1830–1837 *Nucleic Acids Research, 2000, Vol. 28, No. 8*

Determination of the recognition sequence of *Mycobacterium smegmatis* topoisomerase I on mycobacterial genomic sequences

Devanjan Sikder and Valakunja Nagaraja*

Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India

Received November 18, 1999; Revised January 17, 2000; Accepted February 23, 2000

ABSTRACT

***Mycobacterium smegmatis* topoisomerase I has several distinctive features. The absence of the zinc finger motif found in other prokaryotic type I topoisomerases and the ability of the enzyme to recognise single-stranded and duplex DNA are unique characteristics of the enzyme. We have mapped the strong topoisomerase sites of the enzyme on genomic DNA sequences from *Mycobacterium tuberculosis* and *M.smegmatis*. The enzyme does not nick DNA in random fashion and DNA cleavage occurred at a few specific sites. Mapping of these sites revealed conservation of a pentanucleotide motif CG/TCT↓T at the cleavage site (↓ represents the cleavage site). The enzyme binds and cleaves consensus oligonucleotides having this sequence motif. The protein exhibits a very high preference for C or a G residue at the +2 position with respect to the cleavage site. Based on earlier and the present studies we propose that the enzyme functions *in vivo* mainly at these specific sites to carry out topological reactions.**

INTRODUCTION

Topoisomerases are a group of enzymes that catalyse the interconversion of various forms of DNA and hence participate in almost all cellular functions involving DNA transactions. These ubiquitous enzymes have been divided into two classes based on the DNA cleavage pattern during the reaction. Type I enzymes nick one of the strands of the DNA and then pass the intact strand to change the linking number by a step of one. In contrast, type II topoisomerases cleave duplex DNA through which another duplex DNA segment is passed, changing the linking number by steps of two. In both cases, the reaction involves initial binding, cleavage of one strand or duplex DNA, passage of another segment and subsequent religation of the cleaved DNA (1). The topological status of the genome is thus maintained during replication, transcription and other processes by the action of these enzymes at different locations in the DNA. This would entail a high degree of non-specific binding and cleavage by topoisomerases during the supercoiling and relaxation reactions. Thus, it is no surprise that a majority of topoisomerases do not show prominent sequence specificity (1,2). However, many of the enzymes exhibit a certain degree of non-randomness in cleavage site recognition and few show some sequence specificity. Several studies have addressed the sequence selectivity of type I and type II topoisomerases (3–11). These studies have led to the identification of strong topoisomerase sites for some of the enzymes. *Tetrahymena* DNA topoisomerase I has been shown to preferentially cleave a hexadecameric sequence present in the rDNA (12) while vaccinia virus topoisomerase I recognises a pentanucleotide sequence motif (13). DNA gyrase, a type II topoisomerase found only in bacteria, also recognises specific sites referred to as strong gyrase sites (SGS). A quinolone-induced SGS has been identified at position 990 of pBR322 DNA within the tet gene (14). Another SGS has been located towards the centre of the bacteriophage Mu genome (15). However, none of these topoisomerases exhibit a degree of sequence specificity comparable to the site-specific DNA-binding proteins such as restriction enzymes, repressors and transcription activators, with the exception of vaccinia virus topoisomerase I. This enzyme is the smallest topoisomerase known and probably constitutes the minimal functional unit of a type I enzyme (16) and exhibits a high degree of site-specific interaction. Vaccinia virus DNA topoisomerase I catalyses DNA cleavage with the concomitant formation of a 3′ covalent DNA–protein complex, strand passage and religation at the pentameric sequence motif C/TCCTT in a double-stranded context (17). The only other topoisomerase I having comparable sequence specificity was recently characterised from *Mycobacterium smegmatis* (18). In contrast to the enzyme from vaccinia virus, mycobacterial topoisomerase I is a large, single subunit topoisomerase of 110 kDa and belongs to the type IA topoisomerases (19). The protein relaxes negatively supercoiled but not positively supercoiled DNA. The enzyme forms a 5′ covalent intermediate at the newly created nick before strand passage and subsequent religation (19). The purified *M.smegmatis* topoisomerase I has some unusual characteristics: (i) it is not inhibited by spermidine, unlike *Escherichia coli* topoisomerase I; (ii) the enzyme lacks a zinc finger motif, a characteristic of *E.coli* topoisomerase I (19); (iii) it interacts with both single- and double-stranded DNA, a property not shared by other enzymes. The mycobacterial enzyme recognises and cleaves sequence CG/TCTT, a sequence very similar to that of vaccinia virus. Identification of the strong topoisomerase I site (STS) for mycobacterial topoisomerase I relied on the ability of the enzyme to bind to duplex DNA generated

*To whom correspondence should be addressed. Tel: +91 80 360 0668; Fax: +91 80 360 2697; Email: vraj@mcbl.iisc.ernet.in

Table 1. Oligonucleotides used for the study

Sequence of the oligonucleotides	Description	Purpose
GTT TTC CCA GTC ACG AC	Forward primer	Sequencing
CAG GAA ACA GCT ATG AC	Reverse primer	Sequencing
CTG ACG CGC TGC CGC TTC CA	Con-CCA	Non-covalent complex, cleavage
CTG ACG CGC TGC CGC TTC AA	Con-CAA	Non-covalent complex, cleavage
CTG ACG CGC TGC CGC TTG GT	Con-GGT	Non-covalent complex, cleavage
CTG ACG CGC TGC CGC TTA AC	Con-AAC	Non-covalent complex, cleavage
CTG ACG CGC TGC CGC TTG TA	Con-GTA	Non-covalent complex, cleavage
CTG ACG CGC TGC CGC TTT CG	Con-TCG	Non-covalent complex, cleavage
GAC TGC GCG ACG GCG AAG GT	Non-specific	Non-covalent complex, cleavage

by restriction digestion of an *E.coli* plasmid (18). This analysis was limited to a few sites in a 50% GC-rich 2.6 kb plasmid. This would not be a fair reflection of the sites in the GC-rich genome of mycobacteria. Further, since the analysis was on linear DNA, the importance of topological status on cleavage site recognition could not be ascertained. Now we have determined specific cleavage sites for *M.smegmatis* topoisomerase I on a large number of mycobacterial DNA fragments and arrived at a consensus recognition sequence.

