Biochem. J. (1987) 246, 519–527 (Printed in Great Britain)



# Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes

## Characterization, a detailed map of the human ribosomal transcription unit and diversity among clones

B. Edward H. MADEN,* Carolyn L. DENT, Thomas E. FARRELL, Julie GARDE, Fiona S. McCALLUM† and Jane A. WAKEMAN
Department of Biochemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, U.K.

We have isolated several new clones of human ribosomal DNA. Each clone contains part of the external transcribed spacer, a complete 18 S-rRNA gene, the internal transcribed spacers, a complete 28 S-rRNA gene and a short downstream flanking region. We present a detailed map of the human ribosomal transcription unit with the locations of numerous useful restriction sites. In particular, a unique *Nhe*I site in the 5.8 S-rRNA gene enabled this gene to be mapped with respect to the 18 S-rRNA and 28 S-rRNA genes. The human 45 S-rRNA coding region is approx. 13000 nucleotide residues long, of which the external transcribed spacer comprises approx. 3700 nucleotide residues and the first and second internal transcribed spacers comprise approx. 1070 and 1200 nucleotide residues respectively. A partial survey for sites of variation between clones has revealed a single point of variation among 18 S-rRNA gene sequences (a T/C variation at position 140), several sites of length variation in the regions of the transcribed spacers closely flanking the 18 S-rRNA genes, and some sites of length variation among 28 S-rRNA genes. Most of these sites of variation are associated with simple sequence tracts and are in regions that are known to undergo relatively rapid evolutionary divergence. In particular, the sites of variation among 28 S-rRNA genes occur in G+C-rich tracts whose lengths vary among vertebrates and that can be correlated with extensive hairpin structures previously observed by electron microscopy. Each of the clones so far surveyed in detail differs from the others in one or more respects.

## INTRODUCTION

As part of a study on human ribosomes and their biosynthesis we wanted to isolate clones containing the complete 18 S-rRNA and 28 S-rRNA genes in a single tract of DNA. Such clones should be useful for structural characterization of intact extensive tracts of rDNA and also for the development of functional assays on the maturation of human rRNA. In the present paper we report the isolation and characterization of several such human rDNA clones.

Various clones have been used for sequence analyses on human rDNA. Published sequence data encompass the region of initiation of transcription (Financsek *et al.*, 1982; Miesfield & Arnheim, 1982), the 18 S-rRNA-gene region (Torczynski *et al.*, 1985; McCallum & Maden, 1985; Gonzalez & Schmickel, 1986) and the 28 S-rRNA-gene region (Gonzalez *et al.*, 1985). However, none of these clones contains the complete 18 S-rRNA and 28 S-rRNA genes in a single DNA tract. [The clone isolated by Torczynski *et al.* (1985) contains a complete 18 S-rRNA gene.]

Approaches to isolating clones containing the desired rDNA region were constrained by the fact that several restriction endonucleases that are commonly used for DNA cloning cut internally in the human 18 S-rRNA or 28 S-rRNA genes or the ITS (Wilson, 1982; the present work). However, *Sal*I (Fig. 1) cuts at two sites in the ETS and at a site downstream from the 28 S-rRNA gene, but not in the region in between (Wilson *et al.*, 1978; Wilson, 1982). We have therefore used this enzyme to clone large restriction fragments of human rDNA containing the complete 18 S-rRNA and 28 S-rRNA genes.

Analysis of these clones has led to resolution of a number of important details on the structure of the human ribosomal transcription unit. In the present paper we combine this information with available sequence data to give a detailed overview of the transcription unit, including the sizes of the major functional regions, the locations of many useful restriction sites and some data on sequence variation between clones of rDNA.

## METHODS

A human genomic library was constructed by ligating 3 $\mu$g of *Sal*I-digested DNA from a human placenta to 5 $\mu$g of *Sal*I-digested bacteriophage $\lambda$EMBL3 DNA. About $5 \times 10^5$ recombinants were generated, as indicated by plating samples of the diluted packaged library on *Escherichia coli* NM539, which suppresses plaque formation by non-recombinants (Frischauf *et al.*, 1983).

Abbreviations used: rDNA, ribosomal DNA; ETS, external transcribed spacer; ITS, internal transcribed spacer.

* To whom correspondence should be addressed.
† Present address: Department of Genetics, Univeristy of Glasgow, Glasgow G12 8QQ, U.K.

About $3 \times 10^4$ recombinants were plated on *E. coli* NM539, and duplicate nitrocellulose-filter replicas (Benton & Davis, 1977) were screened with the human rDNA clone pHrA (Fig. 1); 24 plaques yielded positive signals on the duplicate filters. This yield was roughly in accordance with expectation, given that there are a few hundred copies of rDNA in the human genome (Wilson, 1982). Plaque purification was achieved for 16 clones by carrying out two further rounds of plating and screening. *E. coli* NM538 was used during plaque purification and for subsequent procedures since this strain yielded larger plaques than did strain NM539. During plaque purification we screened separately for the presence of the 18 S-rRNA and 28 S-rRNA genes in each clone by using the *Xenopus laevis* rDNA clones pX1r14E and pX1r11R (Maden, 1980). These probes contain only internal regions of the respective genes without any flanking sequences, and were used to eliminate the possibility that the ITS region of pHrA might contain sequences that hybridize to non-ribosomal regions of the human genome.

DNA was obtained from confluent plate lysates after growth on *E. coli* NM538, with the use of underlying agarose and top agarose, rather than agar, to obtain DNA of sufficient purity for restriction. After carrying out several analytical restrictions we transferred rDNA inserts from several clones into plasmid pBR322. Large-scale plasmid preparations, restriction, further subcloning into bacteriophage M13 vectors and sequencing by the dideoxy method were carried out by standard procedures. Some of the plasmid DNA preparations were difficult to restrict even after purification by banding in CsCl/ethidium bromide density gradients followed by phenol extraction. This tendency was more marked for some restriction endonucleases than others. However, sufficiently vigorous digestion yielded the required fragments.

