Locations and contexts of sequences that hybridize to poly(dG-dT).(dC-dA) in mammalian ribosomal DNAs and two X-linked genes

Douglas C.Braaten, James R.Thomas, Randall D.Little, Kevin R.Dickson, Ilya Goldberg, David Schlessinger, Alfredo Ciccodicola¹ and Michele D'Urso¹

Department of Microbiology and Immunology, Washington University School of Medicine, St Louis, MO 63110, USA and ¹International Institute of Genetics and Biophysics, Naples 80125, Italy

Received October 29, 1987; Accepted December 21, 1987

ABSTRACT

Sequences located several kilobases both 5' and 3' of the stably transcribed portion of several genes hybridize to radio-labeled pure fragments of the alternating sequence poly (dG-dT)(dC-dA) ["poly(GT)"]. The genes include the ribosomal DNA of mouse, rat, and human, and also human glucose-6-phosphate dehydrogenase (G6PD) and mouse hypoxanthine-guanine phosphoribosyl transferase (HPRT). HPRT has additional hybridizing sequences in introns. Fragments that include the hybridizing sequences and up to 300 bp of adjoining DNA show perfect runs of poly(GT) (>30bp) in all but the human 5' region of rDNA, which shows a somewhat different alternating purine:pyrimidine sequence, poly(GTAT) (36bp). Within 150 bp of these sequences in various instances are found a number of other sequences reported to affect DNA conformation in model systems. Most marked is an enhancement of sequences matching at least 67% to the consensus binding sequence for topoisomerase II. Two to ten-fold less of such sequences were found in other sequenced portions of the nontranscribed spacer or in the transcribed portion of rDNA. The conservation of the locations of tracts of alternating purine:pyrimidine between evolutionarily diverse species is consistent with a possible functional role for these sequences.

INTRODUCTION

Eucaryotic genes contain large amounts of nontranscribed spacer DNA (NTS). For example, ribosomal DNAs encoding 18S, 5.8S, and 28S rRNAs are organized in tandem repeats, two-thirds of which are not transcribed (1,2,3). The function of NTS DNA remains obscure, but some discrete elements have been defined that include transcriptional regulatory elements like enhancers and promoters (4,5,6), sequences involved in the organization of chromatin loop structures (7,8), and repetitive elements such as the human Alu family (9).

Of the highly repetitive elements in spacer DNAs, tracts of alternating purines and pyrimidines have received special attention, since they can promote alternate conformations of DNA (10). The sequence poly(dG-dT)(dA-dC) [called poly(GT) below] can adopt the left-handed Z conformation in appropriate conditions in fixed cells and in pure DNA (11,12), though it is not known if Z-DNA exists in living cells.

We have asked whether any positions or contexts of poly(GT) tracts are conserved in mammalian genes. Our approach has been to detect candidate tracts of poly(GT) by the hybridization of radio-labeled poly(GT) to cloned portions of the genes, then to subclone and sequence the poly(GT) containing fragments.

A complete repeating unit of human ribosomal DNA and portions of mouse and rat rDNAs have been screened for the presence of poly(GT). In addition, cloned portions of two X-linked genes have been studied: human glucose-6-phosphate dehydrogenase (G6PD; 110 kb) and mouse hypoxanthine-guanine phosphoribosyl transferase (HPRT; 60 kb). Our approach is based on the notion that any conservation of the location of tracts of poly(GT) strongly suggests an intrinsic importance, since spacer sequences vary much more rapidly during evolution than do mature RNA sequences.

MATERIALS AND METHODS

Clones employed. The clones of mouse rDNA were as reported in ref. 2, covering 27 kb in three RI-RI fragments (Fig. 1). The human DNA clones employed included a 12 kb subfragment (Clone C) of a lambda clone provided by N. Arnheim (13), and a 7kb fragment (Clone B), 8kb fragment (Clone A), and 16kb (Clone D) fragments provided by Boncinelli et al. (14). Lambda Hurib 8, which contains Clone C and a portion of the neighboring Clone D, was provided by W. Salser (15). Lambda phages and cosmids covering the entire glucose-6-phosphate dehydrogenase gene have been reported (16), and were the sources of the corresponding DNAs. The phages covering portions of the HPRT gene were reported in ref. 17, and were provided by Dr. C.T. Caskey, Baylor College of Medicine.

DNA electrophoresis, blotting and hybridization. Agarose gels contained 0.8-2% agarose, 1X TBE (50 mM tris borate/2.5 mM EDTA, pH 8.3), and were electrophoresed with 1X TBE as running buffer containing 0.5 ug/ml ethidium bromide. DNA in gels was first denatured for 30 min in 0.5 M NaOH/1.5 M NaCl, neutralized for 2 x 30 min in 0.5 M Tris-HCl/1.5 M NaCl, and then blotted to nitrocellulose for 16 hrs in 20X SSC (3M NaCl/ 0.3 M Na citrate). The nitrocellulose filters were dried in an 80°C vacuum oven for 90 min.

