Recognition sites of eukaryotic DNA topoisomerase I: DNA nucleotide sequencing analysis of topo I cleavage sites on SV40 DNA


Department of Physiological Chemistry, John Hopkins Medical School, Baltimore, MD 21205, USA

Received 3 February 1982; Revised and Accepted 5 April 1982

ABSTRACT

Eukaryotic DNA topoisomerase I introduces transient single-stranded breaks on double-stranded DNA and spontaneously breaks down single-stranded DNA. The cleavage sites on both single and double-stranded SV40 DNA have been determined by DNA sequencing. Consistent with other reports, the eukaryotic enzymes, in contrast to prokaryotic type I topoisomerases, links to the 3'-end of the cleaved DNA and generates a free 5'-hydroxyl end on the other half of the broken DNA strand. Both human and calf enzymes cleave SV40 DNA at identical and specific sites. From 827 nucleotides sequenced, 68 cleavage sites were mapped. The majority of the cleavage sites were present on both double and single-stranded DNA at exactly the same nucleotide positions, suggesting that the DNA sequence is essential for enzyme recognition. By analyzing all the cleavage sequences, certain nucleotides are found to be less favored at the cleavage sites. There is a high probability to exclude G from positions -4, -2, -1 and +1, T from position -3, and A from position -1. These five positions (-4 to +1 oriented in the 5' to 3' direction) around the cleavage sites must interact intimately with topo I and thus are essential for enzyme recognition. One topo I cleavage site which shows an atypical cleavage sequence maps in the middle of a palindromic sequence near the origin of SV40 DNA replication. It occurs only on single-stranded SV40 DNA, suggesting that the DNA hairpin can alter the cleavage specificity. The strongest cleavage site maps near the origin of SV40 DNA replication at nucleotide 31-32 and has a pentanucleotide sequence of 5'-TACAT-3'.

INTRODUCTION

Eukaryotic DNA topoisomerase I is a type I DNA topoisomerase which introduces transient single-stranded DNA breaks on DNA (1-3). Compatible with its mechanism, topo I has been implicated in a variety of genetic processes, such as DNA replication, RNA transcription, recombination, and DNA condensation and decondensation by providing a swivel (reviewed in ref. 1 and 4). More recently, eukaryotic topo I has been demonstrated to be able to transfer heterologous DNA strands, implying an additional role of topo I in nucleic acids sequence rearrangements (5, 6). In order to study the biological functions of topo I, we have examined the sequence specificity of topo I on SV40 DNA.
MATERIALS AND METHODS

Enzymes.

HeLa topo I was purified as a homogeneous monomeric protein of 100,000 daltons as described (3). Calf thymus topo I was purified as a partially proteolyzed enzyme of 82,000 daltons (Halligan, B. D., Edwards, K. A. and Liu, L. F., unpublished result). The large fragment of E. coli DNA polymerase I and restriction enzymes were purchased from Bethesda Research Lab. T4 polynucleotide kinase was a gift from Dr. Leonard Klevan.

Isolation of restriction fragments of SV40 DNA.

SV40 virus (small plaque, strain 776) was grown on BSC-1 cells in minimal Eagle's medium (MEM) with 2% fetal calf serum. SV40 DNA was extracted by the Hirt procedure (7) and purified by CsCl/ethidium centrifugation and gel filtration on an A50m column. Fifty μg of the purified SV40 DNA was digested with HindIII restriction enzyme, followed by phenol extractions, other extractions and ethanol precipitation. Ten μg of SV40 DNA was also digested with HpaII restriction enzyme and processed similarly.

3'-end labeling of the SV40 restriction fragments.

Fifty μg of HindIII digested SV40 DNA was dissolved in a 100 μl reaction mixture containing 40 mM potassium phosphate, pH 7.4, 1 mM 2-mercaptoethanol, 6 mM MgCl₂, 10 μM dATP, 10 μM dTTP, 10 μM dCTP, 0.5 μM α-32P-dOTP (2000 Ci/mmol), 6 units of the Klenow fragment of E. coli DNA polymerase I. The reaction was carried out at 10 C for 1 hour and then stopped with 300 μl 5 mM Na EDTA and 7% (w/v) glycerol. 10 μl of the HpaII cut SV40 DNA was labeled similarly except that dATP and dTTP were omitted in the reaction mixture.

