

Human ribosomal RNA variants from a single individual and their expression in different tissues

Bruce A. Kuo[†], Iris L. Gonzalez[§], David A. Gillespie and James E. Sylvester^{*}

Department of Pathology and Laboratory Medicine, Medical College of Pennsylvania and Hahnemann University, Philadelphia, PA 19102, USA

Received June 21, 1996; Revised and Accepted October 7, 1996

DDBJ/EMBL/GenBank accession no. U13369

ABSTRACT

We have investigated the extent of sequence variation in human ribosomal RNA (rRNA) genes and the expression of specific rRNA gene variants in different tissues of an individual. Focusing on the fifth variable region (V5; nt 2065–2244) of the 28S rRNA gene, we find that sequence differences between rRNA genes of a single individual are characterized by differences in number of repeats of simple sequences at four specific sites. These data support and extend previous findings which show similar V5 sequence variation in rRNA genes from a group of individuals. We performed experiments to determine if there is differential gene expression within the rRNA multigene family. From the analysis of data of six variant V5 probes protected from RNase digestion by rRNAs isolated from different tissues of the individual, we conclude that each variant rRNA is present in a similar proportion in these tissues, whereas the actual contributions of variants differ, their relative proportion is maintained from tissue to tissue in an individual. We favor the explanation of a gene dosage effect over that of a regulated gene effect to account for this pattern of rRNA gene expression. In addition, computer generated secondary structure models of each V5 clone structure predict the same three helix structure with the regions of sequence variation contained in one stem–loop structure.

INTRODUCTION

Ribosomal RNA genes contain sequence tracts that are conserved in size, sequence, and secondary structure (1). When expressed as part of the mature rRNAs, these conserved sequences form the enzymatic core of the ribosome which is essential for translation (2–4). While the sizes of the small subunit rRNAs do not vary much from species to species, the large subunit rRNAs vary widely in size: yeast 26S: 3788 nt (5); *Drosophila* 26S: 3945 nt (6); *Xenopus laevis* 28S: 4110 nt (7); mouse 28S: 4712 nt (8); and human 28S: 5035 nt (9). This size disparity is due to the presence

of variable sequence tracts (Fig. 1, V regions) which vary in both size and sequence and are interspersed amid the conserved core sequences (10). These regions have also been referred to as divergent (D) domains (8), and expansion segments (11). The *Escherichia coli* 23S rRNA gene lacks V regions, which is reflected in its short length (2904 nt) (12). Whether V regions evolved in an ancestral rRNA gene or *E.coli* selectively eliminated them, the origin of V regions is unclear. Cross-species comparisons show that the V regions of a species frequently have a similar high G/C content (6,7). Similarities in this sequence bias from one species to another suggests that these regions of the rRNA gene have a common origin (13). Because the size and sequence of V regions appear to be species-specific, any differences between large subunit rRNAs of different species, within a species, and within an individual are most likely to exist in the V regions. Whether size and sequence variation of the V regions affect rRNA activity is not known.

Sequence variation of V regions has been characterized predominantly as differences in the number of simple sequence repeats as is seen in different clones of the V5 region which were isolated from multiple sources (9,14). Slippage during replication (15–17) and unequal recombination between chromosomes (9,18) have been proposed as mechanisms that can generate and propagate this variation. Detection of expressed rRNA heterogeneity has focused on that found in the V regions. In one study, rRNAs were isolated from several individuals and differences were detected by an S1 nuclease protection assay (19). Different band patterns indicated that individuals express different V5 variants while additional minor bands suggested that different variants are expressed within an individual. Another study identified expressed rRNA sequence heterogeneity by sequencing the V8 region in rRNAs isolated from different human primary and established cell lines (20). These results suggest that within the ~400 copies per human genome, a considerable amount of sequence heterogeneity exists among large ribosomal RNAs, predominantly in the variable regions. This raises the questions: (i) how much rRNA sequence variation does actually exist; (ii) is it possible to detect the expression of specific rRNA variants in an individual; (iii) do different tissues of an individual express different rRNA genes?

^{*}To whom correspondence should be addressed at present address: Nemours Children's Clinic, 807 Nira Street, Jacksonville, FL 32207, USA. Tel: +1 904 858 3909; Fax: +1 904 390 3425; Email: sylvesterj@hal.hahnemann.edu

Present addresses: [†]Thomas Jefferson University, 365 JAH, 1020 Locust Street, Philadelphia, PA 19107, USA and [§]Alfred I. DuPont Institute, 1600 Rockland Road, Wilmington, DE 19899, USA

This paper is dedicated to the memory of D. A. Gillespie

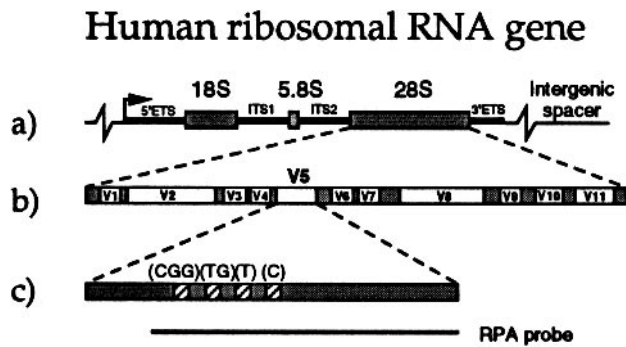


Figure 1. Human rRNA gene map (one repeat 42.9 kb). (a) A single 13 kb transcript of the rRNA gene (thick line and stippled boxes) is processed into the mature 18S, 5.8S and 28S rRNAs (stippled boxes). Tandemly arranged transcriptional units are separated by 30 kb intergenic spacers. (b) Variable regions (V1–V11) of the 28S rRNA are shown. (c) The location of the simple sequence repeats in V5 region are shown. The line below represents the V5 region used as an RNase protection probe.

