Volume 4 Number 7 July 1977 Nucleic Acids Research

Nucleotide sequence relationships between vertebrate 5.8 S ribosomal RNAs

M. S. N. Khan and B. E. H. Maden

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, UK

Received 3 June 1977

ABSTRACT

Nucleotide sequences of the 5.8 S ribosomal RNAs from HeLa cells, Xenopus laevis and chick embryo fibroblasts were compared. Xenopus laevis 5.8 S RNA differs from that of HeLa cells in four internal positions and at the 3' end of the molecule. Chick 5.8 S RNA differs from that of HeLa cells in two positions. Six out of the seven interspecies differences are due to base substitutions. The other difference is due to the presence of an extra nucleotide, internally located, within the Xenopus 5.8 S sequence.

INTRODUCTION

5.8 S ribosomal RNA is a small RNA species (∿ 160 nucleotides) which occurs in the large ribosomal subunit of eukaryotes[1,2]. It is encoded by the major ribosomal DNA (rDNA)[3,4] and is co-transcribed with 18 S and 28 S RNA within ribosomal precursor RNA [5,6]. It is cleaved into a separate entity during ribosome maturation and remains non-covalently attached to 28 S RNA[1,2].

Nucleotide sequence data on 5.8 S RNAs might illuminate questions concerning the secondary structure of the molecule, its processing, its interactions and the evolution of rRNA. The nucleotide sequences of Saccharomyces cerevissiae[7] and rat hepatoma[8] 5.8 S RNAs have been determined. They share 75% homology and a basically common secondary structure has been proposed[8]. HeLa cell and mouse 5.8 S RNAs were reported to be indistinguishable by fingerprinting techniques from that of rat hepatoma except for minor differences at the termini[9-11]. The results of S_1 nuclease digestion of HeLa cell 5.8 S RNA were consistent with features of the proposed secondary structure of the mammalian sequence[11]. Turtle 5.8 S RNA differs from that of mammals in only one position, close to the 5' end of the molecule[12]. In this paper we compare the sequences of the 5.8 S RNAs of two further non-mammalian vertebrates, Xenopus laevis and chick, with that of HeLa cells. These

differ in five and two positions respectively from the mammalian sequence. Some implications of these vertebrate 5.8 S sequence data are discussed.

METHODS

Growth and radioactive labelling of cells were described elsewhere[13]. X laevis cells, from a kidney line, were originally a gift from Dr. K. Jones. ^{32}P-labelled 5.8 S RNAs were prepared as described previously[5,13] and were purified on a second sucrose gradient. Complete digestion by T_1 or pancreatic ribonuclease and fingerprinting were carried out by standard procedures[14]. The sequences of the T_1 or pancreatic oligonucleotides were determined by complete digestion with one or more of the following enzymes:- snake venom phosphodiesterase, pancreatic, T_1 or U_2 ribonuclease[14]. Assay of oligonucleotides for the presence of pseudouridylate was carried out as described previously[5].

RESULTS

Almost the entire mammalian 5.8 S sequence is spanned by distinctive oligonucleotides which are liberated in approximately unimolar yield on complete digestion with T_1 or pancreatic ribonuclease (Figure 1). It should therefore be possible to locate and identify sequence differences between closely related 5.8 S RNAs by fingerprinting, followed by sequence analysis of those oligonucleotides that differ between species. Fingerprints of the RNAs are shown in Figures 2-4. The molar yields of the HeLa products were in agreement with the published sequence[9,11], with minor qualifications mentioned below. Table 1 summarizes the significant qualitative and quantitative interspecies differences between the fingerprints. From these the sequence differences shown in the Table and in Figure 5 were deduced.

X laevis 5.8 S RNA differs from that of HeLa cells at five positions.

(i) A68 → G68. The HeLa cell T_1 oligonucleotide, A-A-U-U-A-A-U-Gp (T24) is replaced in the Xenopus fingerprint by A-A-U-U-A-Gp (T25) (Figures 2 and 3), together with an extra mole of U-Gp (Table 1). This finding, together with the unique pancreatic oligonucleotide A-G-Up (P17a) in Xenopus, and the absence of A-A-Up (P11), signify an A→G substitution at position 68 (Figure 5).