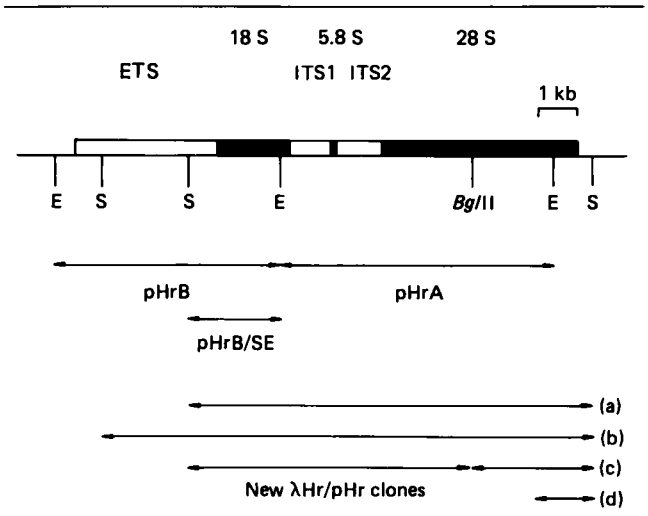



**Fig. 1. Human rDNA clones**

The upper section shows the locations of the *Eco*RI (E) and *Sal*I (S) sites and a unique *Bgl*II site in the human ribosomal transcription unit. The middle section shows the *Eco*RI-digest fragments that were cloned by Wilson and co-workers (Wilson, 1982). Clones pHrB/SE (a subclone of pHrB) and pHrA were used by McCallum & Maden (1985) and also as reference clones during this work. The lower section shows the *Sal*I-digest fragments that were cloned during this work. Most of the clones contain the 10.4 kb fragment: line (*a*). Two clones contain an extended insert: line (*b*). The rDNA insert of one clone (λHr10; see Table 1) was transferred from bacteriophage λ into plasmid pBR322 as two separate fragments divided by the *Bgl*II site: line (*c*). Another clone contains an atypical rDNA insert: line (*d*). Further details on the structures of the new clones are given in the text, Table 1 and Fig. 2.

## RESULTS AND DISCUSSION

### Initial characterization of clones

The λ clones were first characterized by digestion with *Sal*I (Table 1; see also Fig. 2). Most of the clones yielded four bands. These were, in order of size: the λ-clone left arm, a fragment of approx. 10.4 kb that was identified as the rDNA fragment (see below), the λ-clone right arm, and either the 8.4 kb or the 4.8 kb internal λ-clone fragment. Two clones yielded a *Sal*I-digest fragment of 2.25 kb in addition to the 10.4 kb fragment. One clone (λHr24) lacked the 10.4 kb fragment but yielded a 1.5 kb fragment and the 8.4 kb and 4.8 kb vector fragments. The presence of internal vector fragments in the recombinants could have resulted from one or more of the following: incomplete digestion of the vector; re-insertion of a vector fragment (Fig. 2); the need to exceed a critical minimum length for viability of the recombinants.

Positive identification of the rDNA fragments came partly from restriction tests on the λ clones and partly after subcloning the *Sal*I-digest fragments from several of the λ clones into plasmid pBR322. The identifications are summarized here for convenience. The 10.4 kb *Sal*I-digest fragment was shown to contain numerous restriction sites that are diagnostic of the rDNA region encompassing the 18 S-rRNA and 28 S-rRNA genes (Wilson, 1982; McCallum & Maden, 1985). The 2.25 kb band, present in two of the clones, is derived from the ETS (Fig. 1), as also confirmed by the presence of several diagnostic restriction sites (Wilson, 1982; Miesfield & Arnheim, 1982). This fragment is in the 'correct' orientation with respect to the 10.4 kb fragment in the two λ clones. Its presence presumably resulted from incomplete digestion of the respective tracts of rDNA with *Sal*I before cloning. The rDNA insert in clone λHr24 is a truncated but otherwise normal fragment encompassing the 3′ end of the 28 S-rRNA gene, attached at the upstream end to a small amount of unidentified non-ribosomal material. Presumably this rDNA fragment had been sheared before cloning.

To explore the possibility of variation between clones, particularly in the transcribed spacers, the λ clones were digested with a combination of *Sal*I, *Xba*I and *Kpn*I (or the isoschizomer *Asp*718). The digests yielded several fragments, of which three relatively rapidly migrating ones were of interest (see Fig. 3): an *Xba*I/*Xba*I-digest fragment of 1.5 kb from within the 18 S-rRNA gene, a *Sal*I/*Xba*I-digest fragment of approx. 0.9 kb extending from the *Sal*I cloning site in the ETS into the 18 S-rRNA gene, and an *Xba*I/*Kpn*I-digest fragment of approx. 0.65 kb extending from the 3′ end of the 18 S-rRNA gene into ITS1. The 1.5 kb 18 S-rRNA gene fragment was indistinguishable in size between clones. However, both the *Sal*I/*Xba*I-digest fragment and the *Xba*I/*Kpn*I-digest

**Table 1. Human rDNA clones**

Column 1 lists the names of the clones including two reference plasmid clones. Clones in bacteriophage λ are prefixed λ; clones in plasmid pBR322 are prefixed p. The letters Hr designate human ribosomal. Column 2 lists the approximate sizes of the rDNA fragments in the clones (see also Fig. 1). For explanation of column 3 see the text and Fig. 2. In columns 4 and 5 the approximate sizes of the indicated restriction fragments are given (see also the text), – signifies that the relevant fragment is not present in the clone, ? signifies that the fragment was not clearly visualized, and n.d. signifies that the restriction test was not done on the clone. Column 6 lists the plasmid clones that were obtained by subcloning the rDNA fragments from the λ clones. Notes on the plasmid clones are as follows. (*a*) Attempts to subclone the 10.4 kb *Sal*I-digest fragment from clone λHr10 into plasmid pBR322 after first purifying the fragment from an agarose gel were unsuccessful. The purified *Sal*I-digest fragment band was therefore cut with *Bgl*II (Fig. 1) and the two resulting fragments were separately subcloned without repurification. Clone pHr10L contains the large left-hand region of the 10.4kb *Sal*I-digest fragment from clone λHr10, as indicated in line (c) of Fig. 1, cloned between the *Sal*I and *Bam*HI sites of plasmid pBR322. Clone pHr10R contains the right-hand region of the clone λHr10 rDNA insert [line (c) of Fig. 1] also cloned between the *Bam*HI and *sal*I sites of plasmid pBR322. (*b*) *Sal*I-digest fragments from this and further λ clones were introduced into plasmid pBR322 by 'shotgun' subcloning from each λ clone followed by identification of the desired recombinants in mini-preparations (Birnboim & Doly, 1979). Clone pHr12 contains the 10.4 kb *Sal*I-digest fragment from clone λHr12 cloned into the *Sal*I site of plasmid pBR322. Clone pHr12E contains the 2.25 kb *Sal*I-digest ETS fragment from clone λHr12 in plasmid pBR322. (*c*) Each of these clones contains the 10.4 kb *Sal*I-digest fragment from the respective λ clone in plasmid pBR322. In some experiments more than one subclone of the required rDNA region was obtained. Orientation checks on plasmid clones were carried out where necessary. When plasmid pBR322 is drawn with the *Eco*RI site at '12 o'clock' and the tetracyline-resistance gene to the right, the RNA-like strand of the rDNA insert is in the following orientation in the respective clones: pHr12, pHr13, pHr15 and pHr19, clockwise in the clones used (one anticlockwise isolate of pHr13 was obtained); pHr12E, anticlockwise; pHrA, clockwise. (The orientations in clones pHr10L, pHr10R and pHrB/SE are fixed by the restriction sites used for cloning: anticlockwise, clockwise and anticlockwise respectively.)

