Volume 16 Number 21 1988 Nucleic Acids Research

Human 28S ribosomal RNA sequence heterogeneity

Iris Laudien Gonzalez+*, James E.Sylvester+ and Roy D.Schmickel

Department of Human Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA

Received July 7, 1988; Revised and Accepted October 5, 1988

ABSTRACT

DNA sequencing of several cloned human 28S ribosomal RNA gene fragments has revealed sequence heterogeneity (1) but it was not clear whether these are inactive pseudogenes or are active genes that are transcribed and represented in ribosomes. S1 nuclease analysis allowed us to examine the population of ribosomal RNA molecules of a cell, and we found that 28S rRNA is a heterogeneous assortment of molecules in both mono- and polysomal preparations. Sequence variation, although largely concentrated in variable regions of the molecule, apparently also occurs in the conserved regions.

INTRODUCTION

The approximately 400 copies of ribosomal RNA genes of the human genome have been considered to be identical and to produce a homogeneous population of 18S and 28S ribosomal RNAs. Comparative studies of large subunit ribosomal RNA genes of several species have shown that the genes consist of alternating "conserved" and "variable" regions (2,3). The "conserved" regions show similarity of length, sequence, and secondary structure of the RNA. In contrast, the "variable" (V) regions differ not only in length and sequence, but also in secondary structure. The 28S rRNA gene contains 11 variable regions (V1-V11); several of these regions have a very high content of G+C, similar to that of the ribosomal spacers (2,4). The differences between 28S rRNA genes of closely related species, such as human, gorilla and chimpanzee, are concentrated in the variable regions (4).

We and others have demonstrated that several cloned human ribosomal genes contain variation in their sequences (1,5). The sequence heterogeneity in this small group of genes appears to be limited to the variable regions. The frequency of variation is

high. For instance, we found three variants of a short 180 base segment (V5) among six clones. We therefore undertook to test the ribosomal RNA to determine whether variant genes are transcribed and the RNA product is incorporated into ribosomes.

S1 nuclease analysis with defined probes allowed us to test the whole population of cellular RNA at once. We chose probes from a 900 base section of the human 28S rRNA because it included several short variable regions: V3 (30 bases), V4 (29 bases) and V5 (180 bases) (Figure 1). Moreover, V3 and V5 have shown inter-species differences in primates (4), and V5 was known to contain intra-species differences (1,5) at the gene level. If rRNA contained intra-species sequence heterogeneity, it would more likely be found in these regions. Probes covering different regions were synthesized by primer extension and hybridized to total RNA obtained from human leukocytes and to monosome and polysome fractions from tissue culture fibroblasts. Subsequent S1 digestion and gel electrophoresis analysis revealed appreciable sequence heterogeneity in 28S rRNA. We discuss these results in the context of rRNA secondary structure, function and evolution.

MATERIALS AND METHODS

Cells - Leukocytes for total RNA preparation were separated by centrifugation on Ficoll-Paque from 10 ml blood samples obtained from six laboratory members. Fibroblasts for polysome rRNA preparation belonged to a normal human cell line (University of Pennsylvania Cell Center CC91) and were grown in RPMI 1640 medium supplemented with 10% fetal calf serum and 1% PenStrep.

RNA preparation - RNA from fresh leukocytes was prepared by lysing the cells in GuSCN (4 M guanidine thiocyanate, 1 M ß-mercaptoethanol, 0.1 M Na acetate, 0.01 M EDTA), adding 1 g CsCl/2.5 g lysate and layering this solution on a 2.5 ml cushion of 5.7 M CsCl/0.1 M EDTA. Centrifugation was performed overnight in a SW41 rotor at 25,000 rpm, 15°C. The supernatant was quickly decanted, the sides of the tubes were wiped dry, the pellet was resuspended in 200 µl TE (Tris 10 mM, EDTA 1 mM); the suspension was extracted with chloroform/butanol = 4/1, and the organic phase was back-extracted with 200 µl TE. The aqueous phases were

combined and precipitated by adding 1/10 volume of 3 M Na acetate and 2.5 volumes ethanol.