Gel isolation of the end-labeled SV40 restriction fragments.

The reaction mixture containing the end-labeled SV40 DNA restriction fragments was loaded onto a 1.8% agarose gel in a horizontal gel apparatus containing TBE electrophoresis buffer (8) and 0.5 μg/ml ethidium bromide. Electrophoresis was carried out in the dark. Restriction fragments were isolated from the gel by cutting a slot ahead of each gel band (visualized by hand UV light) and electrophoresing the DNA into the slot. A piece of dialysis membrane was inserted into each slot to block the DNA from pitting through the slot. The electrodes were then reversed for 30 seconds in order to avoid any nonspecific binding of the DNA to the membrane. DNA was collected from the slots, made 0.6 M NaCl and extracted three times with n-butanol to remove the ethidium. DNA was then ethanol precipitated, washed and redissolved in a proper buffer for the digestion with another restriction enzyme.
DNA sequencing.

DNA restriction fragments isolated from the agarose gel were redigested with various restriction enzymes. The fragments were then isolated using a native 5% polyacrylamide gel (9). DNA fragments were extracted from the gel as described (9). These double-digested restriction fragments were then used for DNA sequencing (9) and topoisomerase cleavage.

Topoisomerase Cleavage reactions.

DNA restriction fragments were either used as double-stranded or alkali denatured (0.13 N NaOH, 10 min at room temperature and then neutralized) immediately before the reactions. Normally, the reaction mixture (50 μl) contained 40 mM Tris, pH 7.7, 30 mM NaCl, 1 mM MgCl₂, 10 μg/ml bovine serum albumin, labeled DNA and 100 ng of topo I. Incubation was at 37 °C for 30 min. Reactions were terminated either by the addition of 1% SDS, or 1% SDS and 50 μg/ml proteinase K. In the latter case, the reaction was continued for another 30 min.

T4 polynucleotide kinase reactions.

The topo I cleavage products were phenol extracted, ether extracted, ethanol precipitated and redissolved in 70 mM Tris, pH 7.6, 100 mM KCl, 10 mM MgCl₂, 5 mM dithiothreitol, 1.5 mM ATP and 2 units of T4 polynucleotide kinase. The kinase reactions were carried out at 37 °C for 30 min.

RESULTS.

Site-specific cleavage of SV40 DNA by eukaryotic DNA topoisomerase I.

To determine if topo I cleaves DNA at specific sites, a 3'-end labeled HindIII restriction fragment of SV40 DNA (nucleotide number 1708-3476) was cut with Eco RI and the larger fragment (nucleotide number 1782-3476) was isolated. As shown in Fig. 1, lane a, treatment of this end-labeled DNA fragment (in single-stranded form) with HeLa topo I produced a series of DNA bands in an agarose gel containing 0.1% SDS in the electrophoresis buffer (TBE). This result suggested that topo I cleaves DNA at specific sites. Furthermore, the same set of bands was produced whether HeLa topo I or calf thymus topo I was used (Fig. 1, lanes a and c respectively). Previously, we have reported that when 5'-end labeled DNA was used and the topo I cleavage products were analyzed by the same gel system (0.1% SDS in the electrophoresis buffer), all the products showed mobility shifts before and after the proteinase K treatment due to the removal of the covalently bound topo I (6). When 3'-end labeled DNA was used and the products treated with proteinase K (Fig. 1, lane b), no mobility shifts were observed (compare lane a and b). This result is consistent with the previous reports that topo I is linked exclusively to the 3'-ends of the broken DNA strands (6,10,11).
Figure 1. Site-specific cleavage of DNA by eukaryotic DNA topoisomerase I. The largest restriction fragment of HindIII digested SV40 DNA (nucleotide 1708-3476) was labeled at its 3'-ends, gel isolated and redigested with Eco RI. The large fragment of this doubly digested DNA (nucleotide 1782-3476) was isolated, denatured and treated with eukaryotic DNA topoisomerase I as described in the MATERIALS AND METHODS. The products were loaded onto a 2% agarose gel containing TBE buffer and 0.1% SDS. (a) HeLa topo I, (b) same as (a) except stopped by SDS proteinase K treatment, (c) calf thymus topo I. (a) and (c) were stopped by 1% SDS. Gels were dried and autoradiographed. The arrow indicates the position of the unreacted fragment.

