Eukaryotic topoisomerase II preferentially cleaves alternating purine-pyrimidine repeats

J.R.Spitzner, I.K.Chung and M.T.Muller*

Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210, USA

Received December 6, 1988; Revised and Accepted August 1, 1989

ABSTRACT

Alternating purine-pyrimidine sequences (RY repeats) demonstrate considerable homology to the consensus sequence for vertebrate topoisomerase II (Spitzner and Muller (1988) Nucleic Acids Res. 16: 1533 – 1556). This is shown below and positions that can match are underscored.

**RYRYRYRYRYRYRYRYRYRY = alternating purine-pyrimidine 18 bp
RNYNNCNGYNGKTNINY = topoisomerase II consensus sequence**

(R is purine, Y is pyrimidine, K is G or T.) Topoisomerase II cleavage reactions were performed (in the absence of inhibitors) on a plasmid containing a 54 base RY repeat and the single strong cleavage site mapped to the RY repeat. Analysis of this DNA on sequencing gels showed that the enzyme cleaved a number of sites, all within the 54 base pair RY repeat. Topoisomerase II also made clustered cleavages within other RY repeats that were examined. Quantitative analysis of homology to the consensus sequence, as measured by the match of a site to a matrix of base proportions from the consensus data base (the matrix mean), showed that both the locations and the frequencies of cleavage sites within RY repeats were proportional to homology scores. However, topoisomerase II cleaved RY repeats preferentially in comparison to non-RY sites with similar homology scores. The activity of the enzyme at RY repeats appears to be proportional to the length of the repeat; additionally, GT, AC and AT repeats were better substrates for cleavage than GC repeats.

INTRODUCTION

DNA topoisomerases transiently break and then reseat the DNA helix, altering DNA linking number and allowing adjustment of topology. Topoisomerases have been implicated in many important cellular processes (see 1,2 for reviews); in particular,

topoisomerase II is a major component of the nuclear matrix/chromosome scaffold (3–5), it is required for completion of mitosis in yeast (6–9) and subject to cell-cycle regulation (10). To investigate further the cellular role of topoisomerase II, we previously characterized catalytic sites of vertebrate topoisomerase II and derived the following consensus sequence:

5' R N Y N C N G Y ^ N G K T N Y N Y 3'

where R is purine, Y is pyrimidine, K is G or T, N is any base, and cleavage occurs at the caret (^) (11). The topoisomerase II consensus sequence is distributed nonrandomly in eukaryotic DNA and is frequently found in or near transcriptional enhancers (11). We also recognized that any sequence containing alternating purines and pyrimidines matches the consensus sequence quite well (see Table 1) and in the present study, we report on the reactivity of topoisomerase II to alternating purine-pyrimidine DNA.

Alternating purine-pyrimidine sequences (which we refer to as RY repeats) have been identified in naturally occurring DNA and have intriguing properties. Certain RY repeats have been found to exist as non-B form DNA (Z-DNA, cruciforms) (12,13) and may have significant biological consequences by affecting protein binding at RY repeats (14) and proximal sequences (15). Long RY repeats are highly recombinagenic in bacterial hosts (14,16–19) and frequently result in deletions. Additionally, lesions in RY repeats are not efficiently processed by DNA repair machinery (20,21). In three human fetal globin genes the second intervening sequence contains large RY repeats (40–60 bp) which can adopt the Z-DNA conformation in vitro and which are hotspots for recombination and gene conversion in vivo (22). Furthermore, there is a long RY repeat about 550 bp upstream of the human β -globin translation start which is also a hotspot for recombination (23), and mutations in this region are conservative, that is, they maintain the RY repeat. Finally, proteins that specifically bind (CG)_n DNA have been identified from *Drosophila*, SV40 minichromosomes, and wheat germ (24–26). We show in this report that topoisomerase II is extremely reactive towards RY repeat sequences; therefore, topoisomerase II is also a eukaryotic DNA binding protein that preferentially recognizes alternating purine-pyrimidine DNA.

* To whom correspondence should be addressed

Table 1. Match of (GC)_n and (GT)_n to the topoisomerase II consensus sequence.

Vertebrate topoisomerase II consensus^a showing critical nucleotides:

[illegible]

A. Match of (GC)_n to the topoisomerase II consensus sequence^b

-10 -8 -5 -2 -1 2 3 4 6 8
 5' G C G C G C G C G C G C G C G C G C G C G C G C
 3' C G C G C G C G C G C G C G C G C G C G C G C G
 8 6 4 3 2 -1 -2 -5 -8 -10

Matrix Mean = 0.309°

B. Match of (GT)_n to the topoisomerase II consensus sequence^b

-10 -8 -5 -2 -1 2 3 4 6 8
 5' G T G T G T G T G T | G T G T G T G T G T G T
 3' C A C A C A C A C A | C A C A C A C A C A
 8 6 4 3 2 -1 -2 -5 -8 -10

Matrix Mean = 0.318°

^a The site of topoisomerase II cleavage is marked at the arrows. "N" indicates that no nucleotide is preferred at the position, R is purine, Y is pyrimidine and K is G or T. The consensus sequence is from Spitzner and Muller (1988).

