Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA

Michael D.Been[1,3], Richard R.Burgess[2] and James J.Champoux[1]

[1]Department of Microbiology and Immunology, School of Medicine, University of Washington, Seattle, WA 98195, and [2]McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI 53709, USA

Received 16 January 1984; Revised and Accepted 6 March 1984

ABSTRACT

The site specificities of the type 1 DNA topoisomerases (topo 1) from rat liver and wheat germ were investigated. The nucleotide sequence at break sites on duplex SV40 DNA were determined for 245 wheat germ topo 1 sites and 223 rat liver topo 1 sites over a region of 1781 nucleotides. The enzymes from the two different sources show similar, but not identical patterns of DNA strand breakage. The sites occur frequently, but are not broken with equal probabilities. Major sites of breakage occur on the average every one to two turns of the helix, thus if sites of breakage accurately represent topo 1 sites of activity, the DNA sequence alone would appear to place few limits on the access of the enzyme to DNA. Sequences around the strongest sites for both enzymes show a bias in base composition for the four nucleotides immediately 5' to the break site (-4 to -1 positions), but no bias is observed 3' to the site of breakage. Consensus sequences for both enzymes were determined. Variations from the consensus sequence appear to affect the two enzymes differently and may account for the differences observed in the specificity of breakage.

INTRODUCTION

The topological state of DNA can be altered by a class of enzymes known as DNA topoisomerases which break and reseal one or both strands of a duplex molecule (1,2,3). Type 1 topoisomerases break and reseal a single strand of DNA in the absence of an added energy source and are able to relax a superhelical closed duplex DNA by providing the equivalent of a swivel in the DNA. The apparent requirement for a swivel in a number of processes involving DNA has led to postulated roles for a type 1 topoisomerase in DNA replication, recombination and transcription (1,2,3). The ability of eucaryotic type 1 topoisomerase (topo 1) to break single-stranded DNA (4,5) and catalyze the subsequent rejoining of the 3' end of the broken DNA to the 5' hydroxyl end of a DNA fragment different from the one generated by the breakage reaction (6, 7), lends support to a postulated role for a topoisomerase in strand exchange in some recombinational processes. The *int* gene product of phage lambda has topo 1 activity and catalyzes a reaction of this nature (8).

Breakage of duplex DNA by eucaryotic topo 1 can be detected if the reaction is terminated by the addition of a denaturing agent (9). The amount of breakage observed is independent of the length of time of incubation and it has been proposed that the observed nick results from a trapping of the intermediate in the nicking-closing reaction (9,2). Consistent with this suggestion is the finding that the broken strands contain covalently attached enzyme which could serve as the means of storing the energy required for resynthesis of the phosphodiester linkage in the closure step of the reaction (10). As mentioned above, the eucaryotic topo 1 will also break single-stranded DNA (4,5). The breakage of single-stranded DNA by topo 1 proceeds spontaneously and also results in the trapping of the topoisomerase in a covalent complex with the DNA (4,5). This covalent complex appears to represent the same structure as the hypothesized nicked intermediate in the overall reaction with duplex DNA (5,11) and is probably the result of an aborted reaction between topo 1 and single-stranded DNA.

In cases where topoisomerase induced breakage of DNA has been detected, the sites at which breakage occurs are non-random (1,8,12,13,14,15,16). If the sites of breakage reflect the loci at which topoisomerases act, then mapping the topo 1 breakage sites may provide some information on the nature and specificity of the interaction of the enzyme with DNA and, perhaps, on the role of topo 1 *in vivo*. The sequence specificity of topo 1 from HeLa cells on denatured DNA has been reported by Edwards *et al.* (16). Here we report results of studies on the nucleotide specificity of the breakage sites of topo 1 from wheat germ and rat liver using duplex DNA as the substrate for breakage.

MATERIALS AND METHODS

Enzymes.

Restriction endonucleases, Klenow fragment of *E. coli* DNA polymerase I and T4 polynucleotide kinase were purchased from New England Biolabs and Bethesda Research Laboratories. Bacterial alkaline phosphatase was purchased from Worthington and nuclease P1 was purchased from P-L Biochemicals. Rat liver topo 1 (100,000 mol. wt. form) was purified as described previously (17) and wheat germ topo 1 was purified as described by Dynan *et al.* (18).

DNA.

SV40 DNA was prepared as previously described (19). The SV40 strain used here was obtained from Peter Tegtmeyer and was a clone of strain VA45-54 (20). It shows some differences from the standard SV-S strain in restriction endonuclease cleavage patterns (21). A few interesting changes from the published

nucleotide sequence of the SV-S strain (22) were observed in the regions examined for topo 1 breakage and are shown in Figures 1 and 2.

Preparing end-labeled SV40 DNA fragments.

Combinations of restriction endonucleases that cut SV40 DNA at a single site were used to generate pairs of fragments that could easily be separated by sucrose gradient sedimentation. SV40 DNA (2.5 µg) was cut with either EcoRI, BamHI, HpaII, or TaqI under conditions recommended by the suppliers except that 0.2 M NaCl was used in the reactions for EcoRI and BamHI. Reactions were terminated by the addition of EDTA, extracted with phenol and chloroform and precipitated with ethanol. The 3' ends were labeled with ^{32}P in a 15 µl reaction containing 50 mM Tris-HCl (pH 7.4), 10 mM $MgSO_4$, 0.1 mM dithiothreitol, 50 µg/ml BSA, and deoxynucleoside triphosphates as follows: for EcoRI and BamHI ends, 50 µM dGTP, 50 µM TTP, and 1.5 µM [α-^{32}P]dATP; for HpaII and TaqI, 1.5 µM [α-^{32}P]dCTP. Klenow fragment ($\sim$2 units) was added and the reactions were incubated at 24° for 30 min and terminated by adding EDTA to 20 mM and heating to 65° for 15 min. The DNA was ethanol precipitated twice from 2 M ammonium acetate, rinsed with 80% ethanol, dried and resuspended in 10 mM Tris-HCl (pH 7.4), 1 mM EDTA (TE). Fragments labeled at the EcoRI site were recut with BamHI and vice versa. Fragments labeled at the HpaII or TaqI sites were similarly cut with the other enzyme. The products were layered onto 5-20% neutral sucrose gradients and sedimented for 5 hrs at 50,000 rpm in a Beckman SW56 rotor. Twenty-four fractions were collected from the bottom of each tube, and fractions containing the isolated fragments were pooled, ethanol precipitated twice, rinsed with 80% ethanol, dried and resuspended in TE.