This study has focused on characterizing the nature and extent of sequence variation of the V5 region at both the DNA and RNA levels in an individual. Our goal was to isolate V5 variant DNA clones from an individual for sequence comparisons, and then use these clones as RNase protection probes to detect rRNAs containing specific V5 variants in different tissues from the same individual. Because previous attempts to isolate V5 fragments from genomic DNA by PCR consistently generated deletion artifacts, a direct cloning approach was used to isolate V5 fragments for this study. We present a computer generated secondary structure model for the human V5 region which appears to tolerate sequence variation, and is shared with chimpanzee and gorilla. Sequence comparisons between human V5 clones, then with primate and rodent V5 sequences, indicate that variation in the human V5 region may be limited to four specific sites.

MATERIALS AND METHODS

Cloning, sequencing, and secondary structure analysis of the V5 region

The V5 portion of rDNA was isolated from muscle genomic DNA as a size-selected 580 bp *Sau3A* fragment and cloned in M13 RF. An *Apal* V5 fragment was subcloned from the M13 recombinant into the *Apal* site of the pBluescript (SKII⁻) vector (Stratagene). The sequence of each clone was determined and entered into the GCG database (21) for manipulation and analysis. The sequence of each clone was entered into the Mulfold program (version 2.0) (22–24). The structure with the lowest free energy (largest $-\Delta G$) was determined and used for comparison of V5 secondary structures of different human clones as well as that of chimpanzee, gorilla, and mouse. The existing rDNA sequence was annotated to incorporate these new sequences (accession no. U13369).

Probe and target synthesis

Both rRNA-like target and rRNA-complementary RNase protection probes were generated by *in vitro* transcription from V5-containing plasmid templates using either T3 or T7 RNA polymerase. Briefly,

a 20 μ l reaction consisted of 4 μ l of a 5 \times reaction buffer (Promega), 1 μ l RNasin (40 U), 4 μ l rNTPs (final concentration of 400 μ M each), 2 μ g of a linear plasmid template (1 μ g/ μ l), 8 μ l water, and 1 μ l (20 U) of RNA polymerase (Promega) which was then incubated for 30 min at 37°C. This was followed by the addition of 1 μ l (1 U) of RQ1 (DNase1-Promega) to the reaction tube and further incubated for 10 min at 37°C. Transcripts to be used as probe in the RNase protection assay were synthesized in the presence of 50 μ Ci (100 μ M) of [α -³²P]GTP (800 Ci/mmol) (Amersham). Following probe synthesis, the reaction mix was diluted to 500 μ l and stored at -20° C.

RNase protection assays

Portions of nine different tissues were collected from a single human individual at autopsy. To extract total RNA, frozen tissue samples were pulverized and solubilized at a concentration of 100 milligrams (mg) of tissue per milliliter of 6 M guanidinium thiocyanate (GuSCN) (25,26), or extracted by the method of Chomczynski and Sacchi (27). Lysates were stored at -20° C and thawed just prior to use. NIH3T3 mouse cells were dissolved at a concentration of 10^6 cells/ml 6 M GuSCN.

A typical RNase protection reaction consisted of 1 μ l of an RNA target solution (50 ng RNA) in 15 μ l of Chomczynski Buffer (6 M GuSCN, 0.5% N-lauroyl sarcosine, 100 mM Tris pH 7.6, 1% β -mercaptoethanol) which was heated for 15 min at 55°C (28). An *in vitro* synthesized labeled-rRNA probe (2.5 ng probe/ >4000 c.p.m.) complementary to the V5 region was added, and further incubated at 55°C for 2 h. A volume of 400 μ l of 0.75 \times SSC (1 \times SSC = 0.15 M NaCl, 0.015 M sodium citrate) containing RNaseA (20 μ g/ml) was added and incubated at 37°C for 30 min. Carrier DNA (5–10 μ g sheared salmon sperm DNA) was added, followed by the addition of 1 ml of ethanol containing 3% DEPC at room temperature for 15 min before ethanol precipitation. The precipitated RNA was spun at 13 500 r.p.m. (microfuge) for 30–60 min. The supernatant was discarded and the pellet was resuspended in 5 μ l of water. An equal volume of a formamide loading dye (96% deionized formamide, 10 mM EDTA, 0.3% xylene cyanol, 0.3% bromophenol blue) was added to the resuspended pellet. The sample was boiled for 2–3 min, and 7 μ l of the sample was loaded on a 6% acrylamide-urea gel. The gel was dried and exposed to film (Kodak X-AR film) overnight at -80° C, and developed in a Kodak X-O-MAT film processor.

