Mung bean nuclease cleavage pattern at a polypurine·polypyrimidine sequence upstream from the mouse metallothionein-I gene

Albino Bacolla[1,*] and Felicia Y.-H.Wu[1,2]
[1]Department of Pharmacological Sciences, State University of New York at Stony Brook, Stony Brook, NY 11794, USA and [2]Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan, RO China

Received November 8, 1990; Revised and Accepted March 6, 1991 EMBL accession no. X53530

ABSTRACT

Mung bean nuclease, an enzyme specific for single-stranded DNA, was used to probe a non-*B* DNA structure present in the mouse metallothionein-I gene. The region sensitive to the enzyme was constituted by a 128 base-pair long polypurine·polypyrimidine sequence located at 1.2-kb from the start of transcription. A detailed analysis of the mung bean nuclease cleavage pattern revealed that: *(i)* under conditions of supercoiling and low pH a triplex structure was formed, *(ii)* the triplex was flanked by a sequence with the potential of forming a *Z*-DNA structure, *(iii)* most of the enzymatic activity was localized at some of the junctions between double-stranded and triple-stranded DNA and at mismatches in the triplex, *(iv)* no unpaired bases were observed in the loop or outside the triplex, and *(v)* the triplex was present in more than one configuration.

INTRODUCTION

The association of polypurine with polypyrimidine sequences into triple-stranded structures has been studied for more than 30 years (1). It is known, in fact, that when complementary polypurines and polypyrimidines are present in a 1:2 ratio at low pH, the product is a three-stranded molecule in which the purine bases are simultaneously engaged in Watson–Crick and Hoogsteen hydrogen bonds (2–8). The characterization of these *inter*molecular structures has been achieved by means of different methods, including *X*-ray fiber diffraction (9–12), infrared spectroscopy (4–6), circular dichroism (13,14) and NMR (15–17). Recently, however, the ability of polypurine·polypyrimidine (polyR·polyY) sequences to form *intra*molecular triplexes has been observed under the conditions of low pH and negative supercoiling (18–28). *Intra*molecular triplexes are more complex than the *inter*molecular ones: they contain at least one internal loop and four junctions, two between triplex and loop, and two between triplex and vector sequences (Scheme 1b). *Intra*molecular triplexes have been studied so far by enzymatic and chemical probes, since bases in a single-stranded configuration or at junctions have been shown to display hypersensitivity toward these probes. Chemical modification has been proven to be relatively nonspecific: bases outside the triplexes are also hyperreactive, and bases in the loop may retain their hyperreactivity upon release of supercoiling (20,22,24,26,29). Enzymatic probes have been poorly exploited, while biophysical methods have been impeded mostly by the requirement of supercoiling.

In this paper we describe the use of mung bean nuclease (*MBN*), an enzymatic probe, to study a long and heterogeneous polyR·polyY sequence upstream from the mouse metallothionein-I (mMT-I) gene. We chose this enzyme for several reasons: first, like S1 nuclease, it recognizes triplexes with high sensitivity; second, it is a milder enzyme than S1 nuclease (30) and third, it has been used at low temperature to specifically cleave single-stranded loops of hairpins (31). Our results showed that *MBN* activity was confined to the polyR·polyY sequence and that the bases outside and in the loop of the triplex were not in a single-stranded configuration. The specificity of the probe was exploited to derive a hydrogen bonding pattern for the triplex structure.

MATERIALS AND METHODS

DNA and biochemicals

Plasmid pBR-MT harbors a 3.8-kb *EcoRI-EcoRI* fragment containing the mMT-I gene and flanking sequences cloned in pBR322 (Fig. 1a). Phage M13mt1 was constructed as follows. Five μg of plasmid pBR-MT was cleaved with *EcoRI* and *KpnI* and the purified DNA was incubated with 5 U of *Klenow* enzyme (Boehringer/Mannheim) for 5 min at 37°C, followed by 15 min at 23°C in the presence of 2.5 mM of the four deoxynucleoside triphosphates in order to make both ends blunt. The products were separated on agarose gel and the 1.2-kb fragment was recovered with DEAE paper (32). The vector DNA was prepared by dephosphorylating *SmaI*-cleaved M13mp18. Ligation, transformation and large scale DNA preparation were according to standard procedures (33–36). *E. coli* JM101 used to grow

* To whom correspondence should be addressed at Boehringer Ingelheim, MBL, PO Box 368, 90 East Ridge, Ridgefield, CT 06877, USA

phage DNA were maintained and grown on minimal media (37). A *1-kb DNA ladder* was obtained from BRL. Restriction enzymes (New England Biolabs and Boehringer/Mannheim) were used according to the manufacturer's specifications in buffers prepared before each reaction. Radioactive nucleotides were from Amersham and most of the other chemicals from Fluka (MicroSelect grade when available).

MBN reactions and DNA labeling

DNA was stored at 0°C in 10 mM Tris-HCl (pH 8.0), 1 mM EDTA *or* 10 mM Tris-HCl (pH 8.0), 50 mM NaCl and 1 mM EDTA. DNA was first ethanol precipitated to avoid the contribution of Tris-HCl to the final pH, then resuspended in H_2O and freshly made *MBN* buffer (final concentration: 30 mM sodium-acetate (pH 4.5), 50 mM NaCl, 1 mM zinc-acetate and 5% glycerol). When high concentrations of *MBN* were used the glycerol content of the enzyme storage buffer was taken into account. The dilution buffer for *MBN* was 10 mM sodium-acetate (pH 5.0), 1 mM dithiothreitol, 0.1 mM zinc-acetate, and 0.001% Triton X-100. Starting from DNA banding in the CsCl gradient the contamination with metals was kept minimum (25,38). The DNA samples were allowed to equilibrate on ice for 15–20 min and *MBN* (Pharmacia) was added. The reactions were carried out on ice for the period indicated and stopped by the addition of Tris base and EDTA to the final concentrations of 70–90 mM and 40–45 mM, respectively. The DNA was finally phenol/chloroform extracted and ethanol precipitated.

To locate the *MBN* cutting sites in the polyR·polyY sequence, RF M13mt1 was cleaved with 10, 100, or 800 U of *MBN* for 10 min and digested with *EcoNI* or *AlwNI* (Fig. 3). The *EcoNI* site was labeled with *Klenow* and [α^{32}P]dTTP; the 3′ ends of the *AlwNI* cleaved sample were labeled with terminal deoxynucleotidyl transferase (*TdT*) (Pharmacia) and [α^{32}P]ddATP. The products were separated on native 6% polyacrylamide gel (0.3×40 cm) after removal of the unincorporated radioactivity by Sephadex G-50 minicolumn.

To specifically locate the *MBN* cutting sites on both strands of the polyR·polyY sequence, RF M13mt1 was cleaved with 800 U of enzyme for 10 min. The 3′ ends were labeled with *TdT* and [α^{32}P]ddATP; the 5′ ends were labeled with T4 polynucleotide kinase and [γ^{32}P]dATP. After removing the unincorporated radioactivity the DNA was cleaved with *EcoNI* or *AlwNI* and the products were separated on a native 6% polyacrylamide gel, recovered and rerun on a 8% sequencing gel. A radioactive *1-kb ladder* was prepared by filling-in the 3′ ends with *Klenow*, using [α^{32}P]dCTP. After labeling, this marker was 8-bp longer than what was specified by the manufacturer.

Sequencing

Sequencing was carried out on either single-stranded M13mt1 or RF M13mt1 and pBR-MT by the primer extension method. Commercial (Pharmacia) or synthetic HPLC-purified primers were used in an overall walking primer strategy.