Nick-translation of poly(GT) (Pharmacia) was by standard techniques (18). Specific activities of nick-translated poly(GT) were typically $2.5-5.0 \times 10^7$ cpm/ μ g. Hybridizations were carried out in 5X SSPE, 2.5X Denhardt's solution (18) and 0.1 % sodium dodecyl sulfate (SDS) for 12-16 hr at 68°C with

Figure 1. The ribosomal DNA repeat unit of rat, mouse, and human. The uncloned portions of the rat and mouse 40 kb DNA repeat unit are indicated by hatched lines; cloned portions are indicated by solid lines. R indicates the Eco RI restriction sites. The positions of the sequences of mature 18S, 5.8S, and 28S rRNA sequences are shown along the transcription unit, below the DNA. The arrows indicate the approximate locations of sequences that hybridize to nick-translated poly(GT).

10^6 cpm/ml. Filters were washed in 2X SSC/0.1% SDS twice for 15 min and then in 0.2X SSC/0.1 % SDS twice for 15 min at 68°C.

Z-DNA antibody reactions. Goat anti-Z-DNA antibody (courtesy of Dr. David Stollar) was incubated at indicated concentrations with 0.5 ug of supercoiled or relaxed cloned DNAs at 37°C in phosphate buffered saline. After incubation the reaction was filtered through a Millipore nitrocellulose filter HAWP 02500 as previously described (12). Filters were dried at 80°C for 15 min; denatured in 0.5N NaOH/1.5M NaCl for 10 min; neutralized in 0.5 M Tris-HCl, pH 7.5, 1.5 M NaCl; and then baked at 80°C under vacuum. Filters were then hybridized to nick-translated pBR322, autoradiographed and scanned by

densitometry. Intensities as measurements of retention for each clone and pBR322 are given relative to that obtained with no anti-Z-DNA antibody present in the incubation.

DNA sequencing. Subfragments of DNA were isolated by agarose gel electrophoresis and were ligated into appropriate restriction sites in the polylinker of M13mp18. When necessary, fragments were blunt-end ligated into the Sma I site of the polylinker. Single-stranded templates were obtained by standard methods (Amersham) and then sequenced using a modification of the Sanger dideoxy chain termination method with ³⁵S-dCTP (Amersham) as label. Sequencing reactions were incubated at 50°C. Fragments cloned into pUC18 were sequenced by a method using double-stranded DNA (19). Sequencing gels contained 5% polyacrylamide/0.2% bisacrylamide, 7 M urea, 1X TBE and were run at a constant voltage of 1800 volts (50°C). Gels were treated with 10% acetic acid/ 10% methanol for 30 minutes, then dried at 80°C under vacuum for 1hr. Autoradiography was at room temperature for 24 hours using Kodak XAR film.

Computer homology searches for the topoisomerase II cleavage site consensus sequence. A computer program was developed in Pascal that allows a DNA sequence to be searched for a given level of match to the 15 bp consensus topoisomerase II cleavage site reported by Sanders and Hsieh (20). The program identifies all potential matches in a DNA sequence compared to another defined sequence (e.g., the consensus) at a given homology level (e.g., 11/15). Since certain nucleotides are excluded at given positions within the 15 bp consensus, the program excludes the matches containing those nucleotides at the defined positions. Accordingly, the program eliminates from the list of matches any with C or G at position 4 of the consensus, any with C at position 8 or G at 9. The remaining matches are then summed.

Composites of available sequence data for rat, mouse, and human rDNAs were searched with the program. Sequences for rat rDNA included 2501 bp of the 5' nontranscribed spacer (21), 1129 bp around the transcriptional promoter (22,23), 85 bp of the external transcribed spacer (4), the complete coding region for 18S, 5.8S, and 28S rRNAs and the internal transcribed spacers (24,25), and 4025 bp of 3' nontranscribed spacer (21). Mouse rDNA sequences included 1753 bp of the 5' nontranscribed spacer (26), sequences from the promoter region (27,28), the last 320 bp of the external transcribed spacer (1), complete coding regions for the 18S, 5.8S, 28S rRNA's and the internal transcribed spacer (1,29,30,31), and 1180 bp of 3'nontranscribed spacer (32). Human rDNA sequences included 1197 bp around the transcriptional promoter (28), complete coding regions for 18S and 28S rRNAs (25,33,34), 224 bp of the

internal transcribed spacers (33), and 2347 bp of the 3' nontranscribed spacer (14).

RESULTS

Poly(GT) tracts in ribosomal DNA. These studies were initiated after we observed that radio-labeled poly(GT) hybridized to spacer portions of the mouse rDNA, at or near sequences in nucleolar chromatin that were shown previously to be resistant to DNase I (35). Southern hybridization with poly(GT) gave positive signals: two approximately 2 and 3 kb upstream and two approximately 1.0 and 2.5 kb 3' distal to the transcription unit (Fig.1). Subfragments of those regions nearest the 5' and 3' ends of the transcription unit were then sequenced and tracts of poly(GT) confirmed (Fig. 2).

These results suggested that the locations of some poly(GT) tracts within the rDNA repeat might not be random. Southern hybridizations with radio-labeled poly(GT) and comparisons of mouse sequences to those published previously for rat rDNA (21,36) showed that poly(GT) sequences are indeed located at similar positions. In rat rDNA, two positive signals again were found at each end of the transcription unit (2 and 4kb upstream, and 0.5 and 2.0 kb downstream; Fig. 1).