Polysome RNA preparation - Polysomes were prepared from exponentially growing CC91 cells. Approximately 10^7 cells were harvested, washed, lysed (in buffer containing 10 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.25 M sucrose, 0.2% Triton, 0.05% deoxycholate, 0.2 mg/ml heparin), centrifuged to remove nuclei and mitochondria. The supernatant was layered on a 15-50% linear sucrose gradient (with 10 mM HEPES, pH 7.5, 10 mM KCl, 3 mM $MgCl_2$, 1 mM DTT, 0.2 mg/ml heparin) and centrifuged in a SW28 rotor at 27,000 rpm for 3.5 hours. 1 ml fractions were collected while monitoring UV absorbance. Fractions were extracted with phenol/chloroform before precipitating with 1/10 volume 3 M Na acetate and 2 volumes ethanol. Pellets were resuspended in water.

In vitro transcription - In vitro transcripts were prepared from 1 ug of rRNA clones containing Bam_{1406} -Sma_{2304} inserts in Bluescript vectors with the Bluescript transcription kit following manufacturer's instructions (Stratagene).

Probe preparation. - Uniformly labeled single stranded probes and 5' end labeled probes of defined length were synthesized by primer extension on single stranded DNA of M13 clones described previously (4). The synthesis was performed after annealing 0.1 μg template DNA, 2-12 ng primer, 2 μl of 10x 50mM NaCl buffer in a total volume of 13.5 μl, that was allowed to slow-cool from 65° to <40°C. The following were added (for fully labeled probes): 1 μl of 100 mM DTT, 2 μl of dNTP mix (100 μM of dGTP, TTP and dATP, and 15 μM of dCTP), 2 μl of α^{32}P-dCTP (Amersham, specific activity 3000 Ci/mmol), and 1.5 units Klenow fragment (US Biochemicals). Incubation at 50°C for 10 minutes was followed by addition of 2 μl chase mix (2 mM each of dATP, dGTP, dCTP and TTP), and incubation was continued at 50°C for another 10 minutes. 5' end labeled probes were synthesized at 50°C for 10 minutes with only 1 μl of 100 mM DTT, 1 μl of α^{32}P-dCTP, 1 μl of 10 mM dATP, and 1.5 units Klenow fragment, followed by a 10 minute chase (2 μl of 10 mM dCTP, 2 μl of chase mix) that permitted completion of the synthesis. In each case the synthesized double-stranded DNA was cut with the appropriate enzyme (after

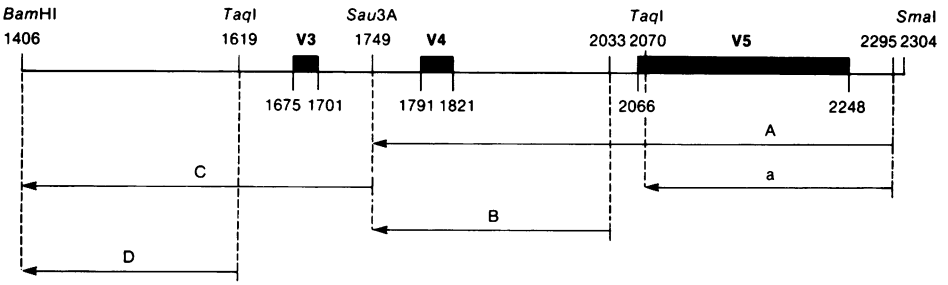

Figure 1. Map of human 28S rRNA region and probes used to test for rRNA sequence heterogeneity. The shaded blocks correspond to variable regions.

adjusting the buffer concentration) to release the probe fragment. To ensure uniformly sized probe fragments, the staggered ends were filled in by adding 1 μl 100 mM DTT, 1 μl chase mix and 0.5 units Klenow and incubating 10 minutes at 50°. The volume was increased to 100 μl, and the reaction mix was extracted once with phenol and once with chloroform. The probe was then mixed with an equal volume of formamide loading dye (0.3% xylene cyanol, 0.3% bromophenol blue, 0.01 M Na_2EDTA, in deionized formamide), boiled for 3 minutes, loaded onto a 4% acrylamide/7 M urea gel and electrophoresed overnight (8mA; 3V/cm). The gel was wrapped in plastic and exposed to X-ray film for 15-30 minutes to localize the probe band, which was then electroeluted in TBE buffer (Tris 0.9 M, Borate 0.89 M, EDTA 0.025 M). The electroeluted probe (in 2-4 ml volume) was extracted with phenol and with chloroform before co-precipitating with RNA.

