence, but which are significantly greater than the actual frequencies of G and C in the DNA fragment; 0.44 for C in the -1 position, 0.38 for G in the +4 position, and 0.44 for C in the +8 position.

Using the principles of statistics applicable to binomial populations, it is possible to estimate the significance of the nucleotide frequencies tabulated in Fig. 5. According to these principles, when the frequency of a nucleotide among a group of randomly aligned sequence segments is determined, a normal distribution of frequencies should be observed. The mean of the distribution is equal to the frequency of the nucleotide in the source DNA, and the standard deviation is equal to $((f)(1-f)/n)^{1/2}$, where f is the calculated frequency of the nucleotide in the source DNA, and n is the number of sequence segments (23). The frequencies of the four nucleotides A, G, C, and T in the source DNA for the topoisomerase cleavage sequences are known exactly, since they are calculated from the sequence data. Thus, based on the characteristics of a normal distribution, the probability for the occurrence of any frequency for any of the four nucleotides is known, simply by reference to the appropriate statistical table (23). The frequencies observed for the nucleotides of the topoisomerase consensus sequence (those boxed in Fig. 5A) all have a probability of less than 0.05 of occurring in a random alignment of sixteen sequences, with one exception. The exception is that the frequency of 0.5 for A (found in position +5), would occur in a random sequence alignment

with a probability of 0.09.

If the consensus sequence derived by nucleotide frequency analysis has validity, then dinucleotide frequencies in the region surrounding the cleavage site should show a nonrandom distribution, in which high frequencies for the dinucleotides of the consensus sequence are found in the appropriate positions of the frequency table. The dinucleotide frequency analysis of the 16 strong topoisomerase cleavage sites is shown in Fig. 5B. Marked with boxes are all dinucleotide frequencies which are at least 3.5 fold greater than the calculated dinucleotide frequencies for the P element and flanking DNA (parentheses at left of figure). The pattern of preferred dinucleotides confirms the pattern of preferred nucleotides. For example, AA and TA are the predominant dinucleotides spanning positions -3 to -2, and AC and AT are the predominant dinucleotides spanning positions -2 to -1. From this one deduces the preferred sequence 5' A/TAY 3'. This matches the consensus sequence deduced by nucleotide analysis. The dinucleotide frequency analysis was extended in the 3' direction from the cleavage site to nucleotide position +49. From position +10 to +49 the occurrence of dinucleotide frequencies satisfying the cutoff of 3.5x the calculated dinucleotide frequencies for the source DNA, was 1 per 60 dinucleotide frequencies tabulated. This contrasts with the occurrence of preferred dinucleotides in the 20 base region near the cleavage site, which is 1 per 12 dinucleotide frequencies, an increase of 4.8 fold.

Weak site analysis

As can be seen in the densitometer tracing of Fig. 3A, weak topoisomerase cleavage sites occur more frequently than strong topoisomerase cleavage sites. In the regions of pπ25.1 analyzed by sequencing gel, 47 weak sites were identified; an average frequency of 1 per 30 nucleotides. The sequences of the weak topoisomerase cleavage sites were analyzed in the following ways:

1) The 47 sites were scored for homology to the consensus sequence derived by nucleotide frequency analysis of the strong cleavage sites. The distribution observed is shown in Fig. 6. On average 50% homology is seen; the bulk of the sites match in either 4, 5, or 6 of the eleven positions of the consensus sequence. A similar analysis of strong topoisomerase cleavage sites was carried out. Fig. 6 shows that strong sites have, on average, a higher level of homology to the consensus sequence (60%) than weak sites. This suggests that the strength of the topoisomerase

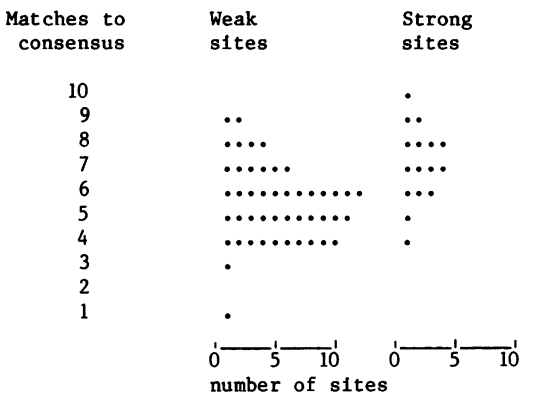

Fig. 6- 47 strong and 16 weak topoisomerase cleavage sites were scored for homology to the consensus sequence 5'GT.A/TAY↓ATT.AT..G 3'. Number of dots indicates number of sites for each level of homology.

cleavage site is related to its homology to the proposed consensus sequence.

2) The regions of p π25.1 insert DNA which have been sequenced (15) were searched by computer for sites with greater than 55% homology to the topoisomerase consensus sequence. The percentage of the sites identified by computer which are actual topoisomerase cleavage sites was then determined. (Only weak sites were scored, since strong sites were the database used to generate the consensus sequence.) 25% of the sites identified are cleaved as predicted. The remaining 75% are not cleaved by topoisomerase. Many of the predicted cleavage sites which are not cleaved occur in close proximity to strong topoisomerase cleavage sites. Steric factors may play a role in determining which members of a set of potential topoisomerase cleavage sites are in fact cleaved by the enzyme.

