
Transcription of two classes of rat growth hormone gene-associated repetitive DNA: differences in activity and effects of tandem repeat structure

Arthur Gutierrez-Hartmann⁺, Ivan Lieberburg*, David Gardner, John D. Baxter and Guy G. Cathala[§]

Metabolic Research Unit and Departments of Medicine and Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Received 1 June 1984; Revised and Accepted 16 August 1984

ABSTRACT

The rat growth hormone (rGH) gene contains two classes of repetitive DNA arranged as clusters within intron B and the 3' flanking region. The major family is equivalent to the CHO type 2 DNA. The second ("truncated repeat", TR) is a truncated version of the first and occurs in certain neural-specific transcripts and genes ("identifier" elements, ID). Here we report, using the HeLa cell-free transcription assay, that RNA polymerase III (Pol III) efficiently initiates at internal promoters within a tandem array of rGH gene repetitive DNA monomers and results in a novel organization of overlapping Class III transcription units. Transcription competition studies revealed that the rat type 2 structures share Pol III transcription factors with a tRNA gene, a human Alu repeat, and a mutant VA1 gene. Also, the rGH type 2 but not the TR DNA efficiently promotes Pol III initiation, yet other TR members, which differ only in flanking DNA, are transcribed. Thus, the rGH gene is strikingly enriched with 10 repetitive DNA monomers; multimeric type 2 elements are actively transcribed; rGH-TR sequences are expressed only as part of larger transcripts promoted by type 2 DNA; and, type 2 DNA uses tRNA gene transcription factors. These studies show that flanking sequences, promoter organization and factor competition may all affect rat repetitive DNA expression.

INTRODUCTION

The major component of mammalian short-length, middle-repetitive DNA is the Alu-equivalent family which consists of 10^5 closely-related sequences comprising 3-5% of the total DNA (1-5). These are mostly interspersed (and not clustered) throughout the genome both flanking and within genes, and are frequently flanked by direct repeats suggesting their recent insertion (1,2). Most individual members of the Alu-repetitive DNA families contain in vitro RNA polymerase III (Pol III) promoter activity, whereas the Chinese hamster ovary (CHO) type 1 repeat family does not have such activity (1-4). In the rat, the Pol III template activities of only one class have been tested directly (3). In spite of the large quantity of this DNA, very little is known about its functions; whereas roles in regulation of gene expression, transcription and recombinational events have been proposed (1-3), more data are needed related to the mechanisms by which this DNA can be expressed.

Alu-like middle repetitive DNA found in association with the rat growth hormone (rGH) gene has unique features of interest. First, unlike most repetitive DNA, many of the rGH repeats occur in a tandem arrangement. This may be of particular interest in terms of expression, since in the cases studied to date in which multiple Pol III promoters occur naturally as tandemly arranged tRNA genes, Pol III transcription is initiated only with the first (i.e., the most 5') gene (16-18). Secondly, sequences of one of these repeat elements have been found in certain rat genes expressed uniquely in the brain and also comprise the major structure of a 160 nucleotide (nt) neural-specific RNA (3,9-11). This led Sutcliffe et al. to label these elements as brain "identifier" (ID) sequences (9-11) and to speculate that these sequences in some way are involved in regulation of brain functions. One of these ID sequences has been shown to have Pol III promoter activity *in vitro* (3). Thus, information about the mechanisms of expression of this DNA may not only be critical for understanding brain- and pituitary-specific expression, but may also help in defining better the biological role of repetitive DNA.

The repetitive DNA associated with the rGH gene has been found both in intron B (I_B) and in the 3'-flanking region (3'FR). The I_B DNA sequence in the Sprague-Dawley rat contains 508 bp of three repeat monomers with a head-to-tail array of two repeats that have 86% homology to the CHO type 2 repetitive DNA class (4,7). The third monomer is truncated after the first 73 bp (which are 66% homologous to the type 2 DNA) by an A-rich segment; we have termed this DNA the truncated repeat (TR) family (Fig. 1). Another single TR is found 300-500 bp beyond the rGH gene polyadenylation signal (Fig. 1) (7,8; G.G. Cathala, M.O. Showers, and J.D. Baxter, unpublished results). The putative "identifier" sequence is equivalent to the rGH TR sequences and therefore is a member of the TR family (3,7-15).

In the current study we used the cell-free transcription system of Manley et al. (19) and the unique organization of the Alu-like sequences associated with the rGH gene to: 1) determine the Pol III activity of the two major rat repetitive DNA families; 2) map the rGH gene Pol III transcription units; 3) determine if the Pol III promoters tandemly arranged within I_B behave as a single transcription unit; and, 4) determine, by competition studies, if the rat Alu-like repeats and tRNA genes share Pol III transcription factors.

MATERIALS AND METHODS

Construction of rGH gene subclones and truncated DNA templates.

The 5.8 kb EcoRI-HindIII rGH gene-containing fragment in prGHeh-5.8 was used (7). The 1250 bp and 1350 bp PstI fragments of prGHeh-5.8 cloned into the PstI site of pBR322 resulted in subclones prGH-1250 (containing I_B) and prGH-1350 (containing 3'FR) (Fig. 1C). The 425 bp fragment containing the single 3'TR was cloned into the HindIII site of pBR322 by digesting prGHeh-5.8 with PvuII and adding HindIII linkers. The *Drosophila* arginine tRNA genomic clone, pArg, and 5' deletion subclones, pArg 5.9 and pArg 5.12, were kindly supplied by Drs. S. Sharp and D. Söll (20). The human Alu-repeat clones, pPD8 and pPD6, were kindly provided by Drs. T. Friedmann and P. Deininger; adenovirus VA1-5'Δ55 DNA was a generous gift of Dr. R. Weinmann (5,21). Plasmid DNAs were purified by banding in CsCl gradients (22,23). Restriction enzymes were used as recommended by the supplier and fragments were phenol extracted and ethanol precipitated prior to use.

Cell-Free Transcription

Whole cell lysates were prepared from cultured HeLa cells as described by Manley et al. (19). The standard assay was performed as previously described (19) in a 25 μl reaction mixture containing 1 U/μl placental RNase inhibitor (Biotec), 5 μCi [³²P]GTP (ICN, 30 Ci/mmol), 15 μl of extract and DNA template as indicated. After 60 min at 30°C, labelled RNA products were isolated by three phenol/chloroform extractions, two ethanol precipitations and a final 70% ethanol rinse of the RNA pellet. The dried RNA pellet was dissolved in 90% formamide sample buffer and electrophoresed through a 5% acrylamide/8M urea sequencing-type gel at 35 watts for 2.5 hr.

S₁ Nuclease Mapping

For S₁ nuclease mapping, unlabeled I_B Pol III transcripts were prepared from a scaled-up (5X) transcription reaction containing 50 μg/ml of prGH-1250 and 2 μg/ml of α-amanitin. After a 2 hr incubation the template was digested with RNase-free DNase I previously purified by affinity chromatography on agarose-coupled 5'-(4 aminophenyl-phosphoryl) uridine 2'(3') phosphate (Miles-Yeds, Laboratories) as described by Maxwell (24). The RNA was co-precipitated in ethanol with either probe 1 (0.4 pMoles) or probe 2 (0.6 pMoles) described below. After drying, the ethanol precipitates were then dissolved in 20 μl of 0.5 M NaCl, 80% formamide, 40 mM Pipes (pH 6.4) and 1 mM EDTA, placed in a boiling water bath for 3 min and hybridized at the indicated temperature for 18 hrs. Each hybridization reaction was diluted 10X with ice-cold S₁ buffer containing 22 U/μl of S₁ nuclease (Miles

Laboratories), 0.25 M NaCl, 30 mM NaOAc (pH 4.6) and 4.5 mM ZnSO₄. The RNA-DNA hybrids were digested for 30 min at 37°C, then they were phenol-chloroform extracted, ethanol precipitated, dissolved in formamide sample buffer, heated, and electrophoresed on 8% polyacrylamide/8M urea sequencing gels. Probe 1 was prepared by digesting prGH-1250 with AvaII and BglIII, and probe 2 by digesting prGH-1250 with PstI and BglIII. The specific fragments were isolated from preparative gels, treated with calf intestinal alkaline phosphatase and 5'-end labelled with [γ -³²P]ATP (3000 Ci/mmol, Amersham) using bacteriophage T₄ polynucleotide kinase (P-L Biochemicals).