^b (GC)_n is cleaved weakly by topoisomerase II, whereas (GT)_n is cleaved much more frequently (see Figs. 4 and 5).

• The "Matrix Mean" corresponds to the sum of nucleotide proportions of all non-N positions in the data base divided by 20 (see Table 4 and Methods).

RESULTS

Alternating purine-pyrimidine repeats show homology to the topoisomerase II consensus sequence

The topoisomerase II consensus sequence derived in this laboratory allows for accurate predictions of enzyme cleavage sites on naked DNA and in chromatin (11,27,30). It is clear that the consensus sequence has a degree of alternating purine-pyrimidine (or RY) character, and further that specific examples of RY repeats show considerable homology to the consensus sequence (see Table 1). In fact, because alternating purine-pyrimidine sequences essentially have a repeated sequence every second base, a sufficiently long RY stretch should contain a clustering of topoisomerase II consensus sequence matches. This suggested the potential for a high density of topoisomerase II recognition sites within RY repeats.

The RY repeat is the preferred topoisomerase II cleavage site on the plasmid pRYG

A computer search of the human β -globin gene region revealed a 54 base RY repeat; this RY repeat was of particular interest

to us because it was shown to be a hotspot for recombination *in vivo* (23). The RY repeat was subcloned into pUC19 to produce pRYG (see Methods and Table 2 below for sequence). Topoisomerase II cleavage reactions were performed on either supercoiled or relaxed pRYG using the purified chicken enzyme (30), and the DNA was repurified and digested with appropriate restriction enzymes. Cleavage sites were identified by indirect end labeling using two different probes. In all cases, a single strong cleavage band (arrows, Figure 1) was detected at the 54 base RY repeat. The intensity of the band was roughly proportional to the amount of enzyme added. The topology of the plasmid (supercoiled or relaxed) did not seem to affect the results (however, the presence of ATP would allow relaxation of the supercoiled substrate). In pRYG, the RY repeat was clearly the preferred topoisomerase II site compared to any of the other potential 3000 or so sites. The results of this experiment were confirmed using different restriction enzymes and different probes to verify that the cleavage site was correctly localized at the RY repeat (data not shown). The experiment shown in Fig. 1 was also repeated in the presence of the topoisomerase II inhibitors

Table 2. Recognition sites and homology scores in pRYG

Recognition sites defined by cleavage and primer extension ^a	
Flanking sequence not cleaved by topo II	
TGATTTGGATTAAAA	CCTTCTGGTAAGAAAAGAAA
ACTAAACCTAATTTT	GGAAGACCATTCTTTTCTTTT
<div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> </div>	
ATATATATATATATATATATGTATATATACACACATACATA	
TATATATATATATATATACATATATATGTGTGTATGTAT	
<div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> </div>	
TACATATATATGCATTCATTTGTTGTTGTTTTCTTAAT	
ATGTATATATACGTAAGTAAACAACAACAAAAGAATTA	
<div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> <div>am</div> </div>	

Homology scores in cleavage and non-cleavage (flanking) sequences of pRYG^b

Flanking sequence not cleaved by topo II (from above)	
Matrix	
Mean	
values:	.259 .220 .248 .279 .168 .266 .245 .159 .242 .320 .235 .209 .243 .262 .266 .240 .226 .248 .275 .255
	C C T T C T G G T A A G A A A A G A A A
pRYG sequences cleaved by topo II (from above)	
Matrix	
Mean	
values:	.182 .302 .186 .359 .189 .347 .182 .285 .187 .366 .193 .330 .175 .323 .197 .344 .191 .307 .216 .282
	C A T A C A T A T A C A T A T A T A T G

^aThe sequence for the RY repeat of pRYG is given. Sites labeled "am" are cleavages mapped in the presence of *m*-AMSA; the boxed sites are those mapped in the absence of drug (from Figs. 2A and 2B). The flanking sequence in brackets represents a non-RY stretch that lacks topo II sites

^bThe top strand of the non-RY sequence (in brackets in the top sequence) and an RY sequence containing a number of topo II cleavage sites were analyzed for matrix mean values. Cleavage sites are marked by vertical lines.

m-AMSA and VM26; with either drug, the major cleavage site was localized to the RY repeat and, although other sites were amplified by the drugs, none were nearly as strong as the RY repeat site (data not shown).

Sequencing topoisomerase II sites on the RY repeat of pRYG

Having determined that topoisomerase II is very reactive towards an RY region in pRYG, we mapped the cleavage sites within

the RY repeat on sequencing gels. Cleavage reactions were performed on a DNA fragment containing the RY repeat plus flanking sequences from pRYG. In the absence of inhibitors (Fig. 2A, lane 2), essentially all cleavages were clustered around the RY repeat. The same is true when *m*-AMSA was included in the reaction (lane 3) although additional sites were also revealed. The sites of cleavage within the RY repeat are indicated in Table 2 along with the homology to the topoisomerase II consensus