RESULTS

Characteristics of RF M13mt1

Phage M13mt1 contains a 1.2-kb fragment from the upstream region of the mMT-I gene cloned in the *SmaI* site of M13mp18 (Fig. 1). The 12 clones tested carried the insert in the same orientation. The transcribed messenger for the disrupted α-peptide

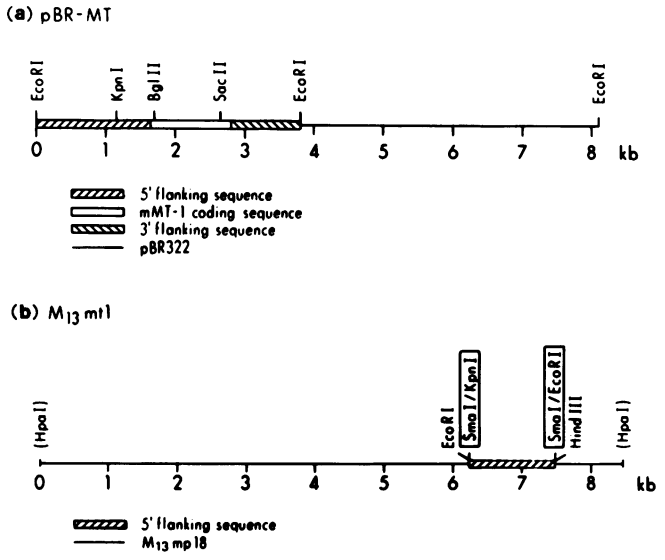


Fig. 1. Map of (a) plasmid pBR-MT and (b) phage M13mt1. Thin line: vector DNA; rectangles: mMT-I sequences. The boxed *SmaI/KpnI* and *SmaI/EcoRI* show the orientation of the *EcoRI-KpnI* fragment of the mMT-I into the *SmaI* site of M13mp18. The restriction sites flanking the polylinker are indicated.

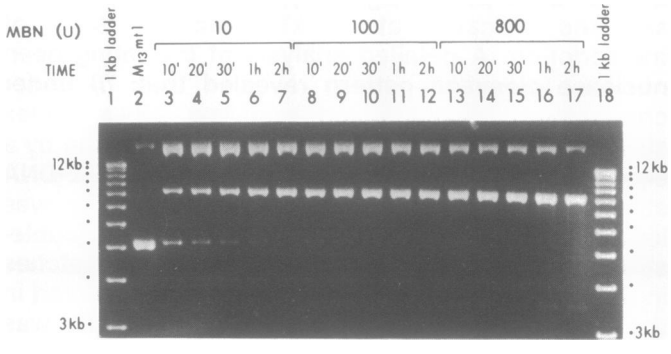


Fig. 2. Agarose gel electrophoretic pattern of *MBN* activity on RF M13mt1. Lanes 1 and 18: *1-kb DNA ladder* (bands from 3- to 12-kb are shown); lane 2: native supercoiled M13mt1, nicked DNA migrates above 12-kb; lanes 3–7, 8–12, and 13–17: 10, 100, and 800 U, respectively, of *MBN* for 10, 20, 30, 60, and 120 min. The reactions were performed at 0–4°C.

bore the sequence of the non-coding strand of the cloned mMT-I fragment (Fig. 1b). The repaired *EcoRI* site of the mMT-I fragment was located toward the hybridization site of the universal direct sequencing primer of M13mp18. The bias observed in cloning fragments containing asymmetric sequences seems to be a relatively common event (39). We did not investigate this observation any further.

MBN activity on RF M13mt1

When supercoiled pBR-MT and RF M13mt1 were incubated with *MBN*, the DNA was readily converted to the circular form and, at a lower rate, to the linear one. Preliminary experiments carried out on pBR-MT and RF M13mt1 located the *MBN* sensitive region between the *EcoRI* and the *KpnI* sites of the mMT-I gene, at about 1.2-kb from the start of transcription (Fig. 1a). Further experiments located the *MBN* sensitive region in a polyR·polyY sequence. This activity was dependent on supercoiling (data not

```
-1774  AATTCTGGTTTTTTTGTTTGTTTGTTTGGTTTTGGTTTTTGGTTTTTCGA
       GACAGGGTTTCTCTATATAGCCCTAGCTGTCCTGGAACTCACTTTGTTGA
-1674 CCAGGCTGGCCTCGAACTCAGAAATCCGCCTGCCTCTGCCTCCTGAGTGC
       TGGGATCAAAGGCGTGCGCCACCAACACCCGGCTTGTTTTTTTTTTTTTT
-1574 TTTTAATCTTTAATCACTTCCTTATTAACTTTTGAAATAGGATCTCATTT
       GTGTGGCCCTGGCTGGCCTTGAATTTTGCAGAGATCATCTGCCTCTGACT
-1474 TCTAAGAGCTGAGATTAAAGAGCTACCCTAAGGCAAGAATTTAACTGTAT
       AGATAAAGACTCAGTTCCAAATGAAGGTTTTTTTTTGGATCTTCCCACGC
                                  EcoN I
-1374  AGAGGGTATTGCTGTGTTGTCTCCTCCAAGAGGAAATATCCGGAGGAAGG
       AGAAACGCCCTGGGTAATATTATACTACCAAACCAGTGGGTTGATAATAG
-1274 AAAGACACTTGTATCTAAACCACCATGAAGGAGGAGGAAGAGGAGGAGGG
       GGGAGGAGGAGAAAAAGGAAGAAGAGAGGAGAAAGAAGGGGAGAAGAACG
-1174 GAGAAGAGGAGGAAGAAGAGGAGGAAGAGGGAGAGGGGGAAGAGGAGGGG
                                  AlwN I
       GAAGTGTGTGTGTGGGGGGGGGGAGTGTCATAGCCCAGCACCTGGTGCCT
-1074 TCTCTGCCTACAGATGAACTTTGTACAGAATCACAGCAAGAGAACGTGGA

       CAAATACAGGAGCGTTAGTGCCAGGCAGTGCTGGTTCACAGCCGTTGTAC
- 974 TACTGACTGGATATATGGCCTGGGCAAGTCACTGAACCTCTGAATCCTCT
       GTCCTTGTGTGTAAATAGAATTTGCATCTCTCTATAGGTATTAAATGAGA
- 874 AATTGGTTTAACTCCTGGTTCTGGCATAATAGACATATTGCTGGTGGTGG
       TTCTACCACAGTTAAGCAAAAACCTGCCCAAAGTCGTATAGCTCAGAAGT
- 774 GATGGTATGCATGAGCCAACAGTGTAGCTGTTCTCTTGCCCAGGACACAA
       GGCCTGCCTGACAGCAGACTCTAATGTTACTCAAAATCAGAAGCATTAAA
- 674 AAAAAAAGAAAGAAAGAAATAGCTGACTTTATTTTTTTTATCTAATGCAG
                                  Kpn I
       TCCCCACCTAAAATGATCTCTCTGGAGGGAGGGTACCTTGGTTTTTAAAA
- 574 CCAGCCTGGAGTAGAGCAGATGGGTTAAGGTGAGTGACCCCTCAGCCCTG
       GACATTCTTAGATGAGCCCCCTCAGGAGTAGAGAATAATGTTGAGATGAG
- 474 TTCTGTTGGCTAAAATAATCAAGGCTAGTCTTTATAAAACTGTCTCCTCT
       TCTCCTAGCTTCGATCCAGAGAGAGACCTGGGCGGAGCTGGTCGCTGCTC
- 374 AGGAACTCCAGGAAAGGAGAAGCTGAGGTTACCACGCTGCGAATGGGTTT
       ACGGAGATAGCTGGCTTTCCGGGGTGAGTTCTC  (-292)
```

Fig. 3. Sequence of the 5′ region of the mMT-I gene. The location of three relevant restriction sites is underlined. The underlined *nt* from *−300* to *−292* overlap with a published sequence (40), to which the numbers in italic refer. The 128-bp long polyR·polyY is shown in boldface characters.

shown). We performed all of the subsequent experiments with *MBN* at 0−4°C. A time course of *MBN* activity on RF M13mt1 at this temperature is shown in Fig. 2. With each enzyme concentration used, most of the linear DNA product was detected within the first 10 min of incubation (lanes 3, 8 and 13). Increasing amounts of linear DNA were accumulated at times greater then 10 min, in part as a result of nonspecific cleavage of the nicked DNA. When 800 U of enzyme were used, some degradation of the linearized DNA was already seen after 10 min (lane 13).