MATERIALS AND METHODS

Purification of topoisomerase I

Topoisomerase I was purified from wild-type *M.smegmatis* SN2 cells as described before (19). One unit of enzyme corresponds to 2 ng of the purified protein catalysing 50% relaxation of 300 ng of supercoiled plasmid at 37°C in 15 min.

Cloning of mycobacterial DNA sequences

Genomic DNA was isolated from *M.smegmatis* SN_2 and *Mycobacterium tuberculosis* H37Ra as described before (20). The DNA was digested with *Bam*HI and *Pst*I and ligated to pUC 19. After transformation, representative colonies were chosen and the recombinant plasmids were purified.

Purification and radiolabelling of oligonucleotides

The oligonucleotides used in this study are listed in Table 1. Oligonucleotides were purified by urea–PAGE and 5′-end-labelled using bacteriophage T4 polynucleotide kinase and [γ-^{32}P]ATP (6000 Ci/mmol). The labelled oligonucleotides were purified from the unincorporated nucleotides using G-25 spin column chromatography.

Mapping topoisomerase I cleavage sites

Different supercoiled plasmids (2 µg) were individually incubated with 50 ng of topoisomerase I for 5 min at 37°C in buffer containing 40 mM Tris–HCl, pH 8.0, and 20 mM NaCl. The reaction mix was then treated with 10 mM NaOH and further incubated for 2 min, after which it was neutralised with 10 mM HCl. The reaction mix was treated with 0.6% SDS and 500 µg/ml proteinase K for 2 h at 55°C. The DNA was subjected to phenol/chloroform extraction and precipitated with 2.5 vol of ethanol. The DNA thus obtained was further purified by spin column chromatography and then annealed to 5′-end-labelled universal forward or reverse primers (Table 1). Primer extension was carried out using the Klenow fragment of DNA polymerase I. The samples were then suspended in loading dye containing 45% formamide, 10 mM EDTA, 0.025% xylene cyanol, heat denatured for 5 min at 90°C and then electrophoresed on 6% denaturing polyacrylamide gels containing 8 M urea. Dideoxy sequencing reactions were carried out using the same labelled primer.

Electrophoretic mobility shift assay

Non-covalent enzyme–DNA complexes were formed using 40 mM Tris–HCl, pH 8.0, 20 mM NaCl, 1 mM EDTA, 100 fmol radiolabelled oligonucleotide and 20 U *M.smegmatis* topoisomerase I by incubation on ice for 5 min. The products were separated in an 8% native polyacrylamide gel (30:0.8) using 0.5× TBE as the running buffer. The radioactivity associated with free DNA and DNA–protein complexes was quantified with a phosphorimager (model BAS 1800; Fuji Film).

Topoisomerase I cleavage reaction

End-labelled oligonucleotides (100 fmol) were incubated with 200 U topoisomerase I in a 50 µl reaction mix containing 40 mM Tris–HCl, pH 8.0 and 20 mM NaCl. The reaction was stopped with 0.6% SDS after 15 min incubation at 37°C, suspended in loading dye and electrophoresed in a 12% denaturing PAGE gel. To map the cleavage site G- and T-specific reactions were carried out on a 20mer consensus oligonucleotide substrate (Con-CCA, Table 1). The end-labelled oligonucleotide (100 000 c.p.m.) was treated with 1% DMS for 1 min at room temperature. DNA was recovered by ethanol precipitation after stopping the reaction with 50 µl of DMS stop buffer. The precipitate was washed with 70% ethanol, dried and dissolved in 90 µl of water. This was followed by the addition of 10 µl of piperidine and incubation at 90°C for 30min. The reaction mix was frozen in liquid nitrogen and dried under vacuum. The pellet was resuspended in 15 µl of formamide dye and electrophoresed in a 16% urea–acrylamide gel for 2 h at 2000 V. For T-specific cleavage, the oligonucleotide substrate was treated with 2 mM $KmNO_4$ for 1 min at 25°C in 10 µl of water. The reaction was stopped with 40 mM β-mercaptoethanol and 0.3 M sodium acetate. The DNA was ethanol precipitated,

suspended in 90 µl of water and heated to 90°C after addition of 10 µl of piperidine. The pellet was suspended in 15 µl of formamide dye and the products were resolved by electrophoresis in a 16% denaturing PAGE gel.