Figure 1 shows locations within the 28S gene of the five probes described in this article. Probes A and a were prepared in both fully-labeled and 5'-labeled forms, to determine the location of S1 cuts. Probes A, a, and B were synthesized using ribosomal sequence specific primers as opposed to the universal M13 primer.

Annealing. Typically, 15 to 100 μl of probe were precipitated with 0.3 μg RNA, 15 μg carrier tRNA, 1/10 volume of 3 M Na acetate and 2 1/2 volumes ethanol. RNA/probe precipitates were resuspended in 10 μl formamide annealing buffer (0.4 M NaCl, 1 mM Na_2 EDTA, 0.04 M PIPES, in 80% formamide), dissociated at 95°C

Table 1

SEQUENCE DIFFERENCES IN REGION V5 OF HUMAN rDNA CLONES

Variant Clone	$[CGG]_n$	$[UG]_n$	$[U]_n$	$[C]_n$
	2132-2152	2158-2163	2170-1	2177-2187
*A1, A6	5	2	2	11
*A2, A3, A5	7	3	1	11
*A4	7	2	2	11
**pHrA	5			11
**pHr12	7			9
**pHr15	7			12

* Ref. 1
** Ref. 5

for 3 minutes, then quickly transferred to the proper temperature for annealing for 45 minutes. [Probes A,a,C,D were annealed at 65°C, probe B at 57°C.]

S1 digestion. The amount of S1 nuclease to be used was determined by pre-tests with each lot of enzyme. 25-200 U S1 nuclease (Boehringer Mannheim) in 200 µl S1 buffer (30 mM Na acetate pH 4.6, 50 mM NaCl, 1 mM $ZnSO_4$, 5% glycerol) were rapidly added to the annealed probe/RNA samples, and digestion was carried out at 37°C for 40-60 minutes. The samples were precipitated with carrier tRNA and 2 1/2 volumes ethanol.

Gels. RNA-protected DNA precipitates were resuspended in 10 µl formamide loading dye, boiled for 3 minutes and kept on ice before loading onto pre-warmed 6% acrylamide/7 M urea sequencing gels. Each set of samples was run next to a sequence of the probe, to allow sizing of the fragments.

All steps from electroelution through the separation of fragments on sequencing gels were carried out in a single day to prevent breakdown artifacts.

RESULTS

Our original six variant ribosomal DNA clones were characterized by sequencing the entire 898 base pair stretch from $BamHI_{1406}$ to $SmaI_{2304}$ and shown to differ only in three regions of V5, as shown in Table 1. The variant with the longest V5 region (clone A3) was chosen for probe preparation. This is also the most prevalent variation among the six clones.

(a) Synthetic Transcript Controls

Control S1 nuclease protection experiments were performed with in vitro RNA transcripts from variants A1 and A3, which had been subcloned in a Bluescript (Stratagene) vector. Transcripts complementary to and similar to rRNA could be obtained by initiating from either the T3 or T7 promoters. As expected, transcripts complementary to rRNA did not protect the labeled probe (data not shown). On the other hand, experiments with rRNA-like transcripts derived from A3 should fully protect the probe, while predictable specific probe fragments should be found when A1-derived transcripts are used. These controls are shown in Figures 2a and 2b, where lanes A1 and A3 contain the heterologous and homologous transcripts respectively. Lane A1 shows a series of bands next to the $[GGC]_n$ segment of the sequence. This is expected because the annealing of $[GGC]_6$ to $[GGC]_4$ can occur in various registers. These bands are appropriately absent from the homologous A3 lane. Lane A1 also shows slightly more intense bands at location 2158, which corresponds to the site where the probe has an extra [UG] pair, and at 2170, where the probe has one less U than the A1 transcript. Since the major mismatch occurs in the $[GGC]_n$ stretch, and since under-digestion conditions are used, there is less digestion at these secondary sites. The homologous transcript A3 in lane A3 does not show these bands, although it shows bands near the top of the gel which are ascribed to artifacts of probe folding (also seen in the mock reaction) and of digestion within 11 bases of the 3′ end of the probe. Not shown are control experiments with mixtures of the two transcripts, which should and did give some fully-protected probe and some digested probe, according to the proportions of transcripts used.