| Clone | rDNA insert(s) (kb) | λ internal *Sal*I-digest fragment | ETS *Sal*I/*Xba*I-digest fragment | ITS1 *Xba*I/*Kpn*I-digest fragment | Transferred to pBR322 | Notes |
|---|---|---|---|---|---|---|
| λHr1 | 10.4 | 4.8 | 0.90 | ? | – | |
| λHr3 | 10.4 | 4.8 | 0.90 | ? | – | |
| λHr4 | 10.4 | 8.4 | n.d. | n.d. | – | |
| λHr5 | 10.4 | 8.4 | 0.90 | 0.65 | – | |
| λHr6 | 10.4 | 4.8 | 0.90 | 0.63 | – | |
| λHr10 | 10.4 | 4.8 | 0.90 | 0.65 | pHr10L, pHr10R | (*a*) |
| λHr11 | 10.4, 2.25 | 4.8 | 0.89 | 0.65 | – | |
| λHr12 | 10.4, 2.25 | 4.8 | 0.90 | 0.65 | pHr12, pHr12E | (*b*) |
| λHr13 | 10.4 | 4.8 | 0.90 | 0.63 | pHr13 | (*c*) |
| λHr14 | 10.4 | 8.4 | 0.90 | 0.65 | – | |
| λHr15 | 10.4 | 4.8 | 0.90 | 0.63 | pHr15 | (*c*) |
| λHr19 | 10.4 | 8.4 | 0.88 | 0.65 | pHr19 | (*c*) |
| λHr20 | 10.4 | 4.8 | 0.90 | ? | – | |
| λHr21 | 10.4 | 8.4 | 0.90 | ? | – | |
| λHr22 | 10.4 | 4.8 | n.d. | n.d. | – | |
| λHr24 | 1.5 | 4.8, 8.4 | – | – | pHr24 | |
| pHrB/SE | 2.38 | – | 0.90 | – | | |
| pHrA | 7.1 | – | – | 0.65 | | |

fragment showed slight but detectable size variation between clones (Table 1). Length variations between the ITS *Xba*I/*Kpn*I-digest fragments were confirmed for some of the clones by subcloning into bacteriophage M13 and partial sequence analysis (see below).

rDNA inserts from several of the λ clones were then transferred to plasmid pBR322 (Table 1) to facilitate further study.

### rDNA region encloding 45 S rRNA

It is useful to give an overview of the 45 S-rRNA coding region based on data obtained in this analysis and recent sequence data. This information is summarized in Fig. 3. The restriction data refine and add to those of Wilson (1982).

The length of the ETS is the sum of the lengths of three contiguous regions whose junctions are defined by the two *Sal*I sites (Figs. 1 and 3). The size of the ETS *Sal*I-digest fragment was determined as 2250 bp by calibrating against appropriate size markers. The size was indistinguishable between clones λHr11 and λHr12. The overall length of the ETS is approx. 3700 nucleotide residues.

The large-scale structure of the ITS was established as follows. Inspection of 5.8 S-rRNA sequences from several vertebrate species (compiled in Erdmann & Wolters, 1986) indicated the presence of a conserved *Nhe*I site. This site was experimentally demonstrated in our human rDNA clones and is unique in the 10.4 kb *Sal*I-digest fragment (Fig. 3). The 5.8 S-rRNA gene was mapped in relation to the 18 S-rRNA and 28 S-rRNA genes by using this enzyme in combination with other enzymes in appropriate double-digestion experiments. The results defined the lengths of ITS1 and ITS2 as approx. 1070 and 1200 nucleotide residues respectively. Confirmatory data were obtained by utilizing an *Asu*II

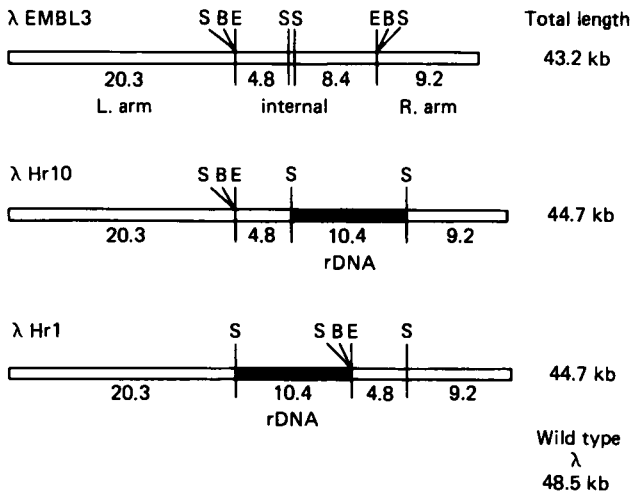



**Fig. 2. Structure of bacteriophage λEMBL3 and two of the rDNA clones**

The upper section shows bacteriophage λEMBL3 with the locations of the sites for *Sal*I (S), *Bam*HI (B) and *Eco*RI (E). The internal region is bounded at each end by an *Sal*I/*Bam*HI/*Eco*RI-site polylinker. The lengths of the *Sal*I-digest fragments are given in kb. The distance between the innermost *Sal*I sites is 0.5 kb. Further details on restriction sites in this vector are given by Murray (1983). The middle and lower sections show the structures of clones λHr10 and λHr1 respectively. Both clones contain an rDNA *Sal*I-digest insert of 10.4 kb (blackened) and the 4.8 kb internal *Sal*I-digest vector fragment. However, in clone λHr1 the internal *Sal*I-digest vector fragment has been re-inserted during cloning as shown. Other clones contain the fragments listed in Table 1. All of the clones are intermediate in length between the vector and wild-type bacteriophage λ. Restriction sites within rDNA are shown in Fig. 3.

site (Fig. 3), which is also evident in the published sequences (Erdmann & Wolters, 1986).