Topo 1 breakage of end-labeled duplex fragments.

The amount of enzyme required in a reaction to get extensive, but not complete breakage of the substrate DNA was determined empirically; the ratios of enzyme to DNA were generally found to be higher than would have been predicted from the results on breakage of Form I SV40 DNA (5,20). Approximately 50 to 100 ng of end-labeled DNA was incubated for 5 min at 24° with approximately 75 ng of either RL or WG topo 1 in 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 0.5 mM DTT, 5% glycerol and 50 mM NaCl in a final volume of 50 µl. The reactions were terminated by addition of either 2 volumes of 1.5% SDS or 2 volumes of 0.225 M NaOH. The mixtures were then neutralized in the cases where the alkali stop was used, aliquoted to 5 tubes and each was ethanol precipitated twice, rinsed with 80% ethanol twice and dried. Pellets were resuspended in 80% formamide, 1 mM EDTA, 0.2% xylene cyanol, 0.2% bromphenol blue and heated to 90° for 2 min before loading onto a sequencing gel.

DNA sequencing reactions and gels.

Base specific chemical cleavage reactions were done according to Maxam and Gilbert (23). DNA sequencing gels contained 89 mM Tris-borate, 2.5 mM EDTA (pH 8.3), 50% weight/volume urea and either 6%, 8%, or 12% polyacrylamide (polyacrylamide:bis-acrylamide, 39:1). Gels (40 cm long and 0.3 mm thick) were run for various times at from 1600 to 1800 volts. The 6% gels were transfered to Whatman 3 MM paper covered with plastic wrap and dried. The 8% and 12% gels were transferred to used X-ray film and covered with plastic wrap. Films were exposed to the gels at -70°.

5' end group analysis.

The analysis of the distribution of 5' nucleotides at breakage sites was done as previously described (14) and is detailed in the legend to Table IV. Thin layer chromatography on PEI plates was done in one dimension using 0.5 M lithium formate (pH 3) as the mobile phase.

DNA sequence analysis.

DNA sequence searches were performed with the aid of a DNA sequence analysis program on an IBM Personal Computer.

RESULTS

Mapping breakage sites.

Upon strand breakage, topo 1 from both rat liver and wheat germ attaches covalently to the 3' phosphate and generates a free 5' hydroxyl group at the site of the break (10,11 and unpublished results). Breakage of a linear duplex fragment, that has been 3' end-labeled with ^{32}P, will therefore generate a labeled fragment with a 5' hydroxyl and an unlabeled fragment with covalently bound enzyme. SV40 DNA restriction endonuclease fragments, 3' end-labeled in one strand, were used as substrates for topo 1 breakage. The restriction fragments were incubated with topo 1 from either rat liver (RL) or wheat germ (WG), the reactions terminated with SDS and the products were analyzed by electrophoresis in a denaturing polyacrylamide gel. The amount of breakage was limited such that only a fraction of the labeled substrate was broken and a series of fragments with common 3' ends were generated. The products from a topo 1 reaction were analyzed in multiple gels: 8% and 12% polyacrylamide to visualize break sites from approximately 20 bases to 80-90 bases from the labeled end, and two or more 6% polyacrylamide gels run for various times to locate sites from about 70 bases to 300-350 bases from the labeled position. The results of two analyses using 6% polyacrylamide gels are shown in Figure 1. Lanes marked RL and WG contained the products of breakage by RL topo 1 and

WG topo 1, respectively. With both enzymes, breakage of duplex DNA occurred at a large number of sites and although the patterns are not identical, many break sites were common to both enzymes.

Electrophoresis of topo 1 breakage products adjacent to lanes containing the same fragment cleaved by the base specific sequencing reactions of Maxam and Gilbert (23), allows the exact nucleotide position of the breakage to be determined. For the shorter fragments, it was necessary to correct for the fact that topo 1 breakage generates labeled fragments with 5' hydroxyl groups whereas chemical cleavage generates labeled fragments with 5' phosphate groups. The charge difference results in a higher mobility for the chemically cleaved fragment (24). This altered mobility is obvious in sequencing gels for fragments less than approximately 100 bases long where the smaller mass makes the charge difference more significant. As described by Tapper and Clayton (24) this charge effect can change the mobility by the equivalent of one base for short fragments in an 8% polyacrylamide sequencing gel. We observed that the relative magnitude of the shift in mobility for short fragments also varied with the percent polyacrylamide in the gel, increasing as the percent polyacrylamide is decreased. The RL topo 1 generated fragment located at position 2606 (left panel Fig. 1, see also Fig. 2E) is 70 bases long and migrated between the A band at 2607 and the T band at 2606, but was the result of breakage between T at 2606 and A at 2605. The enzyme remains covalently attached to the T at 2606. It was demonstrated by Edwards et al. (16) that this altered mobility of topo 1 generated fragments is lost when the 5' hydroxyl group is phosphorylated.

The sequences from six fragments of SV40 DNA used in this study are shown in Figure 2. The breakage sites for RL topo 1 and WG topo 1 were determined over a total region of 1781 nucleotides and are indicated in Figure 2 by marking (above the sequence for RL topo 1 and below the sequence for WG topo 1) the nucleotide to which the topoisomerase is covalently attached. The site of breakage is immediately 3' to the indicated nucleotide. The relative intensities of the bands varied greatly (Fig. 1), indicating that individual sites were broken with differing frequencies. Individual break sites were qualitatively judged as strong, intermediate or weak sites and are indicated as such in Figure 2. From densitometer tracings we estimated that the strong sites, which represent less than 25% of the sites being broken, accounted for a minimum of 80% of the breakage. The weak sites, representing 50% or more of the breakage sites, accounted for less than 5% of the breakage.