RNA Polymerase III Stable-Complex Competition Assay.

The standard cell-free transcription assay was pre-incubated for 5 min at 30°C with either 2 or 10 μ g of the competitor template (0.6 or 3 μ Moles of Pol III promoters, respectively) and "stable" Pol III transcription complexes were allowed to form. Then 1 μ g (or 0.5 μ Moles of Pol III promoters) of the prGH-1250 reference template was added to each reaction and the assay continued for another 55 min. Transcription stop buffer, containing ³²P-labelled U₁ RNA as an internal control for non-specific losses during the extraction and precipitation procedures, was then added. The cell-free transcripts were purified and analyzed on 5% sequencing gels as described in the previous section. Uncut plasmid DNAs were used in these assays. Gel-purified U₁ RNA (prepared as described in 25) was 3' end-labelled with RNA ligase and [γ -³²P] pCp as described by England et al. (26).

RESULTS

1. Pol III transcripts predicted from the rGH gene sequence. The Class III gene transcription unit is defined by two discontinuous and intragenic Pol III promoter elements, the A and B boxes, and a termination signal containing a stretch of at least four thymidine residues (20,21,27-34). Analysis of the Sprague-Dawley rGH gene revealed several putative Pol III promoters in I_B and the 3'FR resulting in multiple, potential Pol III transcription units. Based on data with multimeric tRNA genes, the tandem arrangement of the Pol III promoters found within the I_B repeats (Fig. 1A, boldface nucleotides) predicts a single 720 nt polycistronic transcript initiating at the beginning of the first I_B repeat and extending thru the remaining repeats and exon III to the first strong termination signal (CTCTTTTC) found 37 bp into intron C (I_C). However, a weak Pol III stop signal (ATCTTTA) is found 175 bp into the first and second long I_B repeat (Fig. 1A, underlined nucleotides), which if utilized predicts additional transcripts of 370 and 175 nucleotides. Indeed,

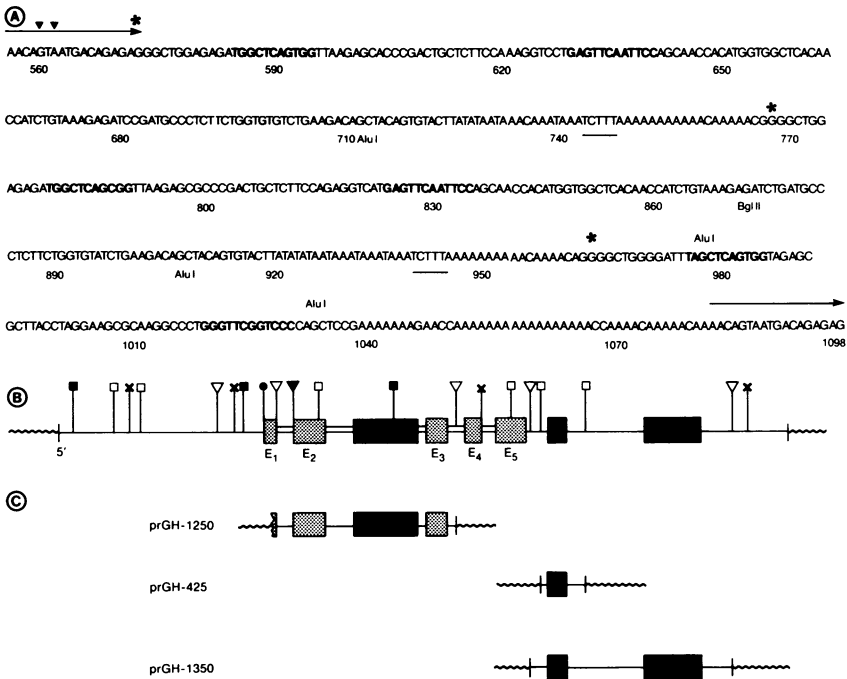


Figure 1 Structures of rGH gene fragments. (A) The nucleotide sequence of the three tandem repeats clustered within I_B. The 18 bp direct repeats are overlined with arrows. The beginning of each repeat is marked with an asterisk (*). The putative Pol III promoter sequences are shown in boldface and the "weak" Pol III stop sequences are underlined. The nucleotide numbering starts as +1 with the rGH gene RNA polymerase II CAP site (7). (B) Relevant restriction sites in the 5.8 kb EcoRI-HindIII rGH gene fragment in prGH_{eh}-5.8 are shown relative to exons (E₁ to E₅), introns (I₁ to I₄), 5' and 3' flanking DNA (-), I_B and 3'FR repetitive DNA (---) and pBR322 (wavy lines). Symbols for restriction sites include: ♯ XhoI, ∇ PstI, ♯ XbaI, ♯ PvuII, ♯ BglII, and ♯ KpnI. (C) The rGH gene subclones used in this study are aligned just below and with the same symbols as in B, to indicate the portions of the 5.8 kb fragment they contain.

Pol III initiation at each putative promoter (Fig. 1A, bold nucleotides) followed by multiple termination events at both the weak I_B (Fig. 1A, underlined nucleotides) and the strong I_C Pol III stop sequences predicts transcripts of 720, 360 and 175 nt initiating at the first repeat; 540 and 175 nt initiating at the second repeat and 330 nt initiating at the third, truncated repeat, as depicted in Fig. 4. The single TR in the immediate 3'FR has the same Pol III promoter sequences as the I_B TR. However, it contains a relatively good stop signal (GTTTTA) as the last 6 nucleotides of its 17 bp

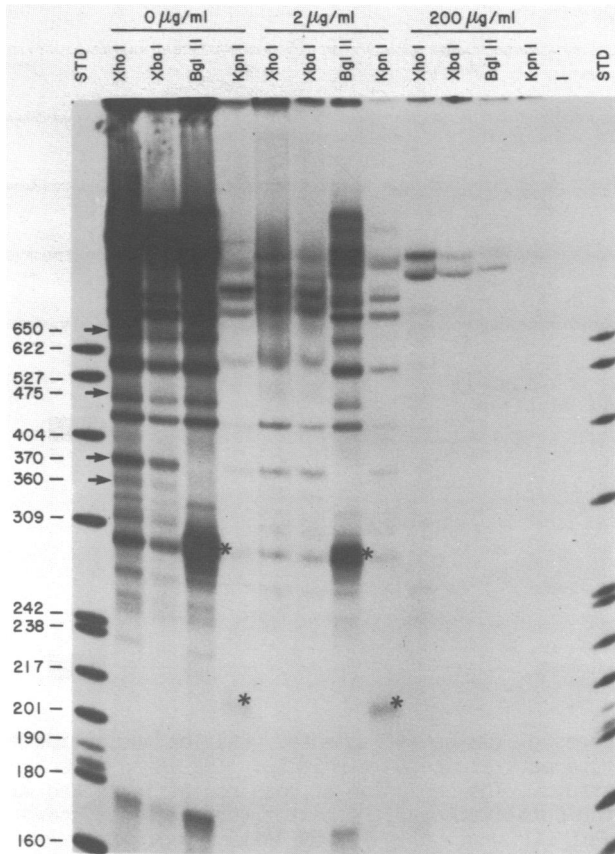


Figure 2 Cell-free transcription of rGH gene fragments. prGHeh-5.8 DNA, cleaved with the indicated restriction endonuclease, was transcribed in the presence of 0, 2, or 200 $\mu\text{g/ml}$ of α -amanitin at a DNA concentration of ca. 40 $\mu\text{g/ml}$ and the RNA products electrophoretically separated as described in METHODS. The lanes marked STD show ^{32}P -labelled, HpaII-cut pBR322 DNA size markers. The new transcripts generated from BglII (300-310 nt) and KpnI (217 nt) digested templates are marked with asterisks, and those deleted as a result of BglII (370 and 360 nt) or KpnI (650 and 465 nt) template digestion are indicated with arrows.