Figure 1. Topoisomerase II cleavages on the plasmid pRYG. Cleavage reactions were performed on pRYG using 0.8 μ g DNA and 10 units topoisomerase II in the absence of inhibitors. Following termination of the reactions, the DNA was purified, digested with Ava II, and electrophoresed on a native 1.6% agarose gel. The DNA was transferred to nitrocellulose and was probed with a 181 base Ava II-Dde I fragment (lanes 1–4) or a 137 base Ava II-Dde I fragment (lanes 5–8). In odd numbered lanes the input plasmid was supercoiled and in even numbered lanes the input plasmid was relaxed. Lanes 1, 2, 5 and 6 received 10 units of topoisomerase II, while lanes 3, 4, 7 and 8 received 20 units. The arrows point to the strongest cleavage site, which corresponds to the location of the RY repeat for either probe. Note that the cleavage band in lane 5 appears weaker than others; this was not a reproducible result. The gel was calibrated with a series of restriction fragment markers (not shown).

sequence. It is apparent that cleavages within the RY repeat were more frequent than at other sites (such as those indicated with arrows, Fig. 2A) within the fragment. While the sites and strengths of VM26 enhanced cleavages were not identical to those with the enzyme alone or with *m*-AMSA added (the latter two showed the same strong sites), the trend was generally the same: the enzyme made an array of cleavages within the RY repeat.

Cleavages were also mapped on the opposite strand of the RY repeat of pRYG. As indicated in Table 2, the cleavages on this bottom strand were very similar to the pattern observed for the top strand, suggesting that most of the cleavages were indeed double strand events with a 4 base 5' overhang, as expected for topoisomerase II. There were several exceptions where unequal cleavage frequencies were detected between the two strands at a site as reported previously (30). The data show that the strong cleavage band detected by indirect end labeling was in reality a cluster of proximal cleavage sites. This conclusion was further supported by data from mapping the topoisomerase II sites on the whole plasmid by primer extension: sites mapped by primer extension of an end labeled oligonucleotide coincided with sites observed in Fig. 2 (data not shown).

Topoisomerase II reactivity toward other RY repeats

To test the generality of our findings, namely clustered topoisomerase II cleavage sites within RY repeats, we performed

cleavage reactions on other DNA fragments containing RY repeats. Fig. 3 shows the reactivity of topoisomerase II at an RY repeat found in the regulatory region of SV40; the 8 base repeat is AT(GT)₃. In general, the RY associated sites were similar for reactions with and without drugs. Although other sites on the fragment were also cleaved, the RY associated sites represented the only clustering of topoisomerase II sites in the fragment. Note that for some of the RY sites the cleavage point was not in the RY repeat; however, the sequence covered by the enzyme extended well into the RY repeat.

Topoisomerase II cleavages were performed on a fragment from pCGTA-C (28), which contains the RY repeat (CG)₆TA(CG)₆ (Fig. 4). The enzyme was clearly less reactive towards this sequence than to the previous RY repeats and this was predicted by the lower topoisomerase II matrix mean score of a GC repeat compared to the matrix means for GT or AT repeats (see Tables 1 and 4). However, the cleavage sites on the pCGTA-C fragment were still within the RY repeat, supporting the hypothesis that alternating purine-pyrimidine sequences are preferentially recognized by topoisomerase II. The results of this experiment and others (not shown) revealed that topoisomerase II was generally much less reactive to GC repeats compared to other RY repeats, especially in the absence of the drugs (see Fig. 4, lane 3). In the presence of VM26 and *m*-AMSA, additional, stronger cleavage sites within the RY repeat were observed (lanes 1, 2).

As noted above, (GT)_n repeats display a higher degree of homology to the topoisomerase II consensus sequence, with a maximum matrix mean of 0.318. Cleavages were performed on a fragment containing a 25 base (GT)_n repeat (TGC(GT)₁₁) from the fifth intron of the mouse thymidylate synthetase gene (29); the results are shown in Figure 5 (panel A shows cleavages on the (GT)_n strand and panel B shows cleavages on the opposite, (AC)_n strand). Cleavage sites on the actual sequence are indicated in Table 3. Again, the RY repeat contained a cluster of strong topoisomerase II cleavage sites, and these sites were essentially identical whether no drug, *m*-AMSA, or VM26 was included in the reaction. Comparison of panels A and B shows that all of the RY repeat cleavage sites appeared symmetric (i.e. relatively equal frequencies of cleavage on both strands of a site), as shown in Table 3. Consensus homology scores across the RY repeat reveal that the cleavage sites possessed the best matches (refer to Table 3). The most frequently cleaved site scored a matrix mean of 0.344 (which is fortuitously higher than a pure GT repeat due to non-GT flanking sequences). However, the cleavages in the RY repeat were, in general, stronger than the cleavages at some of the other sites in the fragment that had comparable scores (not shown); therefore, it appears that while the hierarchy of cleavage strengths within RY repeats is proportional to the matches to the topoisomerase II consensus sequence, the enzyme may be more reactive at sites containing alternating purine-pyrimidine sequences as compared to sites lacking them.

Topoisomerase II cleavages were also performed on a DNA fragment containing the HSV-1 origin of replication (which contains an (AT)₆ repeat); the results are shown in Figure 6. The only strong cleavages by the enzyme (plus or minus drugs) were within the RY repeat. The cleavage sites were nearly the same whether or not inhibitors were included, the only differences being at the 3' end of the fragment (top of the gel). The intense cleavage bands observed in the RY repeat were predicted to be recognition sites by matches of those sites to the topoisomerase

Figure 2. Topoisomerase II cleavages on human β -globin sequence containing 54 base RY repeat.