To test this correlation further, we examined human rDNA for poly(GT) sequences. That choice is more informative than mouse or rat rDNA, since the evolutionary distance is great and all of the DNA representing a tandem repeat of human rDNA is available (in four subclones; see Materials and Methods). Looking at human rDNA enabled us to answer conclusively whether poly(GT) sequences are distributed throughout the NTS or only at preferred locations.

Portions of human rDNA, 5' and 3' of the transcription unit, consistently showed hybridization to poly(GT). As in rat and mouse, two hybridizing sequences map 3' distal to the human transcription unit. The location of the 3' elements varies somewhat among tandem repeats in some human rDNAs because a sequence 700 bp in length is duplicated as many as 4 times distal to the transcription unit in different repeats (14). As a result, the locations of poly(GT) sequences are displaced corresponding distances relative to the transcription unit. In an rDNA unit containing two copies of the repeated element, the poly (GT) signals are located approximately 3kb downstream of the transcription unit.

The hybridization signal from the region 5' proximal to the human rDNA transcription start site (Fig. 3) was much weaker than all the others observed. The hybridizing sequence, about 2kb upstream from the transcription

unit, proved to be a repeat of (GTAT) rather than (GT) (Fig. 2). The relatively weak hybridization of the 5' region to poly(GT) apparently results from base-pairing to poly(GTNT).

Sequences of the 5' and 3' poly GT tracts are given in Fig. 2. The sequenced tracts contain at least 11 consecutive GT dinucleotides, and at least 15 elements with no more than one mismatch. Theoretically, that would be sufficient to promote at least three double-helical turns of Z DNA (37); and some of the sequences extend further into additional poly (purine:pyrimidine) sequences.

Poly(GT) tracts in two X-linked genes. The comparable location of poly(GT) tracts in a number of rDNAs raises the question of whether similar positioning is seen in genes that encode messenger RNAs.

For G6PD, we examined over 110 kb of contiguous DNA. The transcription

MOUSE 5'

```
CCTCTAATTCTCTTTAAAAAAAACAAACAAAAAAAACCAAACAGTATGTATGTATG 60
TATATTTAGAAAGAAATACTAATCCATTAATAACTCTTTTTTCTAAATTCATGTCATTC 120
TTGTTCCACAAAGTGAGTTCCAGGACTTACCAGAGAAACCTGTGTTCAAATTTCTGTGT 180
TCAAGGTCAACCTGGCTTACAAAGTGAGTTCCAAGTCCGATAGGCTACACAGAAAAACCA 240
TATCTCAGAAAAAAAAGTTCCAAACACACACACACACACACACACACACACACACACA 300
CACACACACACACACACACACAGGCGCGCGGGGTCAT
```

MOUSE 3'

```
GGAATTACAGCTACCACCACCCGGCGACCTATTATTATAAGTATTTATAAAACGAGCGCG 60
CGCACGCTGCACGCACACACACACACACGCACACACACACACACACACACAAAGAAAA 120
ATTTTTAAAAAAGAGAGAGAAGAAAAGGGATAATTGGTCGATTTTCGCATTTATTTTAA 180
AAATTTATTTTAAAAAAGTGCATGTTTGTGTACAAAAAACTCGGTTAGGGACAACGTG 240
TCGACCCGCAAAACACAAAACAACAGCAAAACAAAACCAAAACAACCAAAACAACA 300
CAAAACCAAAACAACACAACCCAACCCAACCCAACCCAACCCAACCCAACCCAACC 360
CAACCCAACACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAACAAAAC 420
CAAAACAAAACAACACAAAACAACAAAACGAAAAACAGAAAAACAATCTAGATCAG 459
ATAATTTATCGACCACCAGACTTCTGACATCTGGTCGG
```

HUMAN 5'

TCAGCTCGGCCTCCGAGTAGCTGGATTACAGACAGTACAGCACAGCACAGCGCCCGGCTA 60
 ATGTTGTGTATTGTGAGTAGAGACGGGGTTTCACCATACTGGCCCTGTTGGTCTGACCGC 120
 CTCTTGATCCACCGGCCTTGGCTCCCAAAGTGAGGGGATGACAGGCTTGAGCACCGCGCA 180
 GGGCCCATTTATTTATTTATTTATTTATTTATTTAATTATGTATGTATGTATGTATGTAT 240
GTATGTATGTATGTATTATTAGTTATTA

HUMAN 3' proximal

CCTGTAGTCCCAGCTATTTGGGAGGCTGAGGTACCAGAGCGCAGAGACAGAAGACCAGG 60
 CGGGCCGGCCCCAAAAGAAAAGAAAATAAATAAACGAAAATTATTAATAAATAATGAATG 120
 AAGGAAGGGAGGAAGGATATACATACAGTGTATGTAAATGAAATGGGGCTTCGATACAT 180
 ATTCATCCATTAAGTAACATAATATAAACGTATTAATTATGGAAATATCATTTACATA 240
 GTTTTATCACTACGGGGT 300
GTGTGTGTAGATGTATGTTTCATACACGGCAGCGTTCAGAAAATAAGATTGAAAAAGGAA 360
 GGAATCGGCCGGGCATGGTGGCTCACATCTGTACTCCCAGCAGTCTTTGGGAGGCCGAGG 420
 CGGGCG