3) Weak sites and predicted cleavage sites were then analyzed for the presence of nucleotides which were excluded from strong cleavage sites, as noted above. These nucleotides, although present at low frequencies among the weak topoisomerase cleavage sites, are somewhat enriched in the weak sites with higher homology to the consensus sequence. Among the computer generated list of 84 putative cleavage sites, 26 also contained one of the nucleotides excluded from the strong cleavage sites. 16% (4 of 26) of these were actually cleaved as predicted, whereas 31% (18 of 58) of the predicted cleavages which did not contain the excluded bases were cleaved.

DISCUSSION

A consensus sequence for topoisomerase cleavage sites was derived based on the nucleotide frequencies in a group of sixteen strongly cleaved sites. This consensus sequence is 5'GT.A/TAY↓ATT.AT..G 3'. The validity of the consensus sequence proposed here is suggested by two facts: First, the consensus sequence nucleotides occur among the strong cleavage site sequences at frequencies that would occur by chance with probabilities <0.05 (with one exception, as noted above). Second, strongly cleaved sites have, on average, greater homology to the consensus sequence than weakly cleaved sites. It will be necessary to test the individual nucleotides of the consensus sequence with respect to their functional importance for topoisomerase recognition of cleavage sites. This has not yet been done, and for this reason, it must be emphasized that the stated consensus sequence should be viewed as tentative.

The level of homology of both strong and weak topoisomerase cleavage sites to the consensus sequence varies greatly (Fig. 6). Additionally, no conserved nucleotides occur. Yet the sequence dependence of the reaction is clear. No molecular basis for this type of DNA/protein interaction has been established. Proposals have, however, been made, which suggest that sequence dependent DNA helix structural parameters, instead of nucleotide sequence *per se*, can be recognized by DNA binding proteins (24-26). Preliminary analysis of the strong topoisomerase cleavage site sequences for such structural homology by the method of Nussinov (24) was carried out. By this method, the amount of structural homology observed in the 9 bp region near the cleavage site was similar to the amount of sequence homology in that region. Other methods, such as the analysis of cocrystals of DNA/protein complexes, may be necessary to further our understanding of the complexities of sequence dependent protein/nucleic acid interactions, whether they be of a highly specific nature, or less specific, as in the case of the topoisomerase cleavage reaction.

Previously, similarity in the structure of the cleavage complex generated by *E. coli* DNA gyrase and *Drosophila* topoisomerase II has been described (9). To further extend this comparison, published gyrase cleavage sites (10,11,27,28) were compared, and the homology near the cleavage site was determined. The preferred bases near the site of cleavage are 5' TAT↓G....T 3', similar to what has been reported previously (11,27). All bases match the *Drosophila* topoisomerase II consensus sequence, except the G at position +1. Although the sequence specificities

of the E. coli DNA gyrase and Drosophila topoisomerase II cleavage reactions are not identical, the similarity in the patterns of preferred nucleotides near the site of cleavage is additional evidence of the relatedness of the two activities.

Since Drosophila topoisomerase II is a homodimer, it is of interest to note the absence of a consistent dyad symmetry axis in the cleavage site sequences (several strong sites show some dyad symmetry; most show little or none). If the enzyme in its in vivo functional role can act as a gyrase and catalyze supercoiling in a directional manner, then at some level the DNA/enzyme interaction must be asymmetrical. Asymmetry in the E. coli DNA gyrase DNA/enzyme complex has been previously described (10,27,28).

The analysis of the Drosophila topoisomerase II cleavage site sequences that is presented here shows that DNA sequence near the site of cleavage is a determinant for the specificity of the cleavage reaction. Sequences distal to the site of phosphodiester bond cleavage might also be important for recognition, binding, and/or cleavage of the DNA by topoisomerase. However, sequence comparison of strong cleavage sites was extended from position -60 to +120, and no homologous regions were detected outside positions -10 and +10. A role for flanking sequence in determining sites of cleavage or rates of cleavage at specific sites is not ruled out by this data. If the spacing between sequence segments recognized by the enzyme can vary, then alignment at one point of homology will tend to cause other homologies to be obscured.

Recently, evidence has been accumulating which implicates type II mammalian topoisomerase as the target of several antitumor drugs, including epipodopyllotoxins, ellipticines, acridines, and anthrocyclines (29-35). In tissue culture cells and isolated nuclei these drugs induce double-strand breaks in DNA and covalent protein-DNA linkage (31-35). In vitro studies show enhancement of cleavage efficiency by topoisomerase II in the presence of the drugs (29,30). We have tested the sensitivity of Drosophila topoisomerase II to several antitumor drugs and found it to be less sensitive than the mammalian enzymes which have been studied. Some stimulation of the Drosophila topoisomerase II-mediated cleavage is however seen in the presence of the acridine derivative m-AMSA, the epipodophyllotoxin VM-26, and ellipticine. These drugs will be useful for further study of topoisomerase DNA cleavage in vivo, since they enhance cleavage efficiency. However, it will be necessary to rigorously determine that topoisomerase II uniquely mediates the double-strand DNA cleavage

events, in order to study the site specificity of the *in vivo* cleavage reaction.

