Individual and Evolutionary Variation of Primate Ribosomal DNA Transcription Initiation Regions

Golder N. Wilson, Mechthilde Knoller, and L. Lynne Szura
University of Michigan

Roy D. Schmickel
University of Pennsylvania

A 16-kilobase region surrounding the transcription initiation site for ribosomal DNA and including the entire ribosomal DNA transcription unit has been characterized in man and compared in higher primates. Restriction analysis of ribosomal DNA from the pygmy chimpanzee (Pan paniscus), the common chimpanzee (Pan troglodytes), the gorilla (Gorilla gorilla), the orangutan (Pongo pygmaeus), the gibbon (Hylobates lar), and the rhesus monkey (Macaca mulatta) allows a primate phylogeny to be constructed based on ribosomal DNA structure. Individual variation and methylation are demonstrated in the ribosomal DNA repeats of all primates examined. Restriction analysis with HincII endonuclease suggests cleavage at sites containing methylated CpG and adds the SalI/HincII pair to those enzymes useful for studying DNA modification.

Introduction

Although primate relationships have been reconstructed from morphologic (Andrews and Cronin 1982), immunologic (Sarich and Wilson 1967), protein sequence (Wilson et al. 1977), and karyotypic (Miller 1977; Yunis and Prakash 1982) evidence, an evolutionary scheme that unifies these diverse phylogenetic data is not yet available. Recent studies of repeated DNA families (Gillespie 1977), mitochondrial DNA sequences (Brown et al. 1982), and the β globin gene cluster (Barrie et al. 1981) offer another approach to primate phylogeny, but novel molecular features pose new problems for evolutionary synthesis. The significance of rapid mitochondrial DNA mutation rates (Brown et al. 1979) or the sudden extinction of δ globin gene function (Barrie et al. 1981) will be clarified only by assembling structural data from a wide variety of primate genes.

The tandemly repeated genes for ribosomal RNA (rRNA) are attractive for molecular evolutionary studies because of their universal presence, high copy number, and functional similarity. Within each repeating unit are rapidly evolving...
and highly conserved segments. In *Xenopus*, for example, rDNA nontranscribed spacer (NTS) sequences are widely divergent in sister species (Brown et al. 1972), while certain rRNA gene sequences are homologous to those of *E. coli* (Gourse and Gerbi 1980; Clark and Gerbi 1982). Comparisons between *Xenopus* and human rDNA again show great similarity between rRNA gene sequences but considerable differences between external transcribed spacer (ETS) segments (Wilson 1982). Ribosomal genes, then, should be useful for phylogenetic studies over long or short evolutionary distances and may recveal principles for repetitive DNA evolution that are applicable to diverse sequences and organisms. The detailed understanding of phylogenetic relationship and rDNA variation being assembled in *Drosophila* (Tartof 1979; Coen et al. 1982) may thus be helpful for understanding primate evolution through the medium of rDNA structure. In this report, we characterize the rDNA transcription initiation region of man and use cloned DNA segments from this region to compare its structure in seven primate species.

Human rDNA has been demonstrated by restriction analysis (Arnheim and Southern 1977; Schmickel et al. 1980) and electron microscopy (Wellauer and Dawid 1979) to consist of 44 kilobase (kb) repeating units with four EcoRI fragments labeled A–D (fig. 1). Molecular cloning of the EcoRI B fragment (Wilson et al. 1978), A fragment (Erickson et al. 1981), most of the C fragment (Higuchi et al. 1981; Meisfeld and Arnheim 1982), and a variable region of the D fragment (Erickson and Schmickel, in preparation) has provided probes for more extensive Southern analysis (Wilson et al. 1982) and DNA sequencing (Financsek et al. 1982; Meisfeld and Arnheim 1982). Some variable restriction sites within and among individuals have been defined, including an EcoRI site adjacent to the transcription initiation region (site E, in fig. 1), a HindIII site in the 28S rDNA gene (Arnheim et al. 1980), and length polymorphisms near the EcoRI site E (Wilson et al. 1982) and 3' to the 28S gene (Krystal and Arnheim 1978; Schmickel et al. 1980). The variable EcoRI and HindIII sites as well as the length polymorphism 3' to the 28S gene are distributed throughout the five nucleolar organizer regions in man, suggesting that exchanges between nonhomologous chromosomes may be one mechanism for maintaining rDNA homogeneity (Krystal et al. 1981). Certain of these polymorphisms have also been useful in preliminary comparisons of primate rDNA structure (Arnheim et al. 1980; Nelkin et al. 1980).

**Material and Methods**

**Primate DNA Isolation**

Human DNA was prepared from autopsy spleen or postnatal placenta specimens using the method of Marmur (1963). Nonhuman primate DNA was prepared from spleen, placenta, or blood samples provided by the Yerkes Primate Center. Lymphocytes were purified from primate blood specimens using the Ficoll-Hypaque method (Boyum 1975) prior to lysis and DNA purification (Marmur 1963). All specimens were dialyzed extensively in 10mM Tris-(hydroxyamino)-methane hydrochloride, pH 7.5–0.1mM EDTA and stored over chloroform.