direct repeat which flanks the entire 110 bp repeat monomer (G.G. Cathala, M. O. Showers, and J.D. Baxter, unpublished results). Such a structure predicts a single Pol III transcript of about 120 nt. The several other repetitive DNA regions identified in the distal 3'FR of the rGH gene by hybridization studies (A. Barta, personal communication) have not been sequenced. In summary, there are at least six potential Pol III transcription units in I_B ,

one in the 3' TR and an undetermined number in the distal 3' flanking region.

2. Cell-free transcription of the rGH gene. An autoradiogram of the gel-separated, cell-free transcripts directed by various prGHeh-5.8 fragments (7,19) is shown in Fig. 2. Multiple, discrete transcripts of 175 to ca. 1300 nt are observed. By contrast, no transcripts are found in the absence of added DNA, and transcription of HindIII-linearized pBR322 results in transcripts from ca. 1100 to 1300 nt, most of which are sensitive to 2 µg/ml of α-amanitin (data not shown). Differences in intensities of the lanes are due to variable DNA recoveries after phenol extraction and prior to transcription. The transcription pattern with XhoI or XbaI generated fragments is similar. However, the BglII (which cuts within the I_B repeat cluster) generated DNA fragments produce more transcripts migrating in the 290-310 range (marked with an asterisk) and do not yield the 370-360 nt RNA pair (indicated with arrows). Similarly, the KpnI-cleaved template, which is cut within the distal 3'FR repetitive DNA, directs the synthesis of a new 217 nt product and not the 650 and 475 nt transcripts seen with the other templates. These data imply that Pol III transcription units are present within I_B and the 3'FR since truncation of prGHeh-5.8 by BglII and KpnI, respectively, generates shorter run-offs (4). It is unlikely that the observed transcripts arise from the rGH gene Pol II promoter since the predicted α-amanitin sensitive, Pol II run-off transcripts of 270 and 870 nt for XbaI-, and BglII-truncated prGHeh-5.8, respectively, are not found. In addition, the XhoI and XbaI generated templates yield the same pattern and yet XhoI cleaves the rGH gene at the CAP site. Instead, many of the transcripts are probably generated by Pol III since they are resistant to 2 µg/ml but not 200 µg/ml of α-amanitin. Also, the multiple Pol III rGH gene transcripts are most likely due to multiple discrete initiation and termination events and not processing, since the transcription pattern remains unchanged even after a 2.5 hr. incubation (data not shown).

3. Intron B and the 3'FR of the rGH gene contain multiple Pol III transcription units imbedded within repetitive DNA. Repetitive DNA regions identified within I_B and the 3'FR of the rGH gene by previous studies (7; A. Barta, personal communication) were found to hybridize with [³²P]-labelled cell-free transcripts directed by prGHeh-5.8 DNA (data not shown). These repetitive DNA regions were subcloned (Fig. 1C) and analyzed further.

Shown in Fig. 3 are autoradiograms of the [³²P]-labelled RNA products from the subcloned DNA. Both supercoiled or PstI-cut prGH-1350 (containing the distal 3'FR, lanes 3 and 4, Fig. 3A) and prGH-1250 (containing I_B

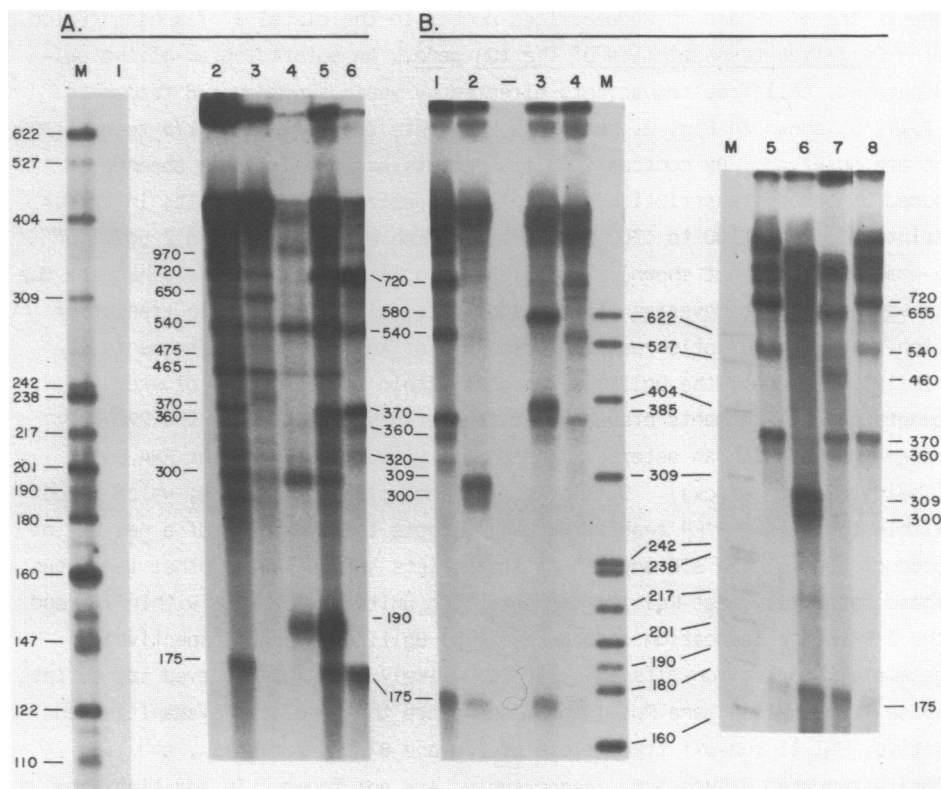


Figure 3 Cell-free transcription of subcloned DNA fragments containing rGH gene-associated repetitive DNA. (A). DNAs (Fig. 1C) were transcribed in the standard system and the RNA products analyzed as described in METHODS. Lanes 1-3: supercoiled prGH-425, prGHeh-5.8, and prGH-1350, respectively. Lanes 4-6: PstI-cut prGH-1350, prGHeh-5.8 and prGH-1250, respectively. (B). Cell-free transcription with circular or restriction enzyme truncated prGH-1250 plasmid DNA at 40 μ g/ml. Lane (1) uncut; (2) BglII; (3) AvaI; (4) Pvu II; (5) uncut; (6) BglII; (7) HpaII; and (8) XbaI. Lanes (M) contain 32 P-labelled HpaII-cut pBR322 DNA used for calibration.

sequences, lane 1, Fig. 3B and lane 6, Fig. 3A) DNA result in multiple discrete transcripts. Furthermore, a summation of the transcripts obtained with PstI-cut prGH-1350 (lane 4) and prGH-1250 (lane 6) results in all of the major transcripts obtained with PstI-cut prGHeh-5.8 (lane 5). Similar conclusions can be derived from data with supercoiled templates. These data indicate that the Pol III activity of the 5.8 kb rGH insert is dominantly confined to the I_B and distal 3'FR repetitive DNA.