The control experiments helped to define the annealing and digestion conditions for the remaining experiments. However, experiments were carried out under incomplete digestion conditions to exclude over-digestion artifacts, as can be seen from the remaining intact probe in lanes A1 and P+S1. Thus, the variability we demonstrate in cellular rRNA is an underestimate in terms of the proportion of variant molecules and probably also in terms of single base mismatches detected.

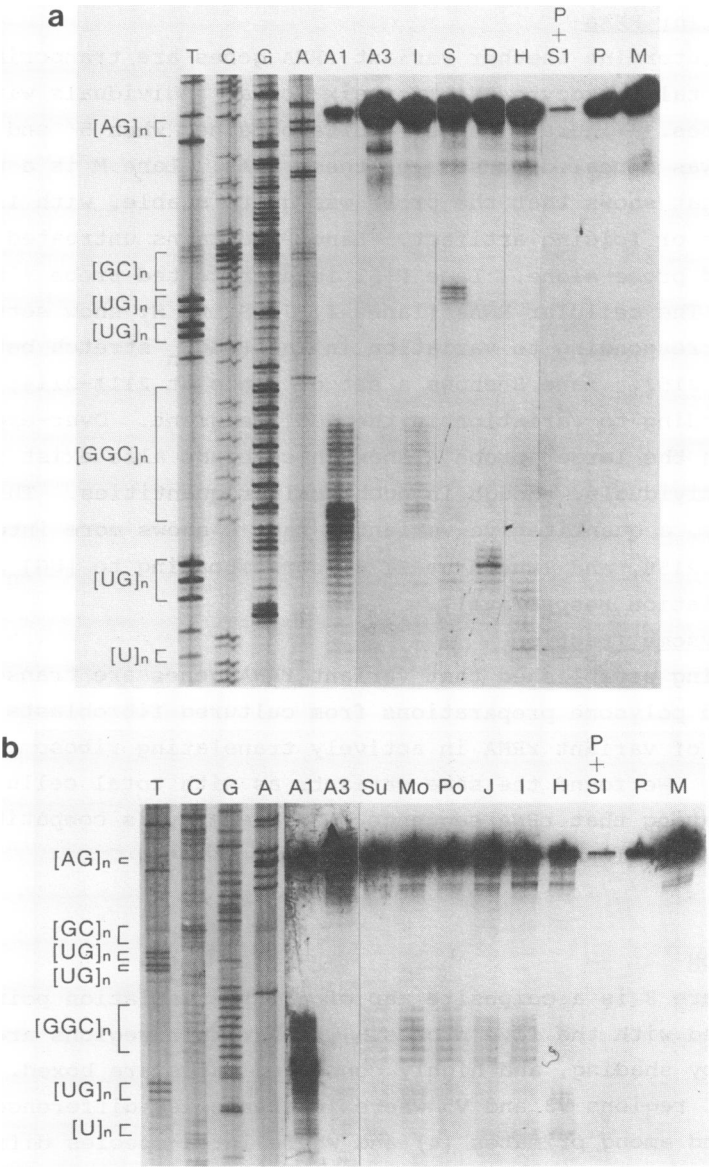

Figure 2. Representative S1 nuclease protection experiments with 5' labeled probe a. The gel shows the fragments generated by S1 nuclease digestion of annealed probe/test RNA, next to a sequence of the probe. Lanes A1 and A3: heterologous and homologous transcripts; lane M: mock reaction; lane P: 1 μl probe; lane P+S1: probe + S1; lanes I, J, S and D: total RNA from normal subjects; lane H: total Hela RNA. In part (b) lanes Su, Mo, Po contained subunits, monosomes and polysomes from cultured fibroblasts.