The conditions of the breakage reaction were varied to determine whether

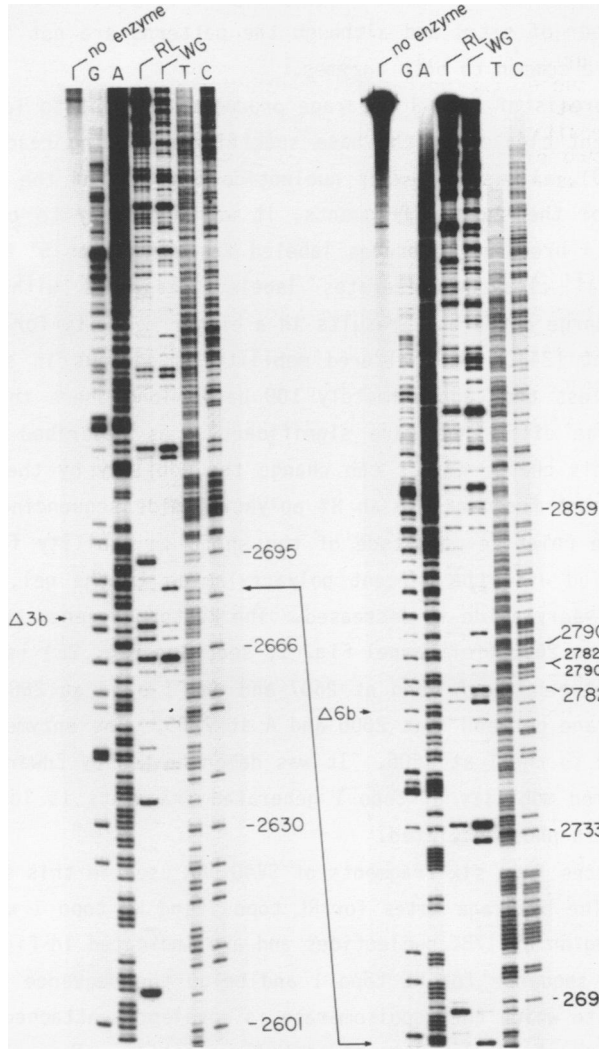


Figure 1. Breakage of end-labeled duplex DNA by RL and WG topo 1. The large EcoRI/BamHI restriction endonuclease fragment of SV40 DNA, 3' end-labeled at the BamHI site, was used as the substrate DNA. Topo 1 (~75 ng) was incubated with DNA (50-100 ng) in a 50 µl reaction at 24° for 5 min and the reaction terminated by adding 100 µl of 1.5% SDS. The reaction was divided into five aliquots and the samples were ethanol precipitated. The same substrate DNA was used in the DNA sequencing reactions of Maxam and Gilbert (23). The first lane in each panel contains the products of a reaction in which enzyme was omitted. Lanes marked G, A, T, and C contained the products of the sequencing reactions for G, G+A, T+C and C, respectively. (There was some reaction with G in the T and C lanes in this experiment.) The lanes marked RL and WG contained the products of breakage by the respective topoisomerases. The results

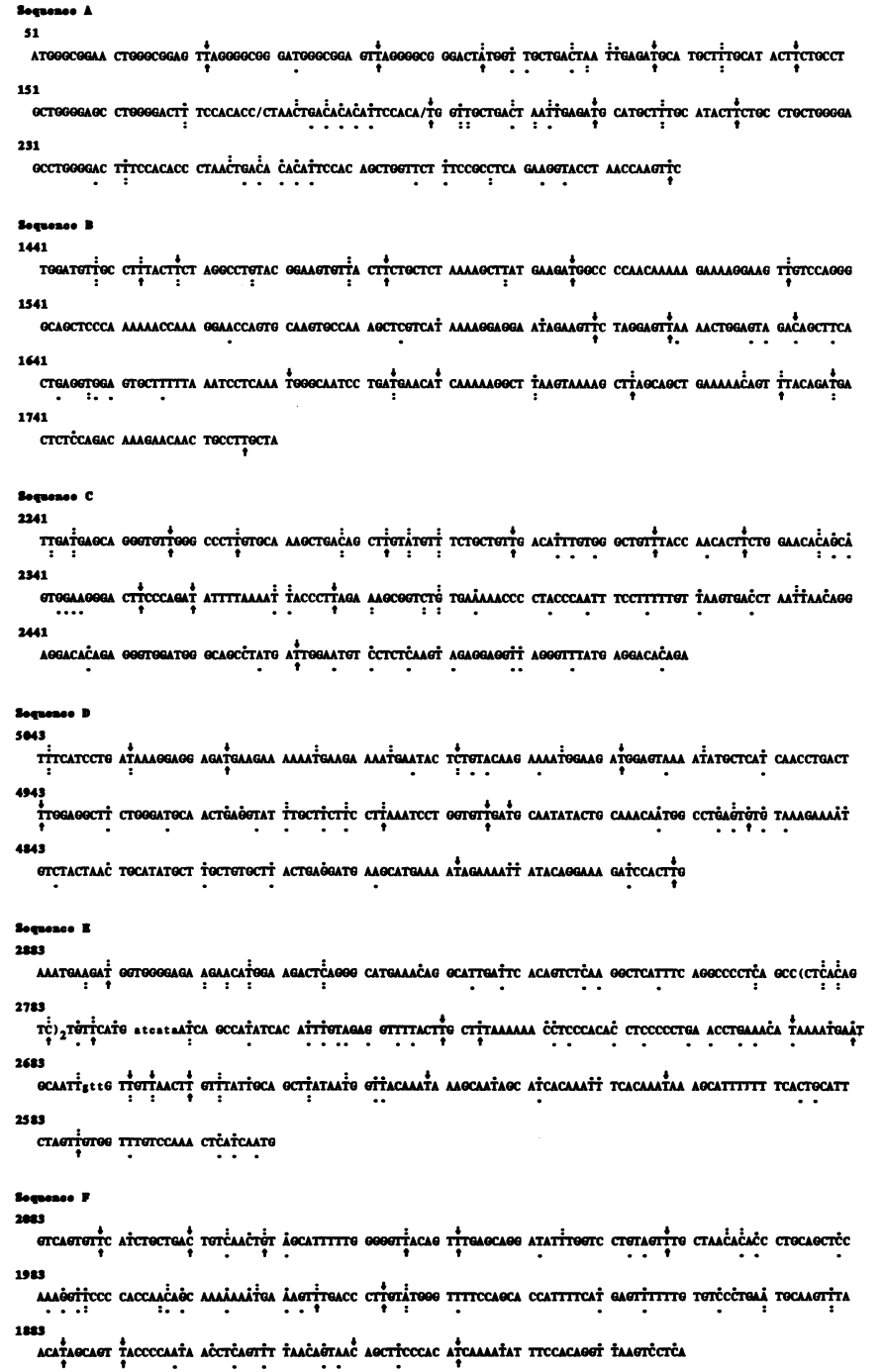
shown are from electrophoresis of the products in 6% polyacrylamide- urea sequencing gels, the panel on the left was run for 2.5 hours the one on the right for 5.5 hours; corresponding positions in the sequence are connected by the arrow. The nucleotide positions of specific fragments in the C lanes are given on the right of each panel for cross-referencing to Figure 2E. Also shown are the positions at which variations occur in the sequence from the published sequence of SV40 (22). The equivalent of a 3 base and a 6 base deletion as well as a 9 base duplication (positions 2782-2790) were found (see also Fig. 2E).

the specificity could be altered (data not shown). Decreasing the enzyme concentration resulted in less overall breakage but no apparent alteration in the pattern of fragments generated. Increasing the salt concentration also resulted in a decreased amount of breakage at all sites, without a preferential loss of the weak sites in the sequences examined. With salt concentrations above 0.25 M NaCl or KCl, breakage could not be detected. Increasing the glycerol concentration, or adding Mg^{++} or Mn^{++} also had no obvious effect on the relative intensities of fragments generated and terminating the reaction with alkali instead of SDS generated the same pattern of fragments.