Densitometry of RPA bands

A SUNSPARC mini-workstation was used to run the Bioimage/Visage 4.6 electrophoresis gel analysis system of the Millipore Corporation (Ann Arbor, Michigan). The integrated band values for the optical density of BAND1, BAND2, and BAND3 were normalized for each band in a gel lane by calculating its percent contribution to the total band signal for each gel lane:

$$[\text{BAND1 O.D.} / (\text{BAND1 O.D.} + \text{BAND2 O.D.} + \text{BAND3 O.D.})] \times 100 = \% \text{BAND1 OF TOTAL SIGNAL}$$

RESULTS

Cloning and sequencing of V5 variants

Seven V5-containing DNA fragments were isolated from an individual. Six unique sequences were found among the seven V5 clones. The redundant variants represent independent clones



Figure 2. V5 region sequence alignment. Sequence alignment of the V5 region of human clones A11–A17 (this study), human clones A1–A7, chimpanzee (CH), gorilla (GO) and mouse (MO) clone (13). Where the human sequences are presented as HU, no variation is known to exist among human clones. Differences among human variants are differences in the number of CCG, TG, T and C repeats (bold underlined). The absence of a base relative to another sequence is noted as a dash (-). The human sequences correspond to the *ApaI* fragment of the 28S coding region, at 2041–2305 (1).

since they were isolated in different orientations in the original M13 vector. The sequences for the six V5 clones are shown in Figure 2 along with those of V5 clones isolated previously from different sources. Each human isolate differs only in the number of repeats (*n*) of simple sequences at four specific sites of the V5 region [(CGG)_{*n*} = 5–8, (TG)_{*n*} = 2–3, T_{*n*} = 1–2, and C_{*n*} = 9–12]. Variation of the V5 region in rDNAs isolated in this study is comparable with that isolated previously from several human sources (9,14). The number of simple repeats (*n*) for all known V5 clones are listed in Table 1. Table 2 shows that some repeat combinations have been isolated more frequently than others, such as variants A1, A6, pHrA versus A12. Whether this reflects rRNA gene distribution in an individual or population is not known. Each complete variant sequence is defined by a specific combination of variant simple sequence repeats. As many as 64 combinations could theoretically exist based on the different known repeat lengths (*n*) at the four sites of V5 variation. However, in all known V5 clones, the TG and T repeats occurred either in a (TG)₃/(T)₁ combination or in a (TG)₂/(T)₂ combination, and never in a (TG)₃/(T)₂ or (TG)₂/(T)₁ combination. Because of this observed linkage for (TG)_{*n*} and (T)_{*n*} repeats, the 3/1 and 2/2 forms are each considered a single form of variation, thus limiting permutations of possible repeat combinations to 32.

Secondary structure models

To determine whether variation in repeat lengths may cause differences in V5 structure, secondary structure models of each of the 64 different V5 variant repeat combinations were generated by the Mulfold program of Zuker (22–24). Variants which were not predicted to exist based on the TG/T linkage were included in the analysis in order to reveal how sensitive V5 secondary structure is to a wider range of variant repeat combinations. All 64 combinations of V5 repeats are predicted to fold into a three

helix model (Fig. 3). The –Δ*G* value of secondary structures predicted for actual V5 sequences range from –96.2 to –100.2 kcal/mol. Structural models of potential V5 sequences containing new combinations of known repeat numbers, have a similar range of –Δ*G* values. Sequences comprising the base of helix I are shared by all eukaryotes and form a helical-stem structure (29). This base stem therefore, must anchor the entire V5 stem-loop structure so that the remaining helices form as an extension of the intact stem. Helix I and helix II are invariant among the 64 predicted structures. The four sites of V5 variation are contained in helix III. Variation at these sites only causes differences in number, size, and location of internal loops and bulges in helix III. Because all 64 repeat combinations had the same general structure, additional unobserved but theoretical V5 sequences (containing repeat numbers other than those isolated), were subjected to the Mulfold program with the intent of perturbing the three helix structure. Even after *n* is increased by 2 at each site, the same structure is still predicted. Theoretical folding using only sequences comprising helix III (nucleotides 2117 to 2201), again predicts the same helix III structure (results not shown). This indicates that even without the constraint of the conserved helix I base, the overall secondary structure of the V5 region is not disrupted by sequence variation. In general, the major structural features predicted for the known human V5 sequences are also predicted for the chimpanzee and gorilla V5 sequences (Fig. 3). The mouse sequences however, are more divergent and contain many gaps relative to the aligned human and primate sequences. Although the mouse V5 sequences are predicted to form smaller helix I and helix II structures relative to human and primate, the mouse V5 sequences are predicted to form a human-like helix III structure (Fig. 3), containing invariant T and C sequences in the terminal loop as well as a similar number and distribution of internal loops and bulges.

Table 1. V5 variant repeats

	(CGG) _n	(TG) _n	(T) _n	(C) _n
A1	5	2	2	11
A2	7	3	1	11
A3	7	3	1	11
A4	7	2	2	11
A5	7	3	1	1
A6	5	2	2	11
pHrA	5	2	2	11
pHr12	7	2	2	9
pHr15	7	2	2	12
A11	7	2	2	12
A12	8	3	1	10
A13	7	2	2	12
A14	7	2	2	11
A15	5	3	1	10
A16	7	3	1	11
A17	6	3	1	9

The number of known simple sequence repeats of each V5 variant are listed. A1–A7, isolated from different human individuals (9); pHrA, pHr12, pHr15, isolated from different human individuals (14); A11–A17, isolated from the same human (this study).

Table 2. V5 variant types

1.	A1, A6, pHrA
2.	A2, A3, A5, A16
3.	A4, A14
4.	A11, A13, pHr15
5.	A12
6.	A15
7.	A17
8.	pHr12

All known V5 variants are grouped according to common sequences.