The cell-free transcription of I_B in uncut vs. PstI-cut prGH-1250 DNA

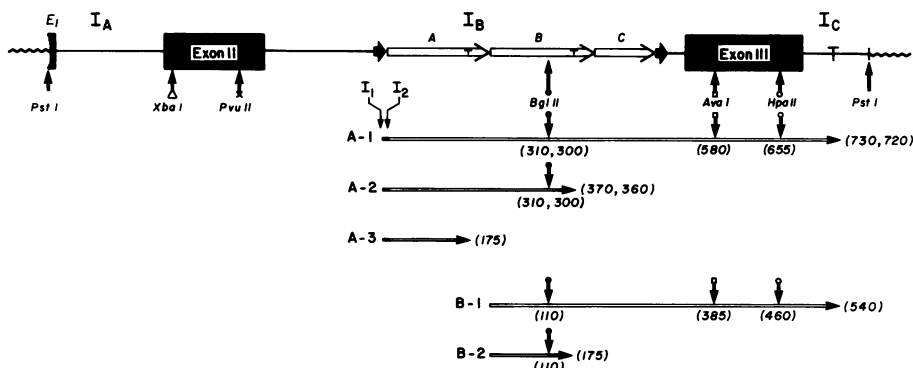


Figure 4 Pol III transcription map contained within the repetitive DNA of the rGH gene I_B . The structure of the PstI-1250 bp fragment of the rGH gene contained in prGH-1250 is presented at the top, including introns (I), exons (E), restriction sites and flanking plasmid DNA (W). The three tandem repeats within I_B are presented as open boxes labelled A, B, and C and the 18 bp direct repeats are shown as short, dark arrows. The weak Pol III termination signal within repeats A and B and the strong Pol III termination signal in I_C are indicated by a (T). The predicted and observed Pol III transcripts are shown as arrows with their lengths in nucleotides indicated in parenthesis at the arrow tip. The run-off transcript lengths generated by digesting prGH-1250 DNA with various endonucleases are indicated in parenthesis just below the restriction site. The transcripts are aligned relative to the rGH region from which they are transcribed with the RNAs labelled A1-3 for those initiating at the first I_B repeat, and B1-2 for those initiating at the second I_B repeat. I_1 and I_2 indicate the two initiation sites for the "A" transcripts mapped by S_1 analysis.

(Fig. 3B, lane 1 and Fig. 3A, lane 6) results in the same six major transcripts migrating at 720, 540, 370, 360, 320 and 175 nucleotides. This finding indicates that both the start and stop sites for these transcripts are contained entirely within the 1250 bp PstI fragment. Also, these are precisely the transcripts that are predicted if all combinations of Pol III promoters and terminators found within I_B and I_C are utilized, as discussed previously and shown in Fig. 4. However, an additional product is seen at 370 nt, and very little, if any, of the 330 nt RNA is present. In order to further substantiate the I_B Pol III transcription map presented in Fig. 4, the prGH-1250 template was truncated with either XbaI, PvuII, BglII, AvaI or HpaII, transcribed *in vitro*, and the lengths of the observed run-off transcripts were correlated with the predicted lengths, as detailed in the next section.

4. Restriction site mapping of the Pol III transcription units within Intron B of the rGH gene. The transcription map presented in Fig. 4 predicts that digesting the prGH-1250 template with either PstI, XbaI or PvuII should

Table 1. Restriction site mapping of the Pol III transcription units within intron B of the rGH gene.*

| Designation* | Transcript size | | | | | | |
|--------------|-----------------|------------|------------|------------|------------|------------|------------|
| | Uncut | PstI | XbaI | PvuII | BglII | AvaI | HpaII |
| A1 | 720 | 720 | 720 | 720† | 310 300 | 580 | 655 |
| A2 | 370 360 | 370 360 | 370 360 | 370 360 | 310 300 | 370 360 | 370 360 |
| A3 | 175 | 175 | 175 | 175 | 175 | 175 | 175 |
| B1 | 540 | 540 | 540 | 540 | 110 | 385 | 460 |
| B2** | 175 | 175 | 175 | 175 | 110 | 175 | 175 |
| C1*** | (330) | (330) | (330) | (330) | (330) | (180) | (260) |

(*)Major transcripts from prGH-1250 either uncut or cleaved with the indicated enzymes as obtained from the data of Fig. 3 are shown. The source of the transcripts are designated as from the first (A), second (B) and third (C) I_B repeat units. (**)With B2, the start site can be documented, but the stop is not proven due to its same size as A3. (***)Shown in parenthesis are the sizes of transcript expected if the putative Pol III promoter of the third repeat (TR) monomer were active that are not observed. (†)A less intense band below 720 nt is also observed.

not affect the transcription pattern, since these restriction sites occur outside of the predicted I_B Pol III transcription units. As shown in Figure 3B and Table I, this is precisely the result obtained. Furthermore, Fig. 3B and Table I also indicate that by truncating the prGH-1250 template with either BglII (lanes 2 and 6), AvaI (lane 3) or HpaII (lane 7), the observed transcripts again match the predicted run-off products as shown in Fig. 4. Thus, the 655 nt, 580 nt, and 310/300 nt run-offs, which replace the 720 nt transcript (labelled A1) for the HpaII, AvaI and BglII truncated templates, respectively, map an initiation site for the 720 nt transcript to the beginning of the first repeat. In addition, since the 370/360 nt RNA doublet observed with uncut templates is replaced by a new 310/300 nt doublet with BglII, another initiation site 10 bp upstream from the first I_B repeat and into the 18 bp DR is mapped, indicating that there are two initiation sites for A-1 and A-2 (Fig. 4) (discussed below). The 720 nt transcript therefore terminates at the CTTTTC sequence, 37 bp into intron C. Similarly, the single initiation site of the 540 nt transcript (labelled B1) can be mapped to the first nucleotide of the second repeat and its termination to the CTTTTC sequence of intron C, since this transcript disappears and new 110

(not shown), 385 or 460 nt run-offs are generated (none of which appear as doublets) with BglII, AvaI or HpaII-cut templates, respectively.

Finally, the data suggest that the TRs associated with the rGH gene are relatively inactive. First, no cell-free transcriptional activity is obtained with prGH-425 containing the TR in the immediate 3'FR (lane 1, Fig. 3A). Secondly, the predicted length of a transcript initiating at the putative Pol III promoter of the TR within I_B and continuing to the first Pol III stop signal ("CTTTTC", I_C) would be 330 nt (C1, Table 1). A transcript of around 320 nt is weakly visible in many of the lanes (lanes 1,2,5 and 6, Fig. 3B) and this transcript is also observed with BglII (lanes 2 and 6, Fig. 3B), but not AvaI (lane 3, Fig. 3B) or HpaII (lane 7, Fig. 3B) digested templates. However, the predicted TR run-offs (C1, Table 1) for AvaI (180 nt) and HpaII (260 nt) truncated templates are not observed. Thus, the TR within I_B may have very weak, if any, Pol III promoter activity.