HUMAN 3' distal

CTCCCATCCCCATTACCTGAGACTATAGGCGCGCACCACCACACCGGCTGACTTTTATGT 60
 TGTTTCTCATGTTTTCCGTAGGTAGGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
TGTGTGTGTGTGTGTGTGTGTGTGTGTATCTATGTATGTACGTATGTATGTATGTATGTG 180
 AGTGAGATGGGTTTCCGGGGTTCATCATGTTGCCCCACGCTGGTCTCGAACTCCTGTCC 240
 TCAAGCAATCCGCCTGCCTCGGCGCCACACTGCTGCTATTACAGGCGTGAGACG 300
 CTGCGCCTGGCTCCTTCTACATTTGCCTGCCTGCCTGCCTGCCTGCCTATCAATCG 360
 TCTTCTTTTTAGTACGGCTGTGCTCTCGCTTTATTGTCAGTCTCTGGGCACAGTGGTCT 420
 CTTTTCAAACCTCTATGATTATTATTATTGTAGGCGTCATCTCACGTGTGCAGGT

Figure 2. The 5' and 3' sequences of segments of rDNA around and including the poly(GT) tracts that border the mouse and human transcription units. For the mouse sequences, a MnlI-MnlI subfragment of the 5' NTS and a TaqI-TaqI subfragment of the 3' NTS were subcloned into the M13mp9 polylinker and sequenced by a modified dideoxy chain termination method. For the human sequences, two Sau3A-Sau3A subfragments of the 3' NTS and a HaeIII-HaeIII subfragment of the 5' NTS were subcloned into the M13mp18 polylinker and sequenced. The poly (pyrimidine:purine) regions are underlined.

Figure 3. Southern hybridization of poly(GT) to clones rat, mouse, and human rDNAs that were digested with appropriate enzymes to show fragments that contained poly(GT). These fragments were further restricted to give small poly(GT) containing fragments that were then sequenced. Panel A shows an ethidium bromide stained gel (1 μ g/ml) that contains human 5' and 3' clones, lanes 1 and 2; mouse 5' and 3' clones, lanes 3 and 4; and rat 5' and 3', lane 5 and lanes 6 and 7, respectively. Panel B shows an autoradiogram of a nitrocellulose filter containing the fragments from the gel in panel A. Fragments that hybridize to poly(GT) are seen in each lane; sizes are determined by comparing the mobilities of the fragments to that of Lambda digested with Hind III. The human 5' clones do not show a hybridization signal under these washing conditions (0.5X SSC, 65° C, 2 x 30 min; see METHODS). A signal is seen when less stringent conditions are used.

unit (18 kb) lies in the middle of this region (Fig. 4) (16). In Southern analyses using nick-translated poly(GT), hybridization to only two fragments was observed in the 110 kb shown in Fig. 4. The signals are located near the ends of the transcription unit. One appears in a 550 bp RI-RI fragment 2.5 kb upstream from the start of transcription. The other is located in another much larger RI-RI fragment 2.8 kb downstream from the poly(A) addition site at the 3' end of the transcription unit. The 5' segment was cloned as an RI fragment, and the 3' segment was cloned as a 600 bp RI-RI* subclone of the 2.8kb RI-RI fragment. Sequences of the 5' and 3' (GT) fragments are shown in Fig. 4. The 5' segment includes a 44 nucleotide sequence of alternating purines and pyrimidines bracketed by tracts containing alternating purines on one strand and pyrimidines on the other (16 bases upstream and 5 downstream). At the 3' end, the purine:pyrimidine motif is extended by 14 alternating A and T residues and 16 alternating A and C residues.

In additional tests with 60kb of DNA that include the transcription unit

for another X chromosome-linked gene, HPRT (17), we again observed limited hybridization of poly(GT) probes. As in G6PD, radio-labeled poly(GT) hybridized proximal and distal to the transcription unit. [Additional signals were seen in two introns within the transcribed portion of the gene. These fragments have not yet been studied in detail. Regions in which the tracts have been partially localized (the HPRT map is not complete) are indicated in Fig. 4.]. The two NTS fragments that hybridized were cloned into compatible sites in pUC18 and sequenced (Fig. 4).

Sequence context of the poly(GT) tracts. The sequences around the various poly(GT) tracts show other repeated motifs. Stretches of poly(A) and poly(G), and regions enriched for AT are seen in various cases. A number of such sequences have been considered as possible bending points (38,39) at which DNA might adopt a conformation distinct from the classical B form.

The sequences near poly(GT) also include sequences specified for *Drosophila* as sites where topoisomerase II binds to DNA (40). Figure 5 shows that segments within 150 nucleotides of each poly(GT) sequence are considerably enriched for the consensus sequence inferred as the probable site of topoisomerase II binding. Homology at the level of 11/15 bases [67%; a figure used in several discussions (41,42)] was sought in analyses using sequences in Figures 2 and 4. Figure 5 shows that the sequences in the vicinity of poly(GT) contain 3 to 7.5 topoisomerase sites per 200 nucleotides in the 5 cases examined, much more than in the available sequences of transcribed (Tx) or other non-transcribed (NTS) DNA. In more detailed analyses of the available sequences of transcribed rDNA (data not shown), one short segment containing a comparable enrichment for putative topoisomerase II sites was detected just downstream of the promoter site in rat, mouse, and human rDNAs. The frequency across the rest of the transcribed portions was essentially low and constant. In fact, it was lower than the value found in random sequences generated with the same base ratios.