RNase protection assay control experiments

To determine whether the cloned V5 variants are expressed in different tissues of the individual from which they were isolated, an RNase protection assay was established. Each variant was used as a template for *in vitro* transcription of an rRNA-like target and an rRNA-complementary probe. Initially, each variant probe was protected from RNase digestion as a homoduplex (using its own complement as a target) or heteroduplex (protection by another cloned variant target) to determine what band patterns arise from single-target protection. This would permit identification of band patterns that arise from protection by a heterogeneous population of related targets, as would be the case when using total tissue RNA as a target. Variant probes (300 bases) and targets (280 bases) can form a 190 bp duplex, spanning the sites of variation and leaving 5' single stranded tails on both ends of the duplex

(Fig. 4). Protection from RNase digestion of this 190 base fragment generates Band 1 which is easily distinguished from undigested probe (not shown). Differences in the number of the simple repeats at the sites of variation should cause local disruptions in the 190 base duplex region. RNase-sensitivity at these sites results in smaller protected fragments. Accordingly, heteroduplexes with different CGG repeat numbers generate a 147 base band upon RNase digestion which is referred to as Band 2. Differences in the number of C repeats should cause RNase sensitivity at this site and generate a 120 base fragment which is referred to as Band 3. All known V5 clones are invariant in the 120 base region defined by the target-probe duplex 3' end and the C repeat region. Therefore, all RNase-resistant fragments (Band 1, Band 2 and Band 3) should contain this region. RNase digestion of each single target-probe homoduplex yielded a 190 base Band 1, whereas different heteroduplex target-probe combinations protected smaller fragments in a sequence-specific manner (Fig. 5). Light bands between Band 2 and Band 3 are considered incomplete RNase digests at the intermediate TG repeat. Since the probe is labeled throughout, all RNaseA resistant fragments should be detectable. However, smaller fragments generated by RNase cutting at two or more sites are not retained on the gel. The possibility of RNase protection due to probe self-folding and duplex formation is discounted since probe alone is degraded to small fragments (not shown).

Band 1 is only generated from probe A16 in the presence of the A16 target and not by other variant targets (Fig. 5, lanes 30–36). In addition, when the A16 probe is protected by an equal-molar mix of the six different variants, roughly one sixth of the total signal is contained in Band 1 (Fig. 6, lane 50), suggesting that only the A16 target in the mix is responsible for the Band 1 protection. Thus, the A16 probe can discriminate itself from a mixture of related rRNAs. Similarly, the A12 probe Band 1 is protected by the A12 target and less so by the A15 and A17 targets (Fig. 5, lanes 12 and 14). The A15 and A17 target sequences differ the most from A12 in the CGG region, where A12 has eight repeats, A15 has five repeats, and A17 has six repeats (Table 1). The nine base bulge in the A12 probe strand, which is expected to be RNase-sensitive at any of the six C residues in the bulge was, instead, protected by the shorter A15 target. This cross-protection indicates that some heteroduplexes assume an RNase-insensitive conformation. Protection of the A12 probe by the mixed target (Fig. 5, lane 20) results in 13–15% of the signal being localized to Band 1. This strongly suggests that the Band 1 signal is generated by the A12 target, which also contributes one sixth of the total target copies. A limited cross-protection by the single A15 or A17 target and apparent lack of protection by the A15 or A17 targets as part of the mixed target, suggests that the A12 probe can also serve as a probe for self-detection. The A11/A13, A14, A15 and A17 probes generate a strong Band 1 by cross-protection by heterologous targets (Fig. 5) and so cannot be used for self-detection.

To control for the specificity of the RNase protection assay, the panel of human V5 probes was protected from RNase digestion by liver tissue lysate rRNAs from four different primates (chimpanzee, gorilla, gibbon, and rhesus), and by mouse NIH3T3 cell lysate rRNAs. The protected bands were smaller than any protected by the human target sources (data not shown). This result is expected because many sequence differences exist between these species and the human probe at the 3' end of the V5 region (Fig. 2). Sequence alignment indicates that differences