Cleavage reactions were performed (as described in Methods) on a 245 bp AccI-SphI fragment (labeled at the AccI site in Panel A and at the SphI site in Panel B) from the upstream region of pRYG (see Methods). Reactions contained 6 units topoisomerase II in the presence or absence of drugs (*m*-AMSA at 50 μ g/ml and VM26 at 0.5 mg/ml), as shown. Chemical sequencing markers are indicated. To facilitate comparison of weak and strong cleavage bands, the amount of isotope loaded in some lanes was adjusted: 10,000 cpm of [32 P]-DNA were loaded onto all lanes except lane 4 of Panel B, which received 20,000 cpm. Brackets delimit the RY repeat.

II recognition matrix (matrix means of 0.311 and 0.325), but these were not the only high scoring sites on the fragment, again suggesting that the enzyme preferentially cleaves RY repeats. Topoisomerase II appears to cleave (AT)_n sequences better than (GC)_n sequences despite similar homology scores, implying that recognition of RY repeats may be more complex than the sequence alone would suggest. In general, topoisomerase II has been found to display extreme reactivity towards certain

alternating purine-pyrimidine sequences in a fashion not entirely predicted by the consensus sequence. The human enzyme has been found to show a specificity for RY repeats that is essentially identical to the chicken enzyme (data not shown). Unexpected behavior of the enzyme towards unusual DNA sequences has been observed before; for example, strong cleavages were found on bent DNA sequences despite only moderate matches to the consensus sequence (unpublished results).

Figure 3. Topoisomerase II cleavages on an SV40 fragment containing an RY repeat. Cleavage reactions were performed on an NcoI-SphI fragment from the control region of SV40 in the absence or presence of inhibitors (concentrations as above). Cleavage sites around the RY repeat are indicated by the bracket.

DISCUSSION

RY repeats are homologous to the topoisomerase II consensus sequence

We have shown that topoisomerase II is very reactive towards alternating purine-pyrimidine sequences. This was anticipated in part, because of homology between RY repeats and the

Figure 4. Topoisomerase II cleavages on pCGTA-C. Cleavage reactions were performed on a HindIII-PvuII fragment of pCGTA-C labeled at the HindIII site, and inhibitors were included where indicated. In pCGTA-C the sequence is (CG)₆TA(CG)₆; however, due to band compression in the chemical sequencing marker lanes, not all CG dinucleotides were resolved.

topoisomerase II consensus sequence. For example, RY repeats of sufficient length (10 bases 5' of the cleavage site on both strands) match a minimum of 4 consensus positions per strand (the -10, -1, +6, and +8 positions will always be matched) and a maximum of 8 consensus positions per strand (all but the -8 and +2 positions). Therefore, many sequences within or proximal to RY repeats will have the necessary homology to the consensus sequence information to be cleaved by the enzyme (a

Figure 5. Topoisomerase II cleavages on GT and AC repeats.

Cleavage reactions were performed on a 200 base pair EcoRI-AccI fragment from the fifth intron of the mouse thymidine synthetase gene (29). Panel A shows cleavages on the GT repeat, using a 5' end label at the EcoRI site, while in Panel B the AccI site was labeled to examine cleavages on the AC repeat. VM26 and *m*-AMSA were included where indicated, at concentrations given above. In Panel A, twice as much isotope was loaded into the lane with enzyme and no drugs as into the other lanes. The cluster of RY repeat cleavage sites is indicated with brackets. In Panel B, there is a separation of the gel (just above the RY repeat); however, nothing was deleted from the affected lanes.

matrix mean of at least 0.28; see Materials and Methods for a description of this measure). Examples of consensus sequence matches for $(GC)_n$ and $(GT)_n$ are presented in Table 1. These matches will occur after every pyrimidine of an RY repeat. Most alternating purine-pyrimidine sequences are not homogeneous, so the homology scores will not be the same at every position.

Furthermore, the sequences flanking the repeats will contribute to the diversity of homology scores of sites that overlap and extend into RY repeats. Previous experiments examining topoisomerase II cleavage at clustered cleavage sites showed that sites with a high degree of homology to the consensus sequence tended to suppress cleavage at nearby sites with lower homology

Table 3. Recognition sites and homology scores in the GT repeat^a

	W		W		S		S		S		M		M		S		M	
Matrix	.299	.313	.344	.313	.318	.318	.318	.285	.315	.285	.301							
Mean:	.199	.184	.187	.187	.182	.187	.172	.183	.194	.209	.199	.201						
	G C	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T	G T
	C G	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A
				W	W	S	S	S	M	M							S	

^aThe cleavage data are from Fig. 5A and 5B. Cleavages are marked by vertical lines and are presented as Strong (S), Medium (M) or Weak (W) based upon band intensities relative to sequencing markers. The average matrix mean is 0.25.

scores (Spitzner and Muller, submitted). For these reasons, it should be expected that topoisomerase II cleavage sites would be found at those sites with the greatest match to the consensus sequence; in general, this was what we found.