Mouse 5' poly(GT) sequences form Z-DNA in vitro. Poly(GT) sequences can assume a Z-DNA conformation in defined conditions. By placing sequences with the potential to form Z-DNA in a supercoiled plasmid, one can induce the transition from B to Z-form (40).

A supercoiled plasmid of the 5' mouse clone in pBR322 was isolated, and its ability to bind to goat anti-Z DNA antibodies was assessed by a filter binding technique described in Materials and Methods. The 5' mouse clone showed significant anti-Z-DNA antibody binding (Fig. 6). Supercoiled pBR322 alone and a portion of rDNA encoding 18S rRNA cloned into pBR322 and

containing no significant alternating purine-pyrimidine stretches gave negligible binding over the same concentrations of antibody. Thus, the Z-DNA antibody binding activity is present in the nontranscribed rDNA sequences and not in the vector sequences. Linearization of the 5' DNA-containing plasmid with a restriction enzyme that cleaved at only one site abolished antibody binding activity. That is consistent with a requirement for supercoiling to induce a B to Z transition in the NTS sequences needed for antibody binding (37).

A

HPRT 5'

CGGACTCTAGAGGATCCCCCAGCATGGAAGTGGGCAGGACTAGGTGGAGACAGAAGAAT 60

TATGCGGCTGTCCAGGTGCCAGACTGCTCCATGATCAGTAAGAGATTCTGTCTTAAACAA 120

GGAATAAAGTAGAGAATGACAGAGCCTTCTCTAGCCCCATGCTCACAGGCACACATATT 180

GAATATGTACAGATACATCACCG 240

TTCCCAAGCGCACACACAGAAGTAAAACTTTTTTTTTTAAAAAAATCAATGTTAAGAAT 300

AAGTGTTTTCTGTAATATTTCTTGTTGAAATATATACGTATATACATCTTTTCTTTTCT 360

TTTTTGAGATAGAGTCTTCTGTAGCCAGCCTGCCCTCAAATTTGCTGC

HPRT 3'

GAAGTCTAGAGGATCATCAGCAAGCTCCAGGCTTAGTAAGGAACTCCAATTCAGTAAATAA 60

AATGAAGATTGAGGAAACCATCAAATGACAAGTCTGTCTCAACAAGCACAAACCACACA 120

CACACACACACACACACACACTACATACACTATGTATGCAAAGTAAAAGTTCTAGATA 180

AAGGACGCCATGTGTGTTACAAGCAAATCTTTGAATGGGGGGCAGCCGGGGCGGGGGTGT 240

TAGGAAAGGAAGTGTGCGGAAAGCTATCAAGTTGAAGTCACTCAGTCACGAGTCAGATG 300

AGGAGCCGTACA

B**G6PD** **G6PD 5'**

GAATTCTGTCCAGTCTCATGCCTTGAACCTCCAATCTTGATGCTTAATTAGACCTCTTTTC 60
TAAACTCCACATTTCATTTATCAAACCTCGACATTTCTACTTGGTTATCTAGTAGCCTTCTC 120
AAAACGAACCTCTCTAAGACCTCCGACAGTCCCCAAAACCCAGATCTACCCAAGGCCTTC 180
TTTGTTCATCAGCTGATGGCGAATCCATCTCCAGCTGTCAAAGCCTTTTGTATTCTCTC 240
TCTCTCTCTCACACACACACACACATACACACACACACACACACACAAGAGCAC 300
ATCTAATCCATCAAAGAAATCCTGTTAAATGTATCCCCAACCGATCTGATCACACTGA 360
GCAATTAACATTAACTATTTTAGGTGTCATAATGATATTNTGGTTGTTTTTCAGAGTCCT 420
GTCTTAGAGAGAGGCACAACCTAAATATTTAAAGATGATGACTGAATTC

G6PD 3'

GAATTCTGGCTTTCTGTAAGGACAGTGAGACTCTATGGCTGGACACAGTGTCTCACTCCT 60
GTAATCCCAGCACTTTGGGCATTTTGCTTGAGCCCAAGAGTTCCAGACCAGCAGAGCAAC 120
ATGATGAAACCTGTTTCTACAATAATACAAAATTAGCCGGGCATGTTGGCACATGCCT 180
GTAGTCCTAGCTACTCAGGAGACTGAGGTGGGAGGTTCACTTGGGCCATGATCACACCAT 240
TGCTATCAGCCGGGCTACACAGCAAGACCCTGTCTCAATATATATATAATATATATATAT 300
ACACACACACACATTATACACACACACACACACACATTTGCAAGGCCAAATTTTCAG 360
AGAAAAGTAGGACCCCATGATACAAATTAACCCCAAGTTGCCTTTGCCACCCAAGAAC 420
TTTTCTTGACAGCCCTATAAAGCGAAGAGGACAGAAATCAAATCTTGAGGCCCCCACT 480
TGCTGCCCCATCTTACTGAGAGGAAACCCAGAATTGCAACTGAGCCAGGGTGCAATTC

Figure 4. Sequences of the poly(GT) tracts of mouse HPRT (A) and human G6PD (B). The approximate locations of the sequences are indicated on the schematics. Cloning strategies are described in RESULTS.