Determination of the topo I cleavage sites by DNA sequencing.

To determine the exact nucleotide positions of topo I cleavage sites, SV40 DNA was cut with either HindIII or HpaII and the fragments were labeled at their 3’-ends by polymerase ‘filling in’ with α-32P-dGTP. All fragments were gel isolated and further cut with the proper restriction enzymes to generate fragments for DNA sequencing analyses. Fig. 2 shows the results of the DNA sequencing analyses of one of these fragments. This particular fragment (nucleotide 5176-348) was produced from the HindIII and HpaII double digest and labeled only at the 3’ end of the HindIII site (nucleotide 5176). The cleavage products produced by HeLa topo I treatment
Figure 2. DNA sequencing analyses of topo I cleavage sites on SV40 DNA. This figure shows part of the sequencing analysis of fragment A (nucleotide 5176-348, HindIII and HpaII double digest). (A) Single-stranded fragment A was treated with HeLa topo I. (B) same as (A) except that the cleavage products were phosphorylated by T4 polynucleotide kinase treatment. (C), (D), (G) and (H) were the sequencing ladders of G, G+A, C+T and C respectively. (E) and (F) were the same as (A) and (B) respectively except that double-stranded fragment A was used.

were run parallel to the sequencing ladders. Because the cleavage products produced by topo I treatment possess 5'-OH ends (6,10,11) rather than 5'-phosphoryl ends which are produced by chemical cleavages (9), the topo I cleavage products were phosphorylated by T4 polynucleotide kinase for accurate size comparisons with the sequencing ladders. Phosphorylation of the 5'-OH ends of the topo I cleavage products increased their mobilities in the gel as expected (compare lanes A and B, and lanes E and F) (12).

When the double-stranded DNA fragment was used as the substrate, one topo I cleavage site (nucleotide 5231-5232) was revealed, (Fig.2, lanes E and F). However, two sites were clearly present when the corresponding single-stranded DNA was used as the substrate (Fig.2, lanes A and B).
site was at exactly the same nucleotide position (nucleotide 5231-5232) as the cleavage site on the double-stranded DNA substrate, and the other site (nucleotide 5241-5242) was only present when the single-stranded DNA substrate was used. The majority (greater than 70%) of the sites, however, located at identical nucleotide positions whether double or single-stranded DNA was used as the substrate. Since the efficiency of topo I cleavage is much lower on double-stranded DNA than on single-stranded DNA (1,6), most of the sites which seemed to be only present on single-stranded DNA substrate can not be unambiguously established. All the topo I cleavage sites on double-stranded DNA substrates, however, were unambiguously identified to be present at exactly the same nucleotide positions as the cleavage sites on the corresponding single-stranded DNA substrates. This result strongly suggested that the DNA sequence was essential for enzyme recognition.

To search for the recognition sequence, 68 sites were determined from a total of 827 nucleotides sequenced (Fig. 3). The strength of each cleavage site was determined by the intensity of each cleavage product in the gel under conditions where very limited cleavage occurred. It was found that the strength of the cleavage sites varied over a 10 fold range. The strongest cleavage site (nucleotide 31-32) and the second strongest site (nucleotide 5231-5232) were all on the same restriction fragment containing the origin of SV40 DNA replication (Fig. 3, A fragment). From the 68 sites sequenced, twenty were relatively 'strong' sites (indicated by long vertical lines) and the rest were relatively 'weak' sites (indicated by short vertical lines) (see Fig. 3).