5. S₁ nuclease mapping of the Pol III initiation sites within intron B of the rGH gene. To map precisely the I_B Pol III initiation sites, nonradioactive prGH-1250 cell-free transcription products synthesized in the presence of 2 µg/ml of α-amanitin were hybridized with the ³²P-labelled DNA probes indicated in Fig. 5A and then treated with S₁ nuclease. The 411 bp AvaII fragment (probe 1) spans the predicted start site within the first repeat. Use of this probe generates fragments of 51-54 and 63-68 nucleotides (Fig. 5B) indicating start sites at the start of the first long repeat (marked with an asterisk, Fig. 1A), and to the fifth (G) and seventh (A) nucleotides within the direct repeat (positions 559 and 561, darkened triangles, Fig. 1A). These findings are consistent with restriction-site transcript mapping data in the previous section. An 804 bp PstI-BglII fragment (probe 2, T_M-50°C) spanning the first two repeats was used to map the initiation site from the second repeat. The hybridizations were done at 49°C and 54°C, in an attempt to map the transcripts predicted from the data of Fig. 3 of 298 nt (48% GC) and 108 nt (56% GC) extending from the first and second repeats, respectively, to the BglII site (Fig. 5A). As Fig. 5B shows, a fragment of about 110 bp is the only one observed at either temperature indicating a single initiation site at the start of the second repeat (nucleotide 767, marked with an asterisk, Fig. 1A). The absence of the 298 bp protected fragment may be due to the lack of formation of the 298 bp hybrid because of the AT-rich cluster of about 55 bp in the middle of this predicted S₁ nuclease resistant fragment as has been found in other similar cases (35).

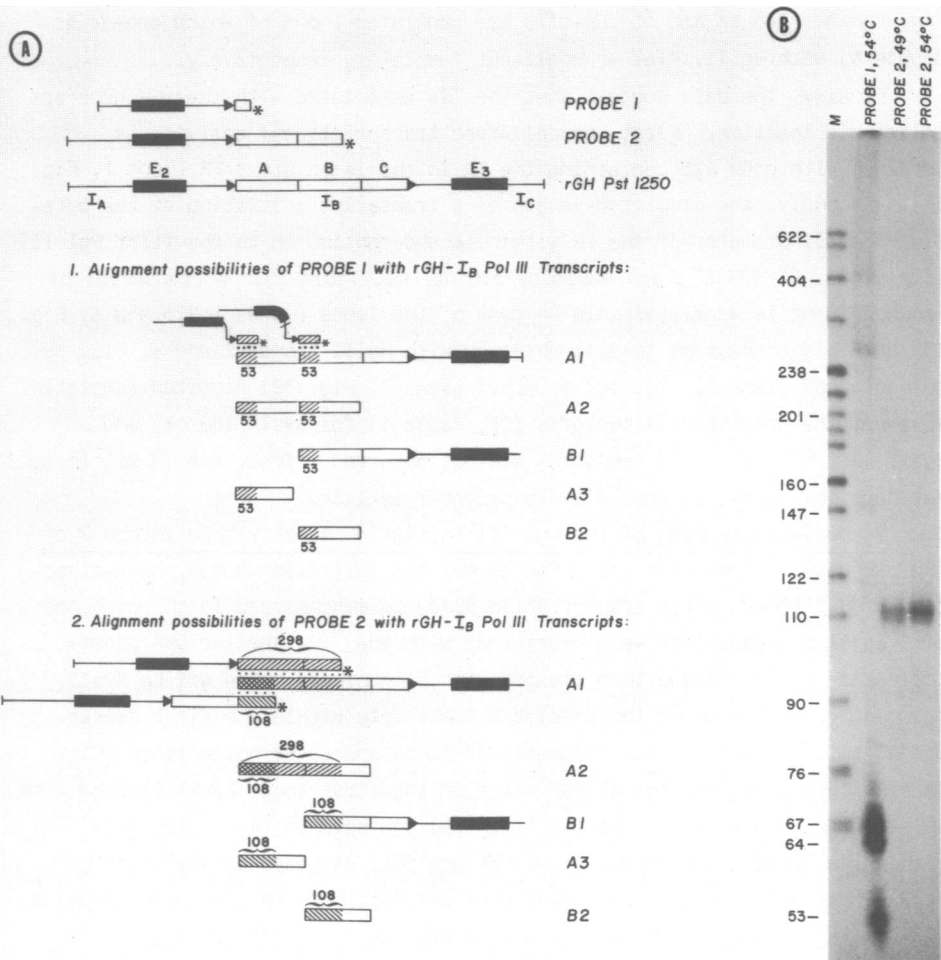


Figure 5 S_1 nuclease mapping of the Pol III initiation sites within intron B of the *rGH* gene. (A) Probes were generated as described in METHODS. Probe 1, 411 bp, contains the 5'-18 bp direct repeat (symbols as in Fig. 4) and 53 bp of the first I_B repeat. Probe 2, 804 bp, contains the 5'-18 bp direct repeat, the entire first I_B repeat and 108 bp of the second I_B repeat. The Pol III *in vitro* transcripts (A_1 - B_2) are aligned with the equivalent structures in *prGH*-1250. The various alignment possibilities of the I_B transcripts and probes 1 and 2 are also schematically shown. The diagonal lines indicate the transcript region homologous to a given probe, and the numbers above or below indicate the size of the predicted hybrid resistant to S_1 nuclease digestion. (B) Alpha-amanitin resistant, unlabelled RNA transcribed *in vitro* by uncut *prGH*-1250 was hybridized to either 5'- ^{32}P labelled probe 1 or 2 at the indicated temperature, digested with S_1 nuclease and analyzed on an 8% polyacrylamide/8M urea sequencing gel as described in METHODS. The size of the protected DNA fragments indicated was estimated by calibrating the gel system with ^{32}P -labelled, *Hpa*II-cut pBR322 (lane M).

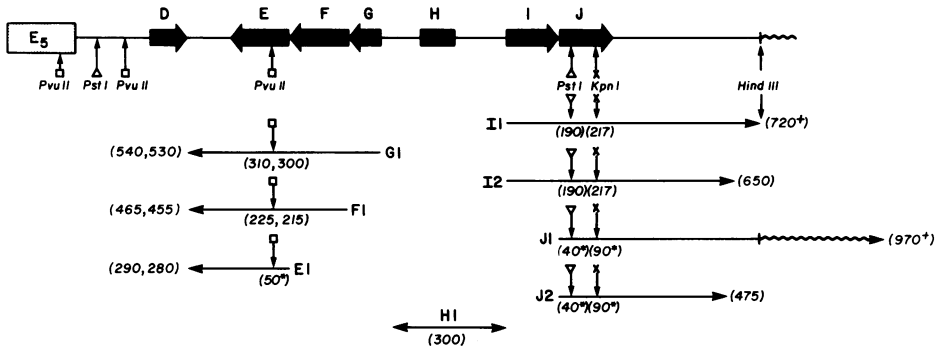


FIGURE 6 Pol III transcription map of the 3'FR of the rGH gene. The structure of the 3'FR including the fifth exon (E₅), restriction sites, and flanking pBR322 DNA (wavy line) is presented at the top. The short dark arrows labelled D-J represent repetitive DNA structures. Repeat "D" represents the 3'FR TR whose position has been determined by sequencing studies (see text), and the position of repeats E-J has been assigned from transcription mapping studies (see text). The Pol III transcripts are shown below, with the legend as in Fig. 4, except that the squiggle portion of the J1 arrow represents transcription read-thru into pBR322 DNA. The transcription orientation for HI is unknown; therefore the RNA is indicated as a double-headed arrow. Numbers in parenthesis marked with an asterisk are predicted transcript run-off lengths which are not seen as explained in Results. Transcripts I1 and J1 are marked with a (+) because although this assignment is most consistent with the data, their conclusive mapping has been complicated as indicated in Results. The TR subclone, prGH-425, contains the D repeat segment bordered by the PvuII sites; and prGH-1350 spans the region between the PstI sites.