(b) Cellular RNAs

To determine whether variant rRNA genes are transcribed, we probed total leukocyte RNA from six normal individuals with our five probes. Figure 2 shows results obtained when 5′ end labeled probe a was annealed to some of these RNAs. Lane M is a mock reaction that shows that the probe was quite stable, with little breakdown or folding artifact. Lane P contains untreated electroeluted probe alone. Lane P+S1 is unprotected probe + S1 nuclease. The cellular RNAs (lanes I, J, S, D, H) show sets of bands corresponding to variation in the $[GGC]_n$ stretch between 2132 and 2152. Lane S shows a set of bands at 2111-2114, corresponding to variation in the $[GU]_2$ segment. Over-exposed gels with the larger probe A show these bands also exist in the other individuals, though in much smaller quantities. This is, therefore, a quantitative variant. Lane D shows more intense bands at 2158, and lane H at 2170, corresponding to $[UG]_n$ and to $[U]_n$ variation respectively.

(c) Polysome fraction

Having established that variant rRNA genes are transcribed, we tested polysome preparations from cultured fibroblasts for the presence of variant rRNA in actively translating ribosomes (Figure 2b). We found the same variants as with total cellular RNA, demonstrating that rRNA sequence heterogeneity is compatible with ribosome function.

DISCUSSION

Figure 3 is a composite map of all the variation points identified with the five probes. The variable regions are denoted by shading, and highly conserved bases are boxed. As expected, regions V3 and V5 where inter-species differences have been found among primates (4) and where intra-species differences have been demonstrated (1,5) at the gene level, also show RNA sequence heterogeneity. At the gene level, differences had been identified at four sites in V5 (Table 1). At the rRNA level, we identify additional sites at $[GA]_n$ 2072-2075, $[CG]_n$ 2099-2114, $[UG]_n$ 2115-2118, $[UG]_n$ 2120-2123. In addition, our results appear to indicate variation in region V4 and in the "conserved"