(ii) Presence of an extra C residue next to C 111. The HeLa cell T_1 oligonucleotide, C-A-C-U-U-Gp (T18) is replaced in Xenopus by the closely related oligonucleotide, C-A-C-C-U-U-Gp (T18a) (Figure 3), whose sequence was determined as described in the legend to Table 1.

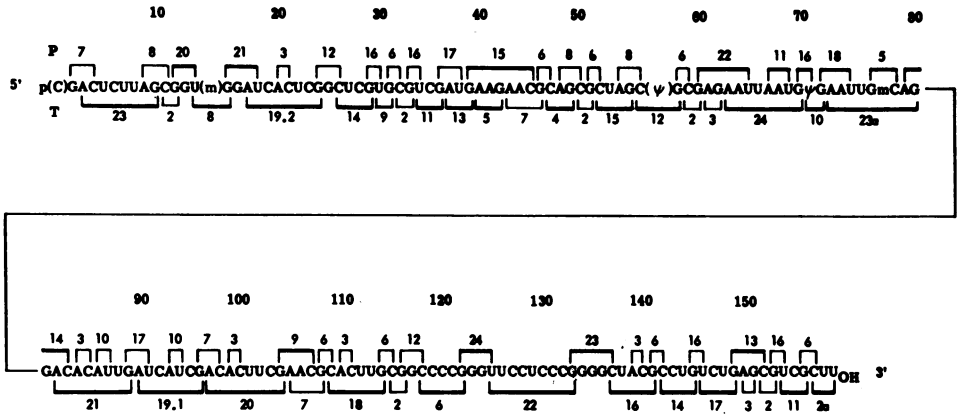

Figure 1. Nucleotide sequence of HeLa cell 5.8 S RNA[9,11]. T_1 ribonuclease products (T) are shown below the sequence and pancreatic ribonuclease products (P) above. Heavy bars indicate oligonucleotides which are liberated in approximately unimolar yield. The numbering system is that used in references 5 and 11.

The extra C residue is next to C 111. Because of this extra nucleotide the Xenopus numbering system is out of phase with that of HeLa by one nucleotide downstream from this position.

(iii) U138 → C139. The HeLa T_1 product C-U-A-C-Gp (T16) is replaced in Xenopus by C-C-A-C-Gp (T6a). This indicates that U138 in HeLa is replaced by C in Xenopus. Both products were obtained in low yields, however, particularly T6a in Xenopus (Table 1). This seems to be due to location of the T_1 ribonuclease cleavage site, adjacent to G136, within a highly GC rich stem (Figure 5). Evidence that this is the cause of the low recovery of this oligonucleotide will be fully described elsewhere (Kelly, Goddard, Khan and Maden, unpublished observations), and may be summarised here as follows:- All oligonucleotides from this stem are recovered in concomitantly low yields, more so in Xenopus than in HeLa. A large, partial digestion fragment containing these oligonucleotides can be recovered from the fingerprint, in which it migrates rapidly in the first dimension, after T_1 ribonuclease digestion under "nominally" complete digestion conditions. (We are grateful to P. Ford for first drawing our attention to this fragment; personal communication). Finally, cytidine residues in the GC rich stem are unreactive towards sodium bisulphite even at 50°, implying high stability of base pairing in this region. The greater resistance of the Xenopus than the HeLa stem to digestion by T_1 ribonuclease may be