Distribution of bases around sites of breakage.

The sequence data (Fig. 2) indicated that topo 1 does not break the DNA at a unique sequence. Also, the frequency of breakage suggested that specificity is likely to be determined by only a few nucleotides. Because the strong and intermediate sites accounted for more than 95% of observed breakage, we assumed that they represented preferred sites for topo 1 breakage. The distribution of nucleotides around the strong and intermediate sites were compiled for the 20 bases spanning the sites of breakage (Tables I and II). The numbering convention of Edwards et al. (16) was used so that the site of breakage was located between nucleotides at positions -1 and 1. The enzyme attaches to the nucleotide at the -1 position. Results for RL topo 1 strong sites (Table IA) indicated a bias at positions -4 to -1 (5' to the site of breakage). The preferred bases at those positions were A or T at -4, G or C at -3, T or A at -2, and T at -1. The WG topo 1 strong breakage sites (Table IIA) showed a similar preference although the position at -4 appeared to be less significant.

The distribution of bases around intermediate sites for both enzymes indicated a similar base preference with the following differences. For RL topo 1 (Table IB) the position at -1 accepted C in addition to T, indicating a possible bias at this position for pyrimidines, and the position at -3 allowed A as well as G or C. For WG topo 1 the position at -4 showed only a slight preference for A and T (Table IIB), and at the -1 position, while a bias for T


Sequence A
51
ATGGGCGGAA CTGGGCGGAG TTAGGGGCGG GATGGGCGGA GTTAGGGGCG GGACTATGGT TGCTGACTAA TTGAGATGCA TGCTTTGCAT ACTTCTGCCT
151
GCTGGGGAGC CTGGGGACTT TCCACACC/CTAACTGACACACATTCCACA/TG GTTGCTGACT AATTGAGATG CATGCTTTGC ATACTTCTGC CTGCTGGGGA
231
GCCTGGGGAC TTTCCACACC CTAACTGACA CACATTCCAC AGCTGGTTCT TTCCGCCTCA GAAGGTACCT AACCAAGTTC
Sequence B
1441
TGGATGTTGC CTTTACTTCT AGGCCTGTAC GGAAGTGTTA CTTCTGCTCT AAAAGCTTAT GAAGATGGCC CCAACAAAAA GAAAAGGAAG TTGTCCAGGG
1541
GCAGCTCCCA AAAAACCAAA GGAACCAGTG CAAGTGCCAA AGCTCGTCAT AAAAGGAGGA ATAGAAGTTC TAGGAGTTAA AACTGGAGTA GACAGCTTCA
1641
CTGAGGTGGA GTGCTTTTTA AATCCTCAAA TGGGCAATCC TGATGAACAT CAAAAAGGCT TAAGTAAAAG CTTAGCAGCT GAAAAACAGT TTACAGATGA
1741
CTCTCCAGAC AAAGAACAAC TGCCTTGCTA
Sequence C
2241
TTGATGAGCA GGGTGTTGGG CCCTTGTGCA AAGCTGACAG CTTGTATGTT TCTGCTGTTG ACATTTGTGG GCTGTTTACC AACACTTCTG GAACACAGCA
2341
GTGGAAGGGA CTTCCCAGAT ATTTTAAAAT TACCCTTAGA AAGCGGTCTG TGAAAAACCC CTACCCAATT TCCTTTTTGT TAAGTGACCT AATTAACAGG
2441
AGGACACAGA GGGTGGATGG GCAGCCTATG ATTGGAATGT CCTCTCAAGT AGAGGAGGTT AGGGTTTATG AGGACACAGA
Sequence D
5043
TTTCATCCTG ATAAAGGAGG AGATGAAGAA AAAATGAAGA AAATGAATAC TCTGTACAAG AAAATGGAAG ATGGAGTAAA ATATGCTCAT CAACCTGACT
4943
TTGGAGGCTT CTGGGATGCA ACTGAGGTAT TTGCTTCTTC CTTAAATCCT GGTGTTGATG CAATATACTG CAAACAATGG CCTGAGTGTG TAAAGAAAAT
4843
GTCTACTAAC TGCATATGCT TGCTGTGCTT ACTGAGGATG AAGCATGAAA ATAGAAAATT ATACAGGAAA GATCCACTTG
Sequence E
2883
AAATGAAGAT GGTGGGGAGA AGAACATGGA AGACTCAGGG CATGAAACAG GCATTGATTC ACAGTCTCAA GGCTCATTTC AGGCCCCTCA GCC(CTCACAG
2783
TC)2TGTTCATG atcataATCA GCCATATCAC ATTTGTAGAG GTTTTACTTG CTTTAAAAAA CCTCCCACAC CTCCCCCTGA ACCTGAAACA TAAAATGAAT
2683
GCAATTgttG TTGTTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA AAGCAATAGC ATCACAAATT TCACAAATAA AGCATTTTTT TCACTGCATT
2583
CTAGTTGTGG TTTGTCCAAA CTCATCAATG
Sequence F
2883
GTCAGTGTTC ATCTGCTGAC TGTCAACTGT AGCATTTTTG GGGGTTACAG TTTGAGCAGG ATATTTGGTC CTGTAGTTTG CTAACACACC CTGCAGCTCC
1983
AAAGGTTCCC CACCAACAGC AAAAAAATGA AAGTTTGACC CTTGTATGGG TTTTCCAGCA CCATTTTCAT GAGTTTTTTG TGTCCCTGAA TGCAAGTTTA
1883
ACATAGCAGT TACCCCAATA ACCTCAGTTT TAACAGTAAC AGCTTCCCAC ATCAAAATAT TTCCACAGGT TAAGTCCTCA

still existed, both A and C occurred frequently (20% and 22%). Neither enzyme showed a bias in the 3' direction from the break site that is consistent with both strong and intermediate sites.