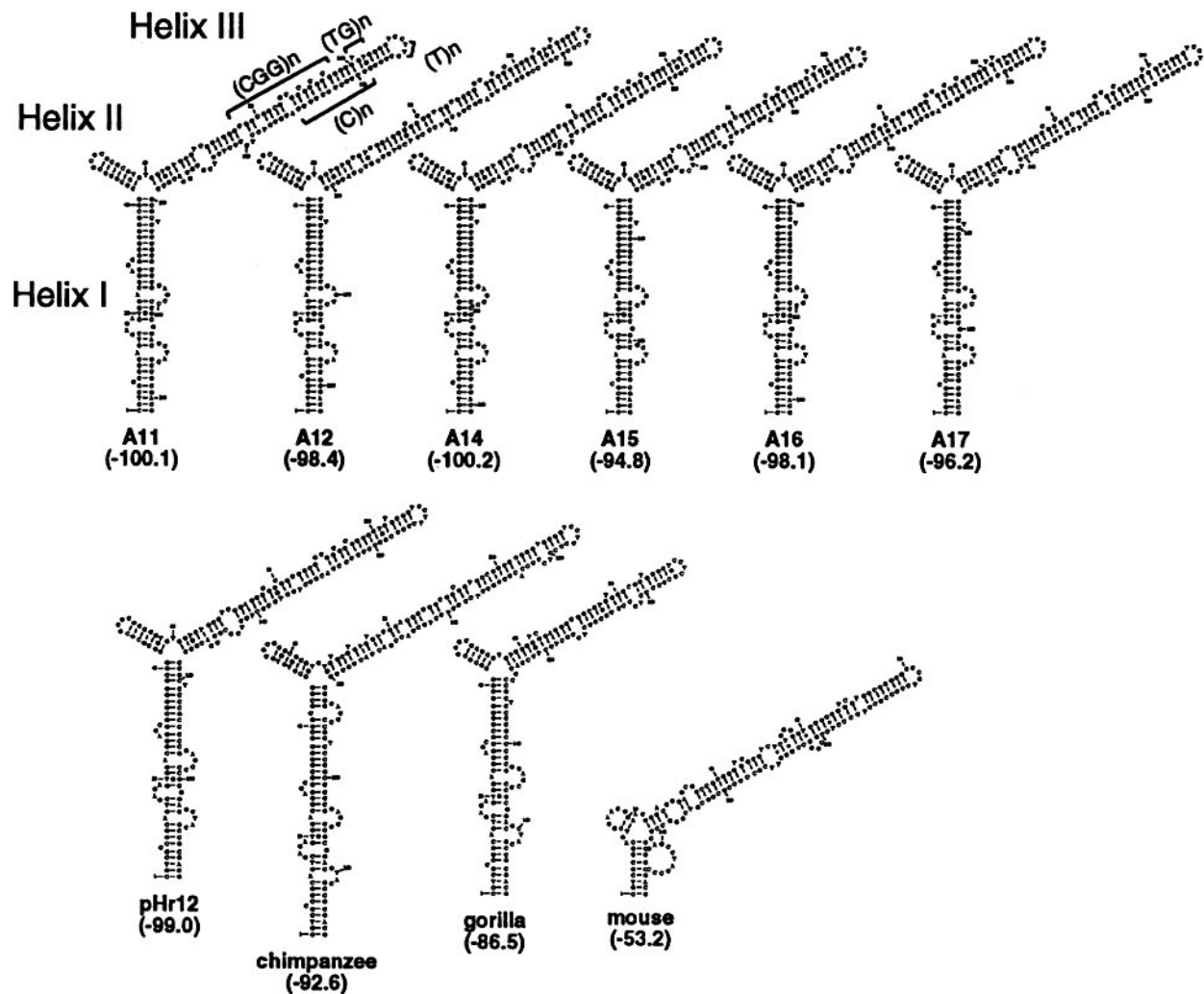


Figure 3. V5 region secondary structure models. The six unique V5 clones (A11–A17) from a single individual were folded by the Mulfold program (22–24). The sequences used for each folding start and end precisely at the sequences forming the base of the V5 region stem. The $-\Delta G$ value for each structure is listed. Also folded by the Mulfold program are V5 sequences from another human variant (pHr12), chimpanzee (Ch), gorilla (Go) and mouse (Mo).

correlate with RNase-sensitive sites in human and primate or human and mouse heteroduplexes.

The results from experiments using the single and mixed (complex) targets and primate and mouse lysate rRNAs, indicate that it is possible to distinguish some individual variants in a population of closely related variant targets. The pattern of protected bands together with their intensity correlate to RNase A activity at specific mismatches in the RNA–RNA duplexes, and relative target abundance. However, it is not possible to calculate the exact contribution of each variant to the overall complex target band pattern due to protection of some V5 probes by heterologous targets.

The amount of a variant is proportionately similar in different tissues

To determine if there is a difference in the expression of a variant in different tissues of an individual, a panel of tissue lysate rRNAs was used to protect each V5 variant from RNase digestion. The same three-band pattern generated by the cloned targets is also

generated by the tissue lysate rRNAs (Fig. 6). In addition, the ratio of band intensities for each probe is identical for each tissue lysate rRNA target. This indicates that the amount of each V5 variant relative to the total amount of rRNA is proportionately similar in different tissues. The fact that some lanes display overall less probe signal is explained by differences in total rRNA amounts in these tissues.

The amount of one variant differs from another

To determine if one variant is expressed in the same proportional amount as another variant, a comparison was made of the proportional amount of each variant probe (A11–A17) that is protected by the rRNAs of a tissue from RNase digestion. In each tissue lysate tested, 10–20% of the A16 probe is protected (as a homoduplex Band 1) (Fig. 6). In contrast, the A12 probe Band 1 is limited to 0–5% in all tissue lysate rRNAs. This indicates that there is relatively more A16 variant than A12 variant in each tissue lysate. As a control experiment, increasing amounts of an *in vitro* transcribed A12 target was added to a constant amount of

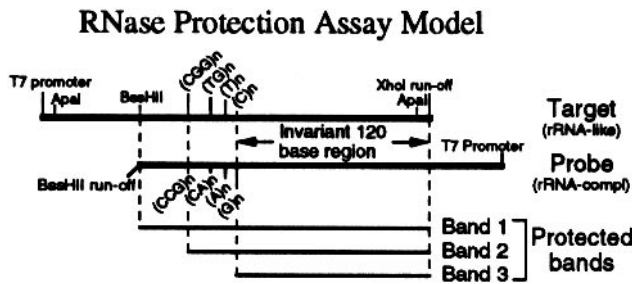


Figure 4. Model of the RNase protection assay. rRNA-complementary probes are protected by rRNA-like targets. The 190 nucleotide region of the target-probe complementarity is defined by the *ApaI* site at the 3' end of the target and the *BssIII* transcriptional run-off site of the probe. Sequences downstream of the 3' *ApaI* site to the *XhoI* run-off site, and upstream of the 5' *ApaI* site to the transcriptional start site are from the pSKII vector. Mismatches internal to this region generate RNase sensitive sites and are detected as smaller bands. Band 2 is generated by RNase sensitivity of mismatches in the CGG repeats and Band 3 is generated by RNase sensitivity of mismatches in the C repeats.