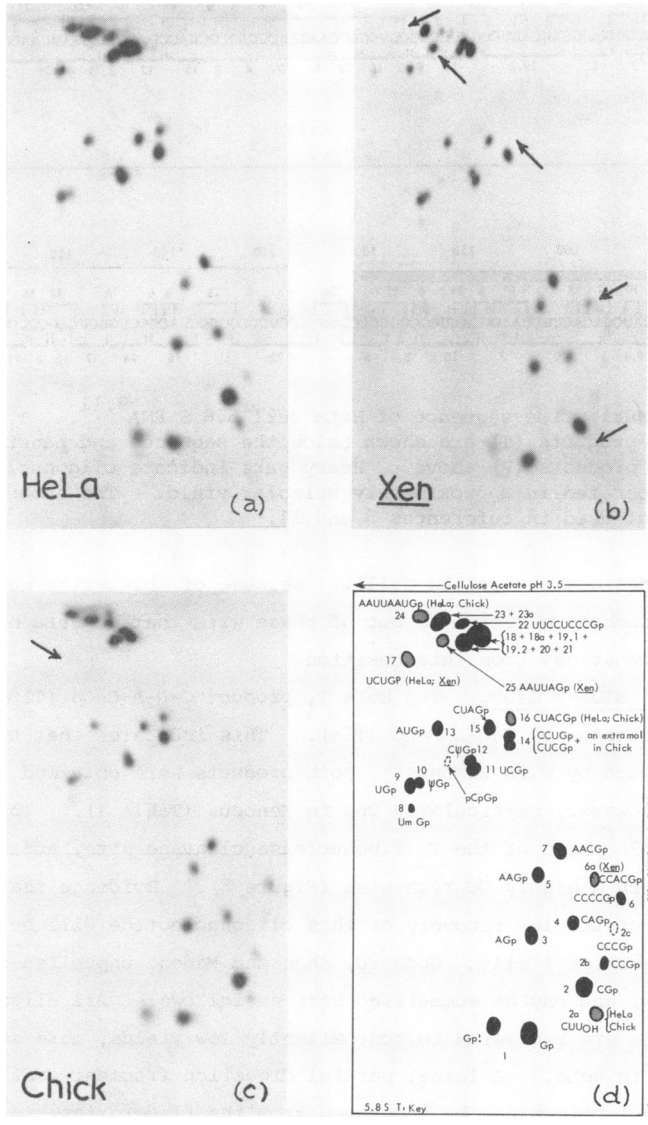

Figure 2. T_1 ribonuclease fingerprints of ^{32}P-labelled 5.8 S RNAs from HeLa Cells (a), Xenopus laevis (b) and chick (c). The arrows indicate qualitative interspecies differences from the HeLa cell fingerprint. Shaded spots in the Key (d) indicate products which are not common to all three species. (The Xenopus 3' end products, CUC$_{OH}$ and CUCC$_{OH}$ remain near the origin in the first dimension and migrate more rapidly than CG in the second dimension.)