The distribution of bases around weak sites from positions -5 to +1 were also tabulated. There appeared to be an additional relaxation of nucleotide preference in the -4 to -1 positions compared with that seen for the intermediate sites. With RL topo 1 there was an apparent exclusion of C at position -4 and T at -3, a bias for A or T at -2, and there was a slight increase in the frequency of purines at the -1 position (Table IC). For WG topo 1 weak sites, only the G and C bias at -3 and the A and T bias at -2 appeared to strongly persist; the preference for T at -1 over the other three bases was small (Table IIC).

Below each of the set of sites tabulated in Tables I and II the most frequently occurring bases in the -4 to -1 positions are given. Nucleotides occurring greater than 10% of the time are listed, in the order of decreasing frequency from top to bottom.

Frequency of detectable breakage at specific sequences.

From the data on the distribution of bases around strong break sites, the preferred or consensus sequences for topo 1 breakage may be taken to be the strong site sequence shown below Tables IA and IIA. The intermediate and weak sites, while similar, did show an increase in the number of variations from this consensus sequence at nearly every position. Because of ambiguities in three of the positions of the consensus sequence there are eight possible per-

Figure 2. Locations of topo 1 breakage sites on duplex SV40 DNA restriction fragments. The sequences (shown 5' to 3') over which topo 1 breakage sites have been located on 6 restriction fragments of SV40 DNA are shown. The position of breakage is identified by marking the nucleotide to which topo 1 binds when it breaks the DNA at that site. Breakage is immediately 3' to the indicated nucleotide. Arrows signify strong sites; colons, the intermediate sites; and dots, the weak sites. Sites for RL topo 1 breakage are above the sequence and sites for the WG enzyme are below. For sequences A, B and C the labeled strand was the one in the sense of the late RNA and for D, E, and F the other strand was labeled. The restriciton fragment used for each sequence was: A, the short TaqI/HpaII(346) fragment; B, the long BamHI/EcoRI(1782) fragment; C, the short EcoRI/BamHI(2533) fragment; D, the short HpaII/TaqI(4742) fragment; E, the long EcoRI/BamHI(2538) fragment; and F, the short BamHI/EcoRI(1787) fragment. The position of the restriction site that was labeled is given. Note that two of the fragments differ in sequence from the published sequence of SV40 (22). These differences have been included but the numbering was maintained according to the published sequence. Sequence A contains an extra 21 bases between 178 and 179, this is a repeat of the sequence from positions 251 to 271 and overall results in a 93 base-pair repeat. Variations in sequence E are described in the legend to Figure 1; the deleted bases are shown in lower case and the repeated sequence is indicated by parenthesis.

Table I. Distribution of bases around RL topo 1 break sites (223 total). For strong and intermediate RL topo 1 break sites, the percent occurrence of each nucleotide for ten bases in either direction of the break site was calculated. For the weak sites percentages were calculated only for positions -5 to +1. The nucleotides which occur greater than 10% of the time in the -4 to -1 positions are shown below the tabulation in each case. Those which occur less than 20% of the time in those positions are in parenthesis.

A. Strong break sites (50 sites)

position	-10					-5			-1		+1				5					10
%G	24	30	21	14	40	24	-	54	-	-	36	36	24	24	26	14	20	26	16	22
%C	16	18	20	24	12	36	4	36	-	8	20	12	18	24	24	22	28	6	22	10
%A	28	24	22	16	24	18	64	10	46	-	30	30	38	28	28	42	22	30	34	30
%T	32	28	38	46	24	22	32	-	54	92	14	22	20	24	22	22	30	38	28	38

5'- A G T T - 3'
T C A

B. Intermediate break sites (62 sites)

position	-10					-5			-1		+1				5					10
%G	28	16	26	29	34	13	8	39	-	5	21	44	15	24	13	16	16	35	19	13
%C	26	31	11	19	23	37	10	31	-	32	6	21	26	27	21	26	15	18	32	21
%A	31	21	29	31	31	34	37	31	45	5	32	24	37	16	29	23	27	23	29	32
%T	16	32	34	21	13	16	45	-	55	58	40	11	23	32	37	35	42	24	20	34

T G T T
5' - A C A C - 3'
A

C. Weak break sites (111 sites)

position	-5				-1	+1
%G	25	13	32	6	9	12
%C	24	7	41	5	29	23
%A	33	47	26	49	10	33
%T	17	33	2	41	52	32

A C A T
5'- T G T C - 3'
(G) A

mutations of the sequence. Not all of the permutations will necessarily be sequences at which topo 1 breaks the DNA, and topo 1 may not break the DNA at every site where such a sequence occurs. In addition, DNA sequences over which the topo 1 breakage sites were mapped are not random; the four nucleotides do not occur with equal frequency (22.6% G, 20.7% C, 29.7% A, 27.1% T in sequences analyzed) and all possible combinations of the four base sequences would not be expected to occur with equal frequencies. The four base consen-

Table II. Distribution of bases around WG topo 1 break sites (245 total). (see legend to table I)

A. Strong break sites (55 sites)

position	-10				-5				-1		+1				5					10
%G	16	22	20	22	36	29	7	58	-	4	35	36	27	22	27	16	24	33	27	16
%C	22	15	16	27	15	31	11	42	-	5	20	15	22	25	24	22	16	9	18	16
%A	35	29	27	16	24	25	65	-	24	4	18	24	27	27	29	36	18	29	31	25
%T	27	35	36	35	25	15	16	-	76	87	27	25	24	25	20	25	42	29	24	42

```
                        A  G  T  T
                5'-(T) C  A     - 3'
                   (C)
```

B. Intermediate break sites (50 sites)

position	-10				-5				-1		1				5					10
%G	16	18	30	22	30	26	16	52	4	4	22	36	14	34	22	24	16	28	20	16
%C	26	26	18	26	34	20	12	48	-	22	10	22	30	20	22	18	32	12	28	24
%A	26	26	26	36	24	26	32	-	32	20	36	30	26	26	32	32	26	24	28	32
%T	32	30	26	16	12	28	40	-	64	54	32	12	30	20	24	26	26	36	24	28

```
                        T  G  T  T
                5'- A  C  A  C
                   (G)         A  - 3'
                   (C)
```

C. Weak break sites (140)

position	-5				-1	1
%G	24	21	45	9	16	17
%C	30	11	44	1	21	19
%A	29	34	7	44	25	33
%T	17	34	4	46	38	31

```
                        T  G  T  T
                5'- A  C  A  A
                    G           C  - 3'
                   (C)         (G)
```

sus sequence may be influenced by a distribution of bases near the break sites which has nothing to do with recognition. For example, the sequence 5'-CG-3' is uncommon in SV40 DNA and, if there is a preference for G at the -3 position of the topo 1 break sites, the frequency of C at -4 is necessarily going to be low.