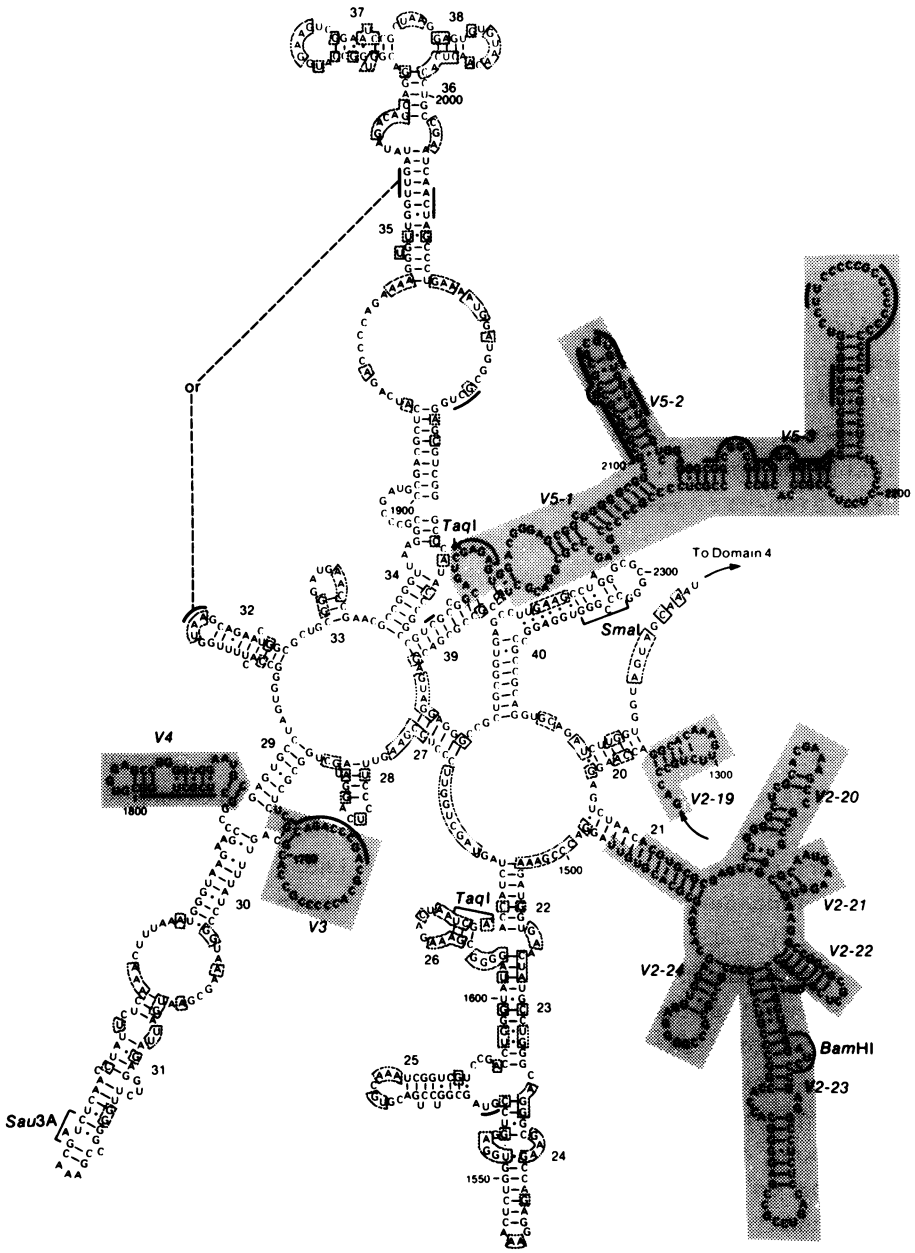

Figure 3. Secondary structure of human 28S rRNA region showing locations of sequence heterogeneity as detected by these S1 protection studies. Variable regions are shaded. Restriction sites that define the probes are also indicated. Bases that are conserved throughout evolution are boxed.

sequences of the gene segment examined, where no intra- or interspecies differences had been detected by our limited gene sequencing. However, S1 nuclease protection assays are more sensitive in the sense that they permit one to characterize the whole population of molecules present in a cell. On the other hand, apparent mismatches at some points could be the result of localized differences in melting temperatures.

Sequence differences can occur in (double-stranded) stem-forming regions and in single-stranded regions. Differences in stem regions could alter the secondary structure if they prevent base-pairing or if they create new base-pairing. The consequences of variable numbers of [GGC] between 2132-2152 have been illustrated in (1). The consequences of variable numbers of [GC] in the 2100-2113 region, or of [GU] in the 2114-2117 and 2119-2122 regions would be to lengthen or shorten the stem in which they are located. Insertions or deletions in the same regions can also add or remove bulged bases, which may be involved in protein-nucleic acid interactions (6). Our S1 nuclease protection experiments also suggest that some mutations may be located in conserved and in variable single-stranded regions as shown in Figure 3. Single-stranded regions in the 28S rRNA molecule often consist of sequences that are highly conserved throughout evolution (2), and in 5S rRNA single-stranded loops are under-represented as sites of mutation (7). It is known that ribosome function is dependent on the integrity of certain single-stranded regions (8). Thus, mutations in conserved single-stranded regions might alter ribosome function or specificity. Figure 3 shows several sites of possible sequence heterogeneity located in single-stranded parts of the molecule.

The experimental conditions used for our S1 assays are such that larger mismatches will be preferentially detected; length differences of one or two bases and single base substitutions will tend to be missed. Thus, the rRNA sequence heterogeneity we have found is an under-representation of the total. Indeed, under full-digestion conditions, we find that only about one-half of the probe is fully protected. Yet, it is instructive to examine the kind of variation found, which apparently consists largely of differences in the numbers of short repeat motifs $[GGC]_n$,

$[GU]_n$, $[GC]_n$, $[AG]_n$, and in the length of homopolymer tracts ($[C]_n$, $[U]_n$). The nature of the variation at other locations is uncertain and must await characterization by sequencing of cDNA clones. cDNA clones will not only clarify the nature of the mutations but will also show whether the variation is randomly distributed or occurs in specific combinations (linked heterogeneity). The rRNA variation found is consistent with the predominant types of inter- and intra-species gene sequence differences of higher primates (1,4,5) and can originate from unequal homologous exchanges among the tandemly repeated genes (9,10) and from slipped-strand mispairing during replication (11,12). The variation may then spread among the many gene copies by the mechanisms of concerted evolution (9,10,13,14,15) and by gene conversion (16).

