Biochem. J. (1988) 249, 459–464 (Printed in Great Britain)

Pseudouridine distribution in mammalian 18 S ribosomal RNA

A major cluster in the central region of the molecule

B. Edward H. MADEN and Jane A. WAKEMAN
Department of Biochemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, U.K.

Human and rodent 18 S rRNA contain about 38 pseudouridine residues. By correlating RNA oligonucleotide data with complete sequence data derived from ribosomal DNA, 30 pseudouridine residues can be located in the RNA sequence, either exactly or to within two or three residues. Pseudouridine and 2′-*O*-methyl groups are interspersed throughout mammalian 18 S rRNA, but not in closely parallel fashion. Whereas the largest cluster of 2′-*O*-methyl groups is in the 5′ one-third of the molecule, the greatest concentration of pseudouridine is in the central one-third of the molecule.

INTRODUCTION

rRNA from most eukaryotes that have been examined contains considerable numbers of pseudouridine residues, as well as many 2′-*O*-methylated nucleotide residues and smaller numbers of base-modified nucleotide residues (Hughes & Maden, 1978; Khan *et al.*, 1978; Maden, 1987, and references cited therein). Knowledge of the locations of the modified nucleotide residues in rRNA is a prerequisite for a detailed understanding of their role.

The exact locations of the methyl groups in 18 S rRNA of *Xenopus laevis* and man have been identified by correlating data on methylated oligonucleotides, derived from enzymic hydrolysis of rRNA, with the complete nucleotide sequences derived from ribosomal DNA (rDNA) (Maden, 1986, and references cited therein). This analysis was facilitated by two features of the methylated nucleotide residues. First, these nucleotide residues can be labelled selectively *in vivo* by using methyl-labelled methionine as a radioactive precursor. This enables them to be detected easily during analytical procedures such as RNA fingerprinting. Secondly, 2′-*O*-methyl groups confer resistance to hydrolysis of the adjacent phosphodiester bond by alkali and some enzymes. This fact can be exploited for determining the locations of 2′-*O*-methyl groups within oligonucleotides (Maden & Salim, 1974; Maden & Khan, 1977).

Pseudouridine does not offer these advantages. However, it can be detected chromatographically in hydrolysates of whole RNA or individual oligonucleotides. Several pseudouridine-containing oligonucleotides were detected during early studies on rRNA from HeLa cells (Amaldi & Attardi, 1968; Maden & Forbes, 1972; Maden & Salim, 1974) and other vertebrate sources (Khan & Maden, 1976). Then Choi & Busch (1978) reported a detailed analysis of the pseudouridine-containing oligonucleotides, as well as other oligonucleotides, in a complete T_1-RNAase digest of rat 18 S rRNA.

We have now been able to map the exact or approximate locations of most of the pseudouridine residues in mammalian 18 S rRNA by matching the various oligonucleotide data to rDNA sequence data. The present paper describes our analysis and findings.

SOURCES OF DATA

Approximate content of pseudouridine in mammalian 18 S rRNA

Hughes & Maden (1978) reported the pseudouridine contents of rRNA from three vertebrate sources (human, mouse and *X. laevis*). The values were obtained by carrying out complete hydrolysis of rRNA to nucleotides and then using a separation system based on electrophoresis followed by two-dimensional paper chromatography. This system gave complete separation of pseudouridine 2′(3′)-phosphates from other products of hydrolysis and, in particular, enabled the ratio of pseudouridine to uridine to be determined. The analytical data yielded initial estimates of approximately 40 pseudouridine residues per 18 S rRNA from human (HeLa) and mouse (L) cells (Hughes & Maden, 1978). The human value was refined to approximately 38 residues by McCallum & Maden (1985) when the gene sequence and hence the exact number of uracil nucleotide residues in the RNA sequence became known. The value obtained from alkaline hydrolysis of mouse 18 S rRNA should be refined similarly to approximately 38 residues. Various minor methodological factors limit the accuracy and precision of these numbers to about $\pm 5\%$ of the stated values (about ± 2 pseudouridine residues for 18 S rRNA); however, a closely similar value has been obtained for rat 18 S rRNA by analysis of oligonucleotides, as described below.

Oligonucleotide data

Choi & Busch (1978) catalogued all of the oligonucleotides that were released by digestion of rat hepatoma 18 S rRNA with T_1 RNAase, and identified and completely or partially characterized the pseudouridine-containing oligonucleotides. Their analysis was facilitated by use of homochromatography in the second dimension of their RNA fingerprinting system. This yielded particularly good resolution of many long U-rich oligonucleotides. The data on all oligonucleotides indicated the presence of about 38 pseudouridine residues in rat 18 S rRNA.