**Cloning and Preparation of Human rDNA**

Plasmid pBR 322 recombinant clones containing the human rDNA segments E, S2, S3, S4, or S5E4 were isolated as described previously (Wilson et al. 1982) and DNA prepared by the method of Bolivar and Backman (1979). Primary clones of the B,B2, BamHI segment of human rDNA were isolated from a human *BamHI*
Fig. 1.—Characterization of the human rDNA transcription initiation region. A single repeating unit of human rDNA is shown in the upper panel with the four major EcoRI (E) fragments A-D, BamHI (B) and SalI (S) sites are also shown. The nontranscribed spacer (thin lines), transcribed spacer (thick lines), and gene regions (boxes) are indicated. The repeating unit is an estimated 44 kb of DNA repeated 100–300 times per haploid genome. The middle panel represents an enlargement of the 1,250-bp E2S2 segment showing relevant restriction sites and the SmaI fragments which have been cloned into bacteriophage M13 mp 9 for dideoxynucleotide sequencing. The DNA sequence of the SmaI B fragment containing the transcription initiation region (TI) as positioned 80 ± 1 bp upstream from a BstE II restriction site is shown in the lower panel. The underlined section is identical with that in the mouse. The DNA sequence supports the data of Meisfeld and Arnheim (1982) over those of Lancette et al. (1982) at positions 34, 53, 126–127, and 192 while supporting Lancette et al. (1982) at positions 5, 135, 175, and 186–188. Data from the latter laboratories agree on the deletion of C at positions 15 and 94, A at position 172, and the insertion of C at positions 84 and 224. The inset demonstrates the 80-bp BstEII-cleaved end-labeled fragment of E2S2 DNA (arrow) protected from S1 nuclease digestion by human 45S rRNA (lane 3). No protected fragment was demonstrated if bacterial tRNA was substituted for human 45S rRNA (lane 2). Lanes 1 and 4 show the end-labeled SmaI fragments A (290 bp), B (234 bp), C (170 bp), D (115 bp), E (60 bp), F (40 bp), G (23 bp), and H (20 bp) as size markers, and lane 5 demonstrates partial cleavage of the SmaI B rDNA fragment by BstEII to yield 130- and 105-bp fragments.

library constructed in Charon 28 (Williams and Blattner 1980), using the procedure of Erickson et al. (1981). Approximately 50,000 recombinants were screened using as probe the plasmid clone E2S2, and the procedure of Benton and Davis (1977) to identify two B,B, recombinants. Bacteriophage DNA was prepared as described previously (Wilson et al. 1982).
Restriction Analysis of Primate DNA

Restriction endonuclease reactions (50μl) containing 2–8 μg of primate DNA and 4–16 units of restriction enzyme (Bethesda Research) were incubated 4 h at 37 C before heating 5 min at 65 C. Electrophoresis was for 16 h at 5 V/cm on 0.8% agarose (Seakem) 15 cm (W) × 20 cm (L) horizontal gels using EcoRI and HindIII restriction fragments of lambda bacteriophage as size markers. Transfer to nitrocellulose (Schleicher and Schuell) was as described by Southern (1975) using 10 × SSC as the transfer medium. Hybridization in Seal-O-Meal bags (Sears Roebuck) was for 6 h at 65 C in 2 × SSC, 0.1% sodium dodecyl sulfate, 500 μg per ml sonicated E. coli DNA (Miles) followed by 15 h at 65 C with the same solution containing 1 μg of nick-translated DNA and 10% dextran sulfate. Nick-translations were as described (Rigby et al. 1977) using [α-32P]dATP (Amersham, 3,000 Ci/mmol) and 1 μg of bacteriophage or plasmid recombinant DNA. Specific activities ranged from 1 to 5 × 107 cpm per μg of plasmid DNA. After hybridization, nitrocellulose membranes were incubated 1 h at 65 C in 2x SSC and 0.1% sodium dodecyl sulfate, then 3 h at 65 C in 3x SSC before autoradiography at room temperature in cassettes containing a fluorescent intensifier screen and XR-5 Royal X-O Mat film (Kodak).