Dinucleotide frequency analysis of the topo I cleavage sites.

Simple analysis of the dinucleotide sequences at the sites of cleavage indicated that the cleavage by topo I was not random (Table 1, column A). The dinucleotide sequence TpT (35%) is the most frequent sequence straddling the cleavage sites. Pyrimidine dinucleotide sequences (TT, TC, CC and CT) at the sites of cleavage represent 69% of the total dinucleotide sequences at the cleavage sites. Certain dinucleotide sequences (AC, AG, AT, GC and GA) showed zero frequency for cleavage by topo I (Table 1, column A). Clearly, topo I prefers pyrimidine dinucleotides over the purine dinucleotides at the sites of cleavage. To demonstrate that the cleavage frequencies of the dinucleotide sequences were statistically meaningful, the dinucleotide frequencies of the sequenced DNA fragments (see Fig. 3) were calculated (Table 1, column B). The dinucleotide frequencies of the sequenced DNA fragments (Table 1, column B) parallels that of the total SV40 DNA (Table 1, column C), but are quite different from that of the topo I cleavage sites (Table 1, column A). We therefore
Figure 3. Topo I cleavage map on SV40 DNA.

Eight fragments of SV40 DNA (a total of 827 nucleotides) were sequenced from their 3'-ends, and their topo I cleavage sites (a total of 68 sites) were mapped (see MATERIALS AND METHODS). The cleavage sites were determined by using both single-stranded and double-stranded restriction fragments. The sites on single-stranded DNA are shown in this figure. Note that sites on double-stranded DNA form a subset of the sites on single-stranded DNA (see text). Long vertical lines indicate the sites of relatively 'strong' cleavage. Short vertical lines indicate the sites of relatively 'weak' cleavage. The eight restriction fragments were all derived from double digests. Fragment A: HindIII-HpaII(*5171-346). Fragment B: TaqI-HindIII(4739-5171*). Fragment C: HindIII-TaqI(*4002-4739). Fragment D: HincII-HindIII(2666-3476*). Fragment E: HindIII-HincII(*1708-2057). Fragment F: HpaII-HindIII(346-1046*). Fragment G: HpaII-HindIII(*346-1046). Fragment H: HindIII-HpaII(5171-346*). The asterisk(*) indicates the 3' end of the fragment and the numbers indicate the nucleotide positions at the termini of the restriction fragments.
Table 1. Dinucleotide frequency distribution at the sites of topo I cleavage. Column A shows the frequency distribution of the dinucleotide sequences straddling the cleavage sites. Column B shows the dinucleotide frequency distribution of the 827 nucleotides sequenced. Column C shows the dinucleotide frequency distribution of entire SV40 genome.

Table: Dinucleotide frequency distribution

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-3'</td>
<td></td>
<td></td>
<td>10.2%</td>
</tr>
<tr>
<td>AA</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0</td>
<td>5.3</td>
<td>5.5</td>
</tr>
<tr>
<td>AG</td>
<td>0</td>
<td>7.4</td>
<td>6.6</td>
</tr>
<tr>
<td>AT</td>
<td>0</td>
<td>5.6</td>
<td>6.7</td>
</tr>
<tr>
<td>CA</td>
<td>3</td>
<td>7.5</td>
<td>8.1</td>
</tr>
<tr>
<td>CC</td>
<td>6</td>
<td>8.0</td>
<td>4.7</td>
</tr>
<tr>
<td>CG</td>
<td>0</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>CT</td>
<td>13</td>
<td>8.0</td>
<td>7.6</td>
</tr>
<tr>
<td>GA</td>
<td>0</td>
<td>5.1</td>
<td>4.6</td>
</tr>
<tr>
<td>GC</td>
<td>1.5</td>
<td>4.4</td>
<td>5.1</td>
</tr>
<tr>
<td>GG</td>
<td>3</td>
<td>4.5</td>
<td>5.2</td>
</tr>
<tr>
<td>GT</td>
<td>1.5</td>
<td>4.5</td>
<td>4.9</td>
</tr>
<tr>
<td>TA</td>
<td>13</td>
<td>5.9</td>
<td>6.2</td>
</tr>
<tr>
<td>TC</td>
<td>15</td>
<td>6.8</td>
<td>5.6</td>
</tr>
<tr>
<td>TG</td>
<td>6</td>
<td>5.5</td>
<td>7.4</td>
</tr>
<tr>
<td>TT</td>
<td>35</td>
<td>12.0</td>
<td>11.0</td>
</tr>
</tbody>
</table>