Sequence of the 5′ flanking region of the mMT-I gene

The sequence of the mMT-I promoter up to nucleotide (*nt*) *−300* from the start of transcription has already been reported (40). We extended the sequencing to the *EcoRI* site at *nt −1775* in order to encompass the polyR·polyY region. The sequence (Fig. 3) starts at the underlined A of the GAATTC *EcoRI* restriction site (*nt −1774*) and it overlaps from *nt −300* to *−292* with the reported sequence (40). The position of the restriction site for *EcoNI* and *AlwNI*, used in subsequent experiments, is underlined. The polyR·polyY stretch responsible for the sensitivity to *MBN* is shown in bold letters. This sequence was 128-bp long, spanning from *nt −1148* to *−1221*, and it was flanked at the 3′ end by a $(TG)_5$ and a $(G)_9$. In Figs. 6−7 and in Scheme 2 only this polyR·polyY sequence, numbered from 1 to 128 shall be reported. Sequencing of the complementary strand excluded the presence of mismatches, at least in the polyR·polyY region. Neither the ability of the $(TG)_5$ to attain a Z-DNA conformation nor the sensitivity of the $(G)_9$ to endonuclease-G (41) was examined.

High resolution mapping of the *MBN* cleavage in the polyR·polyY sequence

RF M13mt1 was incubated with 10, 100, or 800 U of *MBN* for 10 min. Each sample was divided into two aliquots; one was cut with *EcoNI*, the other with *AlwNI*. The *EcoNI* site was filled-in with *Klenow* and [α^{32}P]dTTP, the 3′ ends of the *AlwNI* cleaved sample were labeled with *TdT* and [α^{32}P]ddATP. The purified DNA was run on a native 6% polyacrylamide gel. The result

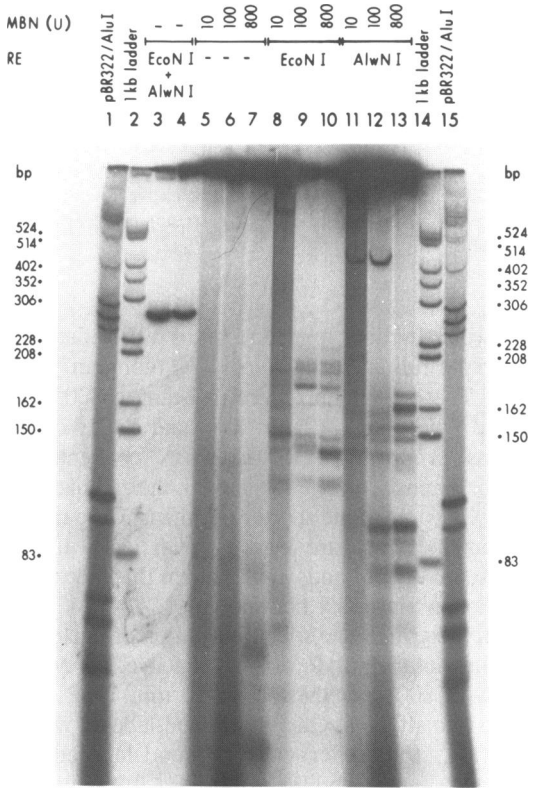


Fig. 4. Native PAGE of the *MBN* cleavage pattern on RF M13mt1. Lanes 1 and 15: *AluI*-cleaved pBR322; lanes 2 and 14: *1-kb DNA ladder* (only the length of these fragments is reported); lanes 3−4: cleavage with *EcoNI* + *AlwNI*; lanes 5−7: 10, 100, or 800 U of *MBN*; lanes 8−10: 10, 100, or 800 U of *MBN* and cleavage with *EcoNI*; lanes 11−13: 10, 100, or 800 U of *MBN* and cleavage with *AlwNI*. Incubations with *MBN* were for 10 min at 0−4°C. *RE*: restriction enzyme(s).

is shown in Fig. 4, lanes 8−13. In lanes 5−7 the restrictions were omitted to verify whether *MBN* alone would release short fragments. Incubation with *MBN* gave rise to a smear whose size decreased at higher enzyme concentrations (lanes 5−13). After

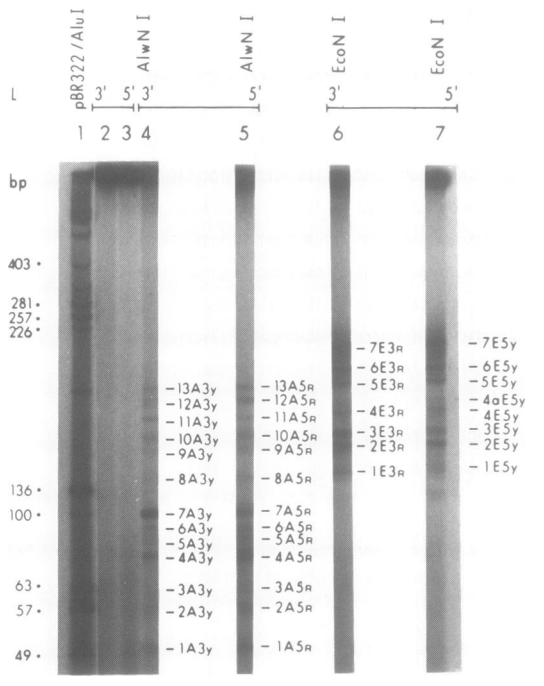
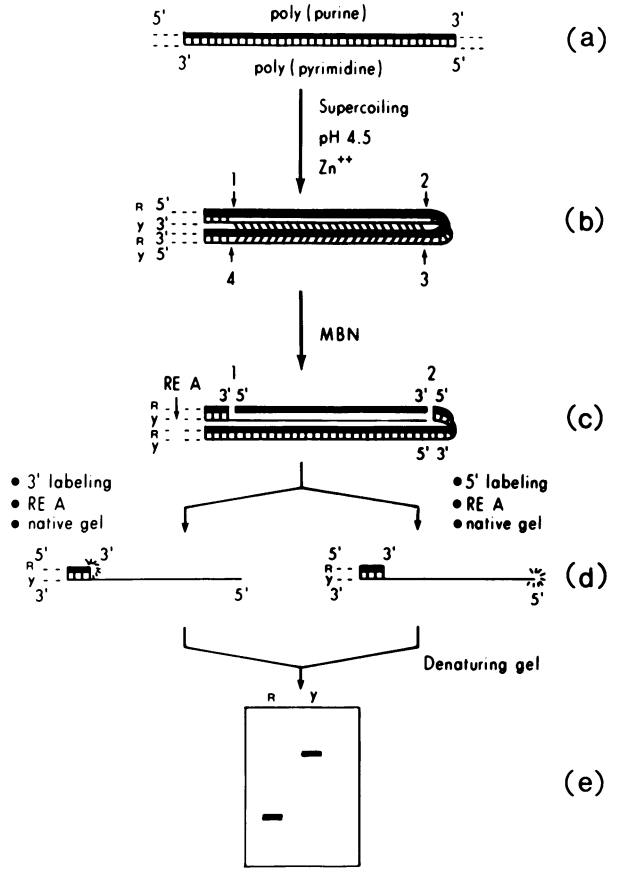