DNA Sequencing of the Human Smal B Fragment

Random Smal subclones in the M13 bacteriophage vector mp 9 (Bethesda Research) were isolated from the E2S2 segment and used as templates for dideoxynucleotide sequencing (Messing et al. 1981). Ligation reactions (50 μl) contained 1 μg of Smal-cleaved mp 9 DNA and 2 μg of Smal-cleaved E2S2 DNA in ligation buffer (Wilson et al. 1978) with 1 unit of T4 DNA ligase (Bethesda Research). After overnight incubation at 4 C, 5-μl aliquots were incubated for 1 h at 4 C with 0.2 ml of shocked cells prepared from E. coli JM 103 (Bethesda Research). Transfectants were plated on yeast-tryptone indicator plates along with control reactions containing no ligase or no target DNA. Of 34 β-galactosidase negative “plaques” obtained, 27 exhibited nitrocellulose membrane hybridization to nick-translated pHrB-ES DNA after isolating single-stranded DNA by the polyethylene glycol method (Messing et al. 1981). Nine clones contained the Smal I fragment (fig. 1) as identified by BsrEII cleavage patterns of the replicative form. Other clones contained the Smal A (five), C (three), E (two), F (two), G (four), or H (one) fragments as identified by nucleotide sequencing, computer overlapping, and comparison to the human rDNA sequences of Financsek et al. (1982). One clone has not been characterized. Dideoxynucleotide sequencing of single-stranded DNA from the M13 Smal B fragment clones (three in each orientation) was performed using a 15-bp primer (P-L Biochemicals) as described (Messing et al. 1981). Sequencing reactions were applied to 40-cm (H) × 34-cm (W) × 0.4 mm 5% acrylamide-urea gels. Electrophoresis was for 3 h (1–120 bp) or 6 h (50–150 bp) at 1,500 V.

S1 Nuclease Protection Experiments

Hela cell suspension cultures in minimal media (GIBCO) with 10% newborn calf serum were harvested at 3–5 × 106 cells per ml. Nuclei and nucleoli were prepared and 45S rRNA was isolated from total cellular RNA by sedimentation on 15%–30% linear sucrose gradients as described (Weinberg et al. 1967).
According to the protocol of Berk and Sharp (1977), E,S, DNA (2–5 µg) was cleaved with *HincII*, *BstEII*, *Xor*, or *SalI* and 5' end-labeled with alkaline phosphatase (Sigma), polynucleotide kinase (Bethesda Research), and [γ-<sup>32</sup>P]-ATP (New England Nuclear). Purified 45S rRNA (1.2 µg) was hybridized with 100 ng (1 × 10^4 cpm) of 5' end-labeled, restricted E,S, DNA. Hybridizations were for 6 h at 60 C in 20 µl of 80% formamide, 0.04 M piperazine-N, N'-bis (2-ethanesulfonic acid), pH 6.4, 1 mM EDTA, 0.4 M NaCl after heating 10 min at 68 C. After diluting to 200 µl with S1 nuclease buffer (30 mM Na acetate, pH 4.6, 0.25 M NaCl, 1 mM ZnSO₄), the hybrids were incubated for 1 h at 37 C with 100 units of S1 nuclease (Bethesda Research). After ethanol precipitation, electrophoresis was on 5% polyacrylamide gels with the end-labeled SmaI fragments A–H of E,S, (fig. 1) as molecular weight standards. The gels were dried under vacuum overnight (Hoefer apparatus); autoradiography was as described above.

Cross Hybridization of Cloned Human rDNA Segments

Since 10 µg of primate DNA should contain approximately 1 ng of rDNA sequences (Schmickel 1973), amounts of recombinant bacteriophage or plasmid DNA calculated to contain 1 and 5 ng of the cloned human rDNA segment were restricted with the appropriate enzymes (i.e., *EcoRI* and *SalI* for the clone E,S,) to separate insert and vector fragments after electrophoresis on 0.8% agarose gels. A panel of human rDNA clones (B₁B₂, E₂S₂, S₂S₃, S₃E₂, and E₁E₄) was included on each gel and hybridized to one of the respective nick-translated clones (five experiments) after Southern transfer as described above. Cross hybridization was striking among the vector fragments but nonexistent between the different rDNA inserts.

Synthesis of a strand-selective probe complementary to the SmaI B fragment (fig. 1) was accomplished using an M13 mp 9 recombinant clone as template that contained the SmaI B fragment in a 5'–3' orientation relative to the primer site (Hu and Messing 1982). Radioactive labeling was accomplished using the standard DNA sequencing reaction (Messing et al. 1981) with 10 µCi of [α-<sup>32</sup>P]dATP (2,000–3,000 Ci/mmol—Amersham) without dideoxynucleotides. After 90 min at 15 C, the reaction was stopped with EDTA. Single-stranded M13 recombinant DNA (0.1 µg) containing the rDNA SmaI fragments A, B, C, E, F, and G in the 5'–3' orientation relative to primer were bound to nitrocellulose in 1 M Na acetate using a dot-blot apparatus (Hybridot-Bethesda Research) and rinsed thoroughly with 2 × SSC. After hybridization with the SmaI B fragment probe under conditions described above, only the homologous SmaI B fragment DNA showed significant hybridization.

Parsimony Analysis of Primate rDNA Restriction Data

Computer-assisted parsimony analysis was kindly performed by Dr. W. Fitch using previously described methodology (Fitch 1977). The primate rDNA restriction maps were first reduced to nine characters (sites) whose states were coded by letters representing present, absent, or alternative restriction sites (see fig. 7). The 935 possible unrooted trees generated for these nine positions were examined in two ways. First, the H₂/S₂, H₃/S₁, and *PvuII a/b/c* alternatives (see fig. 7) were treated as equivalent sites requiring a minimum of 11 character state changes among the primates in the absence of homoplasy (parallel or back-mutations). Two most parsimonious trees were obtained by computer analysis, each requiring
16 changes. A second analysis treated \(H_2/S_2\), \(H_4/S_4\), and \(PvuII\) a/b/c as separate sites requiring a minimum of 13 character state changes. Three most parsimonious trees were found, each involving 25 changes. The single most parsimonious tree common to both analyses is diagrammed in figure 7.