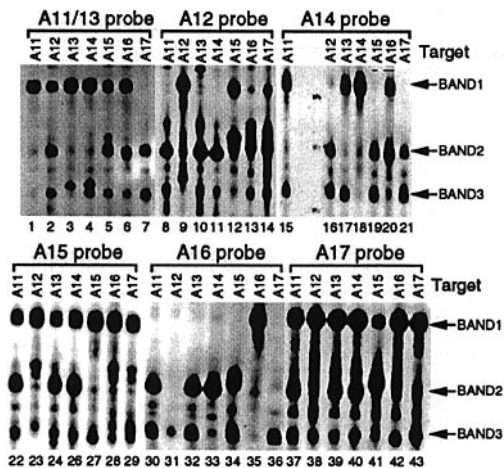


Figure 5. Rnase protection of single variant targets. The protection of one rRNA-complementary probe by one rRNA-like target (each synthesized *in vitro* from A11–A17 cloned V5 templates) from RNase digestion was expected to generate band patterns specific to a target/probe combination. Band 1 would correspond to no RNase sensitivity and Band 2 and Band 3 would correspond to RNase sensitivity at the CGG and C repeats respectively. Band 1 is easily distinguished from the full length probe. No full length probe remains after RNase treatment under normal reaction conditions (not shown).

muscle tissue lysate. This resulted in an A12 target concentration dependent increase in the Band 1 signal (data not shown) proving that if A12 is present, it would be detected. The near absence of an A12 probe Band 1 indicates that the concentration of the cross-protective A15 and A17 variants is also low (Fig. 6). The strong protection by tissue lysate RNAs yielding a Band 1 for the A15 and A17 probes is explained by a cross-protection by other variant rRNAs, as seen in control experiments. RNase protection assay experiments using isolated total RNA rather than tissue lysate RNA as protective targets yielded identical results (data not shown).

We concluded, based on the strength of the control RNase protection results, that the bands generated by human V5 target protection by tissue lysates are real, that other sites in the V5 region contribute little if any to the overall V5 variation, and lastly

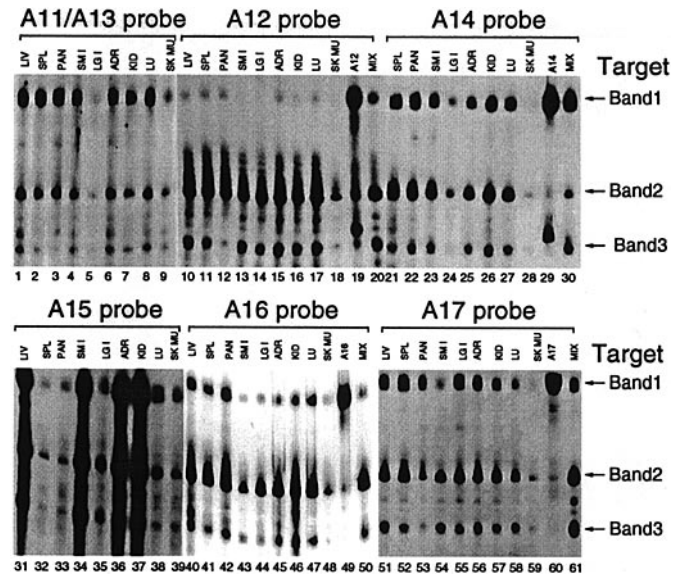


Figure 6. RNase protection of tissue-derived RNAs. Each cloned variant probe (A11–A17) was protected in an RNase protection assay by rRNAs from different tissues of the test individual: liver (LIV), spleen (SPL), pancreas (PAN), small intestine (SMI), large intestine (LGI), adrenal (ADR), kidney (KID), lung (LU), skeletal muscle (SK MU). Equal molar mix of the six different cloned variants (MIX).

that if more V5 sequence variation existed in regions of V5 flanking the four known sites of variation, it would be detected.

DISCUSSION

V5 sequence variation

In this study, we sought to determine the nature and extent of V5 sequence variation in the rRNA genes of an individual and characterize the level of expression of a variant in different tissues of an individual. Our lab has now detected V5 variation by three distinct methods: previous S1 protection studies; current RNase protection studies; direct sequencing of genomic clones in this and previous studies. Because the results derived from the three different methods are in agreement, we strongly believe our results accurately reflect the nature and extent of V5 variation. A study by Leffers and Andersen (20) used an RT-PCR approach to isolate V8 variants. In our hands, isolating V5 clones using this approach proved error-prone and had to be abandoned because it generated numerous deletion artifacts. While the original goal of this study was to isolate many more copies of the V5 region, RNase protection results indicate that the A16 and A12 probes and possibly the A15 probe, comprise up to 25% of the expressed rRNAs without having considered the other variants' contributions. Additional V5 sequences are being determined in another study devoted to chromosome specific variation. It remains to be determined whether this group of seven genomic V5 variants, which detects a significant proportion of expressed rRNAs in this individual, accurately reflects genomic variation at the population level.

From the sequence analysis, it has been noted that despite the presence of a dozen or so distinct simple repeats existing in the human V5 region, only four show variation. The questions arise and are still unresolved: why are some simple sequence repeats invariant while immediately adjacent repeats vary in copy

number? In addition, what mechanisms generate copy number (n) variation and yet limit the range? If each combination of the known copy numbers at each repeat site is considered, there are 32 permutations; this may represent an upper limit to the number of V5 variants. The limit of V5 size variation, and so a limit on sequence variation, is supported by the Southern detection of a narrow 270 bp *ApaI*-digested genomic DNA fragment. If a wider range of (n) exists, forming longer or shorter V5 fragments, a broader band would have been detected by Southern blot hybridization than that seen (not shown). Therefore, if extreme variants exist, they are in low copy number. More V5 sequence variation is likely to exist within this individual, since only one of seven clones is redundant. However, further variation may be limited to combinatorial differences of existing repeats such that the lengths of other V5 repeat combinations do not define variants longer or shorter than those already identified. The extent of V5 variation in an individual's 400 rRNA gene copies would be better characterized by the exhaustive cloning and sequencing of additional variants.