RESULTS

Mapping the STS

The experimental design for mapping the STS in supercoiled DNA is outlined in Figure 1A. The topoisomerase I reaction involves several steps: (i) binding to DNA; (ii) DNA cleavage in one strand and concomitant enzyme–DNA covalent complex formation; (iii) passage of the intact strand through the molecular gate created by the enzyme–DNA complex; (iv) religation of the phosphodiester bond; (v) dissociation of the enzyme. Using protein denaturants such as detergents or alkali, the enzyme reaction can be arrested to obtain the enzyme–DNA covalent complex from the second step. Proteolysis of this complex would leave the plasmid with a nick created by topoisomerase I, which, when subjected to primer extension, would reveal the cleavage site(s). The results of one such experiment are presented in Figure 1B. Plasmid Msp4.4, generated by cloning a 0.4 kb *Pst*I genomic fragment of *M.smegmatis* in pUC19, was used as substrate for the topoisomerase reaction. Specific cleavage is observed at the sequence CTT (Fig. 1B, lane 2). Employing the above strategy, several plasmids containing mycobacterial genomic DNA were analysed for the presence of the STS. Representative data from plasmids MX2R, MX1F, MSal1, MSM5, MBaP and PMSX1 are depicted in Figure 2A–F, respectively. In Figure 2, topoisomerase I-mediated DNA cleavage occurred at the sequence CTT and cleavage appeared to be between the two T bases. Most of the cleavage positions appear to be highly preferred sites as no additional bands were observed, establishing the site specificity of interaction of topoisomerase I with its substrate. Examination of the sequences revealed the presence of a dinucleotide motif CT/G preceding the CTT in all cases (Fig. 2A–F). To gain a better understanding of the sequence requirement for topoisomerase binding and cleavage, we mapped STS from many other plasmids harbouring mycobacterial genomic DNA fragments using the primer extension technique. The results are compiled in Table 2. An analysis of this data reveals certain characteristic features: (i) most of the sites mapped showed sequence similarity at the cleavage position; (ii) a pentanucleotide sequence, CT/GCTT, emerged as the predominant preferred site for *M.smegmatis* topoisomerase I and the cleavage position appeared to be between the two T nucleotides; (iii) in over 97% of the sites mapped, the CTT motif was found, G/T occurred at position –3 (with respect to cleavage) in 86% of cases whereas a C occurred at position –4 in 60% of cases. Sequences upstream to this pentanucleotide element are less well conserved, although there appears to be some preference for C/G residues at positions –5, –8 and –9. A T at position +1 with respect to cleavage is highly conserved. At position +2 (the second nucleotide 5′ of the cleavage site) a high degree of preference for G/C (83%) is seen, while no other sequences downstream of +2 seem to be well conserved. The convenient features of the experimental approach used in identifying the STS include: (i) the sites were mapped on intact plasmids and not on linear DNA, thus the influence of DNA topology (if any) on DNA cleavage could be accounted for; (ii) cloning into a small plasmid such as pUC has the advantages of producing DNA in large quantities and also allowing sequence determination using convenient primers; (iii) since the same set of primers annealing to the vector sequence flanking the different cloned sequences was used, many cleavage sites could be easily determined. It should be noted that the technique is designed to map sites within the primer extension region using a fixed set of primers.

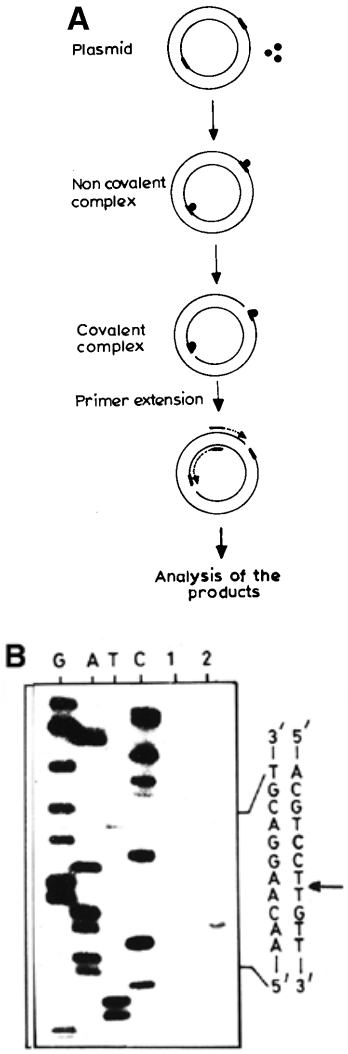

Figure 1. Determination of STS for *M.smegmatis* topoisomerase I. (**A**) Strategy to map the strong topoisomerase sites. STS is depicted in the circular plasmid as filled boxes. Filled circles represent topoisomerase I. Arrows represent extension of annealed primers. The method involves annealing of the same set of primers to the vector sequences flanking the cloned DNA fragment, facilitating both primer extension and sequencing. (**B**) Primer extension reaction. The plasmid Msp4.4 was generated by cloning a 0.9 kb fragment derived from *Pst*I-digested genomic DNA of *M.smegmatis* in pUC19. Lanes 1 and 2 correspond to primer extension reaction products of Msp4.4 untreated and treated with topoisomerase I. Reactions were performed as described in Materials and Methods. Sequencing was performed using the forward primer. Lanes G, A, T and C are the corresponding sequencing of the intact (non-scissile) strand.