This is the only reported study in which all of the pseudouridine-containing oligonucleotides have been

Table 1. Alignment of pseudouridine-containing oligonucleotides with tracts in the complete human 18 S rRNA sequence determined from rDNA

In the first column the spot numbers are as given in Choi & Busch (1978). Where these authors listed more than one T_1-RNAase product co-migrating within a single spot a secondary number is given here, e.g. spot 2 of Choi & Busch (1978) contains U-U-G and (U,Ψ)G; the latter is designated spot 2ii here; their spot 41 is a more complex mixture (see below). In the second column the oligonucleotide sequence reported by Choi & Busch (1978) is given with the preceding G residue, since all products are T_1-RNAase products. Hyphens denote internal phosphates, Ψ denotes pseudouridine and Gm etc. denote 2′-*O*-methylated nucleotides. 18 S rDNA does not contain sequences that match the asterisked (*) oligonucleotide data exactly. However, a closely related rDNA tract is present for each of these oligonucleotides except spot 40 (see below and Fig. 1). In the third column, where incomplete oligonucleotide data permit more than one theoretically possible DNA sequence the tract that is actually present in 18 S rDNA is shown. In the fourth column the 'start of tract' denotes the G residue at the start of the DNA tract corresponding to the RNA oligonucleotide. This notation is used because in several instances the position of Ψ within the oligonucleotide is not known exactly (see Fig. 1).

Spot no.	Oligonucleotide sequence data†	Corresponding DNA tract	Start of tract
1	G-Ψ-U-Gm-G	GTTGG	680, (891, 1345)
2ii	G(U,Ψ)G	GTTG	Many
9ii	G(U,Ψ)A-U-G	GTTATG	108
11ii	G-A-U-Ψ-G	GATTG	(1286), 1294, 1475, 1613, 1718
12	G-U-A-Ψ-G	GTATG	1171
13iii	G(C,U,Ψ)G	G(C,T,T)G	Many
20	G-U-Um-C-C-U-U-Um-G; one Ψ	GTTCCTTTG	114
23	G-A-U(C,U,Ψ)G	GATCTTG	684
24	G-C-A-U(U,Ψ)G	GCATTTG	999
27iii	G-A-A-Ψ-G	GAATG	(90), 559, 1010
30iv	G(C_2,Ψ)G	G(C,C,T)G	Many
34	G-Um-U-Ψ-A-C-U-U-U-G	GTTTACTTTG	798
35	G(C,U,Ψ)A-A-U(U,Ψ)G	GCTTAATTTG	1236
38i	G-A-U-U(or Ψ)A-A-G	GATTAAG	(41), 915, (1674)
40	G-A-U(C_2,U,Ψ)G*	None	
41vii	G(C_3U_2,Ψ)G	GCCCTTTG	1686
44i	G-A-Ψ-C-A-G	GATCAG	1054
51	G-A-A-A-U(C,U_2,Ψ)G	GAAATTCTTG	961
52ii	G(U,Ψ)A-A-U(C_2,U)G	GTTAATTCCG	1366
54	G-A-A-U-A-A-Ψ-Gm-G	GAATAATGG	860
56	G-A-A-Ψ-A-A-Ψ-G		
58i	G(C,U,Ψ)A-A-A-G	GTTCAAAG	821
70i	G-U-U-A-Ψ-U-A-A-A-G*	GTATATTAAAG	646
70ii	G-[U-A,Ψ-A,(C_2,U_2)A]U-G*	GTTATTCCCATG	1110
71	G-C-A-U-C-A-Ψ(U,Ψ)A-G*	GCATTTATCAG	215
76i	G(C_2,$Ψ_2$)A-A-A-G	GTCTCAAAG	33
86	G-m^7G-A-A-U(C_3,Ψ)A-G	GGAATTCCCAG	1638
93	G-C-U-C-Am-U-U-A-A-A-Ψ-C-A-G	GCTCATTAAATCAG	95
94	G-A-A-A-A-A-A-Ψ-Ψ-A-G	GAAAAAATTAG	807
99	G(A-Ψ,A-A-A-C,C,A-C)G	GACCATAAACG	1076
102	G-A-A-C(m^1cap^{3}Ψ-C,A-C)A-C-G*	GACTCAACACG	1245
106	G-C-A-A-U-U-A-U-Ψ-C-C-C-C-A-U-G	GCAATTATTCCCCATG	1617
112	G-U-C-C-A-C(U_2,Ψ)A-A-Am-U-C-C-U-U-U-A-A-C-G	GTCCACTTTAAATCCT-TTAACG	565
113	G-U-C-C-C-C-C-A-A-C-U-Um-C-U-U-A-Gm-A-G; one Ψ	GTCCCCCAACTTCTTA-GAG	1431

† Notes on specific oligonucleotides are given below.

Spot 1. This corresponds to methylated 18 S-rRNA spot 61 of Maden & Salim (1974), who located pseudouridine within the oligonucleotide as described in Appendix 1 of that paper. The corresponding DNA sequence GTTCG occurs at three locations within 18 S rDNA, but the position of the methylated oligonucleotide was located by hybridization in *Xenopus* (Maden, 1980, 1986). The location for Ψ in this oligonucleotide in the human sequence is 681 (Fig. 1).