The overall length of 45 S rRNA from the transcription start site to the 3′ end of the 28 S-rRNA sequence, on the basis of the combination of sequencing and restriction data, is close to 13000 nucleotide residues. This is somewhat shorter than the mean length originally estimated from electron microscopy [4.7 MDa, approx. 14000 bases, in Wellauer & Dawid (1973)]. Because restriction mapping of fragments containing transcribed spacer regions was carried out with reference to size markers of suitable lengths, errors in sizing should not exceed ±2% for any given fragment, or ±120 nucleotide residues in total for the transcribed spacers, on the basis of the most frequently encountered length variants of these regions (Table 1). As is described below, 45 S rRNA harbours length microheterogeneity due to the presence of several regions where individual rDNA units differ slightly in length.

### Survey for variants

Previous work on *Xenopus* rDNA established that the *X. laevis* 18 S rRNA sequence is highly homogeneous (Maden *et al.*, 1982), whereas the transcribed spacers harbour multiple heterogeneities (Stewart *et al.*, 1983). Moreover, the *X. laevis* and *Xenopus borealis* 18 S rRNA sequences differ by only two nucleotides (McCallum & Maden, 1985) whereas the transcribed spacers show extreme divergence (Furlong & Maden, 1983; Furlong *et al.*, 1983). It was of interest to search for corresponding evidence of sequence stability or flux in regions of the human ribosomal transcription unit.

### Minimal variation among 18 S-rRNA genes

Available human 18 S rRNA sequence data now cover three published sequences that were independently determined, and in addition the complete 18 S-rRNA sequence in clone pHr10L, determined during this work, and limited sequence data obtained from other clones during this work (Table 2). The sequences are in very good agreement, but there is one definite site of variation between clones, at position 140. Clone pHrB/SE contains C at this position (McCallum & Maden, 1985; confirmed during the present work) whereas all other human clones that have so far been sequenced contain T at this point, as do other mammalian 18 S-rRNA genes (Torczynski *et al.*, 1983; Chan *et al.*, 1984; Raynal *et al.*, 1984). This is in a small region that shows sequence variation across broader phylogenetic distances (region V2 in the nomenclature of Gonzalez & Schmickel, 1986). Further studies will be required to determine the relative frequencies of the two variants among the human 18 S-rRNA gene population.

Elsewhere we have found no differences among human 18 S-rRNA clones, and we are confident of our assignments at the other points of apparent difference between published sequences. We suggest that the sequence reported by McCallum & Maden (1985) represents the predominant human 18 S-rRNA sequence, except that T may occur more commonly than C at position 140. The sequence in McCallum & Maden (1985) also shows the locations of the RNA methyl groups (detailed in Maden, 1986).

### Variable sequences flanking the 18 S-rRNA genes

In contrast with the stability of the 18 S-rRNA gene sequence, there are multiple variants in the regions of the transcribed spacers flanking the 18 S-rRNA gene. These differences are displayed in the upper parts of Figs. 4 and 5 and summarized in the lower parts of the Figures. There are uncertainties at some points in the respective sequences owing to compression in the sequencing gels but most of the differences were characterized clearly.

In the ETS the first site of variation upstream from the 18 S-rRNA gene is a C/A substitution at position −17, which gives rise to a variable *Taq*I site. (Because this description relates to the region near the 18 S-rRNA gene, nucleotides are numbered negatively with respect to the gene.) Further upstream, starting at position −82 in the autoradiograph, a semi-regular banding pattern reveals a simple sequence region that is a focus of length variation between clones. With reference to the RNA-like strand the most regular feature of the region is a C-C-G-T motif starting at position −115 in clones pHrB/SE and pHr15 and repeated five times in these two clones. There are only four repeats in clone pHr 10 (as shown in parallel sequencing runs on all four clones). Clone pHr19 lacks a more extensive tract of about 24 nucleotide residues with reference to clones pHrB and pHr15, and 20 nucleotide residues with reference to clone pHr10. The deletion in clone pHr19 can be aligned to start immediately after a recurrence of the C-C-G-T element

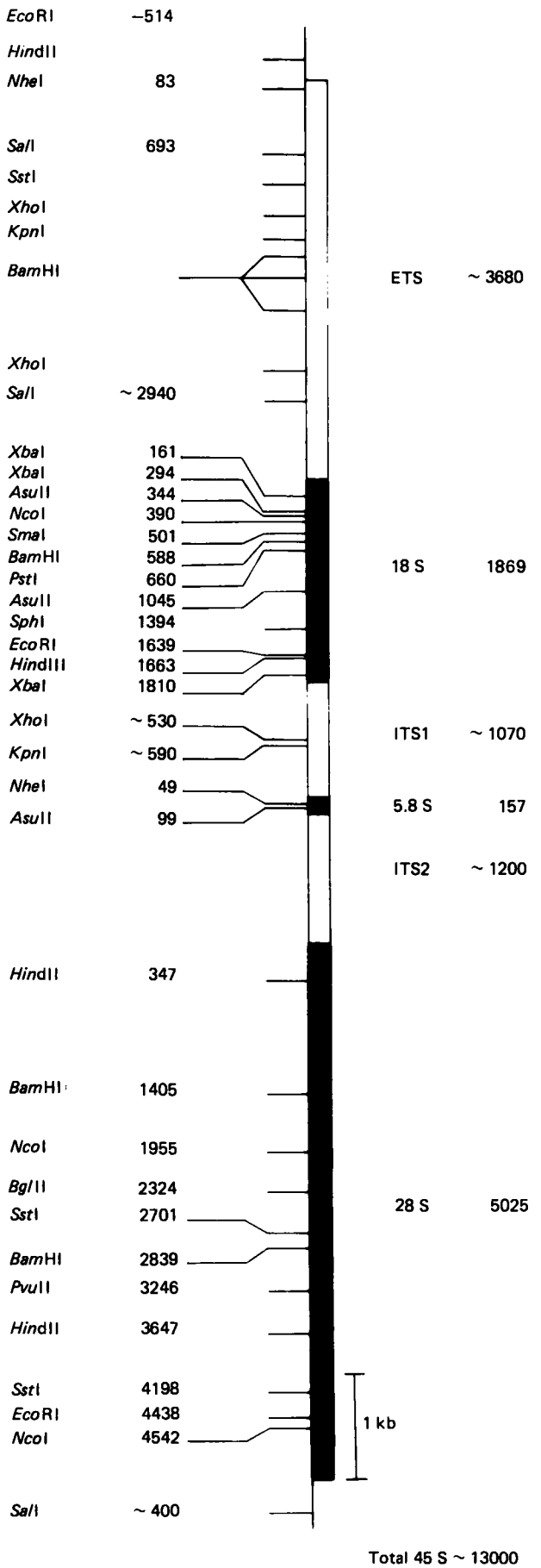



**Fig. 3. Map of the functional regions of the human ribosomal transcription unit with the locations of key restriction sites**