Evolution of V5 sequence

When human V5 sequences are compared with V5 sequences of chimpanzee, gorilla, and mouse (Fig. 2), differences in copy number of simple sequence repeats are noted between each species. Because only one copy of each species' V5 region (other than human) was used for comparison, it cannot be determined which repeat motifs vary in other species' V5 regions. However, a simple sequence repeated in another species and either present in one copy, or not present at all in the human, may represent an expansion relative to the human V5 region. An example of this is the chimpanzee (CCG)₃ repeat (27 bases 3' to the human C repeat, Fig. 2) which is not present in the human, gorilla, or mouse V5 region. This may represent a chimpanzee-specific repeat. Another example is the (ACGGG)₃ repeat in the mouse, located 18 bases downstream of the 5'-V5 border relative to the human sequence. This site does not vary among the human V5 clones and represents a potential site for mouse V5 variation. Each of the human V5 repeat motifs are found in chimpanzee and gorilla, while the CGG and TG repeats are absent in mouse. Of the human V5 simple repeats also found in other species, none are found to be repeated to the same extent as in humans which makes the human V5 region expanded relative to other species' V5 regions. Sequence comparisons of additional V5 clones from many V5-containing species will help define species-specific variable repeats and suggest mechanisms of mutation active in the mammalian-specific V5 region.

V5 region secondary structure

A comparison of the secondary structures predicted for the V5 and V8 sequences, and the nature of sequence variation within these variable regions, indicates that mechanisms which generate variation may differ from variable region to variable region. The study by Leffers and Andersen (20) reported simple sequence variation at two sites (35 differences in 111 cDNAs) in the 700 bp human rRNA gene V8 region (nt 2877–3586). The secondary structure model for the V8 region predicts that the two regions of variation base-pair with each other and is supported by the presence of compensatory mutations (30) within these two sites which preserve the structure. The sites of variation in the smaller V5 region (180 nt) are not predicted to base-pair with each other

as in the V8 model. Although the four hotspots of V5 variation partially align in helix III (Fig. 3), our model does not require compensating mutations to maintain the structure predicted for the 64 possible V5 repeat combinations. The V8 region is much longer than the V5 region and may require a more rigorous maintenance of specific base interactions to maintain proper secondary structure. Perhaps the maintenance of a structure is a result of the selective pressure to maintain or generate specific repeat combinations in the V5 region which do not form aberrant structures (31) as is the case suggested for limiting sequence variation of the V8 region (20). Since V5 clones differ only in the number of simple sequence repeats (Fig. 2), it appears that only slipped-strand mispairing occurs which adds or removes repeat units in the V5 region. This is in contrast to the V8 region where base changes and insertions/deletions occur, which can not be explained by slipped-strand mispairing, in addition to events that cause variation in simple sequence repeat number. The compensatory mutations of the V8 region and apparently not of the V5 region, suggest that variation of these two regions is generated and maintained either by different mechanisms, or by similar mechanisms but with the additional requirement for a compensatory mutation in the V8 region.

Gene dosage or regulated expression?

The results of this study indicate that the relative amount of a variant rRNA is similar in diverse tissues of an individual. This suggests that the expression of an individual's 400 or so rRNA gene copies is similar from one tissue to the next. This study also identified differences in the relative amounts of rRNA variants in diverse tissues of an individual and indicated that these differences are the same in all tissues tested. Table 2 indicates that of the relatively limited pool of all cloned V5 variants, some V5 variants have been isolated more frequently. If the pool of cloned V5 variants reflects the genomic distribution, it is possible that differences in V5 variant expression result from a gene dosage effect. The possibility of differential rRNA gene regulation however, must still be considered based on experimental results which could be interpreted as regulated gene expression. This possibility should be considered in light of detecting active gene clusters (NORs) by selective silver-staining of the nucleolar-protein nucleolin (32) in comparison with inactive gene clusters (33). The presence of nucleolin has been implicated as a marker for transcriptionally active rRNA genes and pre-ribosomal processing (34). Additional results include detecting differential states of rDNA methylation (35), rDNA nucleosome-association (36,37), and nucleolin phosphorylation by protein kinase NII (PKNII) (38); these have all been shown to affect the transcriptional activity of rRNA genes. These experiments however, monitored global regulation of total rRNA and did not make the distinction between rRNA variants.

Do V regions have a function?

Roles suggested for V regions include: not being required for translation in yeast (39); required to prevent lethality as result of the deletion of the V8 region in *Tetrahymena thermophila* (40); functioning as a target for rRNA cleavage in programmed cell death (41,42); and conferring sequence specificity for mRNA translation (19,43).

The large amount of sequence diversity contained in the V regions is becoming increasingly apparent, raising the possibility

that no two 28S rRNA genes are identical. V region sequence variation however, may not have an effect at the level of the ribosome. In addition, since human, chimpanzee, and gorilla V5 sequences each can form a similar basic structure, the search for any V5 function should focus on the entire V5 region and not on subtle differences within the region.