Fig. 5. Native PAGE of the *MBN* cleavage pattern on both strands of RF M13mt1; 800 U of *MBN* for 10 min at 0–4°C were used. Lane 1: *AluI*-cleaved pBR322; lane 2: *MBN* and 3′ end-labeling; lane 3: *MBN* and 5′ end-labeling; lane 4: *MBN*, 3′ end-labeling (pyr. strand) and cleavage with *AlwNI* = A3y fragments; lane 5: *MBN*, 5′ end-labeling (pur. strand) and cleavage with *AlwNI* = A5r fragments; lane 6: *MBN*, 3′ end-labeling (pur. strand) and cleavage with *EcoNI* = E3r fragments; lane 7: *MBN*, 5′ end-labeling (pyr. strand) and cleavage with *EcoNI* = E5y fragments. *L*: end-labeling.



Scheme 1. Scheme for the formation of an *intra*molecular triplex and *MBN* activity. (a) thick line: polypurines; thin line: polypyrimidines; vertical bars: Watson–Crick hydrogen bonds; (b) under supercoiling and low pH the polyR·polyY sequence forms a triplex where the 3′ part of the purine strand (*R*) folds back and engages Hoogsteen hydrogen bonds with the pyrimidine strand (*Y*) running in a parallel orientation; (/////): Watson–Crick hydrogen bonds in the triplex; (\\\\\) Hoogsteen hydrogen bonds; 1–4: junctions between double-stranded and triple-stranded DNA; (c) *MBN* recognizes distortions in the DNA backbone and cleaves at junctions; the event in which *MBN* cleaves the *R* at junction 1 and both strands at junction 2 is considered; *RE A*: restriction site; after *MBN* cleavage the triplex is no longer sustained by supercoiling; (d) 3′ end-labeling occurs on the *R* at junction 1, 5′ end-labeling occurs on the *Y* at junction 2; cleavage 5′ from the triplex with a restriction enzyme (*RE A*) releases a partially single-stranded fragment that can be purified by native PAGE; (e) the length of both strands of this fragment is determined by denaturing PAGE. This scheme corresponds in the text to the set of fragments produced by *EcoNI*. Another set is produced by *AlwNI*.

MBN, the two restrictions released a set of well defined fragments shorter than the full length *EcoNI-AlwNI* (compare lanes 8–13 with 3–4). These fragments were resistant to 800 U of *MBN*. As expected from Fig. 2 (lanes 3, 8 and 13), the amount of product was proportional to the *MBN* concentration used (compare the intensity of the bands with the radioactivity at the top of the gel). Some of the shorter fragments were not seen with 10 U of enzyme (compare lane 11 with lanes 12–13). We sequenced some of the bands and confirm that the termini were at the restriction site (data not shown).

In order to determine the specific location of nuclease activity on both strands of the polyR·polyY sequence, RF M13mt1 was incubated with 800 U of MBN for 10 min. The sample was divided into two aliquots. One was 3′ end-labeled with *TdT* and [α^{32}P]ddATP, the other was 5′ end-labeled with T4 polynucleotide kinase and [γ^{32}P]dATP. Both aliquots were further divided into two parts. One was digested with *EcoNI*, the other with *AlwNI*. Four sets of fragments were therefore obtained on a native 6% polyacrylamide gel (Fig. 5, lanes 4–7). The pattern obtained was very similar to the one in Fig. 4. The bands were designed as follows: labeling of the 3′ end on the pyrimidine strand followed by *AlwNI* cleavage gave 1A3y to 13A3y; labeling of the 5′ end on the purine strand followed by *AlwNI* cleavage gave 1A5r to 13A5r; labeling of the 3′ end on the purine strand followed by *EcoNI* cleavage gave 1E3r to 7E3r and labeling of the 5′ end on the pyrimidine strand followed by *EcoNI* cleavage gave 1E5y to 7E5y. Our assumption, as shown in Scheme 1a,b, was that the formation of an *intra*molecular triplex would be recognized by *MBN* at the site of the unpaired purines and at the junctions between double-stranded and triple-stranded DNA. In the event of a single cut on the purine strand at junction 1 and a double cut at junction 2 (Scheme 1c), the 3′ and 5′ ends of the fragment released by a restriction enzyme (*RE A* in Scheme 1c,d) would not be blunt. The pyrimidine strand would be longer than the purine one, and this difference could be measured on a denaturing gel after the fragment had been purified from a native gel. Fig. 6a,b shows the sequencing gel pattern of the forty fragments eluted from the native gel of Fig. 5. Taking into account the staggered end of the fragments at the restriction site and the addition of the ddATP during labeling, we located all of the bands in the polyR·polyY sequence. Abundant fragments in Fig. 5 (7A, 4A, 6E, 7E) also gave the most prominent bands in Fig 6a,b. Closed squares indicate these bands (Figs. 6 and 7). Likewise, quantitatively less abundant but significant bands (1A, 2E, 3E, 5E in Fig. 5) are denoted by open