The overall length of each region is from the following data. The region from the transcription start site to the first *Sal*I site was sequenced by Financsek *et al.* (1982). Part of this region was also sequenced by Miesfield & Arnheim (1982), with minor differences from Financsek *et al.* (1982). The rest of the ETS was sized by restriction of the ETS *Sal*I-digest fragment and the *Sal*I/*Xba*I-digest fragment, which extends 160 bp into the 18 S-rRNA gene. A number of restriction sites were used to map the unique *Nhe*I site in the 5.8 S-rRNA gene in relation to the 18 S-rRNA and 28 S-rRNA genes, including the *Xba*I site near the 3′ end of the 18 S-rRNA gene and the *Hin*dII site near the 5′ end of the 28 S-rRNA gene. The lengths of the 18 S-rRNA and 28 S-rRNA genes are from sequence data (McCallum & Maden, 1985; Gonzalez *et al.*, 1985). The indicated restriction sites are the only sites for the respective enzymes in this tract of DNA except for *Sma*I: in addition to the single *Sma*I site in the 18 S-rRNA gene (shown) there are multiple sites for this enzyme in the 28 S-rRNA gene and transcribed spacers (not shown). All of the sites in the Figure have been demonstrated experimentally during the present work except for those upstream to the first *Sal*I site in the ETS. The exact or approximate positions are numbered by using separate numbering schemes for each functional region, e.g. there is a *Kpn*I site at approximately position 590 in ITS1. Co-ordinates in the transcribed spacers and the 28 S-rRNA gene differ slightly between clones owing to regions of length variation; see the text. The locations of restriction sites in between the two *Sal*I sites in the ETS have not been assigned precise co-ordinates but are approximately to scale.

further upstream at position −135 in clone pHrB/SE. This suggests that the deletion in clone pHr19 has resulted from an unequal recombination event. The deletion accounts for the difference in length of approx. 20 nucleotide residues between the *Sal*I/*Xba*I-digest fragment of clone Hr19 and the other clones (Table 1). When the variable *Taq*I site and the number of C-C-G-T repeats are taken into account, all four clones differ from each other in the 140 nucleotide residues immediately upstream from the 18S-rRNA gene.

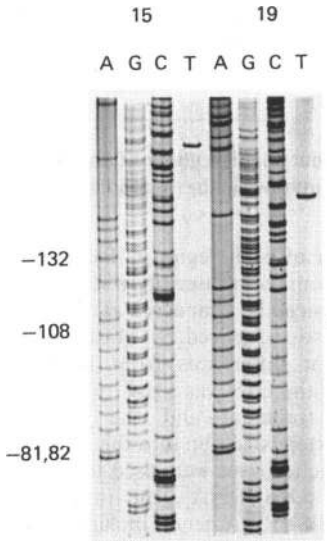
The sequence of this region from a further clone was reported by Gonzalez & Schmickel (1986). Their sequence resembles most closely that of clone pHr15, with C at position 17 and five repeats of the C-C-G-T motif.

The region of ITS1 immediately downstream from the 18 S-rRNA gene is displayed in simplified form in parallel T tracks of clones (Fig. 5). Clones Hr13 and Hr15, which yielded short variants of the *Xba*I/*Kpn*I-digest fragment in the preliminary tests on the λ clones (Table 1), differ in several respects from the other clones. First, they contain a T residue at position approximately 15 in ITS1 that is lacking in the other clones. These two clones are then conspicuously shorter between T-50 and T-98 (numbered according to clone pHrA) than are the other clones, and there are multiple differences between the two classes of clone in this region. Starting just beyond nucleotide 100, clone pHr10 differs from the other clones in a simple sequence region comprising a G-T-T-C, G-C-T-C motif: clone pHr10 lacks one G-T-T-C repeat.

**Table 2. Summary of 18 S-rRNA sequence data**

Notes to Table 2 are as follows. (*a*) This sequence was determined from clones pHrB/SE and pHrA obtained by Wilson (1982), and from a further M13 clone that enabled sequencing to be carried out through the *Eco*RI site as described previously (McCallum & Maden, 1985). (*b*) This sequence was obtained from a bacteriophage λ clone that contained a complete 18 S-rRNA gene, a long upstream region and a downstream extension into ITS1. (*c*) This sequence was obtained from two clones that are homologous to clones pHrB/SE and pHrA. The B clone obtained by Gonzalez & Schmickel (1986) differs slightly in the ETS from clone pHrB/SE, and their clone pHr/A4 differs slightly from our clone pHrA in ITS1, as described in the text. (*d*) The complete 18 S-rRNA gene in clone pHr10L was sequenced. There was no evidence in this clone or in clone pHrB/SE of the extra C following position 967 reported by Gonzalez & Schmickel (1986) (and here designated 967a). Otherwise the 18 S-rRNA sequence in clone pHr10L is identical with that reported by Gonzalez & Schmickel (1986), and to that reported by McCallum & Maden (1985) except for the presence of T instead of C at position 140. Partial 18 S-rRNA sequence data were also obtained from the following clones: pHr13, nucleotide residues 161–590 and 1570–1815; pHr15, nucleotide residues 68–161, 294–480 and 1700–1815; pHr19, nucleotide residues 45–230. The sequences were in agreement with clone pHr10L, with T at position 140 (clones pHr15 and pHr19).

| | Position | | | | | |
|---|---|---|---|---|---|---|
| Source | 140 | 276–277 | 967a | 1772–1774 | Length (residues) | Notes |
| McCallum & Maden (1985) | C | GC | – | CCC | 1869 | (*a*) |
| Torczynski *et al.* (1985) | T | CG | – | CC– | 1868 | (*b*) |
| Gonzalez & Schmickel (1986) | T | GC | Extra C | CCC | 1870 | (*c*) |
| Present work | T | GC | – | CCC | 1869 | (*d*) |
| Suggested predominant sequence | T(C?) | GC | – | CCC | 1869 | |