Results

Characterization of the Human rDNA Transcription Initiation Region

Prior to the comparison of primate rDNA structure, the human rDNA transcription initiation site was defined by molecular cloning, S1 nuclease mapping, and DNA sequencing. The 5.7-kb EcoRI B fragment of human rDNA (fig. 1) has been cloned in lambda bacteriophage vectors (Wilson et al. 1978, 1982) and EcoRI-SalI (\(E_S\), \(S_2\), \(S_3\), \(S_4\), \(S_5\), and \(S_6\)) and SalI-EcoRI (\(S_E\)) segments subcloned in the plasmid pBR 322 (Wilson et al. 1982). The Smal fragments A–F within the \(E_S\) segment have been defined by partial restriction analysis (Wilson et al. 1982) and were end-labeled to provide size markers for S1 nuclease mapping experiments. Nuclease S1 cleavage of a hybrid between human 45S rRNA and the \(E_S\) segment restricted with BstEII yielded a protected fragment which localized the 5’-end of the rRNA 80 bp upstream of the BstEII site (see inset, fig. 1). Also protected from S1 digestion by purified 45S rRNA or nucleolar RNA were 700-bp and 350-bp fragments extending from the SalI or XorII sites, respectively, while no protected fragment was obtained when the \(E_S\) segment was end-labeled at the HincII site \(H_x\) (data not shown). These experiments define a putative initiation site for rRNA transcription which agrees with results from other laboratories (Financsek et al. 1982; Meisfeld and Arnheim 1982). Similar experiments in the mouse (Grummt 1981; Miller and Sollner-Webb 1981) locate the rDNA transcription initiation site 5,200 bp from the EcoRI site \(E_3\) in both mouse and man.

Templates for dideoxynucleotide sequencing of the human rDNA transcription initiation region were isolated by random subcloning of Smal fragments from the \(E_S\) segment into the M13 bacteriophage vector mp9 developed by Messing et al. (1981). The DNA sequence for the Smal B fragment is shown in figure 1 and defines a consensus sequence for the transcription initiation region based on results obtained by two independent techniques and three different laboratories (see legend to fig. 1). As pointed out previously (Financsek et al. 1982; Meisfeld and Arnheim 1982), the 15-bp sequence underlined in figure 1 is identical between mouse and man. It is also identical in the rat (Rothblum et al. 1982).

Before comparing primate rDNA transcription initiation regions, it was important to rule out possible cross hybridizability among the various human rDNA probes that would be used for Southern transfer experiments. Included in this analysis were two primary bacteriophage clones containing the human rDNA segment \(B_2B_3\) which were isolated in the BamHI vector Charon 28 (Williams and Blattner 1980). No cross hybridization among \(B_2B_3\), \(E_2S_2\), \(S_2S_3\), \(S_3E_3\), or \(E_4E_4\) human rDNA segments could be demonstrated using amounts and specific activities of cloned rDNA comparable to those in Southern transfer experiments with primate genomic DNA (data not shown). In addition, a probe constructed by biased labeling of the transcription initiation region (Hu and Messing 1982; see Material and Methods) did not hybridize to other Smal clones from the \(E_S\) segment or to the other rDNA segments mentioned above. These experiments suggest that duplication of promoter-like sequences which occurs in the *Xenopus*
rDNA NTS (Sollner-Webb and Reeder 1979) does not occur in the human rDNA 5'-NTS.

Comparative Restriction Analysis of Primate rDNA with *Bam*HI

The basic topology of the transcription initiation region in several primate species is shown by *Bam*HI restriction in figure 2A. The cloned segment E,S₂ is used as a probe to demonstrate 6.8-kb *Bam*HI segments B₁B₂ (human, pygmy chimpanzee, and gorilla), 7.3-kb B₁B₂ segments (chimpanzee and gibbon), and variable B₁B₂ and/or 9.8-kb B₁B₂ segments in the rhesus monkey. These and subsequent restriction patterns represent analysis of 15 human, four Hela cell, seven chimpanzee, five pygmy chimpanzee, five gorilla, three gibbon, three orangutan, and eight rhesus DNA samples prepared as described in Material and Methods. Only the rhesus shows individual variation in the *Bam*HI patterns, with each monkey having different proportions of the B₁B₂ and B₁B₂ fragments. Orangutan rDNA is unique in having a novel *Bam*HI B₁ site which divides the 6.8-kb B₁B₂