6. Analysis of the Pol III transcription units in the distal 3'FR of the rGH gene. The lack of sequence information for the terminal 1.9 kb of the 5.8 kb rGH genomic insert (Fig. 1) complicated the detailed mapping of the Pol III transcription unit(s) within the distal 3'FR (7). However, the 720 (I1) and 650 nt (I2) transcripts can be tentatively mapped to a single initiation site located 190 bp 5' to the most distal PstI site (Fig. 6), since both of these transcripts are replaced by a single new transcript of 190 nt when prGH-1350 is truncated with PstI (lanes 3 and 4, Fig. 3A). Mapping studies of the 720 nt transcript from the 3'FR utilizing prGHeh-5.8 DNA are complicated by the presence of a similar sized transcript which originates from I_B (lanes 2-6, Fig. 3A, Table I and Fig. 4). The position of the 650 nt transcript shown in Fig. 6 is further substantiated since this transcript is replaced by 190 or 217 nt run-offs when prGHeh-5.8 is cut with PstI (lane 5, Fig. 3A) or KpnI (Fig. 2), respectively. Similarly, the 970 (J1) and 475 nt (J2) transcripts seen with prGHeh-5.8 can both be tentatively mapped as initiating in the region immediately preceding the 3' terminal PstI and KpnI sites since either

PstI (Fig. 3) or KpnI (Fig. 2) digestion of prGHeh-5.8 abolishes the transcription of both the 970 and 475 nt RNAs; and neither transcript is directed by prGH-1350 (Fig. 3A). The 970 nt RNA terminates in pBR322, since this transcript is replaced by a new 560 nt run-off if prGHeh-5.8 is cut with HindIII (data not shown). The 475 nt product appears to terminate at the same site as the 650 nt transcript (Fig. 6). The other transcripts directed by prGH-1350 of about 540, 465, and 300 nt persist after PstI digestion (Fig. 3A, lanes 3 and 4) indicating that these Pol III transcription units are wholly contained in the PstI 1350 bp fragment. Furthermore, RNA bands of 540/530 and 465/455 nt are shortened to 310/300 and 225/215 nt, and a much weaker 270/280 nt doublet disappears, after PvuII digestion of prGHeh-5.8 DNA (not shown). Since the 425 bp PvuII fragment is transcriptionally inactive (Fig. 3A, lane 1), these three transcripts most likely originate 3' to the distal PvuII site, and therefore are transcribed in an orientation opposite to the other Pol III (or Pol II) units. Thus, three Pol III promoter elements occur in tandem (Fig. 6, E-G), each with apparently two initiation sites and a common, strong Pol III stop signal. Consistent with this interpretation is that this stop signal appears to correspond to the oligo-dT sequence that is present on the opposite strand of the poly A segment of the 3' truncated repeat, D. (Fig. 6). The position of the 300 nt transcript resistant to PstI digestion of prGH-1350 (Fig. 3A, lanes 3 and 4) can only be approximated to the DNA region between the G and I repeats (Fig. 6).

7. The rat type 2 repeat elements share Pol III transcription factors with a Drosophila tRNA gene, a human Alu repetitive DNA and a mutant VAL gene. To determine whether the rGH Ig repeats share transcription factors with other Pol III templates, we assessed the ability of various competitor templates to inhibit the formation of a stable Pol III transcription complex by subsequently added prGH-1250 reference DNA (28,29). As shown in Fig. 7, preincubation with an equimolar Pol III promoter amount (relative to prGH-1250) of either the intact Drosophila tRNA gene template (pARG), a Bal 31-generated pArg template deleted of its proximal A box (pARG 5.12), an adenovirus VA-1 gene deleted of its first 55 bp (VAL-5 Δ 55), and a human Alu-repeat sequence (pPD8), does not affect the transcription of any prGH-1250 -specific transcripts (compare the 720, 540, 370, 360, 320 and 175 nt RNA bands in lane 1 with those in lanes 2,4,6 and 8). The overall intensities can be compared with those of the U₁ RNA added as an internal standard. Similarly, pArg 5.12 and pBR322, even at a 6-fold molar excess of Pol III promoters, do not compete (lanes 5 and 10, Fig. 7). At a 6-fold excess, however, the VAL-5 Δ 55

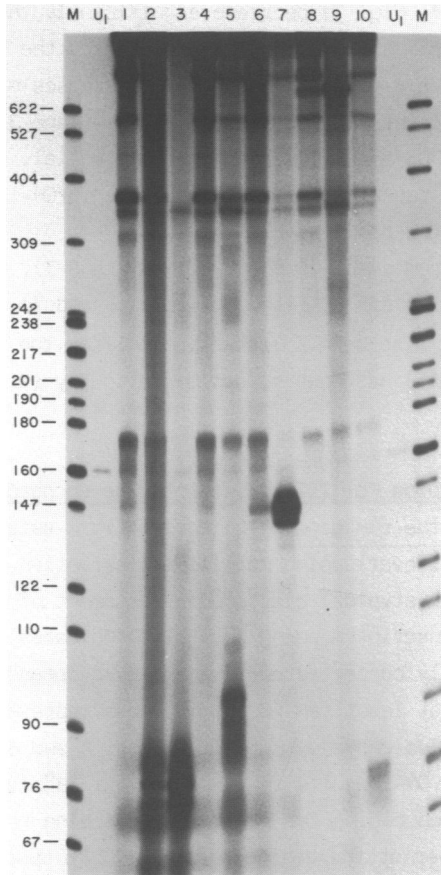


FIGURE 7 Pol III factor competition assay. The assay was performed essentially as described by Sharp et al. and Schaack et al. (28,29) except that pre-incubation was at 30°C for 5 min with either nearly equal or 6 fold Pol III promoter molar excess relative to the reference template (prGH-1250) as described in METHODS. Lane: (1) prGH-1250, 1 µg; (2) pArg, 2 µg + prGH-1250, 1 µg; (3) pArg, 10 µg + prGH-1250, 1 µg; (4) pArg 5.12, 2 µg + prGH-1250; 1 µg; (5) pArg 5.12, 10 µg + prGH-1250, 1 µg; (6) VA1-5'-Δ55, 2 µg + prGH-1250, 1 µg; (7) VA1-5'-Δ55, 10 µg + prGH-1250, 1 µg; (8) pPD8, 2 µg + prGH-1250, 1 µg; (9) pPD8, 10 µg + prGH-1250, 1 µg; (10) pBR322, 10 µg + prGH-1250, 1 µg. The migration of the various transcripts are as follows: the mapped prGH-1250 transcripts are 720, 540, 370, 360, and 175 nt, the pArg precursor RNA at 84 nt and the mature RNA at 76 nt; pArg 5.12 RNA at about 95 nt; the VA1-5'-Δ55 RNA product comigrates with the 147 nt DNA standard; and the pPD8 transcript migrates at about 650 nt (5,21,41). Migration of ³²P-labelled U₁ RNA is shown in lanes "U₁".

and pPD8 (lanes 7 and 9, Fig. 7) decrease significantly the 720, 540 and 370 nt transcripts, and VA1-5' Δ 55, but not pPD8 decreases the 360 and 175 nt transcripts. Finally, a 6-fold excess of pARG decreases markedly the 720, 540, 370 and 175 nt but not the 360 nt transcripts (lane 3, Fig. 7). Thus, pARG shows a marked, pPD8 and VA1-5' Δ 55 show substantial, and pARG 5.12 and pBR322 show negligible capabilities to compete with prGH-1250 promoters for Pol III transcription factors.