Analysis of cleavage sites

Of the RY repeats tested, GC repeats were by far the least reactive. GC repeats, which predictably form a more stable duplex than the other repeats, display homology scores that are lower than the other pure repeats (like GT or AT). The GC repeat still meets the matrix mean threshold for cleavage, yet on pCGTA-C (and also on another, longer GC repeat for which data were not shown), cleavage was minimal and the single site detected in the absence of inhibitors was not at the site showing the best homology to the consensus sequence. However, other RY repeats tested did show cleavage at the sites with the best consensus sequence scores. Pure GT repeats (seen as AC on the other strand) score a matrix mean of 0.318 (Table 1) and the GT associated sites with this score were cleaved strongly or moderately (see also Fig. 5 and Table 3). The strongest RY repeat site in Fig. 5 had a matrix mean of 0.344 due to the sequences just 5' of the GT repeat; clearly, the enzyme capitalized on this site of high homology to make it the most frequently cleaved site in the RY repeat. Likewise, the two strongest cleavage sites on the AT repeat from the HSV-1 origin of replication (Fig. 6) had the two highest matrix means (0.311 and 0.325) around the RY repeat. Thus the matrix mean appears to make accurate predictions as to both the locations and intensities of topoisomerase II cleavage sites within RY repeats, as the method was previously shown to do for non-RY sites (11) (Spitzner and Muller, submitted), although there is clearly a bias favoring the recognition of RY repeats as compared to non-RY sites. Thus, while the relative frequencies of cleavage at sites in each class (RY, non-RY) are proportional to the homology scores, RY repeat sites were cleaved more frequently than non-RY sites of equal, and even higher, consensus sequence scores. One possibility is that alternating purine-pyrimidine sequences attract processive topoisomerase II activity or possibly stalling of the enzyme at the array of sites which are above threshold homology, thereby making it unfavorable for the enzyme to dissociate from the RY repeats. This effect is less apparent on shorter RY repeats as attested by the analysis of cleavage sites on the SV40 DNA fragment containing only an 8 base RY repeat (see Fig. 4). Cleavage sites were found upstream and downstream of the RY repeat, and the strongest sites were just 5' of the RY repeat (although the RY repeat overlaps the ends of these sites). The

frequencies of cleavage were strictly proportional to the homology scores (not shown), and there did not appear to be a significant bias towards cleavage of RY repeats in this particular case.

Analysis of topoisomerase II cleavage of pRYG plasmid

Additional evidence for an exceptional affinity of topoisomerase II towards alternating RY sequences was demonstrated by cleavages on the RY repeat of pRYG. This RY repeat, a mixture of AT and GT (or AC) repeats, was the longest examined, 54 bases of continuous RY character. The strong topoisomerase II cleavages (in the absence of inhibitors) were found only within the RY repeat. There were other, non-RY sites that had above threshold homology scores but were not cleaved. The two strongest sites in Fig. 2 had the highest scores on the RY repeat, matrix means of 0.359 and 0.366 (the mixed RY sequence had sites with scores greater than a pure AT or GT repeat could have). Again, the enzyme preferentially cleaved the RY repeat sites with the greatest match to the consensus sequence.

Comparison of sites cleaved in various RY elements (see Tables 2 and 3) reveals that most of the topoisomerase II sites in RY repeats appeared symmetric, that is, they had relatively equal frequencies of cleavage on both strands of each site. This suggests that the most of the cleavage sites were trapped as double strand events (although examples of single strand cleavage intermediates can occasionally be detected at RY and non-RY sites [30]). Since the majority of cleavages at RY sites display a uniform frequency on both strands of a matched cleavage site (i.e., single strand events separated by 4 bp), we expect that sites identified on native agarose gels, would be the same sites as determined on sequencing gels. Indirect end labeling does not have the resolution of sequencing gels, thus, sites that are close together would appear as a single band at lower resolution. When we mapped the topoisomerase II sites on the whole circular plasmid pRYG, we observed only a single, very strong cleavage site (without use of inhibitors) at the RY repeat (see Fig. 1). Similar results were seen when the analysis was repeated using alkaline agarose gels (data not shown). This single strong band is composed of clustered proximal cleavage sites observed on sequencing gels with end labeled DNA (Fig. 2). The result was confirmed by primer extension experiments to map cleavage sites in the whole plasmid (supercoiled and relaxed) to single base resolution. The data in Fig. 1 suggest that the RY repeat in pRYG is recognized by the enzyme preferentially as compared to the other approximately 3000 potential sites on the plasmid. Clearly, this particular alternating purine-pyrimidine sequence is a very reactive site. Additional data (not shown) suggest that the affinity

Figure 6. Topoisomerase II cleavages of an AT repeat from the HSV-1 origin of replication.

Cleavage reactions were performed in the presence or absence of inhibitors on a 206 base pair *NarI*-*SmaI* fragment from the HSV-1 origin of replication (34) labeled at the *NarI* end. Chemical sequencing markers are indicated and the (AT)₆ repeat is bracketed.

of the enzyme for RY repeats is proportional to the length of the repeats; as noted above with SV40 (Fig. 3), all tested plasmids containing RY repeats were cleaved within the repeats, yet cleavages in shorter repeats were not as strong and were not the exclusive sites when compared to all sites within the plasmid domain.