To determine which sequences are frequently sites of breakage and which positions within the four base consensus sequence may be the most significant, the locations of specific 3 and 4 base sequences were determined for the frag-

Table III. Frequency with which breakage at specific sequences was detected. Column N is the number of times sites with the specified sequence occur in the DNA fragments examined; column RL is the number of those sites for which RL topo 1 breakage is detected; and column WG is the number of those sites for which WG topo 1 breakage is detected.
IIIA, variations of the sequences 5'- X T/A T - 3' and 5'- X X T/A T - 3'.
IIIB, variations of the sequences 5'- X T/A C - 3' and 5'- X X T/A C - 3'.

Sequence				Frequency		
-4	-3	-2	-1	N	RL	WG
	C	T	T	35	30	35
G	C	T	T	15	12	15
C	C	T	T	7	5	7
A	C	T	T	11	11	11
T	C	T	T	2	2	2
	G	T	T	39	34	39
G	G	T	T	10	5	10
C	G	T	T	0	-	-
A	G	T	T	16	16	16
T	G	T	T	13	13	13
	C	A	T	32	20	18
G	C	A	T	12	3	4
C	C	A	T	2	1	1
A	C	A	T	10	10	9
T	C	A	T	8	6	4
	G	A	T	23	17	20
G	G	A	T	6	0	3
C	G	A	T	0	-	-
A	G	A	T	9	9	9
T	G	A	T	8	8	8
	A	A	T	35	24	1
G	A	A	T	5	1	0
C	A	A	T	7	2	0
A	A	A	T	19	17	1
T	A	A	T	4	4	0
	A	T	T	25	10	2
G	A	T	T	2	0	0
C	A	T	T	10	0	0
A	A	T	T	8	7	1
T	A	T	T	5	3	1
	T	T	T	56	2	3
G	T	T	T	14	0	0
C	T	T	T	11	0	0
A	T	T	T	13	1	2
T	T	T	T	18	1	1
	T	A	T	16	0	0
G	T	A	T	3	0	0
C	T	A	T	2	0	0
A	T	A	T	5	0	0
T	T	A	T	6	0	0

A

Sequence				Frequency		
-4	-3	-2	-1	N	RL	WG
	C	T	C	24	12	13
G	C	T	C	6	1	2
C	C	T	C	11	5	6
A	C	T	C	4	3	3
T	C	T	C	3	3	2
	G	T	C	13	7	6
G	G	T	C	2	0	1
C	G	T	C	1	0	0
A	G	T	C	4	3	2
T	G	T	C	6	4	3
	C	A	C	31	12	13
G	C	A	C	1	0	0
C	C	A	C	11	2	2
A	C	A	C	13	10	10
T	C	A	C	6	0	1
	G	A	C	20	9	9
G	G	A	C	6	0	0
C	G	A	C	0	-	-
A	G	A	C	3	1	1
T	G	A	C	11	8	8
	A	A	C	31	14	2
G	A	A	C	7	0	0
C	A	A	C	7	2	1
A	A	A	C	9	4	0
T	A	A	C	11	9	0
	A	T	C	13	0	0
	T	T	C	36	0	0
	T	A	C	17	0	0

B

ments studied. Each of these sites was then examined for detectable breakage. For the data in Table IIIA, sequences were located in which T was assigned to the -1 position, either A or T was allowed at -2, and each of the four nucleotides was tried at positions -3 and -4. The number of times breakage was detected at each of these sites was then determined.

Two sequences stand out in this analysis; 5'-X C T T-3' and 5'-X G T T-3' where X is any nucleotide at the -4 position, together occur 74 times. WG topo 1 breaks at these sites 100% of the time and the significance of bases at the -4 position in the sequence is obscured by the apparent strong preference for the indicated bases in the -3 to -1 positions. However, breakage at 10 of these sites by the RL topo 1 was not detected, and in each of these instances the nucleotide at the -4 position was either G or C. On the other hand, in 22 cases breakage was detectable with G or C at the -4 position so that with this sequence neither G nor C is absolutely excluded at position -4. A similar comparison of sequences with A at position -2 indicates that both enzymes may break at the sequence 5'-X G/C A T -3' more frequently if an A or a T is at position -4. For similar sequences with an A or T at position -3, breakage was detected infrequently. However, for RL topo 1 an A at position -3 appeared to be more acceptable provided an A or T was in the -4 position.

A cytosine occurred frequently at the -1 position in the intermediate and weak sites (Tables I and II). When a C was substituted for T at position -1 for the sequences just examined (Table IIIB) breakage was detected at some of these sites, although less frequently than when a T is at the -1 position. There remained an apparent preference for G or C at -3 and A or T at -4.

Bias at the +1 position.

As noted earlier, the data in Tables I and II indicate that there is no consistent bias for the nucleotide at the +1 position. This was confirmed in an independent type of analysis. Form I SV40 DNA was nicked with RL topo 1 by terminating the reaction with alkali. The reaction was neutralized and the hydroxyl groups at the 5' ends generated by topo 1 breakage were phosphorylated with ^{32}P. The DNA was then sedimented in an alkaline sucrose gradient and the unit length linear molecules and smaller fragments pooled separately. The 5' nucleotide was released by digestion with nuclease P1, the four bases separated by thin layer chromatography, and the relative level of radioactivity in each determined (Table IV). All four of the nucleotides were labeled and, although the distribution was not random with respect to the base distribution in SV40 DNA, there was no strong bias in the distribution. The results for DNase nicked SV40 DNA are presented for comparison.