conclude that topo I prefers certain DNA sequences for its site-specific cleavages.

The 'consensus sequence' of topo I cleavage sites.

To extend the analysis, the nucleotide frequencies at each position extending 10 nucleotides on each side of the cleavage sites were calculated (Table 2). The nucleotide frequencies at positions -10 to -5 and +2 to +10 showed random distribution. The nucleotide frequencies at positions -4 to +1, however, showed quite biased distributions, suggesting that topo I interacts intimately with these nucleotides. Certain nucleotides such as G, T, G, G and A, and G at positions -4, -3, -2, -1, and +1 respectively showed very low frequencies (below 10%). A 'consensus sequence' based on the exclusion of those nucleotides at each nucleotide positions has the following form:

T A T T T
5'-A C A C C-3'
C G C A
Table 2. Nucleotide frequency analysis of the topo I cleavage sites.
Each of the 68 cleavage sites were arranged in a 5' to 3' direction with 10 nucleotides on each side of the cleavage sites. The percentage of each base (A, C, G or T) present at each nucleotide position was calculated for all 68 sites. The arrow indicates the position of cleavage.

|   | 5' | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10' |
|---|----|-----|----|----|----|----|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 18 | 22 | 25 | 24 | 24 | 24 | 37 | 37 | 31 | 31 | 3 | 3 | 3 | 5 | 6 | 7 | 8 | 9 | 10 |   |
| C | 25 | 29 | 25 | 20 | 25 | 34 | 19 | 19 | 19 | 22 | 13 | 16 | 25 | 20 | 19 | 22 | 24 | 19 | 29 |   |
| G | 24 | 18 | 16 | 32 | 32 | 18 | 24 | 6 | 6 | 9 | 25 | 22 | 20 | 18 | 13 | 20 | 18 | 13 | 24 |   |
| T | 34 | 31 | 34 | 22 | 19 | 25 | 38 | 6 | 47 | 69 | 50 | 34 | 33 | 38 | 28 | 35 | 24 | 32 | 43 | 31 |

There are 162 combinations of such a 'consensus' sequence. Since only 68 sites were determined, some combinations certainly were not represented at the cleavage sites we have sequenced. This pattern of site-specificity in the absence of a unique determining sequence seems to be a common feature among DNA topoisomerase. Such sequences are expected to be present at a frequency of once every 6-7 nucleotides. The actual topo I cleavage sites has an average frequency of once every 12 nucleotides (68 sites from a total of 827 nucleotides sequenced).

Effect of a DNA hairpin on the specificity of topo I cleavage.
One peculiar topo I cleavage site (nucleotide 5241-5242, see Fig. 2) which occurred only when a single-stranded DNA substrate was used has a dinucleotide sequence of 5'-GC-3'. Since this cleavage site clearly occurred only on the single-stranded DNA, we looked for secondary structures in this region. Indeed, this region can be organized into a perfect hairpin (Fig. 4) with a 11 base stem and 5 base loop. The cleavage site mapped at the loop region as indicated. This hairpin is quite stable and is most likely part of a larger hairpin at the origin of DNA replication as identified by others (13). Since GpC is not normally cleaved by topo I, we conclude that the hairpin loop structure changed the specificity of topo I cleavage.