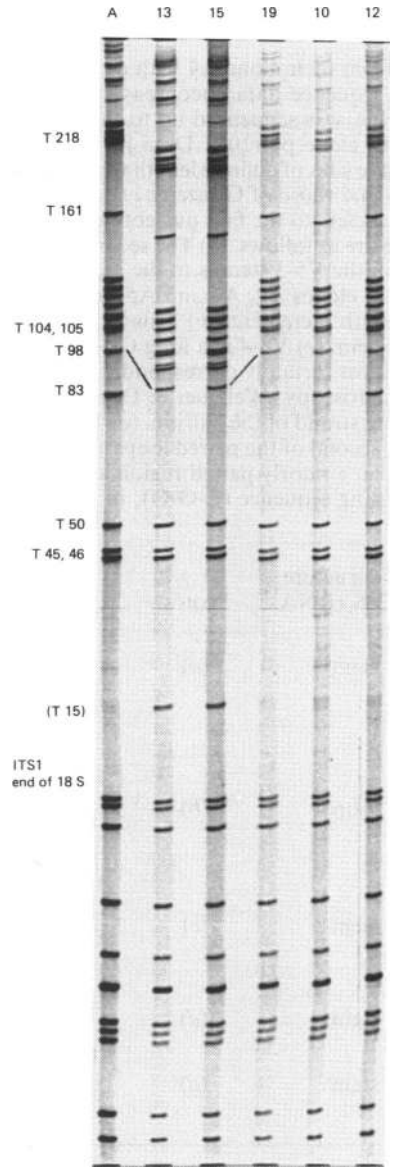

```
pHrB/SE, 10, 15          CCCGCTCGCT CCCTCCCGTC CGGCCCGTCC −121
pHr 19                   CCCGCTCGCT CCCTCCCGT* **********

pHrB/SE, 15    GCGGCCCGTC CGTCCGTCCG TCCGTCGTCC TCCTCGCTTG −81
pHr 10         GCGGCCCGT* ***CCGTCCG TCCGTCGTCC TCCTCGCTTG
pHr 19         ********** ***CCGTCCG TCCGTCGTCC TCCTCGCTTG

All clones     CGGGGCGCCG GGCCCGTCCT CGCGAGGCCC CCCGGCCGGC −41

pHrB/SE, 19    CGTCCGGCCG CGTCGGGGCC TCGACGCGCT CTACCTTACC −1
pHr 10, 15     CGTCCGGCCG CGTCGGGGCC TCGCCGCGCT CTACCTTACC
```

**Fig. 4. Sequence variants in the ETS flanking the 18 S gene region**

The upper section shows detail of sequencing gel in the ETS of clones pHr15 and pHr19. (Sequencing runs on clones pHrB/SE and pHr10 were carried out in parallel.) The templates were rDNA fragments from an *Xba*I site near the 5′ end of the 18 S-rRNA gene (Fig. 3) to an *Xma*I site at position approximately −400 in the ETS, cloned into bacteriophage M13 to allow sequencing leftwards into the ETS. A primer complementary to nucleotides 31–45 of the 18 S rRNA sequence was used. The nucleotide numbers refer to the provisional sequences of clones pHr15 and pHrB/SE (beneath). The deletion in clone pHr19 is from position −131 to position −108 inclusive with reference to clone pHr15. Note also the unusual overall base composition, with only a single T residue (A in the RNA-like strand) in the entire region shown. The lower section shows aligned sequences from four clones in the region of the ETS flanking the 18 S-rRNA gene. The sequences are for the RNA-like strand. A few nucleotide residues in extensive G+C-rich tracts are uncertain owing to secondary structure (indicated by dots). Asterisks indicate deletions.

Further still into ITS1 there is a long simple sequence polypurine tract that was found to differ to some extent among all six clones examined. A sequence up to and including this tract was determined by Gonzalez & Schmickel (1986). Their sequence resembles that of clone pHrA with minor differences, and with an even longer polypurine tract (20 A-G repeats) than in any of the clones examined here.

The overall length difference between the *Xba*I/*Kpn*I-digest fragment of clones pHr13 and pHr15 and that of the other clones, estimated by restriction to be approx. 20 nucleotide residues, is accounted for by the deletions near the 18S-rRNA gene and in the polypurine tract. The rest of ITS1 was screened by T tracking and partial sequencing from the *Kpn*I site (Fig. 3) and leftwards from the *Nhe*I site in the 5.8 S-rRNA gene. The results showed a few base substitutions, and insertions and deletions of single nucleotides. Interestingly, however, most of the length variation between clones in ITS1 is confined to the first 210 nucleotide residues. It is intriguing that the




```
pHrA, 10, 19      ACGGAGCCCG GACGGAGGCC     CGCGGCGGCG CCGCGCGCGC   40
pHr 13, 15        ACGGAGCCCG GAGGTAGGCC     CGCGGCGGCG CCGCGCGCGC
pHrA, 10, 12, 19  GCGCTTCCCT CCGCACACCC     ACCCCCCCAC CGCGACGGCG   80
pHr 13, 15        GCGCTTCCCT CCGCA*****     ***CCCCCGC CGCGACG[
pHrA, 12, 19      CGTGCGCGGC GGGGCCGTGC     CCGTTCGTTC GCTCGCTCGT  120
pHr 10            CGTGCGCGGC GGGGCCGTGC     CCGTTC****  GCTCGCTCGT
pHr, 13, 15       ****  GC  tract  T ]      CCGTTCGTTC GCTCGCTCGT
All clones        TCGTTCGCCG CCCGGCCCCG     CCGGCGCGAG AGCCGAGAAC  160
pHrA              TCGGGAGGGA GAC(G)5 (AG)18      AAAG AAGGGCGTGT  220
pHr10             TCGGGAGGGA GAC(G)3 (AG)15    . AAAG AAGGGCGTGT
pHr12             TCGGGAGGGA GAC(G)5 (AG)15      AAAG AAGGGCGTGT
pHr19             TCGGGAGGGA GAC(G)5 (AG)12  AAAGAAAG AAGGGCGTGT
pHr13             TCGGGAGGGA GAC(G)10 (AG)6      AAAG AAGGGCGTGT
pHr15             TCGGGAGGGA GAC(G)9 (AG)6       AAAG AAGGGCGTGT
```