---

**Fig. 2.**—Restriction analysis of primate rDNA with *Bam*HI. A. Primate rDNA *Bam*HI restriction patterns visualized by Southern transfer and hybridization to the cloned segment E,S₂. DNA from human (lane 1), pygmy chimpanzee (2), gorilla (3), chimpanzee (4), gibbon (5), rhesus (6, 7), and orangutan (8) was restricted with *Bam*HI as described in Material and Methods. Molecular length standards for all experiments were *Eco*RI and *Hind*III restriction fragments of bacteriophage lambda. Positions of primate *Bam*HI sites relative to human restriction sites are summarized in the lower panel and shown individually in fig. 7. B. Primate *Bam*HI restriction patterns visualized by Southern transfer and hybridization to the cloned DNA segment S,S (lanes 1–5) followed by S,E₃ (lanes 6–10). Human (lanes 1 and 6), gorilla (2 and 7), orangutan (3 and 8), gibbon (4 and 9), and rhesus (5 and 10) DNA restriction patterns are shown in comparison to lambda bacteriophage markers.
segment into 5.6-kb \( B_1B_2 \) and 1.2-kb \( B_1B_2 \) portions. The minor bands visible in lane 8 may represent "orphan" sequences (Childs et al. 1981) and have not been analyzed in detail.

The probes \( S_1S_2 \) and \( S_2S_3 \) are used in figure 2B to demonstrate the 2.2-kb \( B_1B_2 \) and 5.3-kb \( B_1B_2 \) BamHI fragments which are conserved in all primates. This and similar experiments (Arnheim et al. 1980; see below) show that the variation among \( B_1B_2 \), \( B_1B_2 \), and \( B_1B_2 \) segments is due to variation in the NTS and not in rRNA gene regions. As shown in figure 1, Higuchi et al. (1981) have defined a cluster of BamHI sites in the human NTS 5' to the \( B_2 \) site. This BamHI cluster is also evident in gorillas and other primates when partial BamHI restrictions are inadvertently obtained (lane 2, fig. 2B). The correspondence of primate sites \( B_1 \) and \( B_2 \) to these human BamHI sites, along with a lack of variation upon EcoRI restriction of the same region (Arnheim et al. 1980), suggests that the BamHI variation is a restriction rather than length polymorphism.

Comparative Restriction Analysis of Primate rDNA with \( Sal_1 \)

Figure 3A shows a major 1.45-kb \( S_1S_2 \) fragment and minor 14-kb \( S_1S_4 \) and 3.7-kb \( S_1S_3 \) fragments after \( Sal_1 \) restriction of several human and Hela cell DNA samples using the cloned \( E_1S_2 \) segment as hybridization probe. Since the \( S_1 \) and \( S_2 \) sites were present in all of 17 independent EcoRI B fragment clones of human rDNA (Wilson et al. 1982), the \( S_1S_2 \) and \( S_1S_3 \) fragments in figure 3A probably represent methylation of CpG nucleotides at the \( Sal_1 S_1 \) or \( S_2 \) sites. The decreased amounts of the \( S_1S_2 \) and \( S_1S_3 \) \( Sal_1 \) fragments in DNA from Hela cells support this hypothesis, since methylation is more variable in cell culture (Kunnath and Locker 1982). The \( Sal_1 \) site \( S_1 \) shows a similar variation in man and other primates, as demonstrated by 6.8-kb \( B_1B_2 \) and 6.1-kb \( B_1S_2 \) fragments visualized after BamHI/\( Sal_1 \) restriction and hybridization to the cloned segment \( E_1S_2 \) (data not shown). Methylation is again suggested as a cause for this variation since the \( Sal_1 \) site \( S_1 \) is present in the two \( B_1B_2 \) clones described above and in the eight clones isolated by Higuchi et al. (1981).

In figure 3B, the various primate species are compared for their \( Sal_1 \) restriction patterns as demonstrated by hybridization to segment \( E_1S_2 \). The chimpanzee, pygmy chimpanzee, and rhesus monkey lack \( Sal_1 S_2 \) sites, while all primates have \( S_1S_1 \) and \( S_1S_3 \) fragments. Additional \( S_1 \) or \( S_3 \) sites are seen in rhesus and orangutan, generating 3.15-kb \( S_1S_1 \) and 2.7-kb \( S_1S_3 \) fragments (lanes O and R, fig. 3B). A 0.55-kb \( S_1S_3 \) \( Sal_1 \) fragment is visible in orangutan and rhesus while a unique \( S_1S_3 \) fragment is demonstrated in the orangutan. Different primate individuals of the same species give identical \( Sal_1 \) restriction patterns except that the amounts of \( S_1S_1 \), \( S_1S_3 \), or \( S_1S_3 \) fragments vary among orangutan and rhesus individuals. Sequential hybridization of the transfer in figure 3B with the cloned segment \( E_1S_2 \), demonstrated additional 2.25-kb \( S_1S_3 \) fragments in those species having an \( S_2 \) site, while additional hybridization with the segment \( S_1E_4 \) showed a 10.3-kb \( S_1S_4 \) segment in all primates (data not shown). These results indicate a high degree of length and sequence conservation of the rDNA transcription unit within and among primate individuals.