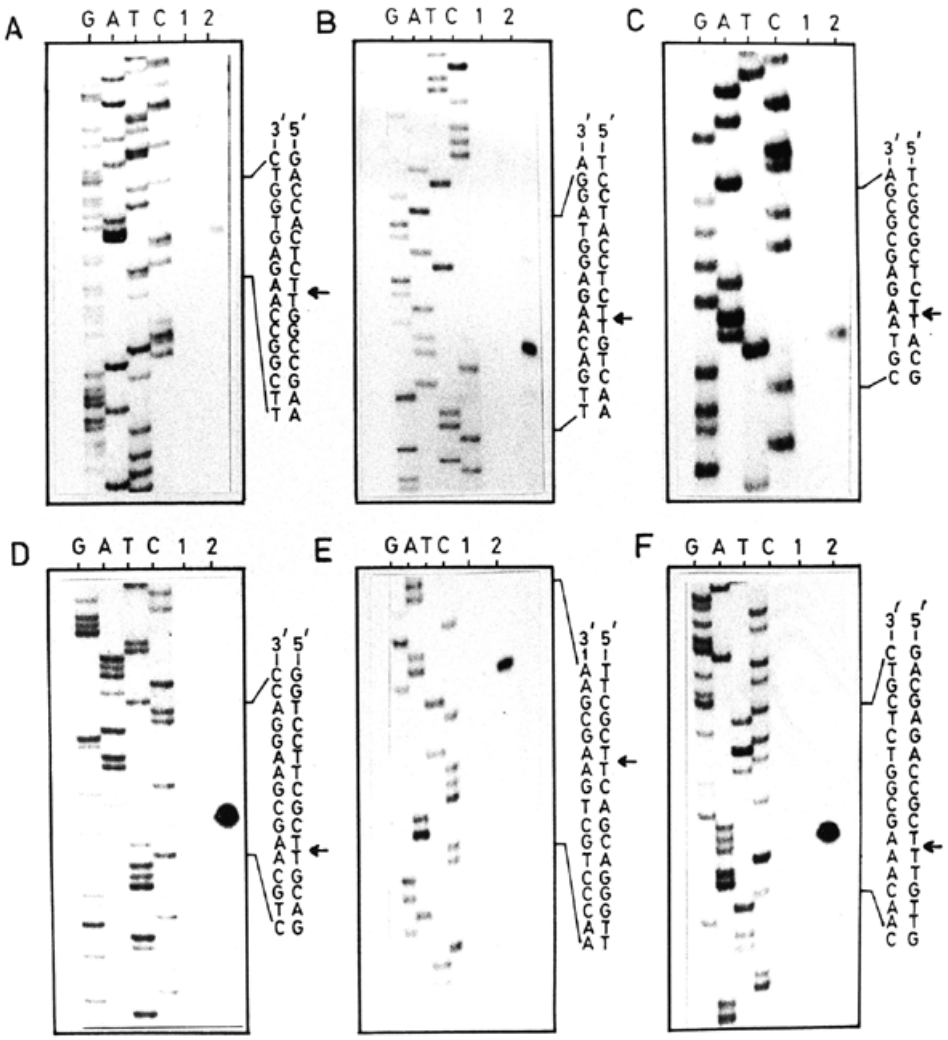

Figure 2. Mapping STS in different mycobacterial sequences. The plasmids containing genomic fragments from mycobacteria were processed as described in Materials and Methods. Lanes 1 and 2 correspond to primer extension reactions of plasmids untreated and treated with topoisomerase I. Lanes G, A, T and C are the corresponding sequencing lanes. (**A**) MX2R carrying a 2.1 kb *Mlu*I–*Xho*I fragment; (**B**) MX1F, a 2.1 kb *Mlu*I–*Xho*I fragment; (**C**) MSa1, a 4.3 kb *Sal*I fragment; (**D**) MSM5, a 1.5 kb *Sma*I–*Mlu*I I fragment; (**E**) MBaP, a 3.2 kb *Bam*HI–*Pst*I fragment; (**F**) sequencing of PMSX1, a 2.4 kb *Sal*I–*Xho*I fragment of *M.smegmatis*.

Topoisomerase I binds specifically to the STS oligonucleotide

A consensus sequence for topoisomerase I action could be derived from the alignment of STS sequences (Table 2). A 20mer oligonucleotide was designed (Con-CCA, Table 1) taking the conserved features into account. An oligonucleotide having the conserved elements should serve as a substrate for topoisomerase I from *M.smegmatis*. We probed binding of the enzyme to the oligonucleotide in an electrophoretic mobility shift assay (EMSA). The topoisomerase formed a complex with the oligonucleotide (Fig. 3A, lane, 2) while no complex was observed with a non-specific 20mer (Fig. 3A, lane, 6). Ability of the protein to bind to the duplex DNA was addressed by annealing the labelled consensus oligomer to the complementary strand and then incubating with topoisomerase I. The enzyme formed a non-covalent complex with the STS-containing oligomer in the double-stranded context (Fig. 3A, lane 4), in agreement with previous results using a 65 bp fragment containing the STS (18). Duplex oligonucleotides without the conserved sequence elements do not serve as substrates for binding (not shown).

Oligonucleotide cleavage by topoisomerase I

In the series of complex steps involved in DNA relaxation, DNA cleavage is the first enzymatic reaction catalysed by topoisomerase I. The ability of the mycobacterial enzyme to cleave DNA substrates having the recognition sequence was addressed. With the same oligonucleotide used in the previous experiment (Con-CCA, Table 1), cleavage at the specific site was observed (Fig. 3B, lane 2). Duplex oligonucleotides (prepared as above) having the STS also served as substrates for topoisomerase I-mediated cleavage (Fig. 3B, lane, 3). An oligonucleotide lacking the conserved element did not support