Whether or not the rRNA sequence heterogeneity is functionally important is not yet known, since the function of the regions affected is unknown. However, comparative alignments of primate 28S DNA sequences have shown what may be compensatory mutations in the variable regions (4). Maden (5) has reported such compensatory mutations in human intra-species comparisons of region V8; we find additional compensatory mutations in that region (Gonzalez, unpublished). This would suggest that these regions, which are thought to be remnants of spacers or introns (1,17) do indeed have a secondary structure requirement. Several reports of sequence complementarity and of cross-hybridization between mRNAs or genes and rRNA involve these variable regions (ferritin H-chain mRNA (18); Herpes Simplex Type I (19); ã2(I) collagen (20), raising the possibility of their involvement in preferential translation.

Heterogeneity among rRNA molecules has been demonstrated using S1 nuclease protection assays. The variation is of the same type reported for rRNA genes, and is found not only in total rRNA but also in rRNA extracted from polysomes. Although not all individuals can be distinguished by using our small array of probes, personal variants can be found. Experiments are under way to examine possible tissue specific or preferential expression of particular variants. We are also exploring the full extent of rRNA heterogeneity by examining other variable regions.

ACKNOWLEDGEMENTS

This investigation was supported by NIH grant HG 16930 to RDS. We thank Vickie Bennet for help with the polysome preparations.

+Present address: Hahnemann University, Pathology Department, Broad and Vine, Philadelphia, PA 19102-1192, USA

*To whom correspondence should be addressed

REFERENCES

1. Gonzalez, I.L., Gorski,J.L., Campen, T.J.,Dorney, D.J.,Erickson, J.M., Sylvester, J.E., and Schmickel, R.D. (1985) Proc. Natl. Acad. Sci. USA 82: 7666-7670.
2. Gorski, J.L., Gonzalez, I.L., and Schmickel, R.D. (1987) J. Mol. Evol. 24: 236-251.
3. Hassouna, N., Michot, B., and Bachellerie, J.-P. (1984) Nucleic Acids Res. 12: 3563-3585.
4. Gonzalez, I.L., in preparation.
5. Maden, B.E.H., Dent, C.L., Farrell, T.E., Garde, J., McCallum, F.S., and Wakeman, J.A. (1987). Biochem. J. 246: 519-527.
6. Peattie, D.A., Douthwaite, S., Garrett, R.A., and Noller, H.F. (1981) Proc. Natl. Acad. Sci. USA 78:7331-7335.
7. Wheeler, W.C., and Honeycutt, R.L. (1988) Mol. Biol. Evol. 5:90-96.
8. Olsnes, S. (1987). Nature 328:474-475.
9. Smith, G.P. (1973). Cold Spring Harbor Symp. Quant. Biol. 38: 507-513.
10. Tartof, K.D. (1973). Cold Spring Harbor Symp. Quant. Biol. 38: 491-500.
11. Levinson, G., and Gutman, G.A. (1987). Mol. Biol. Evol. 4: 203-221.
12. Tautz, D., Trick, M., and Dover, G.A. (1987). Nature 322: 652-656.
13. Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O., and Zimmer, E. (1980). Proc. Natl. Acad. Sci. USA 77: 7323-7327.
14. Coen, E.S., Thoday, J.M., and Dover, G. (1982). Nature 295: 564-568.
15. Liebhaber, S.A., Goosens, M., and Kan, Y.W. (1981). Nature 270: 26-29.
16. Shen, S.-H., Slightom, J.L., and Smithies, O. (1981). Cell 26: 191-203.
17. Clark, C.G. (1987). J. Mol. Evol. 25:343-350.
18. Jain, J., Crampton, J., Gonzalez, I.L., Schmickel, R.D., and Drysdale, J.W. (1985). Biochem. Biophys. Res. Comm. 131: 863-867.
19. Parks, C.L., Jones, T.R., Hyman, R.W., Spector, D.J., Gonzalez, I.L., and Schmickel, R.D. (1986). Virology 154: 381-388.
20. Adams. S., personal communication.