Spot 2. This short pseudouridine-containing oligonucleotide cannot yet be placed in the complete sequence.

Spot 9. The RNA oligonucleotide data match a unique location in 18 S rDNA. This pseudouridine-containing product was also identified in an analysis of oligonucleotides from the 5′ end of *Xenopus* 18 S rRNA (Salim & Maden, 1980).

Spot 11. A partial survey of Hela-cell oligonucleotides revealed G-A-U-U-G and G-A-Um-U-G (methylated spot 43 of Khan *et al.*, 1978), but no pseudouridine was found in association with G-A-U-U-G (B. E. H. Maden, unpublished work).

Spot 12. This pseudouridine-containing product was identified during a partial survey of oligonucleotides from HeLa-cell 18 S rRNA, although the location of pseudouridine within the tract was not determined (B. E. H. Maden, unpublished work). The corresponding DNA pentanucleotide occurs only once in 18 S rDNA (positions 1171–1175, with Ψ at 1174).

Spot 13iii. This pseudouridine-containing oligonucleotide cannot be located uniquely in the 18 S-rRNA sequence.

Spot 20. This corresponds to methylated spot 85 of Khan *et al.* (1978). During an analysis of oligonucleotides from the 5′ end of *Xenopus* 18 S rRNA pseudouridine was identified in this spot and the oligonucleotide was located in the sequence (Salim & Maden, 1980).

Spot 23. The oligonucleotide data are compatible with GATCTTG, GATTCTG or GATTTCG. Only the first of these occurs in 18 S rDNA. This determines the sequence of the oligonucleotide and enables it to be placed (positions 684–690 in Fig. 1). After digestion of this spot with RNAase

A, pseudouridine is released as mononucleotide and not as A-Ψ (Choi & Busch, 1978). Hence pseudouridine is at position 688 or 689. Similar arguments based on the use of RNAase A and U_2 RNAase were used by Choi & Busch (1978) in locating pseudouridine (exactly or approximately) within other U-rich oligonucleotides.

Spot 24. This corresponds to a unique location in the 18 S rDNA sequence as indicated.

Spot 27iii. This pseudouridine-containing product was found during a partial survey of T_1-RNAase oligonucleotides from HeLa-cell 18 S rRNA (B. E. H. Maden, unpublished work). In an analysis of oligonucleotides from the 5′ end of *Xenopus* 18 S rRNA it was found that A-A-U-G does not contain pseudouridine (Salim & Maden, 1980). Therefore pseudouridine is probably at one of the other two locations indicated in the complete seuqence.

Spot 30iv. This oligonucleotide cannot be located uniquely in the 18 S-rRNA sequence

Spot 34. This corresponds to methylated spot 86 of Khan *et al.* (1978). This spot was not analysed for pseudouridine by those authors.

Spot 35. This long oligonucleotide contains two pseudouridine residues each in a separate U_2-RNAase product (Choi & Busch, 1978).

Spot 38. This spot showed only partial modification of uridine to pseudouridine (Choi & Busch, 1978). The corresponding heptanucleotide occurs at three locations in 18 S rDNA. In an analysis of oligonucleotides from the 5′ end of *Xenopus* 18 S rRNA (Salim & Maden, 1980), no pseudouridine was found in the oligonucleotide corresponding to position 41. Hybridization experiments (Maden, 1980, 1986) showed that the methylated spot A-U-U-Am-A-G (spot 48 of Khan *et al.*, 1978) occurs at the 3′ location. Therefore the pseudouridine-containing sequence is assigned by elimination to position 915.

Spot 40. There is no sequence in 18 S rDNA that corresponds to the oligonucleotide data for this spot. Moreover, when all hexanucleotides and larger oligonucleotides listed by Choi & Busch (1978) are matched to rDNA, there are no obvious 'vacancies' that closely match this spot. However, partial analysis of HeLa-cell 18 S rRNA yielded an impure spot with the approximate composition (C_3,U,Ψ,A)G (D. G. Hughes & B. E. H. Maden, unpublished work). Further work is required to correlate the spot with the 18 S rDNA sequence.

Spot 41vii. This product was identified by Choi & Busch (1978) as a component in a mixture of seven co-migrating oligonucleotides, only one of which contains pseudouridine. It is the only oligonucleotide in the mixture which contains no A residues and hence was unchanged after treatment of the mixture with U_2 RNAase. The same spot was obtained fairly pure and was characterized as containing one pseudouridine residue by D. G. Hughes & B. E. H. Maden (unpublished work). Of the various possible arrangements of the pyrimidines within this oligonucleotide only that shown under 'corresponding DNA tract' actually occurs in 18 S rDNA.