Table 2. Compilation of the cleavage sites mapped by primer extension

Plasmid	Sequence Mapped		
Msp4.1	CCCTGGCGTTTACCC	AACTT	AATCGCC
Msp4.1	GCGATCGGTGCGGGC	CTCTT	CGCTATT
Msp4.3	CGTGTAGCGTCCCAA	CGCTT	CTCCTAC
Msp4.3	TCGTTCTCCTACCGT	TACTT	GGCGAAG
Msb12.4	CTACGCTGCGGCACG	TCCTT	GTTGTTG
Msp4.4	CGCTAGGCTCGGATG	CTCTA	GGGCTGA
Msp4.5	GCCCTGATCGCGTAG	CGCTT	GCGTCGC
Msp4.5	GATGCCCAGCTCTGC	CGCTT	CAGGTCG
Msp4.6	CTAACCAGTAGCGAA	CTCTT	GTAGCGT
Msb12.1	TTCTAGCTGTCTTGA	GGCTT	CGAACCA
Msb12.1	GACGTGGCGAGTTCG	GCCTT	GCGGTGG
Msb12.1	CTAGCTCGATACCAT	CGTTT	CGTAAGT
Msb12.2	TTGAGAGCTGGCTTG	CGCTT	CTGGACA
Msb12.2	AGCTACATACGCGAC	CTCTT	CGCCCTA
Msb12.3	AACATAAGCCCAGCC	CTCTT	GGAGCGC
Msb12.4	GCTAGGGCATCGAGT	GCCTT	GACCAAC
Msb12.2	TAGCTTGCCACTTGA	TCCTT	GATATCT
Msb12.3	ATAATCCTTGGCAAT	GCCTT	GCAAGTA
Msb12.3	AATTGTTCATAGGCG	TTCTT	AACTAAC
Msb12.7	TTACCGTGCACGCTT	TGCTT	GACTTTG
Msb12.7	ACTGATCGATCGCGG	CTCTT	GCTACAC
PSN1.1	CGACGTCGCCTGGCC	TGCTT	GACAAGA
PSN1.1	AATATACTAGCCTAC	GTCTT	CCATGAG
PSN1.3	TCTGGTATGTCGCCG	TGCTT	GTCGACG
PSN1.4	TGATCGCAAGAGGCG	TTCTT	AACTAAC
PSN1.5	GAGAGATTCCGCGAC	CGCTT	GGCTCGA
PSN2.2	TCATATTGGGCAGGC	CGCTT	AGTGGGA
PSN2.3	TCGACCGCGAACGCG	GTCTT	CACCTCA
PSN2.4	CTGAAGATCTCGCCG	TGCTT	GTCGACG
PSN2.5	CACGCAGCCGTCTGC	TTCTT	CCACCAA
PSN2.5	AACCTAAGCCTGGTT	CGGTT	CCACTAG
MX2R	ATCGAACTCCGACCA	CTCTT	GGCCGAA
MX1F	AGCGGGTACTCCTAC	CTCTT	GTCAAGC
MSC1	CAGATGTCGGTCGCG	CTCTT	ACGCCGC
MSM5	AATTTGTAGGTCCTT	CGCTT	GCAGAAA
MSM6	GCATTAGTAGGTATT	CGCTT	CAGCAGG
PMSX1	CTGCGCCGACGAGAC	CGCTT	TGTTGAC
PMPS1	TGAACAACTTCCAAG	CGCTT	GATCACG
PXX1	ATCGAATCCCGACCA	CTCTT	GGCCGAA
PXX2	GATCATGCAAGGCCC	AGCTT	CTGGAAG
RvB10.3	AGAGCATAGCCACGC	CTCTT	CCTCGCT
RvB10.5	CCTCTATAAGGCCTC	CACTT	CCGATAG
RvB10.8	GTAGTGTACCACCAG	CTCTT	CGCGGGA

CONSENSUS: NNNNC/G_{70}C/G_{74}NN **C/$G_{67}$$C_{60}$T/$G_{83}$$C_{95}$ T $_{100}$$T_{98}$C/$G_{86}$** NNNNNN

cleavage by topoisomerase I, demonstrating a specific sequence requirement for enzyme function (Fig. 3B, lane 4). That cleavage occurred at the sequence CG/TCT↓T was ascertained by analysing the cleavage product by 18% urea–PAGE with appropriate markers (Fig. 3C). These results show that *M.smegmatis* topoisomerase recognises the same sequence in a linear fragment (18), a circular plasmid and synthetic oligonucleotide substrates. Incidentally, topoisomerase III from *E.coli* is shown to cleave between two T bases in the sequence CTT (Table 3; 21), indicating a resemblance between the two enzymes.

The importance of residues 5′ of the cleavage site

The majority of STS compiled in Table 2 show the occurrence of a G/C residue at position +2 with respect to the cleavage site. The nucleotides at positions +3 and +4, however, do not seem to be conserved. Nucleotide conservation at position +2 has not been reported for any other type I topoisomerases so far. To confirm the requirement for specific residues downstream of the cleavage site, we used defined oligonucleotides containing the specific pentanucleotide element but differing in the sequence downstream of the cleavage site (Table 1). The importance of the downstream sequence on binding of topoisomerase I was addressed by EMSA. The mycobacterial topoisomerase forms a complex with oligonucleotides irrespective of the nucleotide sequence downstream of the cleavage site (Fig. 4A). However, oligonucleotides having either a C or G residue at position +2 were very proficient in forming non-covalent complexes (Fig. 4, lanes 2, 4, 8 and 10). The complexes were 4- to 5-fold weaker when the +2 residue was either A or T (Fig. 4A, lanes 6 and 12, and B).

The influence of the +2 residue on cleavage efficiency of the topoisomerase was studied using DNA substrates with A, G, C or T at the +2 position. A marked difference in cleavage efficiency was observed among the substrates. Oligonucleotides harbouring a C or G residue at position +2 were very effectively cleaved. In contrast, oligomers with an A or a T residue at the same position were poor substrates for the mycobacterial topoisomerase (Fig. 4C, lanes 4 and 7, and D). The order of preference for cleavage by the enzyme can be summarised as C = G >> A > T. Residues at positions +3 and +4 do not seem to influence cleavage efficiency to a significant extent. We have not tested the effect of the sequences downstream of +4 as these sequences do not seem to be conserved (Table 2). Based on all the data, the recognition sequence for *M.smegmatis* topoisomerase I is 5′-CG/TCT↓TC/G-3′.

Table 3. Sequence specificity of type I topoisomerases

Enzyme	Source	Sequence	Reference
Topoisomerase I	*Tetrahymena thermophila*	AAAAAA/GAC↓TTAGAG/AAAAA/TA/TA/T	12
Topoisomerase I	Vaccinia virus	C/TCCTT↓	13
Topoisomerase I	Calf thymus/HeLa cell	TATT↓C	30
		ACAT↓C	
		CGCC↓A	
Topoisomerase I	Human	AAAAAGACTT↓AGAAAAATTTTT	24
Topoisomerase III	*Escherichia coli*	GCAACT↓TCG	21
Topoisomerase I	*Escherichia coli*	CXXX↓X	4
Topoisomerase I	*Bacillus subtilis*	A/TCATA/TTAA/TA/TA	27
Topoisomerase I	*Mycobacterium smegmatis*	CT/GCT↓TG/C	This work

The sequences are in the 5′→3′ direction. Cleavage is indicated by downward arrows.