The cleavage sites in supercoiled and relaxed pRYG were not significantly different either in location or intensity of cleavage. Interpretation of such an experiment is complicated because under the conditions optimal for cleavage, the enzyme will relax the supercoiled substrate and any difference between relaxed and supercoiled DNA is quickly eliminated as DNA linking numbers reach equilibrium. Even in the absence of ATP (in which case linking numbers cannot re-equilibrate) there were no detectable differences in cleavage patterns at RY sites in supercoiled or relaxed DNA (data not shown). These findings were surprising given the fact that RY repeats can adopt non-B DNA structures under conditions of superhelical tension. For example, the RY repeat in pCGTA-C adopts a Z DNA conformation when supercoiled (28). We have also found that pRYG adopts an unusual conformation when supercoiled (Chung and Muller, unpublished data) but because of its sequence (mostly AT), a cruciform structure may be more likely than Z DNA (13). Additionally, formation of non-covalent topoisomerase II/DNA complexes are favored on supercoiled DNA (35) and eukaryotic topoisomerase I also is significantly more reactive toward negatively supercoiled DNA (36). There are a number of possibilities as to why relaxed and supercoiled pRYG gave the same cleavage results. The enzyme might cleave an altered DNA conformation as efficiently as B DNA. Topoisomerase II binding might alter the conformation of an RY repeat to a B DNA form (either locally or from a distance). Since DNA conformation is dynamic, the enzyme may simply bind sequences when the equilibrium shifts towards a B DNA form. Additional studies are in progress to address this problem.

Effects of topoisomerase II inhibitors on cleavage within RY repeats

Drugs such as VM26 and *m*-AMSA are often used to stimulate cleavage events by topoisomerase II, particularly for in vivo experiments (27,32,33). A major concern about use of these drugs is whether they alter the specificity of the enzyme. We addressed this question previously (11) and found that 75% of all sites identified in the presence of *m*-AMSA or VM26 were also topoisomerase II sites in the absence of drugs. Additionally, we have found that sites identified in the absence of inhibitors are indistinguishable from sites identified in the presence of *m*-AMSA in terms of their match to the same consensus sequence (Spitzner and Muller, submitted). Thus, in most cases, the inclusion of *m*-AMSA does not alter the specificity of topoisomerase II at strong cleavage sites and similar results have been reported by other investigators (35). We addressed this question for specific examples by analyzing all of the RY repeat cleavages both in the absence and presence of inhibitors. On a case by case basis one finds that both *m*-AMSA and VM26 may modify the site locations or, more often, the frequencies of topoisomerase II cleavage on a given DNA fragment, and the effect is more pronounced for VM26 than for *m*-AMSA. However, while the drugs caused some variations in cleavage by the enzyme, neither drug affected the general trend we are reporting in this paper, namely, the significant reactivity of topoisomerase II at alternating purine-pyrimidine sequences.

In vivo considerations

RY repeats have been shown to be highly recombinagenic in vivo both in prokaryotes (14,16–19) and in eukaryotes (22,23). Additionally, proteins from several eukaryotic systems have been identified which specifically bind certain RY repeats that can

Table 4. Matrix Mean Calculations on (AT)_n

The Topoisomerase II Consensus Matrix: ^a																		
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
A	.423		.141			.141			.183	.056		.282	.085	.127		.155		.070
C	.141		.352			.521			.141	.366		.113	.113	.056		.282		.352
G	.268		.155			.141			.521	.085		.535	.451	.042		.113		.113
T	.169		.352			.197			.155	.493		.070	.352	.775		.451		.465
5' R	N	Y	N	N	C	N	N	G	Y		N	G	K	T	N	Y	N	Y 3'

The Matrix scores at consensus positions in a 24 bp AT repeat: ^b																		
	.423	.141		.197		.183	.493		.07	.085	.775		.451	.465				
5' A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A
	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A
				.465		.451	.775	.085	.07	.493	.183		.197	.141	.423			

^aThe base proportions shown in the matrix are derived from the data base published by Spitzner and Muller (1988). Positions marked with a broken vertical line are non-consensus bases.

^bThe consensus matrix was applied to the 24 bp AT repeat containing a single topoisomerase II cleavage site marked by arrows on the top and bottom strands. The base proportions shown above each consensus position were summed and divided by 20 to calculate the matrix mean. (The matrix mean is $6.566/20 = 0.328$.)

adopt a Z DNA structure in vitro (24–26). We have shown that eukaryotic topoisomerase II also has a high affinity for RY repeats and specifically the enzyme displayed remarkable activity toward the 54 bp RY repeat in the human globin gene region that has been associated with a hotspot for recombination in vivo (23). The nature of the enzymatic activity of topoisomerase II (cleaving and ligating DNA), suggests a role in the recombination events common to RY repeats if the enzyme can perform transesterification reactions. Studies that address this topic are in progress.

MATERIALS AND METHODS

Materials Restriction enzymes were purchased from Bethesda Research Laboratories, T4 polynucleotide kinase from United States Biochemical Corp., and gamma-[³²P]-ATP from ICN. The ATP analogue AMP-PNP (5'-Adenylylimidodiphosphate) was purchased from Sigma. The topoisomerase II inhibitors *m*-AMSA (4'-(9-acridinylamino)methanesulfon-*m*-anisidine) and VM26 (teniposide) were provided by the National Cancer Institute divisions of synthetic and natural products, respectively. Plasmid DNA isolation, Southern blotting, indirect end-labeling, and 5' end-labeling were performed as described previously (27).