Spot 44i. A pseudouridine-containing product whose electrophoretic mobility in two dimensions matches this spot was found during a partial survey of T_1-RNAase oligonucleotides from HeLa-cell 18 S rRNA (B. E. H. Maden, unpublished work).

Spot 51. Of the possible sequences for this oligonucleotide, only that shown is present in 18 S rDNA.

Spot 52ii. This is one of three co-migrating oligonucleotides. Only this oligonucleotide contains pseudouridine and further analytical data (Choi & Busch, 1978) yielded the partial sequence data shown. Of the three possible solutions at the DNA level, only that shown is present in 18 S rDNA.

Spots 54 and 56. These were listed as separate products by Choi & Busch (1978). However, the corresponding DNA sequence occurs only once in 18 S rDNA. The explanation is as follows. Spot 69 of Khan *et al.* (1978), although not fully sequenced by those authors, corresponds on the basis of its electrophoretic mobility to spot 54 in this Table (see Maden, 1986). Spot 69 of Khan *et al.* (1978) shows interspecies differences in extent of methylation, being almost fully methylated in HeLa-cell 18 S rRNA, partly methylated in mouse and hamster 18 S rRNA and unmethylated in *Xenopus* 18 S rRNA. Partial methylation in rat would account for the finding of both products by Choi & Busch (1978), since T_1 RNAase cleaves at G but not at Gm. [The non-methylated form (spot 56) yields 2 mol of A-A-Ψ/mol (Choi & Busch, 1978). A-A-Ψ in this tract and in spot 27iii accounts for 2.5–3 mol of A-A-Ψ found in combined T_1-RNAase-plus-pancreatic-RNAase digests of 18 S rRNA (Maden & Forbes, 1972).]

Spot 58i. Of the three possible sequences at the DNA level, only the one shown occurs in 18 S rDNA.

Spots 70i and 70ii. These two products were unresolved from each other by Choi & Busch (1978), and the respective sequences were inferred from the products of further digestion by RNAase A and U_2 RNAase. In fact the inferred oligonucleotide sequences do not match the 18 S rDNA sequence, but rearrangement of the RNAase-A and U_2-RNAase products gives matches with unique regions of the 18 S rDNA sequence as shown. However, it is not possible to conclude from the analytical data whether one pseudouridine residue occurs in each oligonucleotide or both pseudouridine residues are present in one of the oligonucleotides.

Spot 71. This inferred oligonucleotide sequence does not match the 18 S rDNA sequence. However, the rDNA tract at positions 216–226 encodes an isomeric and closely related sequence. Hence we infer that the correct sequence is as shown in Fig. 1.

Spot 76i. Choi & Busch (1978) report that this spot contains two pseudouridine residues. The spot was also identified as containing pseudouridine during analysis of oligonucleotides from the 5′ end of *Xenopus* 18 S rRNA (Salim & Maden, 1980), although it is not certain whether one or both uracil nucleotides in *Xenopus* are pseudouridine.

Spot 86. This corresponds to methylated spot 49 of Khan *et al.* (1978). The spot was obtained impure by the latter authors and was not analysed for pseudouridine. The oligonucleotide contains within it the sequence encoded by the unique *Eco*RI site in 18 S rDNA (GAATTC), and this contributed to its initial mapping (Maden & Reeder, 1979; Maden, 1986).

Spot 93. This long oligonucleotide corresponds to methylated spot 87 of Maden (1986, and references cited therein). A-A-A-Ψ was found within this spot during analysis of oligonucleotides from the 5′ end of *Xenopus* 18 S rRNA (Salim & Maden, 1980).

Spot 94. The product A-A-A-A-A-A-Ψ was identified previously in combined T_1-RNAase-plus-pancreatic-RNAase digests of 18 S rRNA from HeLa cells (Maden & Forbes, 1972; Khan & Maden, 1976). The presence of a second pseudouridine residue in the T_1-RNAase product was established by Choi & Busch (1978).

Spot 99. There are 24 possible sequences that are compatible with the oligonucleotide data on this spot, but only the DNA sequence that is shown actually occurs in 18 S rDNA.

Spot 102. $m^1cap^3Ψ$ designates the hypermodified nucleotide 1-methyl-3-γ-(α-amino-α-carboxypropyl)pseudouridine (Saponara & Enger, 1974), previously abbreviated amΨ in papers from this laboratory. The oligonucleotide containing this compound was identified by Maden *et al.* (1975), but the sequence deduced by those authors, and also given by Choi & Busch (1978), was slightly incorrect (discussed by Maden, 1986). The correct sequence is listed for DNA in the 'corresponding DNA tract' column and shown for RNA in Fig. 1.

Spot 106. Pseudouridine in this long oligonucleotide was placed from RNAase A and U_2 RNAase data.