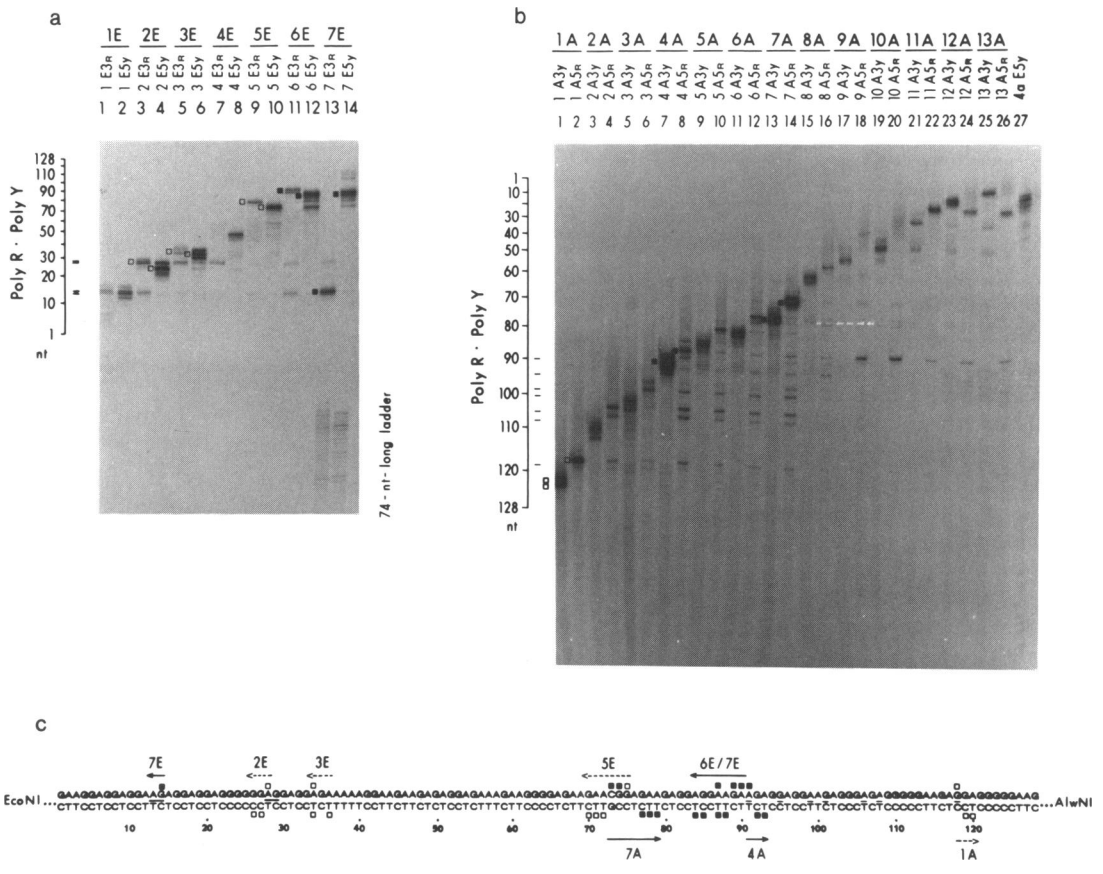


Fig. 6. Sequencing gel of the forty fragments eluted from the native gel of Fig. 5; (a): bands eluted from the *EcoNI* cleaved sample (lanes 6−7 in Fig. 5); the 74-*nt*-long ladder in lanes 13−14 is indicated; (b): lanes 1−26, bands eluted from the *AlwNI* cleaved sample (lanes 4−5 in Fig. 5); lane 27: fragment 4aE5y from lane 7 of Fig. 5; (c): the polyR·polyY sequence is reported, where *nt* 1 corresponds to *nt* −*1248* of Fig. 3. Due to the staggered ends at the *EcoNI* and *AlwNI* restriction site and to the addition of a ddATP during the 3′ end-labeling, the correspondence between bands and numbering in (a) and (b) is only approximative. Closed squares denote the most prominent bands and their location on the polyR·polyY sequence; less intense bands are denoted by open squares; solid and dotted arrows (for closed and open squares) point toward the respective restriction site; underlined purines in (c) and short traits in (a) and (b) denote the 'consistently cut sites'.

squares. The name of these fragments is reported in Fig. 6c, where solid (for closed squares) and dotted (for open squares) arrows point toward the respective restriction site. In Fig. 7, the position of all of the significant *MBN* cleavages is reported as lower case letters.

'Consistently cut sites' on the purine strand

In Fig. 6a,b the purine lanes contained constant bands (denoted by short traits). We called these bands 'consistently cut sites' to indicate that in all of the fragments the DNA was regularly recognized at some position. In the latter gel this pattern was uniform for the first 7 fragments, while for the longer fragments the lower bands were very weak. When the 'consistently cut sites' were located in the polyR·polyY sequence they corresponded to A13, G14, A28, G29 (Fig. 6a), and to G118, A108, A106, G101, A99, G95, A91 (Fig. 6b), respectively (Fig. 6c, underlined purines). Other constant bands were also present (see for example at *nt* 70−72 in lanes 10, 12, 14 of Fig. 6a).

Fragments with an overhanging pyrimidine strand

In fragments 7E, 4E (Fig. 6a), 9A and 10A (Fig. 6b) the purine strand was considerably shorter than the pyrimidine one. These fragments are marked by a star in Fig. 7. In fragment 7E the purine gap started around A13−G14, two of the 'consistently cut sites'. The pyrimidine strand was cut at several *nt*, the more distal of which corresponded to T88 (Figs. 7a and 6a, lanes 13−14). The abnormally slow migration of fragment 7E on the native gel was ascribed to its partial single-stranded character (compare fragments 6E and 7E in Figs. 5, 6a and 7a). Fragment 4E (Figs. 7a and 6a, lanes 7−8) had an overhanging pyrimidine strand whose terminus mapped at T49; the purine strand was mainly cut at the 'consistently cut sites' A28−G29. In fragments 9A−10A (Fig. 6b, lanes 17−20), the purine strand was predominantly cut at the more distal of the 'consistently cut site' (A91) while the pyrimidine strand was cut between C45 and T54 (Fig. 7b).

Shortest and longest fragments

After *MBN*, restriction with *AlwNI* gave consistently more bands than the one with *EcoNI*. The cleavages with *MBN* at the 3′ end of the polyR·polyY sequence (compare the termini of fragments 6E, 7E with those of 1A, 2A, 3A in Fig. 7a,b) were not revealed by the restriction with *EcoNI*. The size of fragments 11A−13A closely matched with the length of fragment 1E (Fig. 7a,b).

DISCUSSION

Most of the studies conducted on cloned polyR·polyY sequences using chemical or enzymatic probes support the formation of triple-stranded DNA but few of them provide information on the