REFERENCES

- Gorski, J.L., Gonzalez, I.L., and Schmickel, R.D. (1987) *J. Mol. Evol.* **24**, 236–251.
- Noller, H.F., Hoffarth, V., and Zimniak, L. (1992) *Science* **256**, 1416–1419.
- Noller, H.F. (1993) In Gesteland, R.F. and Atkins, J.F. (eds.), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 137–156.
- Purohit, P. and Stern, S. (1994) *Nature* **370**, 659–662.
- Otsuka, T., Nomiya, H., Yoshida, H., Kukita, T., Kuhara, S., and Sakaki, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3163–3167.
- Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., and Dover, G.A. (1988) *Mol. Biol. Evol.* **5**, 366–376.
- Ware, V.C., Tague, B.W., Clark, C.G., Gourse, R.L., Brand, R.C., and Gerbi, S.A. (1983) *Nucleic Acids Res.* **11**, 7795–7817.
- Hassouna, N., Michot, B., and Bachellerie, J.-P. (1984) *Nucleic Acids Res.* **12**, 3563–3583.
- Gonzalez, I.L., Gorski, J.L., Campen, T.J., Dorney, D.J., Erickson, J.E., Sylvester, J.E., and Schmickel, R.D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7666–7670.
- Hancock, J.M. and Dover, G.A. (1988) *Mol. Biol. Evol.* **5**, 377–391.
- Clark, C.G., Tague, B.W., Ware, V.C., and Gerbi, S.A. (1984) *Nucleic Acids Res.* **12**, 6197–6220.
- Brosius, J., Dull, T.J. and Noller, H.F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 201–204.
- Gonzalez, I.L., Sylvester, J.E., Smith, T.F., Stambolian, D., and Schmickel, R.D. (1990) *Mol. Biol. Evol.* **7**, 203–219.
- Maden, E.H., Dent, C.L., Farrell, T.E., Garde, J., McCallum, F.S., and Wakeman, J.A. (1987) *Biochem. J.* **246**, 519–527.
- Schlotterer, C. and Tautz, D. (1992) *Nucleic Acids Res.* **20**, 211–215.
- Levinson, G. and Gutman, G.A. (1987) *Nucleic Acids Res.* **15**, 5323–5338.
- Dover, G.A., and Flavell, R.B. (1984) *Cell* **38**, 622–623.
- Erickson, J.M., and Schmickel, R.D. (1985) *Am. J. Hum. Genet.* **37**, 311–325.
- Gonzalez, I.L., Sylvester, J.E., and Schmickel, R.D. (1988) *Nucleic Acids Res.* **16**, 10213–10224.
- Leffers, H. and Andersen, A.H. (1993) *Nucleic Acids Res.* **21**, 1449–1455.
- Devereux, J., Haerberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–396.
- Zuker, M. (1989) *Science* **244**, 48–52.
- Jaeger, J.A., Turner, D.H., and Zuker, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7706–7710.
- Jaeger, J.A., Turner, D.H., and Zuker, M. (1989) *Methods Enzymol.* **183**, 281–306.
- Thompson, J. and Gillespie, D. (1987) *Anal. Biochem.* **163**, 281–291.
- Haines, D.S. and Gillespie, D. (1992) *Biotechniques* **12**, 736–741.
- Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
- Kuo, B.A. (1994) Thesis: Human Ribosomal RNA Variation in a Single Individual. Hahnemann University.
- Raue, H.A., Klootwijk, J., and Musters, W. (1988) *Prog. Biophys. Mol. Biol.* **51**, 77–129.
- Hancock, J.M., and Dover, G.A. (1990) *Nucleic Acids Res.* **18**, 5949–5954.
- Linares, A.R., Hancock, J.M., and Dover, G.A. (1991) *J. Mol. Biol.* **219**, 381–391.
- Roussel, P., Belenguer, P., Amalric, F., and Hernandez-Verdun, D. (1992) *Exp. Cell Res.* **203**, 259–269.
- de Capoa, A., Felli, M.P., Baldini, A., Ba, M., Archid, N., Alexizandre, C., Miller, O.J., and Miller, D.A. (1988) *Hum. Genet.* **79**, 906–999.
- Lapeyre, B., Bourbon, H., and Amalric, F. (1987) *Nucleic Acids Res.* **84**, 1472–1476.
- Flavell, R.B., O'Dell, M., Thompson, W.F., Vincenz, M., Sardana, R., and Barker, R.F. (1986) *Phil. Trans. R. Soc. London* **314**, 385–397.
- Conconi, A., Widmer, M., Koller, T., and Sogo, J.M. (1989) *Cell* **57**, 753–761.
- Dammann, R., Lucchini, R., Koller, T., and Sogo, J.M. (1995) *Mol. Cell. Biol.* **15**, 5294–5303.
- Belenguer, P., Baldin, V., Mathieu, C., Prats, H., Bensaid, M., Bouche, G., and Amalric, F. (1989) *Nucleic Acids Res.* **16**, 6625–6636.
- Musters, W., Venema, J., van der Linden, G., van Heerikhuizen, H., Klootwijk, J., and Planta, R.J. (1989) *Mol. Cell. Biol.* **9**, 551–559.
- Sweeney, R., Chen, L., and Yao, M.-C. (1994) *Mol. Cell. Biol.* **14**, 4203–4215.
- Houge, G., Doskeland, S.O., Boe, R., and Lanotte, M. (1993) *FEBS Lett.* **315**, 16–20.
- Houge, G., Robaye, B., Eikhom, T.S., Goldstein, J., Mellgren, G., Gjersten, B.T., Lanotte, M., and Doskeland, S.O. (1995) *Mol. Cell. Biol.* **15**, 2051–2062.
- Gerbi, S.A. (1996) In Zimmermann, R.A. and Dahlberg, A.E. (eds), *Ribosomal RNA—Structure, Evolution, Processing and Function in Protein Synthesis*, CRC Press, Boca Raton, pp. 71–87.