The relationship of \( Sal_1 \) and \( HincII \) restriction endonuclease recognition sequences diagrammed in figure 4 suggested that \( HincII \) restriction of primate rDNA might clarify the results of \( Sal_1 \) restriction. Six restriction fragments are visualized by hybridization to the segment \( E_1S_2 \) after partial restriction with \( HincII \) (lane 2,
Fig. 3.—Restriction analysis of primate rDNA with *SacI* and *HincII*. A. Human rDNA *SacI* restriction patterns demonstrated by hybridization to the probe E,S; Human spleen (lanes 1–3), human placenta (4–6), and Hela cell lines (Schmickel et al. 1980) S3 (7), Hep-2 (8), 222 (9), Detroit 96 (10). Demonstrate the *SacI* rDNA fragments diagrammed in figs. 3 and 7. B. *SacI* restriction patterns of human (H), chimpanzee (C), pygmy chimpanzee (P), gorilla (Go),ibbon (G), orangutan (O), and rhesus (R) rDNA after hybridization to E,S. C. *HincII* restriction of primate rDNA after hybridization with E,S. Lane 2 shows a partial *HincII* restriction of human rDNA with 3.15-kb H,S, 2.7-kb H,S, 1.8-kb H,H, 1.45-kb S,S, 0.9-kb H,S, and 0.55-kb S,H, fragments (see lower right). A complete human digest (lane 1) again demonstrates the 1.8-kb H,H, restriction fragment even after long periods of restriction and the expected H,S, and S,H, fragments. Partial restriction of chimpanzee rDNA (lane 3) yields the partial restriction fragments S,S, and H,H, as well as the S,H, fragment which remains after complete digestion. Complete digestion yields S,H, and S,H, fragments for pygmy chimpanzee (lane 4), S,S, S,S, and S,S, fragments for orangutan (lane 5), S,H, and S,S, (faint) for rhesus (lane 6).

fig. 3C). These fragments are labeled at the left margin and serve as reference markers for the other *HincII* patterns. Complete *HincII* restriction of human rDNA yields the minor fragment H,H, and the major fragments H,S, and S,H, (lane 1, fig. 3C). The human 3.15 H,S, fragment visualized in the partial *HincII* digest is absent from the complete digest (lane 1, fig. 3C), indicating the *SacI* site S, that
Fig. 4.—The relationship of SalI and HincII restriction endonuclease recognition sequences.

Fig. 5.—Hybridization of primate SalI restriction fragments to different human rDNA clones. SalI restrictions of human (H), chimpanzee (C), and rhesus (R) rDNA transferred in parallel and hybridized separately to the human rDNA clones E;F1 (panel A), S;S1 (panel B), and S;E1 (panel C). Note that the S;S1 fragment (R) hybridizes only to E;S1.

appeared absent or methylated in certain rDNA repeats (yielding an S;S1 fragment by SalI digestion) is cleaved by HincII. Given the previous analysis of the human EcoRI B clones, the simplest explanation is that HincII can cleave methylated SalI sites. Absence of the S;S1 fragment in pygmy chimpanzee rDNA (lane 4) and of the S;S2 or S;S3 fragments in orangutan (lane 5) or rhesus (lane 6) rDNA also
supports this conclusion. Of interest is the nonrandom \textit{Hinc}II cleavage of primate rDNA exemplified by the H$_1$H$_1$ fragment in the human rDNA complete digest (lane 1), the absent H$_2$S$_1$ fragment in the chimpanzee rDNA partial digest (lane 3), or the absent H$_2$H$_2$ fragment in the pygmy chimpanzee and rhesus restriction digests. This may reflect differences in rates of restriction or actual differences in restriction site combinations within individual rDNA repeating units (see Discussion). The positions of primate \textit{SalI} and \textit{Hinc}II sites are summarized in figures 3 and 7.

Results of two confirmatory experiments supporting the restriction maps in figures 3 and 7 are presented in figures 5 and 6. In figure 5, triplicate \textit{SalI} digests of human, chimpanzee, and rhesus DNA are hybridized separately to the probes E$_2$S$_2$ (A), S$_2$S$_3$ (B), and S$_3$E$_3$ (C). The rhesus rDNA fragment S$_2$S$_4$ (lane R, fig. 5C) is detected only by the probe E$_2$S$_2$, confirming the position of site S$_4$ as shown in figure 3. The lack of S$_2$S$_4$ fragments in all three primates using S$_2$S$_3$ as probe

![Image]

**Fig. 6.**—Restriction of primate rDNA with \textit{EcoRI} and \textit{SalI}. Primate rDNA samples are designated as in fig. 3B. Hybridization was with the DNA segment E$_2$S$_2$, followed by S$_2$S$_4$. The numbers refer to human rDNA restriction fragments 1-E$_1$E$_1$ (18 kb), 2-E$_1$E$_1$ (5.7 kb), 3-S$_2$E$_1$ (4.55 kb), 4-S$_3$S$_1$ (3.7 kb), 5-E$_1$S$_1$ (3.5 kb), 6-S$_2$S$_3$ (2.25 kb), 7-S$_4$S$_2$ (1.45 kb), and 8-E$_2$S$_2$ (1.25 kb). \textit{EcoRI}-\textit{SalI} restriction patterns of chimpanzee (C), gorilla (Go), gibbon (Gi), orangutan (O), and rhesus (R) are consistent with the maps in fig. 7.
demonstrates nonrandom methylation patterns within individual repeating units, as discussed below.