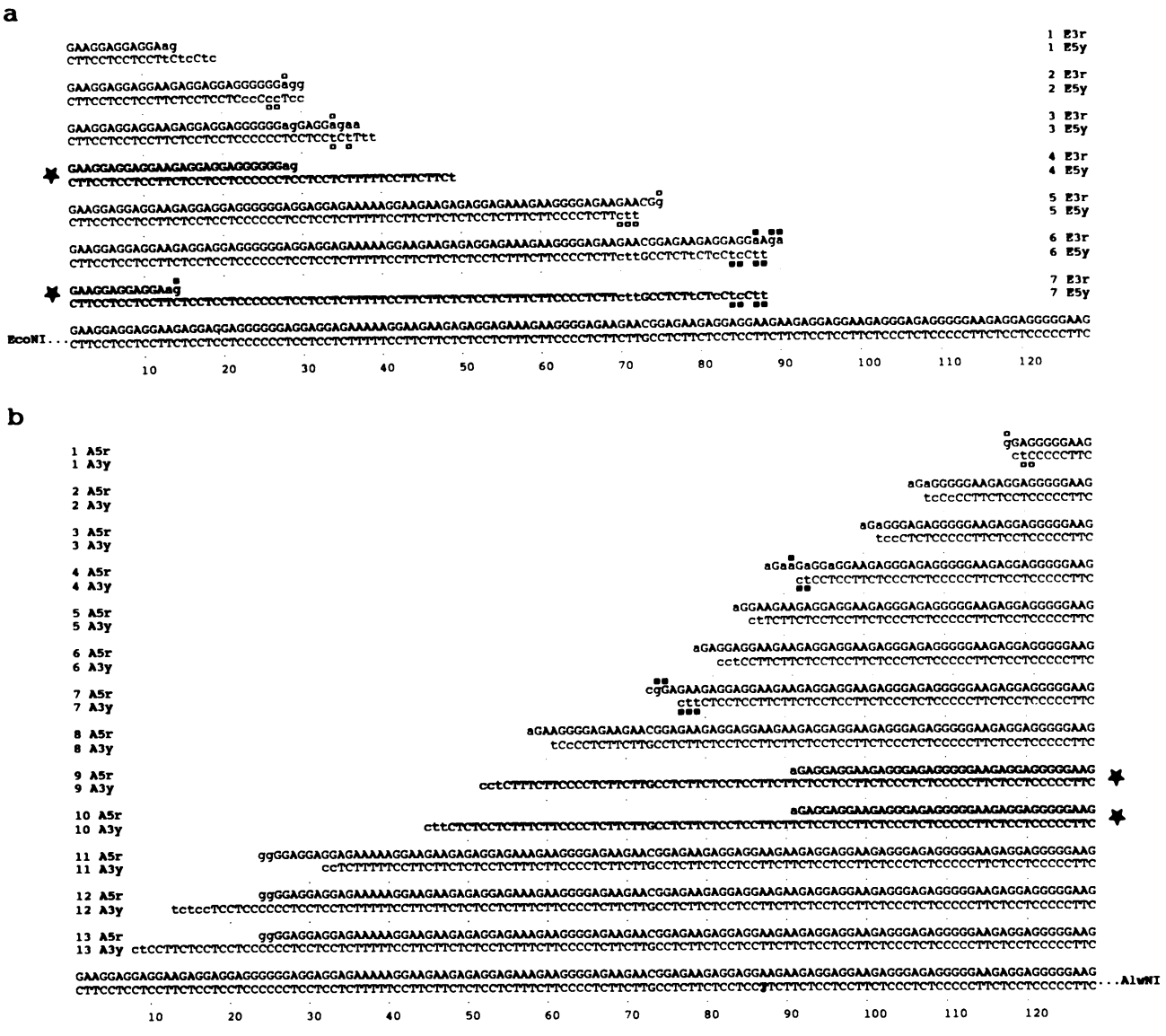
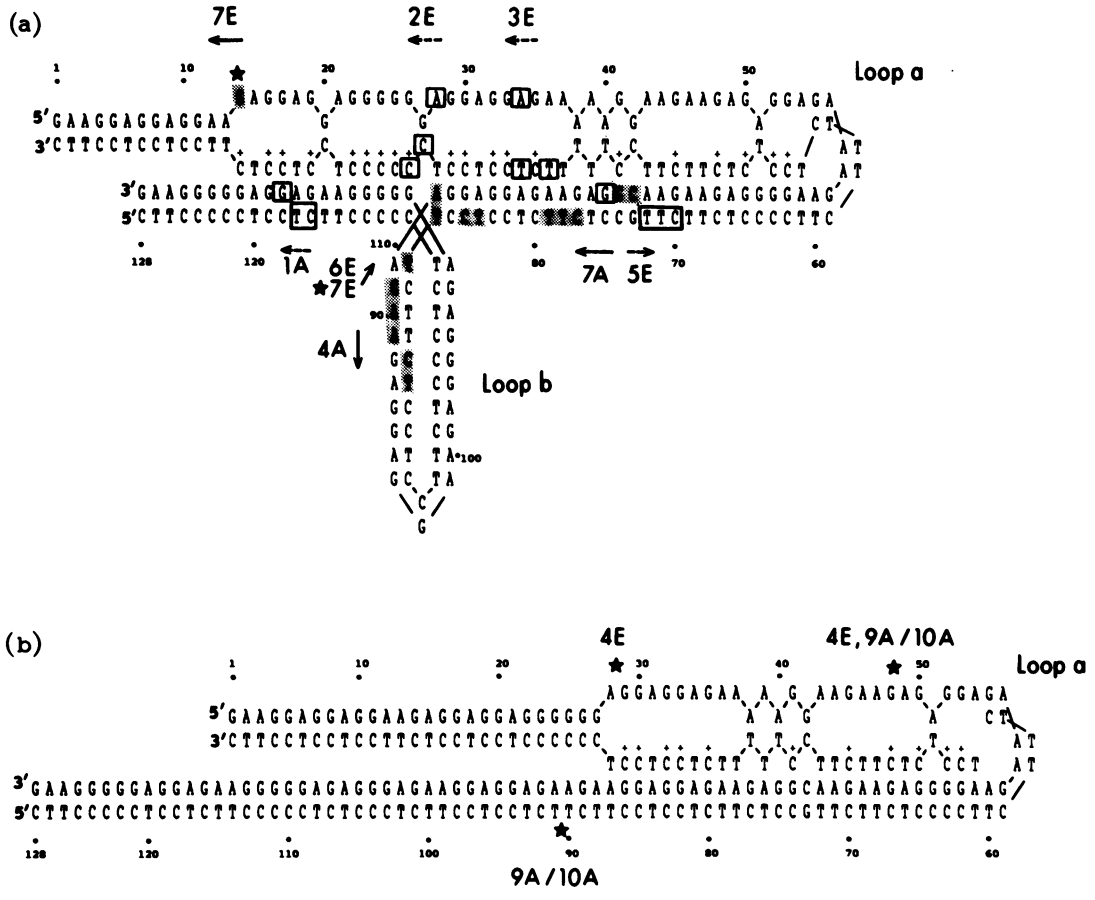


Fig. 7. Location of the termini of the fragments from Fig. 6a,b on the polyR·polyY sequence; (a): fragments 1E−7E; (b): fragments 1A−13A; lower case letters correspond to significant bands on the sequencing gels; closed and open squares are as in Fig. 6; fragments with a significant overhanging pyrimidine strand are shown in bold letters and marked by a star.

actual base-pairing, unless the sequences are simple repeats or contain a mirror symmetry (25,28). Many of the naturally occurring polyR·polyY sequences have a relatively heterogeneous composition (39,42).

We found a 128-bp long heterogeneous polyR·polyY sequence at about 1.2-kb from the start of transcription of the mMT-I gene. The exquisite sensitivity toward *MBN*, the requirement of supercoiling and the retention of these characteristics upon subcloning in another vector (43) suggested that this sequence was able to form a triplex structure. Our aim was, therefore, to test this hypothesis and to gain insight into the base-pairing pattern. To avoid the accumulation of DNA fragments that arises from nuclease activity on nicked molecules we chose *MBN* (30) instead of S1 nuclease. *MBN* has been used at 0°C to selectively digest loops of hairpin structures without affecting the helix stem (31). At low temperature, the stability of a preformed triple helix structure (3−6,15,44) and base stacking (45) increase, while breathing of the base-pairs is reduced. Under the experimental condition we chose (800 U *MBN* for 10 min), a degradation of the linear M13mt1 was apparent on agarose gel (Fig. 2, lane 13). On polyacrylamide gel, however, a smear was usually seen only with lower enzyme concentrations (Fig. 4, lanes 5−13). This may indicate that, while *MBN* could cleave at structures other than the triplex, the triplex products were resistant to further enzyme degradation. The 'consistently cut sites' in the purine lanes of Fig. 6a,b, on the other hand, suggest that all of the fragments were the result of enzyme activity in the polyR·polyY sequence. The possibility that the pattern at 800 U of enzyme (Figs. 4−5) could include secondary cuts from 'nibbling' at a structure other than the triplex, while it cannot be ruled out, seems unlikely. The overhanging pyrimidine strand of fragments 7E, 4E, 9A and 10A clearly indicates the formation of an *intra*molecular triplex. The 'consistently cut sites' in the purine lanes of Fig. 6a,b suggest that all of the fragments, being recognized at the same positions, belong to a common DNA structure.



Scheme 2. Hydrogen bonding schemes for the triplexes in the polyRμpolyY region. Shaded and boxed bases correspond to closed and open squares, respectively, of Fig. 6–7; solid and dotted arrows and the name of the fragments are as in Fig. 6c. A star indicates, with approximation, the termini of the fragments with an overhanging pyrimidine strand; protonated, Hoogsteen-hydrogen-bonded cytosines are marked with (+); see text for details.