Plasmids pRYG is a pUC19 derivative in which the 206 bp SphI-NarI fragment was replaced with the 245 bp SphI-AccI fragment upstream of the human β -globin promoter; this insert contains a 54 bp alternating purine-pyrimidine repeat (see sequence in Table 2). The Z DNA forming plasmid pCGTA-C is described elsewhere (28).

Topoisomerase II Cleavage Reactions Cleavage reactions (final volume of 20 μ l) contained cleavage buffer (30 mM Tris-HCl (pH 7.6), 60 mM KCl, 8 mM MgCl₂, 15 mM 2-mercaptoethanol, 3 mM ATP, 30 μ g/ml BSA), and the specified end-labeled DNA fragment (usually $1-3 \times 10^4$ DPM or about 1–5 ng). Topoisomerase II inhibitory drugs were added as specified in each experiment and the reactions were initiated by addition of the indicated number of units of purified chicken topoisomerase II (purification method and unit definition in ref. 30), incubated 10 to 30 min at 30° C and terminated by addition of SDS to 1% followed by incubation with 50 μ g/ml proteinase K at 56° C for 30 min. Samples were phenol extracted, ethanol precipitated, vacuum dried, and resuspended in sequencing gel loading buffer prior to loading onto 8% or 12% sequencing gels.

Maxam and Gilbert chemical ladders were run in adjacent lanes as markers (31).

Homology to the topoisomerase II consensus sequence: the matrix mean When the sequences from 71 strong topoisomerase II sites (obtained in the presence of *m*-AMSA) were analyzed (11), we determined which positions relative to the cleavage site contained non-random base proportions; from these data we derived the consensus sequence shown in Table 1. Consensus sequences are easy to use, but they might not be appropriate for representing the DNA binding sites of proteins because they do not convey all of the information present in the data base from which they are derived. We found that topoisomerase II cleavage sites were predicted more accurately by making use of the matrix of base proportions of the consensus positions (see Table 4) than by using conventional consensus sequence comparisons (Spitzner and Muller, submitted). Additionally, we found that a new set of strong cleavage sites identified without the use of inhibitors showed virtually identical matches to the consensus sequence information matrix as the original data set used to derive the consensus sequence (Spitzner and Muller, submitted); therefore, it was appropriate to use this information to analyze any topoisomerase II sites. The single most valuable method we found is called the matrix mean; this method yielded the greatest discrimination between cleavage sites and nonsites, and also showed statistically significant differences in scores for different strengths of cleavage (Spitzner and Muller, submitted). Application of the matrix mean procedure is indicated below.

Sites are analyzed by positioning the base proportion matrix over the potential topoisomerase II cleavage site, as shown in Table 4; the enzyme cut site is between positions -1 and +1, with minus positions 5' and plus positions 3' of the cut site. First, the (arbitrarily chosen) 'top strand' is analyzed. The base at the -10 position (10 bases 5' of the cleavage site) is examined and given the proportion value of that base in the matrix; in the (AT)_n sequence in Table 4, the -10 position is an A, which scores 0.423. Next, the -9 position contained (in the original data base) essentially a random distribution of nucleotides, so it is ignored (no score). The -8 position is an A and this matches 0.141 in the matrix. The -7 and -6 positions are ignored and the -5 T is given the value of the -5 T in the matrix, 0.197. After the matrix values have been determined for each of the 10 nonzero positions in the matrix, the same operation is performed for the bottom strand of the cleavage site, 4 bases 5'. The same base proportion matrix is used and is aligned 5' to 3' along the bottom strand. The matrix values are determined for each of the 10 nonzero positions as above; in this example (Table 4), the top and bottom strands are identical, so the matrix values are also the same. Once all 20 values have been determined they are summed (in this case to 6.566) and then divided by 20 to yield the average match to the base proportion matrix (in this case 0.328), which is called the matrix mean. The average matrix mean for a nonsite is 0.25, as would be expected for a random sequence of nucleotides. Typically, sites with a matrix mean of at least 0.28 are candidates for topoisomerase II cleavage. The strong sites used to derive the matrix averaged a matrix mean of 0.315, while an independent sample of strong sites obtained in the absence of inhibitors averaged 0.314. Given that these values are practically identical, it seems valid to apply the matrix mean procedure to analyze topoisomerase II sites in general. Matrix means were calculated using the EDEN Genesys software program (TEAM Associates, 487 Catawba Ct., Westerville, Ohio 43081).

ACKNOWLEDGEMENTS

This work was supported by Public Health Service Grant GM31640 from the NIH. The authors are grateful to the following individuals who provided plasmids: L. Johnson (mouse thymidylate synthetase intron 5 clone), E. Vanin (human globin gene region), R.R. Sinden (pCGTA-C). We thank R.R. Sinden for critical reading of the paper and for helpful suggestions.