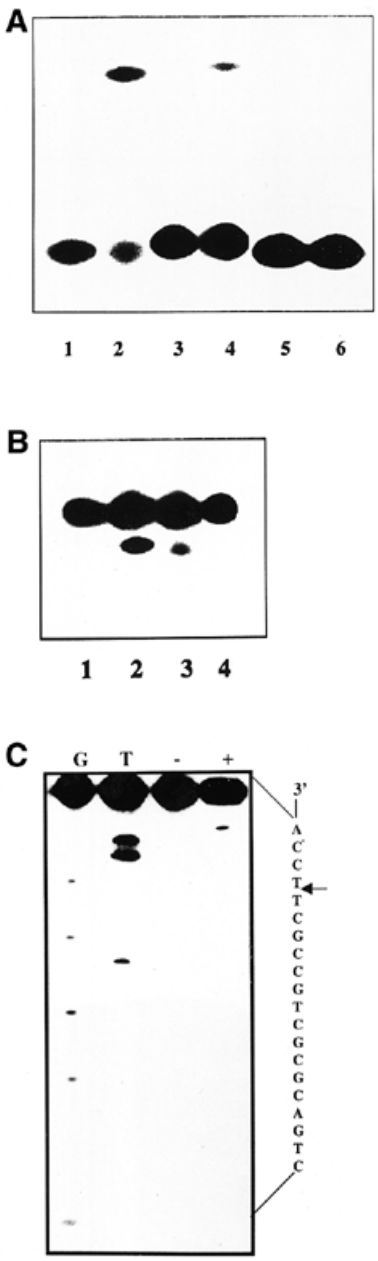

Figure 3. Analysis of consensus site binding and cleavage. (**A**) Binding. Synthetic substrates (100 fmol) containing the conserved pentanucleotide element (CTGACGCGCTGC**CGCTT**CCA, Con-CCA, Table 1) in a single- (lanes 1 and 2) and double-stranded context (lanes 3 and 4) were incubated with 20 U of topoisomerase I at 4°C (lanes 2 and 4). Oligonucleotide lacking the consensus motif (GACTGCGCGACGGCGAAGGT) was incubated in the absence and presence of topoisomerase I (lanes 5 and 6, respectively) and electrophoresed in an 8% non-denaturing PAGE gel at 150 V. (**B**) Cleavage. Lane 1, free lane. Single-stranded consensus oligonucleotide (lane 2), double-stranded consensus oligonucleotide (lane 3) and non-specific single-stranded oligonucleotide (lane 4) were incubated with protein and analysed by 12% urea–PAGE before processing as described in Materials and Methods. (**C**) Mapping of the cleavage site in the consensus oligonucleotide. G- (lane 1) and T-specific (lane 2) reactions were performed as described in Materials and Methods. The oligonucleotide (Con-CCA, CTGACGCGCTGC**CGCTT**CCA) was incubated in the absence (lane 3) or presence of topoisomerase I (lane 4). The differences in mobility of the reaction products of chemical cleavage and of topoisomerase cleavage were taken into consideration while mapping the cleavage site.

DISCUSSION

We have determined the sequence specificity of mycobacterial topoisomerase I on mycobacterial DNA as a prelude to addressing its *in vivo* function and to a study of the mechanistics of the enzyme reaction. The enzyme exhibits remarkable sequence specificity. Synthetic oligonucleotides containing the consensus sequence motif serve as substrates for enzyme binding and catalysis.

Statistical evaluation of the sites mapped from genomic sequences show the conservation of specific residues at the recognition site. Positions +2, +1, –1, –2, –3, –4, –5, –8 and –9 appear to have the most biased nucleotide distribution. A striking feature is that the cleavage almost always occurred at CTT, even though the probability of finding two consecutive T bases at any given position is <0.03 in a 65–67% GC-rich genome. Interestingly, the sequence CTT appears to be conserved in the recognition sequence of many of the type I topoisomerases (Table 3). Topoisomerase III from *E.coli* (22), the only other prokaryotic enzyme displaying some sequence specificity, also harbours the same trinucleotide motif in its cleavage sequence GCAA**CT**↓**T**CG (21). Sequence conservation at the recognition/cleavage site of topoisomerase I suggests that these residues in the scissile strand and/or the complementary non-scissile strand are important for interaction with the enzyme. A support for this premise comes from extensive footprinting and crosslinking studies with vaccinia virus topoisomerase I (23). The protein was shown to establish contact with all three G residues within the pentamer 3′-GGGAA-5′ in the non-scissile strand (17). Base substitution experiments of the T residues in the sequence CCCTT had a drastic effect on cleavage efficiency, underlining their contribution to the protein–DNA interaction (17). Human topoisomerase I was recently shown to contact the T residue at position –1 with respect to the cleavage site (24). Preliminary studies with *M.smegmatis* topoisomerase I seem to further substantiate this view (D.Sikder and V.Nagaraja, unpublished results). The sequence determinants for other topoisomerases I from prokaryotes characterised so far are less stringent than that for the mycobacterial enzyme. The *E.coli* and *Micrococcal luteus* topoisomerases cleave at the sequence CXXX↓ (4,25). In addition, double-stranded DNA is refractory to cleavage in both cases (26). The topoisomerase I from *M.smegmatis*, while displaying recognition/cleavage specificity, forms complexes with single- and double-stranded DNA. In accordance with this difference in interaction with DNA, the mycobacterial enzyme lacks the zinc finger motif found in *E.coli* and other type I enzymes (19,27–29). A stringent sequence requirement would restrict the number of sites at which the mycobacterial protein can bind and carry out catalysis. However, the ability to function on both single-stranded and duplex DNA probably enables the enzyme to act on a larger number of sites *in vivo* which otherwise would have been inaccessible if it were to function only on single-stranded DNA.