We associate the cleavages at A13–G14 of fragment 7E (Figs. 6 and 7a) with junction 1 (in Scheme 1b), and those at T87–T88 with a 'type 4' junction. The cleavages at T87–T88 cannot locate junction 2, because more than 74 additional *nt* would be needed to fold back and form the triplex, and the polyR·polyY sequence is not so long. We interpret fragment 6E as a double cut across 'type 4' junction, and 4A as its complementary fragment (Fig. 7). The 'heavy' cuts at the end of the purine strand of fragment 6E (Fig. 6a and A87–A90 in Figs. 6c and 7a) and the strong 'consistently cut site' at A91 (Figs. 6b,c and 7b) mapped adjacent positions, suggesting that this was a very sensitive *MBN* site. We infer that the same cuts were likely to be present in the purine strand of fragment 7E, therefore allowing a stretch of 74 purines to be labeled at both the 3′ and the 5′ end (see also Scheme 1c). This reasoning is supported by the following observations: a 74-*nt*-long ladder was clearly visible along with fragment 7E (Fig. 6a, lanes 13–14); both the 3′ and 5′ ends of this ladder were labeled and in the 3′ end-labeling (lane 13) there was a shift of about one *nt* as expected from the addition of the terminally labeled ddATP. In order to comigrate with 7E on the native gel, this ladder would have been partly involved in the triplex structure. We, therefore, interpret the 74-*nt*-long ladder as comprising the purines from junction 1 to 'type 4'. The fact that *nt* 87–90 do not locate junction 4, as depicted in Scheme 1b, is supported by the following evidence: 4A (complementary to 6E/7E) was not the shortest fragment, and in its purine strand A91 was not the only 'consistently cut site' (Fig. 6a,b). Fragments 1A–3A, being cut at the 'consistently cut sites' (like 4A), were also to be part of the triplex; we conclude that junction 4 is identified by fragment 1A.

In the other staggered fragments, 4E, 9A and 10A, the 'consistently cut sites' at A13–G14 and from G118 to G95 were extremely weak and unlikely to be involved in any triplex structure. For these fragments, the 'consistently cut sites' A28–G29 (of fragment 4E) and A91 (of fragments 9A and 10A) appear to locate junction 1 and 4, respectively. This interpretation implies the presence of two triplex structures: the first with junction 1 at *nt* 28–29 and junction 4 at *nt* 90–91, the second in which the triplets from 28–29 to 90–91 are extended to *nt* 13–14 and 118–119, respectively. We propose, for these two structures, a hydrogen bonding pattern as outlined in Scheme 2. In Scheme 2b, junction 1 is at *nt* 27–28 and junction 4 at *nt* 87–88. A 21-bp-loop (loop b) allows this triplex to be extended and have junctions 1 and 4 at *nt* 13–14 and 121–122, respectively (Scheme 2a). In Scheme 2a the location of the most prominent bands of Fig. 6a,b are reported as shaded (for closed squares) and boxed (for open squares) bases. As in Fig. 6c, solid and dotted arrows indicate the termini of the fragments and point toward the restriction site. In both schemes a star locates, approximately, the termini of the four fragments with an overhanging pyrimidine strand. The models do not envision regions of single-strandedness other than the purines from G14

(A28) to A54. There are mismatches at positions 20/115 and 27, at *nt* around 40/75 and 51/64. According to Scheme 2a, fragment 7E (Figs. 6a and 7a) would be cut (at *nt* 87−90) at a junction between triple-stranded and double-stranded DNA and the simultaneous activity at A13−G14 would allow the 74 purines to comigrate in the native gel and appear as the ladder seen in Fig. 6a (lanes 13−14). This ladder would represent random cleavage at the single-stranded bases. A double cut across junction 1 *or* at loop b would give fragments 1E and 6E, respectively (Fig. 7a). The termini of 7A, a quantitatively well represented fragment, are located in Scheme 2a in a region where the triplex could be destabilized by extensive mismatches. Fragments 3E and 5E, which are abundant as well, would be related to *MBN* activity in the same region.

According to our interpretation, the main nuclease activity would be associated with perturbations in the triplex structure whereas cleavages on the single-stranded purines would be, comparatively, weaker. If this is true, it implies that a distortion in torsion angles of nucleotides at junctions or at mismatches in a triplex represents the most sensitive target for *MBN* activity. Sugar pucker flexibility plays a key role at the junctions between *B*- and *Z*- (46) as well as *A*- and *B*-DNA (47), while association of nuclease activity with highly distorted parameters in the DNA structure has been proposed (48). Scheme 2a seems to account for the observation that the termini of fragment 6E did not coincide with those of fragment 1A (Fig. 7a,b). *MBN* cleaved at both *nt* 87−88 and 118−119, but preferred pyrimidines 87−88 (compare lanes 7−8 with 1−2 in Fig. 6b). This occurrence, and the possibility of molecules being cut at both positions, could have prevented the accumulation of a product complementary to fragment 1A. In the absence of loop b (Scheme 2b), most of the 'consistently cut sites' would be located too far from the triplex to be recognized by *MBN*. In fragments 4E and 9A/10A the pyrimidine strand was cut between C45 and T55, with some preference around C50. These pyrimidines were weakly recognized in the other fragments. This could be due to a different stability between the two triplexes. In this regard, the fragments that account for Scheme 2a are quantitatively much more abundant than 4A, 9A and 10A. Loop b could also induce *nt* around position 87 to be highly distorted, thus switching the priority of nuclease activity, or influence the triplex in some other ways.

The presence of more than one configuration could be a result of the DNA being at two different superhelical densities (24), but an equilibrium between the two configurations is also possible (21). The fact that we did not see a family of different triplexes, each one giving rise to a set of fragments, indicates that the whole polyR·polyY sequence, rather than short stretches of mirror symmetry, participated in attaining a minimum energy state. In fragments 11A, 12A, and 13A, cleavages were located at G24 and G25 (Fig. 7b). Due to the dimension of the sequencing gel and the distance of these bands from the reference ladder, it is possible that these cleavages coincide with A28−G29. The facts that the nuclease activity was limited to the polyR·polyY sequence and that, except for the weak fragment 8A (Fig. 7b and Scheme 2), there were no additional fragments or 'consistently cut sites' mapping at loop a, suggest that there were no single-stranded bases other than the purines G14−A54.

The *nt* from 77 to 100 contain a 12-bp direct repeat (GAAGAGGAGGAA). We noticed a ladder in some of the pyrimidine lanes at these *nt* (Fig. 6b, lanes 7, 9, 11 and 13), but we do not know whether these two direct repeats can alternatively be involved in the triplex. Configurations involving two loops are, therefore, speculative. The lack of 'consistently cut sites' at the purines around position 40 seems to exclude a triplex-duplex junction in that region. In the configurations we have been discussing, the 5′ portion of the purine strand is single-stranded, while the 3′ half is engaged in the triplex. In the reverse configuration, the 3′ half of the purine strand would be partly single-stranded and partly double-stranded at loop b. In Fig. 6b no single-stranded fragments were observed 3′ of A91. A 41-*nt*-long ladder comprising, for example, purines G121−A78, was also not seen. Loop b would have allowed this ladder to comigrate with some of the fragments on the native gel, thus appearing on the sequencing gel 6b as it was for Fig. 6a (lanes 13−14). Although these arguments are against a reverse configuration, this might still be possible. None of the other biased sequences in pBR-MT or M13mt1 were sensitive to *MBN*. In Scheme 2a the triplex is brought very close to a $(TG)_5$, a sequence that can adopt a *Z*-DNA conformation. Association of a polyR·polyY with *Z*-DNA forming sequences or short direct repeats has been reported for several genes (18,22,38,49−52), at least in one event of unequal sister chromatid exchange (27) and in gene conversion events (53).

CONCLUSION

In this paper we describe our approach to the study of a structure adopted by a polyR·polyY sequence present in the promoter region of the mouse metallothionein-I gene. The data support the formation of an *intra*molecular triplex. The possibility of multiple configurations and a plausible scheme for hydrogen bonding shall guide in designing further experiments to better characterize the structural features and to search for its possible functional role.