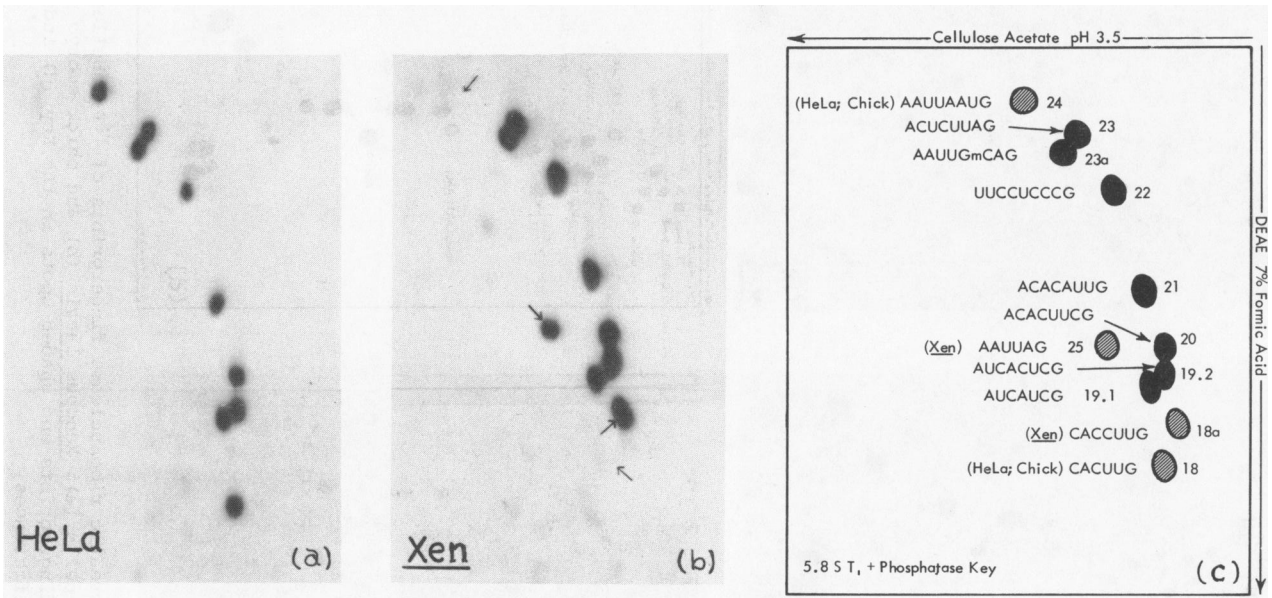

Figure 3. T_1 plus phosphatase fingerprints of ^{32}P-labelled 5.8S RNAs from HeLa cells (a) and Xenopus laevis (b). Products T18-25 are resolved more clearly than in Figure 1. Arrows indicate interspecies differences, also indicated by shaded spots in the Key (c). (Faint products are due to overdigestion or impurities).

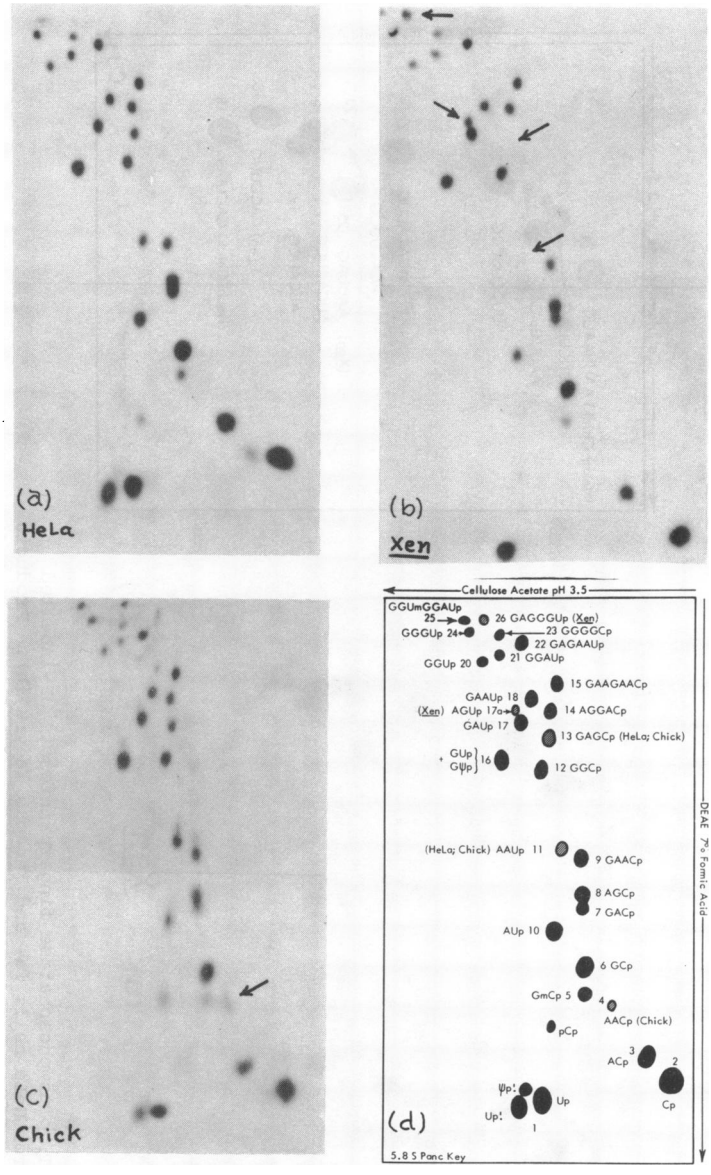

Figure 4. Pancreatic ribonuclease fingerprints of ^{32}P-labelled 5.8 S RNAs from HeLa cells (a), Xenopus laevis (b) and chick embryo (c). Arrows in the fingerprints and shaded spots in the Key (d) indicate interspecies differences.