**Fig. 5. Sequence variants in ITS1 flanking the 18 S gene region**

The upper section shows T tracks of human rDNA clones extending from the 3′ end of the 18 S-rRNA gene into ITS1. The templates were rDNA fragments from the *Xba*I site at position 1810 in the 18 S-rRNA gene to the *Kpn*I site at approximately position 590 in ITS1, cloned into bacteriophage M13 to allow sequencing rightwards into ITS1. Lines drawn between T residues indicate sequence alignments following tracts of length variation between clones. The lower section shows the provisional sequences from the start of ITS1 to position 220. The nucleotide numbering is approximate but is given for clone pHrA in both parts of the Figure to facilitate description. Dots signify uncertainties as to either the order or the precise number of nucleotides. Asterisks signify deletions. The region in square brackets in clones pHr13 and pHr15, corresponding to positions 78–100 in clone pHrA, consists entirely of G and C residues that were irregularly spaced. This tract has not yielded a reliable sequence but appears to be approximately four nucleotide residues shorter in clones pHr13 and pHr15 than in the other clones. This interpretation is supported by the fact that the T tracks of clones pHr13 and pHr15 are approximately 12 nucleotide residues out of register with the other clones beyond position 100, eight of these nucleotide residues being accounted for between positions 55 and 63. Since clones pHr13 and pHr15 also lack the T residue corresponding to T-83 in clone pHrA, a deletion of four nucleotide residues is tentatively inferred at this site. The polypurine tracts starting at position 174 are designated in abbreviated form with realignment at the distal ends; for example the tract in clone pHrA, $(G)_5(A\text{-}G)_{14}A\text{-}A\text{-}A\text{-}G$, is the same length as that in clone pHr19, $(G)_5(A\text{-}G)_{12}A\text{-}A\text{-}A\text{-}G\text{-}A\text{-}A\text{-}A\text{-}G$, but 12 bases longer than that in clone pHr15, $(G)_9(A\text{-}G)_6A\text{-}A\text{-}A\text{-}G$.

regions of the transcribed spacers that closely flank the 18 S-rRNA gene appear to be particularly variable.

### 28 S rRNA sequence

28 S rRNA, unlike 18 S rRNA, differs considerably in length between different groups of organisms (Loening, 1968). The length differences reside in several discrete regions, variously known as 'divergent domains' or 'D domains' (Hassouna *et al.*, 1984), 'expansion segments' (Clark *et al.*, 1984) and 'variable regions' (Chan *et al.*, 1983). The nucleotide sequences of these regions differ substantially between quite closely related species such as mouse and rat (Hassouna *et al.*, 1984; Chan *et al.*, 1983).

Gonzalez *et al.* (1985) reported the human 28 S-rRNA gene sequence and found variation between clones in a region between nucleotide residues 2120 and 2200. We have also found variation among clones in this tract, and also at three other sites in an analysis that extends from this region to the 3′ end of the 28 S-rRNA gene (Table 3). The variants are characterized by expansion or contraction of simple sequences in phylogenetically variable regions of the gene.

Much of the region between nucleotides 2880 and 3570 in the 28 S rRNA sequence can be arranged as two long G+C-rich hairpins (Gorski *et al.*, 1987). These almost certainly correspond to the characteristic paired loops seen in electron micrographs just to the right of the centre of 28 S rRNA (Wellauer & David, 1973; Wellauer *et al.*, 1974; Schibler *et al.*, 1975). Three of the tracts where variation occurs between clones are in the arms of these hairpins, as detailed in the legend to Table 3. Expansion or contraction of simple sequences in these regions would afford a mechanism for generating the observed inter-species length variation between these loops (Wellauer *et al.*, 1974; Schibler *et al.*, 1975).

**Table 3. Sites of length variation detected among 28 S-rRNA genes**

The 28 S-rRNA sequence reported by Gonzalez *et al.* (1985) was obtained mainly from their clone A4 with data on all clones A1–A6 in the region between the *Bam*H1 sites at positions 1405 and 2805. Our sequence data encompass the part of the 28 S-rRNA gene region from position 2120 to the 3′ end (position 5025). Clone pHrA was sequenced up to the *Eco*RI site at position 4438; the region through the *Eco*RI site to the 3′ end was sequenced with clone pHr10R. Data from other clones encompass nucleotides 2120 to 2324 and 2839 to approx. 3070. The table summarizes the sites of definite length variation detected in this search, but excludes the small number of minor differences between our data and those of Gonzalez *et al.* (1985), which may possibly reflect differences in the interpretation of sequencing gels. The position refers to the first nucleotide in a repetitive sequence or other clearly defined tract where length variation occurs. Notes to Table 3 are as follows. (*a*) The secondary-structure regions are numbered in accordance with Gorski *et al.* (1987). The variable region in the V5-3 stem is in the ascending strand, i.e. the strand whose 5′ to 3′ direction in 28 S-rRNA approaches the hairpin loop. In clones A2, A3 and A5 the extra G-G-C is due to a G–C substitution with respect to A4 and hence does not represent a length increment. (*b*) This variable string of C residues starts in the loop of V5-3 and extends into the descending strand of the stem. (*c*) V8-4 is a long G+C-rich hairpin comprising some 240 nucleotide residues, extensively but imperfectly paired, and almost certainly corresponding to the first of two long loops slightly to the right of the centre of 28 S rRNA seen by electron microscopy (Wellauer & Dawid, 1973). The variable C-T tract (C-U in RNA) occurs at about 80% of the distance up the ascending strand of the hairpin. (*d*) V8-9 is another long G+C-rich imperfectly paired hairpin, and it almost certainly corresponds to the second of the paired loops seen by electron microscopy (Wellauer & Dawid, 1973). The variable sites occur opposite each other in a poorly paired region about one-third of the way up the hairpin from its base. The deleted G-C-A is at the end of a repeating sequence $(C\text{-}G\text{-}G)_5$ in the descending strand.

| rDNA source | Position | Variants | Secondary-structure region in 28S rRNA | Notes |
|---|---|---|---|---|
| A4 | 2129 | $(G\text{-}G\text{-}C)_7$ | V5-3 stem | (*a*) |
| A1, A6 | | $(G\text{-}G\text{-}C)_5$ | | |
| A2, A3, A5 | | $(G\text{-}G\text{-}C)_8$ | | |
| pHrA | | $(G\text{-}G\text{-}C)_5$ | | |
| pHr12, pHr15 | | $(G\text{-}G\text{-}C)_7$ | | |
| A1–A6 | 2173 | $(C)_{11}$ | V5-3 loop | (*b*) |
| pHrA | | $(C)_{11}$ | | |
| pHr12 | | $(C)_9$ | | |
| pHr15 | | $(C)_{12}$ | | |
| A4 | 3023 | $(C\text{-}T)_8$ | V8-4 stem | (*c*) |
| pHrA, pHr19 | | $(C\text{-}T)_9$ | | |
| pHr10, pHr13 | | $(C\text{-}T)_8$ | | |
| A4 | 3317 | C-A-C-G | V8-9 stem | (*d*) |
| pHrA | | deleted | | |
| A4 | 3474 | G-C-A | V8-9 stem | (*d*) |
| pHrA | | deleted | | |

## Note on the 5.8 S rRNA sequence

The human 5.8 S rRNA sequence reported by Khan & Maden (1977) has been corrected in this work by deletion of two nucleotide residues that were erroneously assigned immediately following nucleotide 50. Other vertebrate 5.8 S-rRNA sequences have been similarly corrected at this point after DNA sequencing (summarized in Erdmann & Wolters, 1986).