The transcripts from pARG (76-83 nt, lanes 2 and 3), pARG 5.12 (70-95 nt, lanes 4 and 5), VA1-5' Δ 55 (145-150 nt, lanes 6 and 7) and pPD8 (650 nt, lanes 8 and 9) are also seen. Transcription from the deletion templates (pARG 5.12 and VA1-5' Δ 55) has been reported previously (21,40).

Discussion

1. Multiple, overlapping Pol III transcripts are encoded by repetitive DNA within I_B and 3'FR of the rGH gene. The current work establishes a novel organization of several overlapping Pol III transcription units in both I_B and the 3'FR due to the atypical clustered arrangement of multiple repetitive DNA monomers. The transcript mapping data also indicates that repetitive DNA sequences are strikingly concentrated in the 5.8 kb Sprague-Dawley rGH gene fragment resulting in at least ten repeat monomers representing two different subclasses (type 2-equivalent and TR family, Figs. 4 and 6).

The I_B repetitive DNA results in two ("A" and "B" family) Pol III transcription units each coding for several overlapping transcripts. The "A family" initiates either at the beginning or 11-13 bp upstream from the first I_B repeat and then uses each of three termination signals resulting in transcripts of 720-730 (A1), 360-370 (A2) and 175 (A3) nt (Fig. 4). The upstream start site is 23-25 bp 5' from the canonical A box of the first repeat, which is one of the longest reported distances between the start of transcription and the promoter element involved in positioning it (27). However, the upstream start site may be positioned by an 11 mer, the GAGAGGGCTGG sequence, that spans the start of the first repeat (Fig. 1A) and which has significant homology with the Pol III A box promoter consensus sequence (20,21,30-32,47). Alternatively, initiation site selection by Pol III may depend on the 5' flanking sequence, as previously shown for tRNA genes (40,41). By contrast the canonical A box promoter region of the B family (Fig. 1A) initiates transcription from only one site, at the beginning of the second I_B repeat, and then each of the two remaining termination sequences are used thus resulting in transcripts of 540 (B1) and 175 (B2) nt

(Fig. 4). Furthermore, the correlation between the observed transcript size and the predicted size from Pol III initiation and termination signals in I_B, and a lack of further changes with prolonged incubation rule out post-transcriptional processing in the HeLa cell extracts as the basis for the complex transcription pattern. It has previously been suggested that Manley-type HeLa cell extracts may allow weak stop signals to be utilized by Pol III (3,29-34). Here we show that the first two Pol III termination sites (ATCTTIA), which belong to the "very weak" class of Pol III termination signals of Bogenhagen and Brown (34), can also be effective in HeLa cell extracts. Although we found that these stop signals were used, they were also found to be weak, since many transcripts do read through this signal (A1,A2,B2; Fig. 4). However, none were found to read through the strong termination sequence (CTTTTC) in intron C. The transcripts which terminate within I_C (Fig. 4, A1 and B1) are hybrid RNAs containing repetitive DNA followed by single-copy rGH structural gene (exon III) and intervening (I_B and I_C) sequences. This lack of an efficient termination signal within the repetitive DNA transcription unit, with resultant transcription of non-repetitive sequences, is common with the transcribed human Alu elements and also occurs in other rodent type 2 repeats (4,5).

Initiation of Pol III transcription at each of the paired I_B promoters contrasts with Pol III transcription of similar tandemly aligned tRNA genes. The latter result in a single, polycistronic transcript which initiates with the first tRNA gene (16-18). Downstream genes initiate transcription only after removal of the upstream tRNA gene sequence (18). The inability of Pol III to initiate at internal genes of a tandem array is probably not due to lack of specific factor binding by the distal B box regions (30,36,37). However, factor binding to all B box regions might conceivably interfere with subsequent A box factor binding by downstream A box regions that are only 23-27 bp away from the B box of the 5' gene. This might not occur in the case of the two rGH gene I_B Pol III transcription units, since the B box in the upstream promoter (bold face, position 628, Fig. 1A) is 138 bp from the A box of the downstream promoter (bold face, position 779, Fig. 1A). Alternatively, the lack of efficient Pol III terminators or the presence of Pol III inhibitory sequences between paired tRNA genes might also lead to the exclusive formation of a polycistronic transcript. An indication that distance alone may not be inhibitory has been obtained recently by Ciliberto et al. (46) who found that internal initiations could occur for three closely linked tRNA genes constructed in vitro. Parameters such as these may repre-

sent different mechanisms involved in Class III gene regulation.

2. Repetitive DNA elements within I_B are very analogous to tRNA genes.

Previous transcription competition studies have shown that VAI and tRNA genes share Pol III transcription factors and that the VAI gene B box alone is sufficient, since it appears to bind transcription factor(s) relatively tightly (21,28,29,32,36,37). Our studies confirm and extend these results since we show that a tRNA gene (pArg), a human Alu repeat (pPD8), a VAI gene containing only the B box (VAI-5'Δ55), and rat type 2 DNA (prGH-1250) all share Pol III transcription factors (Fig. 7). The differential competitive ability we observed for the various templates is most likely due to the relative stability of the transcription complex formed (28,29,37). Thus, templates with either both A and B boxes intact (pArg and pPD8), or with a high-affinity B box (VAI-5'Δ55) were able to form stable complexes, whereas the pArg B box region alone (pArg 5.12) was not. Of particular interest, the rat type 2 and TR elements have conserved most of the invariant nucleotides specifying tRNA structure and promoter function in the same relative position (38,44). This striking structural similarity may not only explain why pArg is more effective than pPD8 in competing for prGH-1250 transcription, but also suggests that these rat repeats and tRNA genes have evolved from a common, ancestral precursor gene. Finally, since several of these Class III templates share transcription factors, it is likely that differential promoter strength and factor accessibility may affect their competitive ability, and thus ultimately their expression.

3. Diminished Pol III cell-free activity of the rGH-TR DNA. The mechanisms of expression of the TR family is of special interest since a 160 nt RNA containing TR (or "identifier") sequences occurs specifically in the brain and pituitary and could have some important biological role (3,9-11).

Interestingly, our results show that the rGH type 2 repeats efficiently initiate Pol III transcription, whereas the TR elements apparently do not.

Indeed, the rGH-TR sequences appear only as part of larger transcripts directed by upstream type 2 repeats (e.g., A1 and B1, Fig. 4). An examination of the Pol III promoter sequences revealed that the rGH TR, but not the type 2 DNA, contains a significant A₅₈→G change at an invariant position in the B box promoter region. Surprisingly, we found this change to be conserved in all eleven TR elements sequenced so far (3,7,9-15). It could be argued that this "nonpermissive" change of the invariant A₅₈ to a G is responsible for the lack of TR Pol III activity in vitro. Yet, in contrast to our studies, Sutcliffe et al. (3) have found that an equivalent TR

sequence (ID clone p2A120) contains strong Pol III promoter activity in vitro. More recently we have also transcribed these sequences (clones p2A120 and p1B337) and confirmed their strong Pol III promoter activity in vitro (data not shown). Thus, as Traboni et al. (47) have so elegantly shown, the strongly conserved A₅₈ may be more important in tRNA structure or transcript stability rather than promoter function, and therefore mutations at this position affect cell-free transcription only minimally (3,42,47). However, since the TR sequences of the rGH gene and ID cDNA clones are essentially the same, the basis for the striking difference in their Pol III activity is most likely due to either position effects and/or the presence of inhibitory or stimulatory flanking sequences. Although such influences have been reported, they are difficult to predict in cell-free studies (39-41). If the flanking DNA in some other way regulated the activity of the TR RNA polymerase III promoter, this could provide a mechanism whereby the activity of "identifier" or TR sequences is regulated by being dependent on the site of insertion. Hopefully, other studies can be performed to address this issue by exchanging flanking regions between a functional and an inactive TR subclone.