DISCUSSION

Because genomic DNA contains so much poly(GT), it is not surprising to find copies within the rDNA repeat unit. Our results indicate, however, that

Figure 5. DNA fragments were screened for the number of matches to the consensus binding sequence for topoisomerase II. Specifications for defining a match, and the description of sequences used, are described in Materials and Methods. G6PD=glucose-6-phosphate dehydrogenase; HPRT=hypoxanthine guanine phosphoribosyl transferase; Rat, Mouse, and Human=rDNAs. Tx=transcribed sequences; NTS=non-transcribed spacer sequences leaving out the poly(GT) sequences shown in Figs. 2 and 4; 5'3'GT=5' and 3' poly(GT) sequences reported in Figs. 2 and 4 (see text); for the case of rat, poly(GT) sequences used were taken from reported sequences that included 5' and 3' nontranscribed spacer DNA (ref. 36).

the locations of poly(GT) sequences are similar among the ribosomal DNAs of rat, mouse, and human. The results are more striking at the 3' ends of the transcription units (Figure 1); but even in the exceptional case of the 5' end of human rDNA, the observed alternative sequence of poly(GTAT) retains the alternating purine:pyrimidine character. We infer that the similar locations of alternating purines and pyrimidines are maintained evolutionally, and are thus likely to have functional significance.

In the three ribosomal DNAs and two X-linked genes, poly(GT) tracts are found in comparable locations upstream and downstream of the transcription units. The most suggestive distributions detected are in the G6PD and human rDNA units, where 110 and 40 kb, respectively, have been screened for poly(GT) sequences, and the only signals are those that border the transcription unit. Since the X-linked genes were chosen only on the basis of the availability of

Figure 6. Binding of anti-Z-DNA antibody to the mouse 5' NTS clone. A filter binding assay was used as described in Materials and Methods to detect DNA-antibody complexes. Equivalent molar concentrations of supercoiled plasmids of clones 1, 5, 9, and pBR322 were assayed against increasing concentrations of goat anti-Z-DNA (courtesy of D. Stollar). The clones are as defined in Thomas et al. (37), with clone 1 and 9 containing the 5' and 3' poly(GT) segments, and clone 5 containing part of the external transcribed spacer sequence and the adjoining initial 918 nucleotides of 18S rDNA. The binding of each clone was measured at given concentrations of antibody and expressed relative to that obtained with no antibody. Panel A shows a typical filter binding assay for supercoiled clone 1 plasmid. Upon linearization of the plasmid with Eco RI it shows a marked decrease in anti-Z-DNA antibody binding (panel B).

cloned DNA, it is clear that many genes are likely to have poly(GT) sequences near the borders of the transcription unit. It is suggestive that a number of other sequences which could promote alternate conformations of DNA are found nearby. Alternating purine:pyrimidine sequences, especially poly(GT), might occur in specialized regions that serve some function of gene punctuation, bordering a unit of potentially active chromatin. That would be analogous to the way in which poly(AT) (44) or CpG-rich clusters (45,46) have been suggested to define the "edges of a complete gene".

GenBank contains about 100 cases in which poly(GT) sequences have been found in intronic or spacer sequences of genes encoding mRNAs (our results add the case of HPRT to those with poly(GT) in introns). It is of particular interest that recent studies have revealed poly(GT) tracts at similar locations in several genes, including globins (47) and protamines (48). As expected, the number reported increases sharply if one relaxes the search to score for any poly (pyrimidine:purine) sequence (48).

Pardue et al. (49) have recently produced a low resolution map of the general distribution of poly(GT) along *Drosophila* chromosomes. Considerable enrichment is seen in characteristic regions, and is conserved in various types of *Drosophila*. Our results suggest that conservation of certain locations and sequence contexts of poly(GT) may extend from the chromosomal level to the level of individual genes.

Are other repetitive sequences conserved in rDNA NTS? Studies of rDNAs have progressed far enough to show that some sequences unique to rDNA are repeated within the tandem unit. One thought to be involved in the termination of transcription has been found in multiple copies at the 3' end of the transcription unit of rat, mouse, and human rDNA (14,21,32). Another similar element is repeated upstream of the promotor in mouse, rat, and possibly human rDNA, and may also have a role in transcription (26).

Hadjiolov (21) and Rothblum (36) have found that other sequences as repetitive as poly(GT) in the genome, including the "B2" and "ID" sequences, are located in an ordered array outside of the poly(GT) segments at the edges of the transcription unit in rat rDNA. Mouse and human rDNA seem to deviate from this pattern [for example, in work in progress we find that mouse rDNA contains B1 sequences, not B2 as in rat, and human rDNA contains segments that hybridize to ALU1 5' of the upstream poly(GT) region (15), but no signals are detected in the 3' region]. More complete sequence data, however, may show greater conservation of positions of repetitive sequences.