## Concluding comments

The findings described in the present paper indicate that human 45 S rRNA comprises a population of closely related molecules rather than a homogeneous molecular species. Each clone studied differs from any of the others at one or more sites (Figs. 4 and 5, Table 3). The variants occur in regions that undergo relatively rapid evolutionary change. Several years ago Smith (1976) argued that regions of DNA that are not under selective pressure to remain unchanged will tend to accumulate changes in such a way as to generate tracts of simple sequences. More recently, Tautz *et al.* (1986) have argued that such simple sequence regions will in turn be foci for rapid evolution owing to the operation of mechanisms such as slippage during DNA replication. Most of the variants described in the present paper are located within or in close association with simple sequence tracts, both in the transcribed spacers and in the rapidly evolving regions of the 28 S-rRNA gene, in agreement with the above predictions. Further studies will be required to characterize the full extent of variability in the human ribosomal transcription unit, but further variants can be expected in the transcribed spacers, parts of the 5′ region of the 28 S-rRNA gene and possibly the region that encodes the paired loops in 28 S rRNA.

In addition to its intrinsic interest for studies on evolution, the occurrence of variants in the transcribed regions of rDNA may have practical consequences for the analysis of rRNA structure and processing. In particular, DNA probes from single cloned genes might give rise to artifacts if used indiscriminantly in certain types of hybridization experiments (such as $S_1$ nuclease mapping) involving the slightly heterogeneous population of cellular ribosomal precursor RNA molecules.

The availability of clones containing the complete 18 S-rRNA and 28 S-rRNA genes in a single tract of DNA should prove to be useful for studies on the

transcription and processing of human ribosomal precursor RNA *in vitro*, such as those that have been initiated for *Xenopus* rDNA (Akhtar & Maden, 1986).

We thank R. D. Schmickel for sending us a preprint of the paper by Gorski *et al.* (1987) in advance of publication. This work was supported by the Medical Research Council and the Science and Engineering Research Council.

## REFERENCES

Akhtar, Y. & Maden, B. E. (1986) Biochem. Soc. Trans. **14**, 269–270
Benton, W. P. & Davis, R. W. (1977) Science **196**, 180–182
Birnboim, H. C. & Dolý, J. (1979) Nucleic Acids Res. **7**, 1513–1523
Chan, Y. L., Olvera, J. & Wool, I. G. (1983) Nucleic Acids Res. **11**, 7819–7831
Chan, Y. L., Gutell, R., Noller, H. F. & Wool, I. G. (1984) J. Biol. Chem. **259**, 224–230
Clark, C. G., Tague, B. W., Ware, V. C. & Gerbi, S. A. (1984) Nucleic Acids Res. **12**, 6197–6220
Erdmann, V. A. & Wolters, J. (1986) Nucleic Acids Res. Suppl. **14**, r1–r59 (see r54–59)
Financsek, I., Mizumoto, K., Mishima, Y. & Muramatsu, M. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3092–3096
Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) J. Mol. Biol. **170**, 827–842
Furlong, J. C & Maden, B. E. H. (1983) EMBO J. **2**, 443–448
Furlong, J. C., Forbes, J., Robertson, M. & Maden, B. E. H. (1983) Nucleic Acids Res. **11**, 8183–8196
Gonzalez, I. L. & Schmickel, R. D. (1986) Am. J. Hum. Genet. **38**, 419–427
Gonzalez, I. L., Gorski, J. L., Campen, T. J., Dorney, D. J., Erickson, J. M., Sylvester, J. E. & Schmickel, R. D. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7666–7670
Gorski, J. L., Gonzalez, I. L. & Schmickel, R. D. (1987) J. Mol. Evol. **24**, 236–251
Hassouna, N., Michot, B. & Bachellerie, J. P. (1984) Nucleic Acids Res. **12**, 3563–3583
Khan, M. S. N. & Maden, B. E. H. (1977) Nucleic Acids Res. **4**, 2495–2505
Loening, U. E. (1968) J. Mol. Biol. **38**, 355–365
Maden, B. E. H. (1980) Nature (London) **288**, 293–296
Maden, B. E. H. (1986) J. Mol. Biol. **189**, 681–699
Maden, B. E. H., Forbes, J. M., Stewart, M. A. & Eason, R. (1982) EMBO J. **1**, 597–601
McCallum, F. S. & Maden, B. E. H. (1985) Biochem. J. **232**, 725–733
Miesfeld, R. & Arnheim, N. (1982) Nucleic Acids Res. **10**, 3933–3949
Murray, N. (1983) Lambda Vectors, Appendix III, pp. 677–684, Cold Spring Harbor Laboratory, Cold Spring Harbor
Raynal, F., Michot, B. & Bachellerie, J. P. (1984) FEBS Lett. **167**, 263–268
Schibler, U., Wyler, T. & Hagenbüchle, O. (1975) J. Mol. Biol. **94**, 503–517
Smith, G. P. (1976) Science **191**, 528–535
Stewart, M. A., Hall, L. M. C. & Maden, B. E. H. (1983) Nucleic Acids Res. **11**, 629–646
Tautz, D., Trick, M. & Dover, G. A. (1986) Nature (London) **322**, 652–656
Torczynski, R. M., Bollon, A. P. & Fuke, M. (1983) Nucleic Acids Res. **11**, 4879–4890
Torczynski, R. M., Fuke, M. & Bollon, A. P. (1985) DNA **4**, 283–291
Wellauer, P. K. & Dawid, I. B. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 2827–2831
Wellauer, P. K., Dawid, I. B., Kelley, D. E. & Perry, R. P. (1974) J. Mol. Biol. **89**, 397–407
Wilson, G. N. (1982) in The Cell Nucleus (Busch, H. & Rothburn, L., eds.), vol. 10, pp. 287–318, Academic Press, New York
Wilson, G. N., Hollar, B. A., Waterson, J. R. & Schmickel, R. D. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**, 5367–5371

Received 29 December 1986/16 March 1987, accepted 26 May 1987