Abbreviations: SDS, sodium dodecyl sulfate; *m*-AMSA, 4'-(9-acridinylamino)methanesulfon-*m*-anisidide; RY, alternating purine pyrimidine sequences in the form of d(RY)_n-d(RY)_n.

REFERENCES

1. Vosberg, H.P. (1985) in Current Topics in Microbiology and Immunology Vol. 114, 19-101, Springer Verlag, NY
2. Wang, J.C. (1985) Annu. Rev. Biochem. **54**, 669-697
3. Earnshaw, W.C. and Heck, M.S. (1985) J. Cell. Biol. **100**, 1716-1725
4. Berrios, M., Osherooff, N., and Fisher, P.A. (1985) Proc. Natl. Acad. Sci. USA **82**, 4142-4146
5. Gasser, S.M., Laroche, T., Falquet, J., Tour, E.B., and Laemmli, U.K. (1986) J. Mol. Biol. **188**, 613-629
6. Uemura, T. and Yanagida, M. (1986) EMBO J. **5**, 1003-1010
7. Holm, C., Goto, T., Wang, J.C., and Botstein, D. (1985) Cell **41**, 553-563
8. Uemura, T. and Yanagida, M. (1984) EMBO J. **3**, 1737-1744
9. DiNardo, S., Voelkel, K., and Sternglanz, R. (1984) Proc. Natl. Acad. Sci. USA **81**, 2616-2620
10. Heck, M.S., Hittelman, W.N., and Earnshaw, W.C. (1988) Proc. Natl. Acad. Sci. USA **85**, 1086-1090
11. Spitzner, J.R. and Muller, M.T. (1988) Nucleic Acids Res. **16**, 5533-5556
12. Wang, A.H.-J., Quigley, G.J., Crawford, J.L., van Boom, J.H., van der Marel, G., and Rich, A. (1979) Nature **282**, 680-686
13. Haniford, D.B. and Pulleyblank, D.E. (1985) Nucleic Acids Res. **13**, 4343-4363.
14. Jaworski, A., Hsieh, W.-T., Blaho, J.A., Larson, J.E., and Wells, R.D. (1987) Science **238**, 773-777
15. Hsieh, W.-T. and Wells, R.D. (1987) J. Biol. Chem. **262**, 14576-14582
16. Klysik, J., Stirdivant, S.M., and Wells, R.D. (1982) J. Biol. Chem. **257**, 10152-10158
17. Klysik, J., Stirdivant, S.M., Larson, J.E., Hart, P.A., and Wells, R.D. (1981) Nature **290**, 672-677
18. Murphy, K.E. and Stringer, J.R. (1986) Nucleic Acids Res. **14**, 7325-7340
19. McLean, M.J., Blaho, J.A., Kilpatrick, M.W., and Wells, R.D. (1986) Proc. Natl. Acad. Sci. USA **83**, 5884-5888
20. Lagravere, C., Malfroy, B., Leng, M., and Laval, J. (1984) Nature **310**, 798-800
21. Boiteux, S., de Oliveira, R.C., and Laval, J. (1985) J. Biol. Chem. **260**, 8711-8713
22. Kilpatrick, M.W., Klysik, J., Singleton, C.K., Zarling, D.A., Jovin, T.M., Hanau, L.H., Erlanger B.F., and Wells, R.D. (1984) J. Biol. Chem. **259**, 7268-7274
23. Semenza, G.L., Malladi, P., Surrey, S., Delgrosso, K., Poncz, M., and Schwarz, E. (1984) J. Biol. Chem. **259**, 6045-6048
24. Nordheim, A., Resser, P., Azorin, F., Kwon, Y.M., Moller, A., and Rich, A. (1982) Proc. Natl. Acad. Sci. USA **79**, 7729-7773
25. Azorin, F. and Rich, A. (1985), Cell **41**, 365-374
26. Lafer, E.M., Sousa, R., Rosen, B., Hsu, A., and Rich, A. (1985) Biochemistry **24**, 5070-5076
27. Muller, M.T. and Mehta, V.B. (1988) Mol. Cell. Biol. **8**:3661-3669
28. Sinden, R.R. and Kochel, T.J. (1987) Biochemistry **26**, 1343-1350
29. Deng, T., Li, Y., and Johnson, L.F. (1989) Nucleic Acids Res. **17**, 645-658
30. Muller, M.T., Spitzner, J.R., DiDonato, J.A., Mehta, V.B., Tsutsui, K., and Tsutsui, K. (1988) Biochemistry **27**, 8369-8379
31. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA **74**, 564-569
32. Sander, M., Hsieh, T., Udvardy, A., and Schedl, P. (1987) J. Mol. Biol. **194**, 219-229
33. Minford, J., Pommier, Y., Filipinski, J., Kohn, K.W., Kerrigan, D., Mattern, M., Michaels, S., Schwartz, R., and Zwelling, L.A. (1986) Biochemistry **25**, 9-16
34. Elias, P., O'Donnell, M., Mokarski, E., and Lehman, I. (1986) Proc. Natl. Acad. Sci. USA **79**, 4917-4921
35. Pommier, Y., Kerrigan, D. and Kohn, K. (1989) Biochemistry **7**, 995-1002
36. Muller, M.T. (1985) Biochim. Biophys. Acta. **824**, 263-267