The relation of the variable human rDNA EcoRI site E₂ (Wilson et al. 1978; Arnheim et al. 1980) to modified or absent Sal₁ sites is examined by EcoRI/Sal₁ restriction in figure 6. In lane 1, various human rDNA segments are demonstrated by sequential hybridization to the cloned segments Eₛₐₛ and Sₛₛₖ. The occurrence of 1.45-kb Sₛₛ and 1.25-kb Eₛₛₖ segments (numbered 7 and 8 in fig. 6), 3.7-kb Sₛₛₖ and 3.5-kb Eₛₛ₆ segments (numbered 4 and 5 in fig. 6), and the 4.55-kb Sₛₑ₆ segment (numbered 3 in fig. 6) suggests random combinations of variable E₂, Sₛ, and Sₛ sites among individual human rDNA repeats, as would be expected if methylation were independent of the E₂ restriction site polymorphism. Similar patterns are shown for other primates in figure 6, and the results support the restriction maps diagrammed in figure 7.

Discussion

Molecular characterization of the human rDNA transcription initiation region (fig. 1) has provided a panel of rDNA segments for the analysis of primate restriction polymorphisms and has suggested that major rearrangement or duplication of promoter sequences has not occurred in the 75 Myr since divergence of man and mouse. Considerable length conservation of ETS and adjacent NTS regions can be demonstrated both within individuals and among seven primate species (figs. 2, 3). Even greater structural similarity of primate rDNA is observed within the rDNA transcriptional unit. Despite the overall picture of primate rDNA conservation, some specific mutations can be documented by comparing the human rDNA restriction maps with those of other primates.
man rDNA sequence with certain primate rDNA restriction sites. The SalI site GTCGAC labeled S, (figs. 3B, 7) in certain primates has become the HincII site H, in man (figs. 1, 3C, 7) with the sequence GTTGAC (Financsek et al. 1982; Meisfeld and Arnheim 1982). Similarly, the SalI site S, in orangutan, gibbon, gorilla, and man (figs. 3B, 7) has become the HincII site H, in rhesus, chimpanzee, and pygmy chimpanzee (figs. 3C, 7). As shown in figure 4, both changes represent transitions. Also, the BamHI site B, in the orangutan (figs. 2, 7) has a homologous sequence GGGTCC in the human rDNA sequence (G. N. Wilson, unpublished; Financsek et al. 1982; Meisfeld and Arnheim 1982) and represents a transition. No homologue for the Sall S, (figs. 3B, 7) is apparent in the human rDNA sequence.

The novel BamHI sites B,, B,, and B, (fig. 2) in nonhuman primates appear to represent restriction polymorphisms since EcoRI restriction shows no evidence of major length variation in primate 5' NTS regions (Arnheim et al. 1980). Also, the B, site seems to be present in most primates, as demonstrated by our own partial BamHI digests and by the cloning experiments of Higuchi et al. (1981). Yet some length heterogeneity in the 5'-NTS is undoubtedly present, since the primate B,B,, B,, and B,B, BamHI fragments shown in figure 2 appear as doublets after long periods of electrophoresis (not shown). Length variation on the order of 15-30 bp multiples surrounding the EcoRI site E, in man has been suggested by partial restriction analysis of E,E, clones (Wilson et al. 1982). Since these restriction and length variations have not been completely characterized, only the B, site is considered in the phylogeny.

A phylogeny was derived from the nine well-characterized primate restriction site differences diagrammed in figure 7. Using computer-assisted analysis of 935 possible unrooted trees (Fitch 1977), two most parsimonious trees involving 16 character state changes were defined counting the H,/,S,, H,/,S,, and PvuII a/b/c alternatives as equivalent sites, while three most parsimonious trees involving 25 character state changes were defined counting the alternatives as separate sites (see Material and Methods). The single most parsimonious tree common to both treatments is shown in figure 7. It agrees with primate phylogenies derived from comparisons of albumin (Sarich and Wilson 1967), mitochondrial DNA sequences (Ferris et al. 1981; Brown et al. 1982), or the β-globin cluster (Zimmer et al. 1980; Barrie et al. 1981). The rDNA phylogeny also supports the chromosomal data of Yunis and Prakash (1982) over those of Miller (1977) in suggesting that the chimpanzee is more closely related to man than is the gorilla.