ACKNOWLEDGMENTS

We thank Richard Palmiter for providing the metallothionein gene; Barbara Birshtein, Johanna Griffin, Kenneth Breslauer, Robert Wells, William Bauer, Howard Federoff and our colleagues for helpful discussions; Charles Giardina for reading the manuscript; Giovanni Gaudino for assistance. This work was supported by the Grants from the National Institute of Health (GM 28057) and National Science Foundation (DBM 8703990) to F.Y.-H.Wu at SUNY at Stony Brook.

REFERENCES

1. Felsenfeld,G., Davies,R.D. and Rich,A. (1957) *J. Amer. Chem. Soc.*, **79**, 2023−2024.
2. Rich,A. (1958) *Nature*, **181**, 521−525.
3. Lipsett,M.N. (1964) *J. Biol. Chem.*, **239**, 1256−1260.
4. Miles,H.T. and Frazier,J. (1964) *Biochem. Biophys. Res. Commun.*, **14**, 21−28.
5. Miles,H.T. and Frazier,J. (1964) *Biochem. Biophys. Res. Commun.*, **14**, 129−136.
6. Howard,F.B., Frazier,J., Lipsett,M.N. and Miles,H.T. (1964) *Biochem. Biophys. Res. Commun.*, **17**, 93−102.
7. Miller,J.H. and Sobell,H.M. (1966) *Proc. Natl. Acad. Sci. USA*, **55**, 1201−1205.
8. Morgan,A.R. and Wells,R.D. (1968) *J. Mol. Biol.*, **37**, 63−80.
9. Arnott,S. and Bond,P.J. (1973) *Nature New Biol.*, **244**, 99−101.
10. Arnott,S. and Bond,P.J. (1973) *Science*, **181**, 68−69.
11. Arnott,S. and Selsing,E. (1974) *J. Mol. Biol.*, **88**, 509−521.
12. Arnott,S., Bond,P.J., Selsing,E. and Smith,P.J.C. (1976) *Nucleic Acids Res.*, **3**, 2459−2470.

13. Lee,J.S., Johnson,D.A. and Morgan,A.R. (1979) *Nucleic Acids Res.*, **6**, 3073–3091.
14. Antao,V.P., Gray,D.M. and Ratliff,R.L. (1988) *Nucleic Acids Res.*, **16**, 719–738.
15. De los Santos,C., Rosen,M. and Patel,D. (1989) *Biochemistry*, **28**, 7282–7289.
16. Rajagopal,P. and Feigon,J. (1989) *Nature*, **339**, 637–640.
17. Rajagopal,P. and Feigon,J. (1989) *Biochemistry*, **28**, 7859–7870.
18. Caddle,M.S., Lussier,R.H. and Heintz,N.H (1990) *J. Mol. Biol.*, **211**, 19–33.
19. Christophe,D., Cabrer,B., Bacolla,A., Targovnik,H., Pohl,V. and Vassart,G. (1985) *Nucleic Acids Res.*, **13**, 5127–5144.
20. Hanvey,J.C., Shimizu,M. and Wells,R.D. (1988) *Proc. Natl. Acad. Sci. Usa*, **85**, 6292–6296.
21. Hanvey,J.C., Shimizu,M. and Wells,R.D. (1989) *J. Biol. Chem.*, **264**, 5950–5956.
22. Htun,H. and Dahlberg,J.E. (1988) *Science*, **241**, 1791–1796.
23. Htun,H. and Dahlberg,J.E. (1989) *Science*, **243**, 1571–1576.
24. Johnston,B.H. (1988) *Science*, **241**, 1800–1804.
25. Usdin,K. and Furano,A.V. (1989) *J. Biol. Chem.*, **264**, 15681–15687.
26. Voloshin,O.N., Mirkin,S.M., Lyamichev,V.I., Belotserkovskii,B.P. and Frank-Kamenetskii,M.D. (1988) *Nature*, **333**,475–476.
27. Weinreb,A., Collier,D.A., Birshtein,B.K. and Wells,R.D. (1990) *J. Biol. Chem.*, **265**, 1352–1359.
28. Wells,R.D., Collier,D.A., Hanvey,J.C., Shimizu,M. and Wohlrab,F. (1988) *FASEB J.*, **2**, 2939–2949.
29. Hanvey,J.C., Klysik,J. and Wells,R.D. (1988) *J. Biol. Chem.*, **263**, 7386–7396.
30. Kroeker,W.D. and Kowalski,D. (1978) *Biochemistry*, **17**, 3236–3243.
31. Baumann,U., Frank,R. and Blöcker,H. (1986) *Eur. J. Biochem.*, **161**, 409–413.
32. Greene,J.R. and Guarente,L. (1987) *Methods Enzymol.*, **152**, 512–522.
33. Cobianchi,F. and Wilson,S.H. (1987) *Methods Enzymol.*, **152**, 94–110.
34. Miller,H. (1987) *Methods Enzymol.*, **152**, 145–170.
35. Fritz,H.-J. (1985) In Glover, D.M. (ed.), DNA Cloning-A Practical Approach. IRL Press, Oxford, Vol.I, pp. 151–163.
36. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
37. Zoller, M.J. and Smith, M. (1987) *Methods Enzymol.*, **154**, 329–350.
38. Kohwi,Y. (1989) *Nucleic Acids Res.*, **17**, 4493–4502.
39. Shehee,W.R., Loeb,D.D., Adey,N.B., Burton,F.H., Cavasant,N.C., Cole,P., Davies,C.J., McGraw,R.A., Schichman,S.A., Severynse,D.M., Voliva,C.F., Weyter,F.W., Wisely,G.B., Edgell,M.H. and HutchisonIII,C.A. (1989) *J. Mol. Biol.*, **205**, 41–62.
40. Glanville,N., Durnam, D.M. and Palmiter,R.D. (1981) *Nature*, **292**, 267–269.
41. Côté,J., Renaud,J. and Ruiz-Carrillo,A. (1989) *J. Biol. Chem.*, **264**, 3301–3310.
42. Margot,J.B., Demers,G.W. and Hardison,R.C. (1989) *J. Mol. Biol.*, **205**, 15–40.
43. Weintraub,H. (1983) *Cell*, **32**, 1191–1203.
44. Moser,H.E. and Dervan,P.B. (1987) *Science*, **238**, 645–650.
45. Cantor,C.R. and Schimmel,P.R. (1980) Biophysical Chemistry Part I: The Conformation of Biological Macromolecules. Freeman,W.H. and Company, New York, U.S.A., pp. 328–337.
46. Sheardy,R.D. and Winkle,S.A. (1989) *Biochemistry*, **28**, 720–725.
47. Saenger,W. (1984) Principles of Nucleic Acid Structure. Springer-Verlag, New York.
48. Nussinov,R., Shapiro,B., Lipkin,L.E. and Maizel Jr.,J.V. (1984) *J. Mol. Biol.*, **177**, 591–607.
49. Mace,H.A.F., Pelham,H.R.B. and Travers,A.A. (1983) *Nature*, **304**, 555–557.
50. Lyamichev,V.I., Mirkin,S.M. and Frank-Kamenetskii,M.D. (1987) *J. Biomol. Struct. Dyn.*, **5**, 275–282.
51. Johnson,A.C., Jinno,Y. and Merlino,G.T. (1988) *Mol. Cell. Biol.*, **8**, 4174–4184.
52. Lichtler,A., Stover,M.L., Angilly,J., Kream,B. and Rowe,D.W. (1989) *J. Biol. Chem.*, **264**, 3072–3077.
53. Fitch,D.H.A., Mainone,C., Goodman,M. and Slightom,J.L. (1990) *J. Biol. Chem.*, **265**, 781–793.