TABLE 1. Interspecies differences between HeLa, Xenopus laevis and chick 5.8 S RNAs.

Spot No.	Sequence	Molar yield		Sequence difference	Notes
		HeLa	Xenopus		
T24	(G)-A-A-U-U-A-A-U-Gp	0.98(1)	-	A68 ↔ G	
T25	(G)-A-A-U-U-A-Gp	-	1.00		
T9	(G)-U-Gp	2.16(2)	2.88		
P11	(Py)-A-A-Up	1.01(1)	-		
P17a	(Py)-A-G-Up	-	1.09		
T18	(G)-C-A-C-U-U-Gp	1.03(1)	-	C insert, Xen.	(a)
T18a	(G)-C-A-C-C-U-U-Gp	-	0.84		
T16	(G)-C-U-A-C-Gp	0.60(≤1)	-	U138 ↔ C	(b)
T6a	(G)-C-C-A-C-Gp	-	0.34		
P13	(Py)-G-A-G-Cp	0.98(1)	-	C152 ↔ G	(c)
P26	(Py)-G-A-G-G-G-Up	-	0.61		
T2a	(G)-C-U-U$_{OH}$	0.86(1)	-	U159 ↔ C	
	(G)-C-U-C$_{OH}$	-	0.55		
	(G)-C-U-C-C$_{OH}$	-	0.34		
		HeLa	Chick		
T17	(G)-U-C-U-Gp	1.08(1)	-	U146 ↔ C	
T14	(G)-C-C-U-Gp	1.08(1)	2.08		
T23	(G)-A-C-U-C-U-U-A-Gp	0.98(1)	0.35	G2 ↔ A	(d)
	C_{3-4},A_3,U_3,Gp	-	0.35		
P7	(Py)-G-A-Cp	1.82(2)	0.86		
P4	(Py)-A-A-Cp	-	0.82		

The molar yields shown are the means of at least three separate determinations. T_1 fingerprints with long separations in both dimensions or T_1 plus phosphatase fingerprints were used to separate products T18-25. Notes: (a) The base composition of T18a was C_3, A, U_2, G. Digestion with venom phosphodiesterase confirmed C as the 5' terminus. Pancreatic ribonuclease digestion liberated C(++), AC(+) and U(++), suggesting the sequence C(AC,C,U_2)G. Digestion with U_2 ribonuclease, using 0.1 unit enzyme/ml (Ref. 14), liberated CA, CC and CU. The latter two products must result from secondary splits; the identities of these products were confirmed by alkaline hydrolysis. The sequence was therefore deduced as C A C C U U G. (b,c,d) These molar yields are discussed in the text.

due to the extra GC pair, permitted by the U138 → C139 substitution (Figure 5).

(iv) C152 → G153. The Xenopus pancreatic ribonuclease fingerprint lacks the HeLa product G-A-G-Cp (P13), but yields instead G-A-G-G-G-Up (P26). This indicates the change, HeLa C152 → Xenopus G153. The somewhat low yield of Xenopus product P26 may be due to incomplete transfer of this G-rich product from the first dimension to the second dimension of the fingerprint. However, on digestion of Xenopus spot P21 (nominally G-G-A-Up) with T_1 ribonuclease evidence was found for a small amount of co-migrating G-A-G-Up (0.15 mole). This might possibly signify a (G,U) heterogeneity at position 153 in the Xenopus sequence, with G

predominating. However, it is difficult to rule out the alternative possibility of an impurity in the 5.8 S preparation.