Topoisomerases function to modulate torsional stress in DNA. Thus, these proteins should be able to bind to DNA wherever the topology demands it. Consistent with this view, many topoisomerases are not known to display high sequence specificity, with the exceptions of topoisomerases I from vaccinia virus and *M.smegmatis* (13,18,23; this work). In spite of their distant phylogenetic relationship, both vaccinia virus

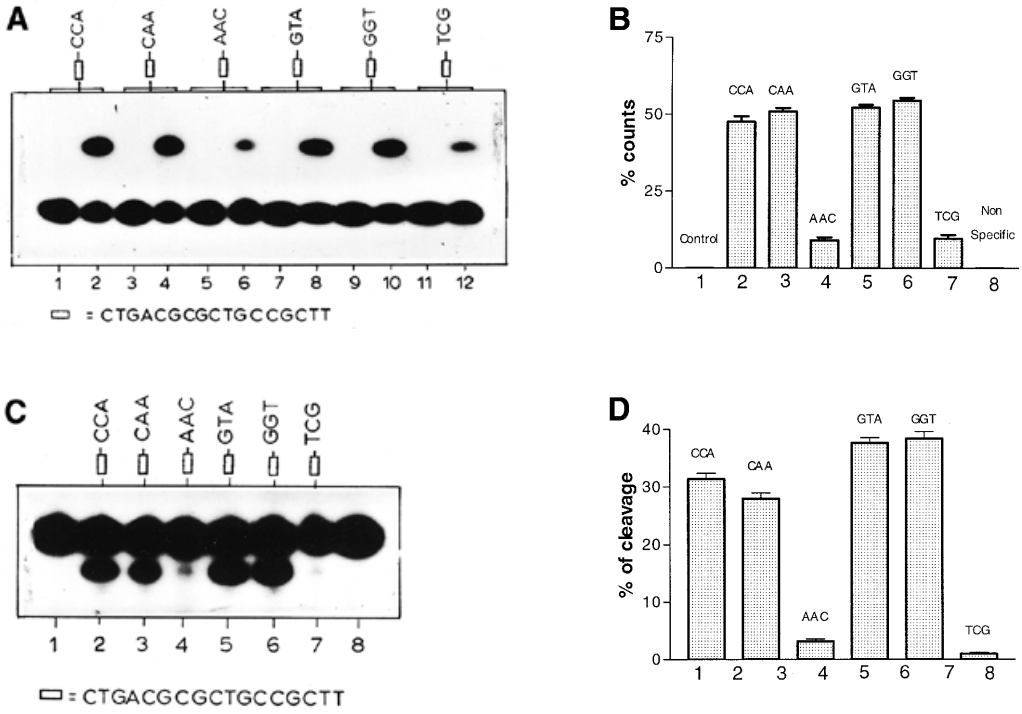

Figure 4. Effect of residues 5′ of the cleavage position on binding and cleavage. Aliquots of 100 fmol of 5′-end-labelled 20mer oligonucleotides, differing only in sequence downstream of CTT, were used as substrates for the gel shift assay. Open rectangles in all the cases represent the sequence 5′-CTGACGCGCTGCCGCTT-3′. The downstream sequences are inscribed above the lanes. (**A**) Effect on DNA binding. The oligonucleotides were incubated with (lanes 2, 4, 6, 8, 10 and 12) or without (lanes 1, 3, 5, 7, 9 and 11) topoisomerase I and electrophoresed in an 8% native PAGE gel to resolve the oligonucleotide–topoisomerase I complex. (**B**) The counts associated with the free DNA and the DNA–protein complexes were quantified using a phosphorimager, expressed as a percentage of the total counts and plotted to obtain the bar diagram. (**C**) Effect on cleavage. Aliquots of 100 fmol of end-labelled oligonucleotides were incubated with topoisomerase I and electrophoresed by 12% urea–PAGE to resolve the cleavage products. Lane 1, CCA oligonucleotide incubated in the absence of protein; lanes 2–7, oligonucleotides differing only in the downstream sequence incubated with the purified protein; lane 8, a non-specific oligonucleotide (GACTGCGCGACGGCGAAGGT) incubated along with the protein. (**D**) Relative efficiency of cleavage.

and *M.smegmatis* topoisomerases I are remarkable in their specificity in recognising very similar STS. Vaccinia virus topoisomerase I binds and cleaves at the sequence C/TCCTT↓, while the mycobacterial topoisomerase I recognises the sequence CG/TCT↓TC/G. However, there are certain contrasting features: (i) the 33 kDa viral protein is the smallest topoisomerase known to date while the 110 kDa mycobacterial topoisomerase I is one of the largest prokaryotic type I enzymes; (ii) vaccinia virus topoisomerase I recognises the consensus sequence only in duplex DNA, whereas the mycobacterial topoisomerase I catalyses cleavage in both the single- and double-stranded contexts; (iii) the +1 and +2 residues next to the cleavage site are not conserved in the recognition sequence of vaccinia virus topoisomerase I, as opposed to the mycobacterial topoisomerase. This might be due to mechanistic differences between the two proteins. The viral topoisomerase I catalyses formation of a 3′-phosphotyrosine covalent adduct with the conserved T residue at the 3′-terminus, while the mycobacterial enzyme forms a 5′-phosphotyrosine covalent intermediate with the conserved T at position +1.