It will be interesting to correlate the phylogeny in figure 7 with the dramatic changes in number and location of rDNA loci on primate chromosomes demonstrated by in situ hybridization and silver staining (Wilson 1982). Our preliminary results suggest that mechanisms for unequal crossing-over which are postulated to maintain rDNA homogeneity within individuals and species (Smith 1973) may lead to sudden changes in rDNA location and dosage without affecting molecular structure. Further conclusions regarding the tempo and mode of primate rDNA evolution will, of course, require restriction analysis of the entire repeating unit and DNA sequence comparisons of selected regions so that bias can be minimized (Adams and Rothman 1982). Also important, as will now be discussed, is better knowledge of the extent and mode of rDNA variation within primate individuals and species.
A notable feature of the phylogeny in figure 7 which may relate to individual variation is the transition from the SalI site S₁ to the HincII site H₁ that occurs in three separate primate branches. These changes suggest parallel evolution which may be related to variable modification of this site as demonstrated in figure 3. Approximately 10%–20% of human rDNA repeating units from placenta or spleen are not cleaved by SalI at the S₁ site, as compared with 2%–5% of Hela cell repeating units (fig. 3A). In all cases, the SalI S₁ site is cleaved by HincII (fig. 3C), which suggests two possibilities for the SalI variation. Certain human rDNA repeats may have developed the HincII site H₁ which is present in other primates (fig. 7). This is unlikely, however, since 17 cloned E₁E₂ segments from two individuals all had the S₁ site (Wilson et al. 1982). Also, there is variable cleavage by SalI at sites S₁, S₂ (all primates—see Results), S₃, or S₄ (orangutan, rhesus); complete HincII restriction occurs at each of these sites. An explanation more consistent with these results is that the SalI variations represent methylation of CpG which is cleaved by HincII.

Although direct evidence examining the ability of HincII to cleave methylated CpG has not been published to our knowledge, this explanation seems reasonable, since the sequence GTTGAC with a similarly positioned thymine methyl group is recognized by HincII (fig. 7). Assuming that methylation is responsible for the SalI variation, then the results in figures 3 and 5 demonstrate nonrandom methylation within the S₁S₄ region of individual rDNA repeats. Figure 2A demonstrates that site S₁ alone or both S₁ and S₄ may be methylated in individual human rDNA repeats. Methylation of site S₃ alone, however, as would have been indicated by the presence of a S₁S₃ fragment, apparently does not occur (fig. 5). Similar selectivity of methylation within individual rDNA repeats of the chimpanzee and rhesus monkey is demonstrated in figure 5. It should be noted that the apparent absence of SalI site S₁ in chimpanzee or pygmy chimpanzee might represent 100% methylation, although this seems unlikely. These nonrandom patterns of rDNA methylation will be interesting to examine in rapidly proliferative versus slowly growing tissues and in individuals with amplified and highly methylated nucleolar organizer regions (Tantravahi et al. 1981).

Another type of heterogeneity among individual rDNA repeating units, which in at least one case is quite distinct from rDNA methylation (fig. 6), is revealed by restriction analysis of primate rDNA. The EcoRI sites E₁ and E₂ vary among individual humans, pygmy chimpanzees, or chimpanzees; the BamHI sites B₁ and B₂ vary within rhesus individuals; and a HindII site in the 28S rRNA gene varies among all primates (Arnheim et al. 1980). The subset of predicted HincII cleavage fragments shown in figure 3C may also represent nonrandom restriction site combinations, although different cleavage rates at certain sites is also possible. These preliminary data suggest that the rDNA of a primate individual comprises a large number of structurally different repeating units which must be augmented or corrected in germ cells to explain the concerted evolution (Arnheim et al. 1980) demonstrated in figure 7. We propose the term “repertype” to describe alternate repeating unit structures within an individual.

If one considers the variable EcoRI sites E₁ and E₂ (fig. 6), the HindII site in the 28S rRNA gene (Arnheim et al. 1980), and the 4–5 length polymorphisms caused by a variable insertion in the 3’-NTS (Schmickel et al. 1980), then as many as 80 repetypes may be present in the chimpanzee, an organism having an estimated 488 rDNA repeating units (Wilson et al. 1982). Since relatively few detailed
restriction studies of primate rDNA have been performed, it is possible that every rDNA repeating unit has different combinations of restriction sites reflecting primarily variation in spacer regions. Further enumeration of rDNA repetypes in somatic and germ-line tissues will be required before the role of intra- (Smith 1973) or interchromosomal (Arnheim et al. 1980) crossover in maintaining rDNA homogeneity can be quantitatively understood. Restriction analysis under conditions where large DNA fragments can be resolved and the cloning of multiple “identical” rDNA segments from a single tissue source (Erickson et al. 1981; Wilson et al. 1982) are two approaches which can be followed. The multiplicity of rDNA repetypes in different cell types of an individual, their inheritance, and their distribution among populations will be important to define before a thorough understanding of rDNA evolution can be achieved.

Acknowledgments

We thank Dr. Walter Fitch for helpful criticism and for performing the parsimony analysis. Grant support from National Institutes of Health (PO1 HDGM 13506) and the National Foundation–March of Dimes (5-169 to G.N.W.) is also gratefully acknowledged.

LITERATURE CITED


WALTER M. FITCH, reviewing editor

Received July 29, 1983; revision received November 14, 1983.