ACKNOWLEDGMENTS:

We thank Drs. A. Barta, N.L. Eberhardt, D. Colby, and G. Bell for helpful discussions; and Dr. G. Sutcliffe for providing us with ID plasmid DNA and manuscripts prior to publication. AG-H was supported by a Giannini Fellowship during the initial phase of this work and more recently by a New Investigator Research Award (HD 16878). I.L. is a recipient of an NRSA Fellowship. This research was supported by N.I.H. Grant (AM 19997) to D.G. and J.D.B. and a grant from California Biotechnology, Inc. We are grateful to Dorothea Faber for typing the manuscript.

+To whom reprint requests should be sent

*Current Address: Albert Einstein College of Medicine, Department of Medicine, 1300 Morris Park Ave., Bronx, NY 10461, USA

§Current Address: USTL, Laboratoire de Biologie Moleculaire, Place Eugene Bataillon, 34060 Montpellier, France

REFERENCES

1. Jelinek, W.R. and Schmid, C.W. (1982) *Ann. Rev. Bioch.* 51, 813-844.
2. Sharp, P.A. (1983) *Nature* 301, 471-472.
3. Sutcliffe, J.G., Milner, R.J., Gottesfeld, J.M. and Lerner, R.A. (1984) *Nature* 308, 237-241.
4. Haynes, S.R. and Jelinek, W.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6130-6134.

5. Fuhrman, S.A., Deininger, P.L., LaPorte, P., Friedmann, T. and Geiduschek, E.P. (1981) *Nucl. Acids Res.* 9, 6439-6456.
6. Cooke, N. and Baxter, J.D. (1982) *Nature* 297, 603-607.
7. Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4867-4871.
8. Page, G.S., Smith, S. and Goodman, H. (1981) *Nucl. Acids Res.* 9, 2087-2104.
9. Sutcliffe, J.G., Milner, R.J., Bloom, F.E. and Lerner, R.A. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4942-4946.
10. Milner, R.T., Bloom, F.E., Lai, C., Lerner, R.A. and Sutcliffe, J.G. (1984) *Proc. Natl. Acad. Sci. USA* 81, 713-717.
11. Fox, J.L. (1983) *Science* 222, 828-829.
12. Lemischka, I. and Sharp, P.A. (1982) *Nature* 300, 33-335.
13. Schuler, L., Weber, J.L. and Gorski, J. (1983) *Nature* 305, 159-160.
14. Watanabe-Nagasu, N., Itoh, Y., Tani, T., Okano, K., Koga, N., Okada, N. and Oshima, Y. (1983) *Nucl. Acids Res.* 11, 1791-1801.
15. Dhar, R., Ellie, R.W., Shih, T.Y., Oroszlan, S., Shapiro, B., Maizel, J., Lowy, D. and Scolnick, E. (1982) *Science* 217, 934-437.
16. Mao, J-i, Schmidt, O. and Söll, D. (1980) *Cell* 21, 509-516.
17. Schmidt, O., Mao, J-i, Ogden, R., Beckmann, J., Sakano, H., Abelson, J. and Söll, D. (1980) *Nature* 287, 750-752.
18. Gruissem, W., Prescott, D.M., Greenberg, B.M. and Hallick, R.B. (1982) *Cell* 30, 81-92.
19. Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Geter, M.L. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3855-3859.
20. Sharp, E., DeFranco, D., Dinger, T., Farrell, P. and Söll, D. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6657-6661.
21. Guilfoyle, R. and Weinmann, R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3378-3382.
22. Clewell, D.B. and Helinski, D.R. (1969) *Proc. Natl. Acad. Sci. USA* 62, 1159-1166.
23. Ish-Horowicz, D. and Burke, J.F. (1981) *Nucl. Acids Res.* 9, 2989-2998.
24. Maxwell, I.H., Maxwell, F. and Hahn, W.E. (1977) *Nucl. Acids Res.* 4, 241-246.
25. Savouret, F., Cathala, G.G., Eberhardt, N.L., and Baxter, J.D., in preparation.
26. England, T.E., Bruce, A.G. and Uhlenbeck, D.O. (1980) *Meth. Enz.* 65, 65-74.
27. Hall, B.D., Clarkson, S.G., and Tocchini-Valentini, G. (1983) *Cell* 29, 3-5.
28. Sharp, S., Dinger, T., Schaack, J., DeFranco, D. and Söll, D. (1983) *J. Biol. Chem.* 258, 2440-2446.
29. Schaack, J., Sharp, S., Dinger, T. and Söll, D. (1983) *J. Biol. Chem.* 258, 2447-2453.
30. Newman, A.J., Ogden, R.C. and Abelson, J. (1983) *Cell* 35, 117-125.
31. Folk, W.R. and Hofstetter, H. (1983) *Cell* 33, 585-593.
32. Fowlkes, D.M. and Shenk, T. (1980) *Cell* 22, 405-413.
33. Bogenhagen, D.F. and Brown, D.D. (1981) *Cell* 24, 261-270.
34. Jat, P. and Arrand, J.R. (1982) *Nucl. Acids Res.* 10, 34067-3425.
35. Zassenhaus, P.H., Farrelly, F., Hudspeth, M.E.S., Grossman, L.E. and Butow, R.A. (1983) *Mol. Cell. Biol.* 3, 1615-1624.
36. Klemenz, R., Stillman, D.J. and Geiduschek, E.P. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6191-6195.
37. Lassar, A.B., Martin, P.L., and Roeder, R.G. (1983) *Science* 222, 740-748.
38. Sprinzl, M. and Gauss, D.H. (1982) *Nucl. Acids Res.* 11, r1-r53.
39. Raymond, G.D. and Johnson, J.D. (1983) *Nucl. Acids Res.* 11, 5969-5988.

40. Dingermann, T., Burke, D.J., Sharp, S., Schaack, J. and Söll, D. (1982) *J. Biol. Chem.* 257, 14738-14744.
41. DeFranco, D., Sharp, S. and Söll, D. (1981) *J. Biol. Chem.* 256, 12424-12429.
42. Murphy, M.H. and Baralle, F.E. (1983) *Nucl. Acids Res.* 11, 7695-7700.
43. Koski, R.A., Clarkson, S.G., Kurjan, J., Hall, B.D. and Smith, M. (1980) *Cell* 22, 415-425.
44. Cathala, G., Eberhardt, N.L., Lan, N.C., Gardner, D.G., Gutierrez-Hartmann, A., Mellon, S., Karin, M. and Baxter, J.D. (1983) In *Perspectives on Genes and the Molecular Biology of Cancer*, Robberson, D.L. and Saunders, G.F., Eds., pp. 169-181, Raven Press, New York.
45. Duncan, C.H., Jagadeeswaran, P., Wang, R.R.C. and Weissman, S.M. (1981) *Gene* 13, 185-196.
46. Ciliberto, G., Traboni, C. and Cortese, R. (1984) *Nucl. Acids Res.* 12, 1277-1285.
47. Traboni, C., Ciliberto, G. and Cortese, R. (1984) *Cell* 36, 179-187.