DISCUSSION
Because eukaryotic DNA topoisomerase I can transiently break one strand of double-stranded DNA and spontaneously cleave single-stranded DNA (5,6), the sites of such cleavage can thus be conveniently studied by DNA sequencing methods to probe the protein nucleic acid interaction. It can be reasonably assumed that the nucleotide sequences around the cleavage
sites may be the recognition sequence of topo I. We have determined 68 topo I cleavage sites by sequencing 827 nucleotides of SV40 DNA. It perhaps is not surprising to find that topo I does not recognize a unique DNA sequence but nonetheless cleaves DNA at specific sites, since prokaryotic DNA topoisomerases also cleave DNA at specific sites in the absence of a unique recognition sequence (14,15). A statistical analysis of the sequences surrounding the 68 cleavage sites, however, has clearly demonstrated a biased nucleotide distribution. The strongest bias is the presence of pyrimidine nucleotides at the 3'-ends of the cleaved DNA strands (91% pyrimidine nucleotides at the -1 position). Bias of nucleotides at positions -4 to +1 (see Table 2 for the designation of the nucleotide positions relative to the site of cleavage) are quite noticeable. Except for position -1, each of these nucleotide positions seems to exclude one particular nucleotide. G, T, G and G seem to be excluded at positions -4, -3, -2 and +1 respectively.

The nucleotide biases at these five nucleotide positions suggest their intimate interactions with topo I. However, nucleotides outside this pentanucleotide core must also influence the site of cleavage. For example, one of the strong sites, at nucleotide 5231-5232, has a pentanucleotide sequence of 5'-CCICT-3' (the cleavage site is between the C and T near the 3' end). The same sequence CCTCT also occurs once in
fragment B at nucleotide 5139-5143 and once in fragment H at nucleotide 310-311 (see Fig. 3). None of the latter two sequences were cleaved by topo I(cleavage sites seemed to be shifted by two nucleotides towards the 3'-ends in both cases). Clearly, this pentanucleotide sequence alone is not sufficient to determine the site of topo I cleavage.

Mapping of topo I cleavage sites on SV40 DNA has the advantage of correlating the cleavage sites with various important landmarks on SV40 DNA. It is interesting to note that the strongest cleavage site occurs between nucleotide 31 and 32 (the pentanucleotide sequence is TGACT) which is located near the origin of SV40 DNA replication (16) and between the AT rich region and the large T antigen binding site III (17). The second strongest site maps between nucleotide 5231-5232 which is located in the large T antigen binding site II (17). Whether or not these correlations are significant has to await further studies. It should be pointed out that the cleavage sites of topo I were determined only from one strand. Additional sites are expected to be present on the complementary strand in this interesting region of SV40 DNA.

Because eukaryotic topo I can spontaneously break down single-stranded DNA and covalently transfer heterologous DNA strands (5,6), it has been implicated to be the key enzyme in illegitimate recombination in eukaryotic cells (5,6). The sites of topo I cleavage may thus reflect the junctions at the sites of illegitimate recombination before rearrangement. Unfortunately, the exact crossing-over points of all the sequenced host-SV40 novel junctions cannot be unambiguously determined because the host DNA sequences before integration are not known (18-20). Any short homologous sequences between the host DNA and the SV40 DNA at the sites of integration will obscure the exact crossing-over points. Rearrangement within SV40 DNA can be determined accurately at the nucleotide sequence level and thus the crossing-over points can be identified unambiguously (21). However, the number of the available rearrangement sites is small (21). A meaningful statistical analysis of the SV40 DNA rearrangement sites is not yet possible. It is also interesting to speculate that the movement of certain transposable elements in eukaryotic cells may be promoted via topoisomerase mediated illegitimate recombination at sites flanking each transposable element (22). Further studies are necessary to establish the role of topoisomerase I in various genetic processes.

ACKNOWLEDGMENTS
This work was supported by an NIH grant and the Chicago Community Trust/Searle Scholars Program. B. D. Halligan and N. L. Nivera are NIH predoctoral trainees.
REFERENCES