DISCUSSION

Conditions have been found under which topoisomerases from both procaryotic and eucaryotic sources break DNA and the distribution of the break sites suggest that the sequences involved in determining sites of DNA breakage are

Table IV. 5' end group analysis of topo 1 nicked duplex DNA. One μg of SV40 Form I DNA was incubated with RL topo 1 (0.5 μg in 10 μl) for 10 minutes at 24° and the reaction terminated with alkali. The reaction was neutralized and the conditions adjusted for labeling with ^{32}P by polynucleotide kinase and [γ-^{32}P]ATP. Incubation was for 3 hours at 0° and the reaction terminated by proteinase K digestion. The products were ethanol precipitated from 2 M ammonium acetate and fractionated on an alkaline sucrose gradient. The fractions containing linears (pool a) and smaller fragments (pool b) were pooled separately and analyzed in Experiment A; only the unit linears were analyzed in Experiment B. In Experiment C, DNase nicked SV40 DNA was denatured, dephosphorylated, and then labeled with ^{32}P as above. The labeled products were digested with nuclease P1 and analyzed by thin layer chromatography. Spots were located by autoradiography, excised and counted for radioactivity. Results are presented as percent of the total counts recovered. Radioactivity amounting to less than 2% of the total remained at the origin. SV40 DNA is 41% G + C.

Experiment	pool	%G	%C	%A	%T	recovered cpm	number of determinations
A.(topo)	a	32	20	25	23	10,400	3
	b	28	20	25	27	6,900	2
B.(topo)	a	33	21	25	22	12,200	5
C.(DNase)		26	19	20	35	64,800	3

not random (12,13,14,15,16,25,26). A relationship between the breakage sites and DNA binding sites has been demonstrated for DNA gyrase by using DNase protection studies (26,27,28). Such a relationship has not been demonstrated for the eucaryotic type 1 topoisomerase, but it is likely that sites at which breakage occurs reflect some feature in the DNA that is recognized by the enzyme. As a minimum, it may be assumed that sites at which breakage occurs are also binding sites for the enzyme, the reverse need not be true. For the procaryotic topo I a unique recognition sequence is not required for strand breakage at discrete sites (15) although a C residue is usually observed four nucleotides 5' to the break site. Studies using denatured DNA as a substrate for topo 1 from HeLa cells also show no requirement for a unique sequence although a bias at certain nucleotide positions near the site of breakage is observed (16).

In this study, 245 sites of breakage by wheat germ topo 1 and 223 sites of breakage by rat liver topo 1 were mapped on duplex SV40 DNA. The frequency of breakage varies greatly from one site to another with 25% of the sites accounting for greater than 80% of the breakage. These strong sites show a distinct bias in the -4 to -1 positions, with positions -3 to -1 apparently exerting the strongest influence and the -4 position possibly modulating the

effect. A consensus sequence for both enzymes would be 5'- A/T G/C T/A T -3'. However, the enzymes do show some differences. RL topo 1 breakage sites frequently have an A at position -3, and WG topo 1 sites appear to be influenced less by the -4 and -1 positions, especially in the intermediate and weak sites. Inspection of the sequences at break sites (Fig. 2) indicates that many of the sites, especially the weak ones, do not conform to the consensus sequence at other positions as well. It is possible that under the conditions used here, a relaxed specificity of the enzyme is being observed. A number of restriction enzymes have this property (29,30,31,32). Increasing the concentrations of salt or glycerol, or including Mg^{++} or Mn^{++} in the topo 1 reaction did not have any obvious effects on the specificity of breakage. While it is possible that conditions will be found to "tighten" the specificity, the current data indicate a rather "loose" sequence requirement for the enzyme.

The nature of the ambiguities seen in the -4 to -2 positions of the consensus sequence suggest that the recognition of this sequence by topo 1 may involve interactions with the minor groove of the DNA. Seeman et al. (33) have pointed out that, within the minor groove, the location of groups capable of forming hydrogen bonds would make it difficult for a protein to discriminate an A-T base pair from a T-A base pair, and a similar situation exists for G-C and C-G base pairs. As a result, an enzyme that associates with the minor groove of DNA would be expected to exhibit ambiguities with respect to the orientation of the base pairs.

The consensus sequence for RL and WG topo 1 is related to that found for HeLa topo 1 by Edwards et al. (16)(T,A, or C at -4; A,C, or G at -3; T,A, or C at -2; T or C at -1 and T,C, or A at +1), but differs in that the consensus determined in this study contains fewer ambiguities in the -4 to -1 positions and there is no nucleotide bias at the +1 position. These differences in the consensus sequences could have several origins. First, the enzymes were isolated from different sources. However, similarities between RL and WG topo 1 suggest this may be of minor importance. Second, active type 1 topoisomerases isolated from eucaryotes come in a variety of sizes apparently as a result of proteolysis during purification (34,35,17,18). The various sizes might show differences in specificity. A comparison of breakage specificity using RL topo 1 purified as either the 70K form (36) or 100K form (17) shows only subtle differences (unpublished results), so this is probably not a large factor in generating the observed differences in specificity. In addition, the HeLa enzyme was purified as the large form (16,34). Third, although most of

the sequences examined in this study were different from those studied by Edwards et al. (16), the sequences did overlap in two regions covering 95 bases. In those regions, 13 rat liver topo 1 break sites were detected and six sites for the HeLa topo 1 were reported (16), three of these sites were common to both enzymes. Therefore, differences were observed even when the same sequences were examined. Fourth, the consensus sequence determined by Edwards et al. (16) was based on data obtained using denatured DNA as a substrate for topo 1. We chose duplex DNA as a substrate because it appears that the specificity of breakage sites on single-stranded DNA by topo 1 is strongly influenced, both quantitatively and qualitatively, by secondary structure in the single-stranded DNA (Been and Champoux, submitted for publication). Finally, the consensus sequence which was obtained in this work is based only on the strongest breakage sites rather than on all of the detectable sites of breakage. While the sequences that fit the consensus are nearly always broken (75/77 for RL and 72/77 for WG, Table III), not all are strong sites. Likewise strong sites exist that do not fit the consensus, so this choice of weighting the strong sites may not be justified. On the other hand, even restriction endonucleases show large variabilities in the rates of cutting recognition sequences (37,38,39,40,41,42). Apparently flanking sequences can influence the interaction of these enzymes with DNA and this is likely to be the case for topo 1 as well. It should be emphasized that the observed breakage of duplex DNA is thought to result from a trapping of the nicked intermediate in the topoisomerase reaction. As a result, the frequency with which a site is broken could reflect factors that affect the probability of trapping the nick, and could be relatively independent of the strength of enzyme binding to that specific sequence. One possibility is that flanking sequences at a break site might affect the length of time that the DNA is in the nicked form versus the closed form.