The remarkable sequence specificity exhibited by *M.smegmatis* topoisomerase I provides an opportunity to study the interaction of the enzyme with its target site. The unique ability of the enzyme to recognise both single- and double-stranded DNA having the specific sites raises intriguing possibilities about the interaction of the enzyme with its substrates. Binding and cleavage of synthetic oligonucleotide substrates allows the design of suicide substrates and also assessment of the religation step in the topoisomerase reaction. It would be of particular interest to determine the specific contacts between nucleotide residues in the recognition sequence and amino acid residues in the protein. Since position +1 is highly conserved and position +2 also affects the efficiency of cleavage to a significant extent, the protein very likely contacts these residues. If so, this is in striking contrast to any other type I topoisomerase. The significance of the remarkable sequence specificity of *M.smegmatis* topoisomerase I, a distinctive feature uncommon amongst the prokaryotic type I enzymes, is unclear to us at this stage. The refractivity to cleavage of non-specific sites even with large excess of the enzyme suggests that the enzyme could be functioning at the specific sites *in vivo*. The DNA relaxation reaction catalysed by *M.smegmatis* topoisomerase I is predominantly processive under ionic conditions similar to the intracellular environment (19). Thus it appears to be advantageous to the organism to have a topoisomerase recognising specific sequences and functioning processively at these sites *in vivo*.

ACKNOWLEDGEMENTS

We thank Tisha Bhaduri and T. K. Bagui for initial help in enzyme purification and D. N. Rao for critical reading of the manuscript. Our thanks are due to one of the referees for the corrections and critical comments. The work is supported by a grant from the Department of Biotechnology, Government of India.

REFERENCES

1. Wang,J.C. (1996) *Annu. Rev. Biochem.*, **65**, 635–692.
2. Kirkegaard,K., Pflugfelder,G. and Wang,J.C. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 411–419.
3. Morrison,A. and Cozzarelli,N.R. (1979) *Cell*, **17**, 180–184.
4. Tse-Dinh,Y.-C., Kirkegaard,K. and Wang,J.C. (1980) *J. Biol. Chem.*, **255**, 5560–5570.
5. Edwards,K., Halligan,B., Davis,J., Nivera,N. and Liu,L. (1982) *Nucleic Acids Res.*, **10**, 2565–2576.
6. Bonven,B.J., Gocke,E. and Westergaard,O. (1985) *Cell*, **41**, 541–551.
7. Tse-Dinh,Y.-C., McCarron,B.G.H., Arentzen,R. and Chawdhri,V. (1983) *Nucleic Acids Res.*, **11**, 8691–8701.
8. Liu,L.F., Rowe,T.C., Yang,L., Tewey,K. and Chen,G.L. (1983) *J. Biol. Chem.*, **258**, 15365–15370.
9. Been,M.D., Burgess,R.R. and Champoux,J.J. (1984) *Nucleic Acids Res.*, **12**, 3097–3114.
10. Capranico,G. and Binaschi,M. (1998) *Biochim. Biophys. Acta*, **1400**, 185–194.
11. Borgnetto,M.E., Tinelli,S., Carminati,L. and Capranico,G. (1999) *J. Mol. Biol.*, **285**, 545–554
12. Andersen,A.H., Gocke,E., Bonven,B.J., Nielsen,O.F. and Westergaard,O. (1985) *Nucleic Acids Res.*, **13**, 1543–1557.
13. Shuman,S. and Prescot,J. (1990) *J. Biol. Chem.*, **265**, 17826–17836.
14. Fisher,L.M., Barot,H.A. and Cullen,M.E. (1986) *EMBO J.*, **5**, 1411–1418.
15. Pato,M.L., Howe,M.M. and Higgins,N.P. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8716–8720.
16. Shuman,S. and Moss,B. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 7478–7482.
17. Shuman,S. and Turner,J. (1993) *J. Biol. Chem.*, **268**, 18943–18950.
18. Bhaduri,T., Sikder,D. and Nagaraja,V. (1998) *Nucleic Acids Res.*, **26**, 1668–1674.
19. Bhaduri,T., Baugui,T.K., Sikder,D. and Nagaraja,V. (1998) *J. Biol. Chem.*, **273**, 13925–13932.
20. Madhusudan,K., Ramesh,V. and Nagaraja,V. (1994) *Curr. Sci.*, **66**, 664–667.
21. Srivenugopal,K.S., Lockshon,D. and Morris,D.R. (1984) *Biochemistry*, **23**, 1899–1905.
22. Zhang,H.L. and DiGate,R.J. (1994) *J. Biol. Chem.*, **269**, 9052–9059.
23. Sekiguchi,J. and Shuman,S. (1994) *J. Biol. Chem.*, **269**, 31731–31734.
24. Stewart,L., Redinbo,M.R., Qiu,X., Hol,W.G.J. and Champoux,J.J. (1998) *Science*, **279**, 1534–1541.
25. Dean,F., Krasnow,M.A., Otter,R., Matzuk,M.M., Spengler,S.J. and Cozzarelli,N.R. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 769–773.
26. Depew,R.E., Liu,L.F. and Wang,J.C. (1978) *J. Biol. Chem.*, **253**, 511–518.
27. Miema,R., Haan,G.-J., Venema,G., Bron,S. and Jong,S. (1998) *Nucleic Acids Res.*, **26**, 2366–2373.
28. Tse-Dinh,Y.-C. and Beran-Steed,R.K. (1988) *J. Biol. Chem.*, **263**, 15857–15859.
29. Fouet,A., Sirard,J.C. and Mock,M. (1994) *Mol. Microbiol.*, **11**, 471–479.
30. Edwards,K.A., Halligan,B.D., Davis,J.L., Edwards,K.A. and Liu,L.F. (1985) *J. Biol. Chem.*, **257**, 3995–4000.