Spot 112. This very long T_1-RNAase oligonucleotide also appears in slightly shortened form in fingerprints of rat 18 S rRNA, apparently as a result of nicks appearing in some molecules immediately following the last U in the sequence (Choi & Busch, 1978; their spot 105). The sequence corresponds to methylated oligonucleotide 92 of Khan *et al.* (1978) and Maden (1986).

Spot 113. This long T_1-RNAse oligonucleotide also occurs in sightly shortened form due to the fact that some molecules are not methylated at Gm (spot 107 of Choi & Busch, 1978). The sequence corresponds to spot numbers 89 (G not methylated) and 94 (Gm) in Khan *et al.* (1978) and Maden (1986). Choi & Busch (1978) reported pseudouridine within the sequence C-U-Um-C-U-U-A (footnote c of their Table II).

---

studied. However, several of the same oligonucleotides have been characterized during partial analyses of 18 S rRNA from related species, as noted in Table 1. Moreover, the methylated oligonucleotides of human and mouse 18 S rRNA (Khan *et al.*, 1978) and rat 18 S rRNA (Choi & Busch, 1978) show complete corre-

spondence. As discussed below, the respective 18 S rDNA sequences are extremely similar. On the basis of these data it seems likely, and is assumed here, that the complete secondary modification patterns of 18 S rRNA from the three species are closely similar or identical.

rDNA sequence data

18 S-rDNA sequences have been published for rat (Torczynski *et al.*, 1983; Chan *et al.*, 1984), mouse (Raynal *et al.*, 1984) and human (McCallum & Maden, 1985; Torczynski *et al.*, 1985; Gonzalez & Schmickel, 1986) genes. The three published human sequences are almost identical, differing from each other at only two or three points. Comparison of these sequences with each other and with the other mammalian sequences suggests that the version given by McCallum & Maden (1985) is correct with one minor exception: position 140 is commonly T rather than C, which was present in the first clone that was sequenced in this laboratory (see Maden *et al.*, 1987). The two rat sequences differ from each other in several minor respects. Comparison of the mammalian aligned sequences suggests that at least some of the apparent differences between the rat sequences may be due to sequencing errors (discussed by McCallum & Maden, 1985). There are only 13 points at which both versions of the rat sequence differ from the human sequence, and most of these are clustered in a phylogenetically variable region near the 5′ end. The mouse 18 S-rDNA sequence (Raynal *et al.*, 1984) is also extremely similar to the rat and human sequences. The human 18 S-rDNA sequence with the definite differences in rat (represented as for rRNA in Fig. 1 below) was used as the data-base for locating the pseudouridine-containing oligonucleotides in the complete 18 S rRNA sequence.

LOCATIONS OF PSEUDOURIDINE RESIDUES

Table 1 lists the pseudouridine-containing oligonucleotides that were reported by Choi & Busch (1978), together with the matching sequences in 18 S rDNA, identified during the present work. Table 1 includes annotations on the individual oligonucleotides, including some additional data from this laboratory.

In many instances there is a unique fit between an rRNA oligonucleotide whose sequence had been fully determined and a corresponding tract in rDNA. Thus the oligonucleotide can be definitely located. In some of these instances pseudouridine had been precisely located within the oligonucleotide, either because there is only one uracil nucleotide (for example spot 44i) or because secondary digestion by RNAase A or U_2 RNAase had provided the relevant information (for example spot 93). In other instances it is not yet possible to distinguish between two or three possible locations for pseudouridine within a given oligonucleotide (for example spot 24). In a second class of oligonucleotides the sequence had not been fully determined, but when the possible solutions were listed it was found that only one solution matched the rDNA sequence. For example the analytical data on spot 51 are consistent with four theoretically possible DNA sequences, but only the one shown in the DNA column actually occurs in 18 S rDNA. Thus spot 51 can be identified with this tract in the sequence. A few oligonucleotides could not be matched exactly with any part of the complete 18 S-rDNA sequence. However, closely similar sequences were found in all except one of these instances, as detailed in the notes to Table 1. Since these were large oligonucleotides in which technical difficulties could have arisen during characterization, we infer that the sequences given in the DNA column are the correct ones corresponding to these oligonucleotides. Finally, in a few instances pseudouridine was recovered in small oligonucleotides that are multiply represented in 18 S rDNA. These pseudouridine residues have not yet been located.

Table 2. Summary of data on pseudouridine residues in 18 S rRNA

The pseudouridine residues that were located exactly were in products 1, 12, 34, 38i, 44i, 56 (2 Ψ), 76i (2 Ψ), 86, 93, 94 (2 Ψ), 99, 102 (m^1cap^3U) and 106, and those located approximately were in products 9ii, 20, 23, 24, 35 (2 Ψ), 41vii, 51, 52ii, 58i, 71 (2 Ψ), 112 and 113. The pseudouridine residues that were not located were in products 2ii, 11ii, 13iii, 27iii, 30iv, 40, 70i and/or 70ii; see the notes to Table 1 on products in this category, particularly 11, 40, 70i and 70ii.