Functional tests of poly(GT). Recent studies suggest possible direct

roles for poly(GT) in the organization and function of genes. For example, Laemmli et al. (41) have pointed out that consensus sequences for topoisomerase II are often nested with putative signals for association with the nuclear matrix and other sequences that can promote changes in DNA conformation. Also, they show both for mRNA-(7,41) and rRNA-(7) encoding genes that the transcription unit in chromatin is often exposed selectively to cleavage by DNase I and restriction enzymes. Such studies have led to models in which loops of the transcription units in DNA are organized by adjacent regions that bind to a structural matrix.

In related studies (7,8, and work in progress), we have found that segments of mouse rDNA neighboring the poly(GT) are more protected against DNase I action than is the stably transcribed portion. One speculation is that the action of topoisomerase II near the poly GT sequences might indeed promote a transition of poly(GT) to the Z conformation, and this might in turn promote a loop structure containing segments at its base protected from nuclease action. However, at present there is no evidence to indicate that poly(GT) adopts the Z conformation *in vivo* or has a function in the formation of chromatin loops. The organization of chromatin more likely depends on additional sequences, perhaps including the topoisomerase II binding sites, as suggested by several groups (8,41,42,50).

Other functions for poly(GT) have been suggested by direct assays that could also be relevant to its location near the borders of transcription units. First, short stretches of alternating purine:pyrimidine can bind factors that promote mRNA formation early in SV40 viral development (51). A possible general role for poly(GT) in transcription, perhaps as part of an "enhancing" function, has therefore been considered, yet no direct evidence for such a function has been reported. Moreover, in work in progress, ribosomal DNA transfected into cultured mammalian cells shows the same high level of transcription with constructs containing or lacking the 5' promoter-proximal poly(GT) sequence.

A second type of functional assay has shown that alternating purine:pyrimidine sequences may create hot spots for recombination. Such sequences increased recombination in the experiments of Treco and Arnheim (52) and of Bullock et al (53), and an enzyme implicated in recombination in *Ustilago* binds specifically to poly(GT). In addition Pardue et al. (49) have noted that poly(GT) is found in *Drosophila* chromosomal regions that undergo meiotic recombination frequently. The notion that highly repetitive elements may be

hot spots for recombination is also supported by the direct observation of one such event (54,55,56).

A general mechanism might include selective pressure that would operate to locate both binding sites for topoisomerase II and tracts of poly(GT) (or other highly repetitive sequences) at the border of the stably-transcribed portion of a gene. Matrix-binding sequences there might help to organize a loop of chromatin; poly(GT) could aid in transferring a unit intact during recombination. This could lessen the incidence of addition-deletion mutations that often accompany recombination events within a gene; and in the case of ribosomal DNA, it could help to maintain the homogenization of the tandem rDNA repeats in nucleolar organizers.

REFERENCES

1. Goldman, W.E., Goldberg, G., Bowman, L.H., Steinmetz, D. and Schlessinger, D. (1983) *Mol. Cell. Biol.* 3, 1488-1500.
2. Bowman, L.H., Rabin, B. and Schlessinger, D. (1981) *Nucl. Acids Res.* 9, 4651-4666.
3. Cory, S. and Adams, J.M. (1977) *Cell* 11, 795-805.
4. Cassidy, B.G., Subrahmanyam, C.S. and Rothblum, L.I. (1982) *Biochem. Biophys. Res. Commun.* 107, 1571-1576.
5. Reeder, R.H. (1984) *Cell* 38, 349-351.
6. Hamada, H., Seidman, M.V., Howard, B.H. and Gorman, C.M. (1984) *Mol. Cell. Biol.* 4, 2622-2630.
7. Keppel, F. (1986) *J. Mol. Biol.* 187, 15-21.
8. Gasser, S.M., Laroche, T., Falquet, J., Boy de la Tour, E., and Laemmli, U.K. (1987) *J. Mol. Biol.* 188, 613-620.
9. Schmid, C.W. and Jelinek, W.R. (1982) *Science* 216, 1065-1070.
10. Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6465-6469.
11. Nordheim, A. and Rich, A. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1821-1825.
12. Nordheim, A., Lafer, E.M., Peck, L.J., Wang, J.C., Stollar, B.D. and Rich, A. (1982) *Cell* 31, 309-318.
13. Brownell, E., Krystal M. and Arnheim, N. (1983) *Mol. Cell. Biol.* 1, 29-37.
14. La Volpe, A., Simeone, A., D'Esposito, M., Scotto, L., Fidanzo, V., de Falco, A., and Boncinelli, E. (1985) *J. Mol. Bio.* 183, 213-233.
15. Higuchi, R., Stang, H., Browne, J., Martin, M., Huot, M., Lipeles, J., and Salser, S. (1981) *Gene* 15, 177-186.
16. Martini, G., Toniolo, D., Vulliamy, T., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M., and Persico, M. (1986) *EMBO* 5, 1849-1855.
17. Melton, D., Konecki, D., Brennand, J., and Caskey, C. (1984) *Proc. Natl. Acad. Sci. USA* 81, 2147-2151.
18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Lab., Cold Spring Harbor, N.Y.
19. Hattori, M. and Sakalei, Y. (1986) *Analyt. Bioch.* 152, 232-238.
20. Sander, M. and Hsieh, T. (1985) *Nucl. Acids Res.* 13, 1057-1072.
21. Yavachev, L.P., Georgiev, O.I., Braga, E.A., Ardonina, T.A., Bogomolov, A.E., Zhurkin, V.B., Nosikov, V.V. and Hadjiolov, A.A. (1986) *Nucl. Acids Res.* 14, 2799-2810.
22. Rothblum, L.I., Reddy, R. and Cassidy, B. (1982) *Nucl. Acids Res.* 10, 7345-7362.