Topo 1 breaks duplex DNA at sites that occur, on the average, more frequently than 1 per 10 bases. Even if one considers only the strong sites, which occur on the average every 30 to 35 bases on one strand, a random sequence of duplex DNA will contain a site every 1 to 2 turns of the helix. If the break sites actually represent those sequences with which topo 1 normally interacts, then DNA sequence alone would appear to place few limits on the access of the enzyme to DNA. The lack of a requirement for a unique sequence for topo 1 is important when one is considering implications of alterations in the topological state of DNA (1) and processes or mechanisms that might depend on such topological states (43). If the access of topo 1 to DNA is controlled

in vivo, then it would appear to require the involvement of other proteins and may be controlled at the level of chromatin structure.

ACKNOWLEDGEMENTS

We thank Dayle Hager for providing copious quantities of purified wheat germ topo 1, Doug Nicoll for the DNA sequence analysis program, and Kathy Spangler for preparation of the manuscript. This work was supported by a National Science Foundation Grant PCM-8300944 to J.C. and National Institutes of Health Grants GM23224 to J.C., and CA07175 and CA23076 to R.B.

[3]Current address: Department of Chemistry, University of Colorado, Boulder, CO 80309, USA

REFERENCES

1. Gellert, M. (1981) Ann. Rev. Biochem. 50, 879-910.
2. Champoux, J.J. (1978) Ann. Rev. Biochem. 47, 449-479.
3. Wang, J.C. and Liu, L.F. (1979) Molecular Genetics, Part III (S.H. Taylor, ed.) pp. 65-88 Academic Press, New York.
4. Prell, B. and Vosberg, H.P. (1980) Eur. J. Biochem. 108, 389-398.
5. Been, M.D. and Champoux, J.J. (1980) Nucleic Acids Res. 8, 6129-6142.
6. Been, M.D. and Champoux, J.J. (1981) Proc. Natl. Acad. Sci. USA 78, 2883-2887.
7. Halligan, B.D., Davis, J.L., Edwards, K.A. and Liu, L.F. (1982) J. Biol. Chem. 257, 3995-4000.
8. Craig, N.L. and Nash, H.A. (1983) Cell 35, 795-803.
9. Champoux, J.J. (1976) Proc. Natl. Acad. Sci. USA 73, 3488-3491.
10. Champoux, J.J. (1977) Proc. Natl. Acad. Sci. USA 74, 3800-3804.
11. Champoux, J.J. (1978) J. Mol. Biol. 118, 441-446.
12. Sugino, A., Peebles, C.L., Kreuzer, K.N. and Cozzarelli, N.R. (1977) Proc. Natl. Acad. Sci. USA 74, 4767-4771.
13. Gellert, M., Mizuuchi, K., O'Dea, M.H., Itoh, T., Tomizawa, J.-I. (1977) Proc. Natl. Acad. Sci. USA 74, 4772-4776.
14. Champoux, J.J., Young, L.S., Been, M.D. (1979) Cold Spring Symp. Quant. Biol. 43, 53-59.
15. Tse, Y.C., Kirkegaard, K. and Wang, J.C. (1980) J. Biol. Chem. 255, 5560-5565.
16. Edwards, A.K., Halligan, B.D., Davis, J.L., Nivera, N.L., and Liu, L.F. (1982) Nucleic Acids Res. 10, 2565-2576.
17. Martin, S.R., McCoubrey, W.K., McConaughy, B.L., Young, L.S., Been, M.D., Brewer, B.J. and Champoux, J.J. (1983) Methods Enzymol. 100, 137-144.
18. Dynan, W.S., Jendrisak, J.J., Hager, D.A. and Burgess, R.R. (1981) J. Biol. Chem. 256, 5860-5865.
19. Champoux, J.J. and McConaughy, B.L. (1975) Biochemistry 14, 307-316.
20. Tegtmeyer, P., Dohan, C. and Reznikoff, C. (1970) Proc. Natl. Acad. Sci. USA 66, 745-752.
21. Rundell, K., Collins, J.K., Tegtmeyer, P., Ozer, H.l., Lai, C.J. and Nathans, D. (1977) J. Virol. 21, 636-646.
22. Buchman, A.R., Burnett, L. and Berg, P. (1981) in DNA Tumor Viruses, Part II/Revised (J. Tooze, ed.) pp. 799-841 Cold Spring Harbor Laboratory.
23. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
24. Tapper, D.P. and Clayton, D.A. (1981) Nucleic Acids Res. 9, 6787-6794.

25. Morrison, A. and Cozzarelli, N.R. (1979) Cell 17, 175-184.
26. Fisher, L.M., Mizuuchi, K., O'Dea, M.H., Ohmori, H. and Gellert, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4165-4169.
27. Kirkegaard, K. and Wang, J.C. (1981) Cell 23, 721-729.
28. Morrison, A. and Cozzarelli, N.R. (1981) Proc. Natl. Acad. Sci. USA 78, 1416-1420.
29. Polisky, B., Greene, P., Garfin, D.E., McCarthy, B.J., Goodman, H.M. and Boyer, H.W. (1975) Proc. Natl. Acad. Sci. USA 72, 3310-3314.
30. Hsu, M. and Berg, P. (1978) Biochemistry 17, 131-138.
31. Malyguine, E., Vannier, P. and Yot, P. (1980) Gene 8, 163-177.
32. Heininger, K., Horz, W. and Zachau, H.G. (1977) Gene 1, 291-303.
33. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) Proc. Natl. Acad. Sci. USA 73, 804-808.
34. Liu, L.F. and Miller, K.M. (1981) Proc. Natl. Acad. Sci. USA 78, 3487-3491.
35. Tricli, J.V. and Kowalski, D. (1983) Biochemistry 22, 2025-2031.
36. Champoux, J.J. and McConaughy, B.L. (1975) Biochemistry 15, 4638-4642.
37. Thomas, M. and Davis, R.W. (1975) J. Mol. Biol. 91, 315-328.
38. Forsblum, S., Rigler, R., Ehrenberg, M., Pettersson, U., and Philipson, L. (1976) Nucleic Acids Res. 3, 3255-3269.
39. Brown, N.L. and Smith, M. (1977) Proc. Natl. Acad. Sci. USA 74, 3213-3216.
40. Armstrong, K. and Bauer, W.R. (1982) Nucleic Acids Res. 10, 993-1007.
41. Armstrong, K.A. and Bauer, W.R. (1983) Nucleic Acids Res. 11, 4109-4126.
42. Gingeras, T.R. and Brooks, J.E. (1983) Proc. Natl. Acad. Sci. USA 80, 402-406.
43. Weintraub, H. (1983) Cell 32, 1191-1203.