No. of Ψ residues located exactly	16
No. of Ψ residues located approximately	14
No. of Ψ residues not located	~8*
Total no. of Ψ residues	~38*

* Analysis of oligonucleotides from HeLa cell rRNA (B. E. H. Maden, unpublished work) suggests 1 mol each of Ψ-G and A-Ψ-G/mol which were not reported by Choi & Busch (1978).

In all, 30 of the 38 pseudouridine residues in the oligonucleotides listed by Choi & Busch (1978) can be located exactly or approximately in the complete 18 S-rRNA sequence inferred from rDNA, as summarized in Table 2. The locations of these pseudouridine residues in the complete sequence are shown in Fig. 1.

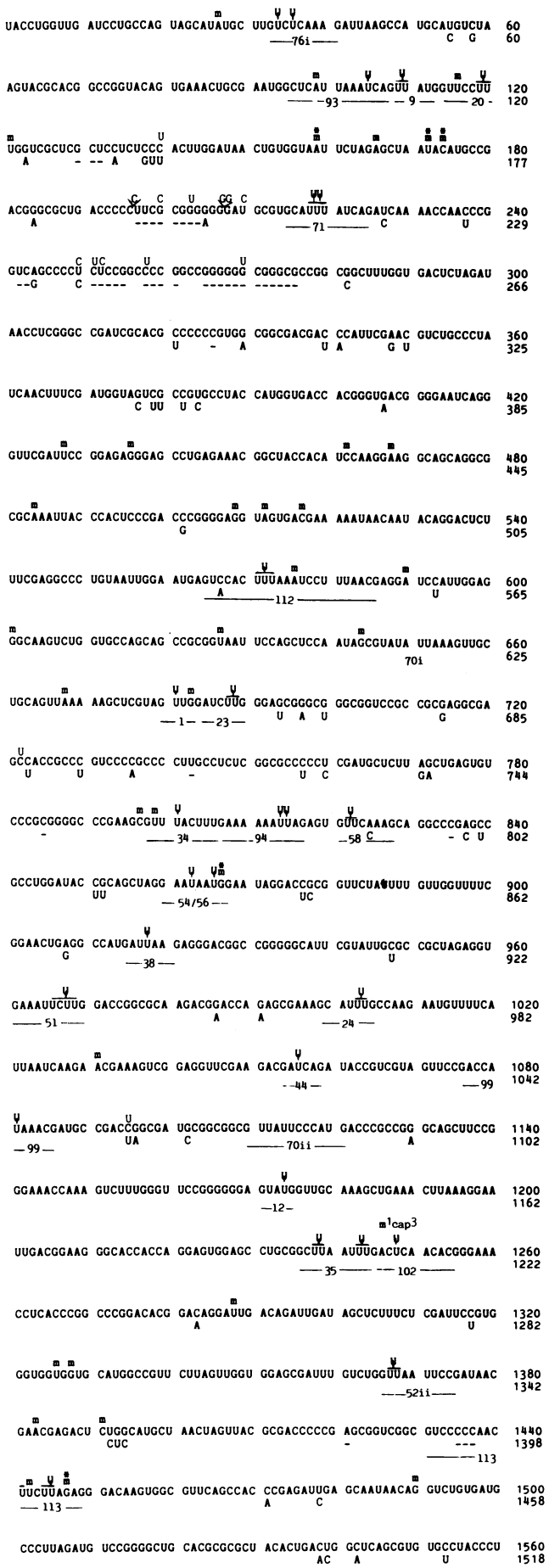
DISCUSSION

Pseudouridine, methylation and rRNA maturation

The pseudouridine residues in Fig. 1 are widely distributed along the 18 S-rRNA sequence, from positions 34/36 near the 5′ end to position approximately 1690/1692 near the 3′ end. Pseudouridine residues are interspersed with methylated nucleotide residues, but the respective distributions do not parallel each other closely.

This information is summarized in Fig. 2, which also shows the boundaries of the proposed domains in 18 S rRNA. The domains are defined by putative long-range interactions between tracts of nucleotides. There is general agreement from phylogenetic comparisons on the sites of the long-range interactions, as indicated in the legend to Fig. 2, but some details of the internal structures of the domains remain to be resolved.