23. Harrington, C.A. and Chikaraishi, D.M. (1983) *Nucl. Acids Res.* 11, 3317-3332.
24. Hadjiolov, A.A., Georgiev, O.I., Nosikov, V.V. and Yavachev, L.P. (1984) *Nucl. Acids Res.* 12, 3677-3692.
25. Torczynski, R., Fuke, M., and Bollon, A. (1985) *DNA* 4, 283-291.
26. Kuehn, M. and Arnheim, N. (1983) *Nucl. Acids Res.* 11, 211-224.
27. Bach, R., Grummt, I. and Allet, B. (1981) *Nucl. Acids Res.* 9, 1559-1566.
28. Urano, Y., Kominami, R., Mishima, Y., and Muramatsu, M. (1980) *Nucl. Acids Res.* 8, 6043-6058.
29. Hassouna, N., Michot, B. and Bachellerie, J.P. (1984) *Nucl. Acids Res.* 18, 3573-3583.
30. Michot, B., Bachellerie, J.P. and Raynal, F. (1982) *Nucl. Acids Res.* 10, 5273-5283.
31. Raynal, F., Michot, B. and Bachellerie, J.P. (1984) *FEBS Lett.* 167, 263-268.
32. Grummt, I., Maier, U., Ohrlein, A., Hassouna, N. and Bachellerie, J.P. (1985) *Cell* 43, 801-810.
33. Gonzalez, I., and Schmickel, R. (1986) *Am. J. Hum. Genet.* 38, 419-427.
34. Gonzalez, I., Gorski, J., Campen, T., Dorney, D., Erikson, J., Sylvester, J., and Schmickel, R. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7666-7670.
35. Thomas, J.R., Bolla, R.I., Rumblyrt, J.S. and Schlessinger, D. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7595-7598.
36. Yang-Yen, H., Subrahmanyam, C., Cassidy, B., and Rothblum, L. (1985) *J. Mol. Biol.* 184, 389-398.
37. Rich, A., Nordheim, A. and Wang, A.H.-J. (1984) *Ann. Rev. Biochem.* 53, 791-846.
38. Koo, H., Wu, H. and Crothers, D. (1986) *Nature* 320, 502-506.
39. Cockerill, R.N. and Garrard, W.T. (1986) *Cell* 44, 273-282.
40. McClellen, J., Palecek, E. and Lilley, D. (1986) *Nucl. Acids Res.* 14, 9291-9309.
41. Gasser, S.M., Laemmli, U.K. (1987) *Trends in Genetics* 3, 16-22.
42. Gasser, S.M., Laemmli, U.K. (1986) *Cell* 46, 521-530.
43. Lafer, E.M., Sousa, R. and Rich, A. (1985) *EMBO J.* 4, 3655-3660.
44. Moreau, J., Marcaud, L., Maschat, F., Kejzlarova-Lepesant, J., Lepesant, J.-A., Scherrer, K. (1982) *Nature* 295, 260-262.
45. Gardiner-Garden, M. and Frommer, M. (1987) *J. Mol. Biol.* 196, 261-282.
46. Bird, A.P. (1986) *Nature* 321, 209-213.
47. Miesfeld, R., Krystal, M. and Arnheim, N. (1981) *Nucl. Acids Res.* 9, 5931-5947.
48. Aiken, J., Miller, F., Hagen, F., Nckenzie, Krawetz, S., van de Sande, J., Rattner, J.B., Dixon, G. (1985) *Biochemistry* 24, 6268-6276.
49. Pardue, M.L., Lowenhaupt, K., Rich, A. and Nordheim, A. (1987) *EMBO J.* 6, 1781-1789.
50. Udvardy, A., Schedl, P., Sander, M., and Hsieh, T. (1986) *J. Mol. Bio.*, 231-246.
51. Nordheim, A., and Rich, A. (1983) *Nature* 303, 674-679.
52. Treco, D. and Arnheim, N. (1986) *Mol. Cell. Biol.* 6, 3934-3947.
53. Bullock, P., Miller, J. and Botchan, M. (1986) *Mol. Cell. Biol.* 6, 3948-3953.
54. Lehrman, M.A., Goldstein, J.C., Russell, D.W. and Brown, M.S. (1987) *Cell* 48, 827-835.
55. Hyrien, O., Debitasse, M., Buttin, G., and Robert de Saint Vincent, B. (1987) *EMBO* 6, 2401-2408.
56. Stringer, J.R. (1985) *Mol. Cell. Biol.* 5, 1247-1259.