ACKNOWLEDGEMENTS

We gratefully acknowledge K. O'Hare and G. Rubin for sharing the P element sequence data with us prior to publication. G. Vergara was of assistance to us with the photographic work, and Dr. C. S. Chang helped with the computer sequence analysis program. This work was supported by grant GM29006 from the National Institute of General Medical Sciences. TH is a recipient of an American Cancer Society Junior Faculty Research Award. MS is partly supported by an NIH predoctoral training grant.

REFERENCES

1. Gellert, M (1981) Annu. Rev. Biochem. 50, 879-910.
2. Wang, J.C. (1981) in The Enzymes, Boyer, P. Ed., Vol. XIV, pp 332-344, Academic, New York.
3. Gellert, M (1981) in The Enzymes, Boyer, P. Ed., Vol. XIV, pp 345-366.
4. Gellert, M., Mizuuchi, K., O'Dea, M.H., Itoh, T., and Tomizawa, J. (1977) Proc. Natl. Acad. Sci. USA 74, 4772-4776.
5. Champoux, J.J. (1972) Proc. Natl. Acad. Sci. USA 74, 3800-3807.
6. Depew, R.E., Liu, L.F. and Wang, J.C. (1978) J. Biol. Chem. 253, 511-518.
7. Tse, Y-C., Kierkegaard, K., Wang, J.C. (1980) J. BIol. Chem. 255, 5560-5565.
8. Liu, L.F., Rowe, T.C., Yang, L., Tewey, K.M., and Chen, G.L. (1983) J. Biol. Chem. 258, 15365-15370.
9. Sander, M. and Hsieh, T. (1983) J. Biol. Chem. 258, 8421-8428.
10. Kierkegaard, K., and Wang, J (1981) Cell 23, 721-729.
11. Morrison, A. and Cozzarelli, N.R. (1979) Cell 17, 175-184.
12. Dean, F, Krasnow, M.A., Otter, R., Matzuk, M.M., Spengler, S.J., and Cozzarelli, N.R. (1983) Cold Spring Harb. Symp. Quant. Biol. 47, 769-777.
13. Been, M.D., Burgess, R.R., and Champoux, J.J. (1984) Nucleic Acids Res. 12, 3097.
14. Edwards, K.A., Halligan, B.D., Davis, J.L., Nivera, N.C., and Liu, L.F. (1982) Nucleic Acids Res. 10, 2565-2576.
15. O'Hare, K. and Rubin, G.M. (1983) Cell 34, 25-35.
16. Spradling, A.C., and Rubin, G.M. (1982) Science 218, 341-347.
17. Hsieh, T. (1982) Methods Enzymol. 100, 161-170.
18 Katz, L., Kingsbury, D.T., and Helinski, D.R. (1973) J. Bacteriol. 114, 577-591.
19. Richardson, C.C. (1965) Proc. Natl. Acad. Sci. USA 54, 158-165.
20. Challberg, M.D., and Englund, P.T. (1980) Methods Enzymol. 65, 39-43.
21. Johnson, R.A. and Walseth, T.F. (1979) Adv. Cyclic Nucleotide Res. 10, 135-167.
22. Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 489-560.
23. Li, J.C.R. (1964) in Statistical Inference I, Statistics, Inc.
24. Nussinov, R. (1984) J. Biol. Chem. 259, 6798-6805.
25. Lomonosoff, G.P., Butler, P.J.G. and Klug, A (1981) J. Mol. Biol. 149, 745-760.
26. Drew, H.R. and Travers, A.A. (1984) Cell 37, 491-502.

27. Morrison, A. and Cozzarelli, N.R. (1981) Proc. Natl. Acad. Sci. USA 78, 1416-1420.
28. Fisher, M., Mizuuchi, K., O'Dea, M.H., Ohmori, H., and Gellert, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4165-4169.
29. Chen, G.L., Yang, L., Rowe, T.C., Halligan, B.D., Tewey, K.M., and Liu, L.F. (1984) J. Biol. Chem.259, 13560-13566.
30. Tewey, K.M., Chen, G.L., Nelson, E.M., and Liu, L.F. (1984) J. Biol. Chem. 259, 9182-9187.
31. Pommier, Y., Kerrigan, D., Schwartz, R., and Zwelling, L.A (1982) Biochem. Biophys. Res. Comm. 107, 576-583.
32. Zwelling, L.A., Michaels, S., Kerrigan, D., Pommier, Y. and Kohn, K.W. (1982) Biochem. Pharmacol. 31, 3261-3267.
33. Filipski, J. and Kohn, K.W. (1982) Biochim. Biophys. Acta 698, 280-286.
34. Ross, W.E. and Bradley, M.O. (1981) Biochim. Biophys. Acta 654, 129-134.
35. Ross, W.E., Glaubiger, D., and Kohn, K.W. (1979) Biochim. Biophys. Acta 562, 41-50.