The heaviest clustering of methyl groups is in domain 1a, with 21 2′-*O*-methylation sites in this part of mammalian 18 S rRNA. There are also several pseudouridine residues in domain 1a, particularly in the first 220 nucleotide residues, but the highly 2′-*O*-methylated region of residues 421–645 contains only one identified pseudouridine residue (with possibilities among short oligonucleotides for two or three others). The pattern is reversed in domain 1b, which contains 14 identified pseudouridine residues and only six 2′-*O*-methyl groups. In domain 2 the pattern is again reversed, with methyl



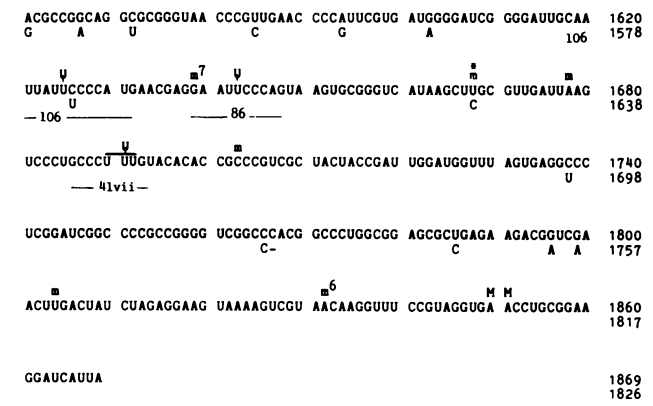


Fig. 1. Locations of identified pseudouridine residues in mammalian 18 S rRNA

The principal sequence is the human 18 S-rRNA sequence in McCallum & Maden (1985). Standard letters above this sequence indicate definite differences between the human and rat 18 S-rRNA sequences, from the data compilation in McCallum & Maden (1985). (Most differences are substitutions; insertions are denoted by V signs). Letters below the sequence indicate differences between the human and *Xenopus laevis* sequences; dashes indicate deletions in *Xenopus* with respect to human 18 S rRNA. (The *Xenopus* sequence is included to indicate the pattern of conserved and variable tracts in 18 S rRNA from vertebrates.) Ψ denotes pseudouridine, located from evidence summarized in the notes to Table 1. A bar above the sequence indicates that one of the indicated uracil nucleotides is pseudouridine. Below the sequence the numbers and bars denote the oligonucleotide numbers in Table 1; see the Table for notes on oligonucleotides 70i, 70ii and 71. The locations of methyl groups are given for comparison: m denotes 2′-*O*-ribose methylation in mammalian and *Xenopus* 18 S rRNA; * denotes 2′-*O*-ribose methylation in mammalian but not in *Xenopus* 18 S rRNA; $m^1cap^3\Psi$ denotes the hypermodified nucleotide 1-methyl-3-γ-(α-amino-α-carboxypropyl)pseudouridine; m^7G denotes 7-methylguanosine; m^6A denotes *N*-6-methyladenosine; M denotes *N*-6-dimethyladenosine. The 2′-*O*-ribose methylation patterns of human and rodent 18 S rRNA are apparently identical (on the basis of oligonucleotide data; Khan *et al.*, 1978; Choi & Busch, 1978), but there are differences between the mammalian and *Xenopus* 18 S rRNA patterns, not readily explainable from the sequences. Only a few pseudouridine residues have been located in *Xenopus* 18 S rRNA (see the notes to Table 1), and it is possible that there may be differences among some pseudouridine sites between mammals and *Xenopus*.

groups exceeding pseudouridine residues. The hypermodified nucleotide $m^1cap^3\Psi$ occurs at position 1248; there are seven further identified pseudouridine residues compared with 12 2′-*O*-methylated nucleotide residues and five base-methylated nucleotide residues in this domain.

Even allowing for the fact that a few pseudouridine residues remain to be located, it is clear that the greatest concentrations of pseudouridine residues and 2′-*O*-methyl groups are in different regions, with prominent clustering of pseudouridine residues in domain 1b.

Eukaryotic rRNA comprises a mosaic of core sequences and expansion segments (Gerbi, 1985). The core sequences are homologous at the level of secondary

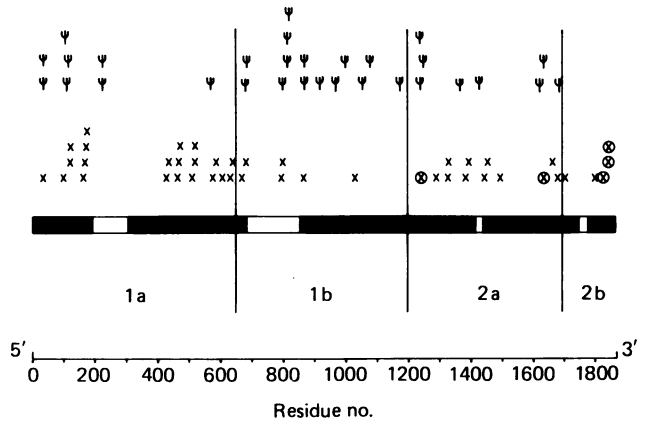


Fig. 2. Summary diagram showing the locations of methyl groups and pseudouridine residues in mammalian 18 S rRNA, from the data in Fig. 1

In the upper section, × symbols denote 2′-*O*-methyl groups, ⊗ symbols denote base methyl groups and Ψ symbols denote identified pseudouridine residues. In the lower section white regions represent eukaryotic 'expansion segments', as defined by the sequence alignments of Nelles *et al.* (1984). Vertical lines indicate domain boundaries. The putative long-range interactions that define the domains involve conserved tracts, which occur at the following locations in the human sequence: domain 1, residues 12–15 with residues 1196–1199; domain 1a, residues 22–31 with residues 643–652; domain 2a, residues 1207–1214 with residues 1685–1692. These interactions are the same as in the secondary-structure models for rat (Chan *et al.*, 1984), *Xenopus* (Atmadja *et al.*, 1984) and *Artemia* (Nelles *et al.*, 1984) 18 S rRNA.

structure with prokaryotic rRNA, whereas the expansion segments comprise additional material that is not present in prokaryotic rRNA. In Fig. 2 the core sequences are shown in black and the expansion segments in white. The expansion segments are relatively small in 18 S rRNA (but larger in 28 S rRNA, especially in vertebrates). A relatively high content of pseudouridine and 2′-*O*-methyl groups is characteristic of eukaryotic and not prokaryotic rRNA. Nevertheless, most of the modified nucleotides in 18 S rRNA are in the core sequences, with only a few in the expansion segments (Fig. 2).

The 2′-*O*-methyl groups of rRNA in eukaryotes are added to 45 S rRNA in the nucleolus, as shown in detail for HeLa cells (Maden & Salim, 1974). Moreover, methylation of ribosomal precursor RNA is essential for ribosome maturation (Vaughan *et al.*, 1967). It was found during early studies that pseudouridine is also present in 45 S rRNA (Jeanteur *et al.*, 1968; Maden & Forbes, 1972). The information given in the present paper provides a starting point for examining the distribution of pseudouridine in 45 S rRNA and the role of pseudouridine in the assembly and function of eukaryotic ribosomes.

Received 3 July 1987; accepted 22 September 1987

We thank Elaine Mayers for help with computer searches. This work was supported by the Medical Research Council and by a Science and Engineering Research Council studentship to J.A.W.

REFERENCES

Amaldi, F. & Attardi, G. (1968) J. Mol. Biol. **33**, 737–755
Atmadja, J., Brimacombe, R. & Maden, B. E. H. (1984) Nucleic Acids Res. **12**, 2649–2667
Chan, Y. L., Gutell, R., Noller, H. F. & Wool, I. G. (1984) J. Biol. Chem. **259**, 224–230
Choi, Y. C. & Busch, H. (1978) Biochemistry **17**, 2551–2560
Gerbi, S. A. (1985) in Molecular Evolutionary Genetics (MacIntyre, R. J., ed.), pp. 419–517, Plenum Press, New York
Gonzalez, I. L. & Schmickel, R. D. (1986) Am. J. Hum. Genet. **38**, 419–427
Hughes, D. G. & Maden, B. E. H. (1978) Biochem. J. **171**, 781–786
Jeanteur, P., Amaldi, F. & Attardi, G. (1968) J. Mol. Biol. **33**, 757–775
Khan, M. S. N. & Maden, B. E. H. (1976) J. Mol. Biol. **101**, 235–254
Khan, M. S. N., Salim, M. & Maden, B. E. H. (1978) Biochem. J. **169**, 531–542
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. (1987) in Chromatographic and other Analytical Methods in Nucleic Acids Modification Research (Gehrke, C. W. & Kuo, K. C., eds.), Elsevier, Amsterdam, in the press
Maden, B. E. H. & Forbes, J. (1972) FEBS Lett. **28**, 289–292
Maden, B. E. H. & Khan, M. S. N. (1977) Biochem. J. **167**, 211–221
Maden, B. E. H. & Reeder, R. H. (1979) Nucleic Acids Res. **6**, 817–830
Maden, B. E. H. & Salim, M. (1974) J. Mol. Biol. **88**, 133–164
Maden, B. E. H., Forbes, J., De Jonge, P. & Klootwijk, J. (1975) FEBS Lett. **59**, 60–63
Maden, B. E. H., Dent, C. L., Farrell, T., Garde, J., McCallum, F. S. & Wakeman, J. A. (1987) Biochem. J. **246**, 519–527
McCallum, F. S. & Maden, B. E. H. (1985) Biochem. J. **232**, 725–733
Nelles, L., Fange, B.-L., Volckaert, G., Vandenberghe, A. & De Wachter, R. (1984) Nucleic Acids Res. **12**, 8749–8768
Raynal, F., Michot, B. & Bachellerie, J.-P. (1984) FEBS Lett. **167**, 263–268
Salim, M. & Maden, B. E. H. (1980) Nucleic Acids Res. **8**, 2871–2884
Saponara, A. G. & Enger, M. D. (1974) Biochim. Biophys. Acta **349**, 61–77
Torczynski, R., 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
Vaughan, M. H., Soeiro, R., Warner, J. R. & Darnell, J. E. (1967) Proc. Natl. Acad. Sci. U.S.A. **58**, 1527–1534