(v) U → C at the 3' terminus. The HeLa 3' terminal oligonucleotide, C-U-U_{OH}, is replaced in Xenopus by C-U-C_{OH} (0.55 molar) and C-U-C-C_{OH} (0.34 molar). The sum of these yields is 0.9 molar, that is, close to unity. (See legend to Figure 2 for mobility of these products).

Chick 5.8 S RNA differs from that of HeLa cells at only two positions.

(i) U146 → C146. The HeLa cell T_1 oligonucleotide U-C-U-G (T17) is replaced by an extra mole of C-C-U-G (T14) in chick. T14 is a complex spot, consisting in the HeLa cell 5.8 S fingerprint of one mole each of C-U-C-G and C-C-U-G. The two components are just resolved in long separations. It is the upper component, C-C-U-G, which is reduplicated in the chick fingerprint. These findings indicate the U → C change at position 146.

(ii) G2 → A2. The pancreatic ribonuclease product, G-A-Cp (P7) occurs twice in the HeLa sequence but only once in chick, in which one mole appears to be replaced by A-A-Cp (P4) (Figure 4 and Table 1). On the basis of this evidence alone the G → A change could be at position 2 or 96 (Figure 1). However, the presence of T_1 products 19.1 and 20 in the chick fingerprint exclude the latter possibility (Figure 1). Product T23, which follows immediately after G2 in the HeLa sequence, was recovered in low yield in chick (Table 1). A product of composition C_{3-4}, A_4, U_3, G was present in 0.35 molar yield. There was also some streaky material near the origin of the second dimension. We suggest the following explanation for these findings:- G2 is largely replaced by A in chick 5.8 S RNA, as in turtle[12]. However, the substitution may be incomplete. Also, as in HeLa, there is heterogeneity at the 5' end with regard to the presence both of the terminal C and the 5' phosphate. Therefore in chick the nonadecanucleotide commencing at the 5' end is recovered in the form of two or three related T_1 products near the origin of the second dimension.

Nucleotides 30 - 36. The region, U-G-C-G-U-C-Gp is the only part of the mammalian sequence which is not spanned by unique T_1 or pancreatic oligonucleotides (Figure 1). We can exclude the possibility of changes affecting G residues in this region from the fact that there are no qualitative differences in the fingerprints which we have not

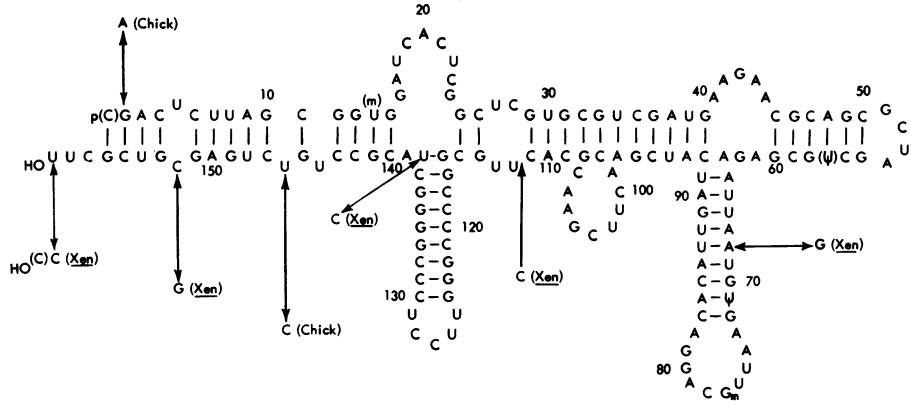

Figure 5. Proposed secondary structure of HeLa 5.8 S RNA[8,11]. The arrows indicate the position of substituted and inserted bases in chick and Xenopus 5.8 S RNAs.

already explained. We cannot rule out the possibility of changes affecting pyrimidine residues (particularly double changes such as U-G-C-G ⟷ C-G-U-G), but such changes would be improbable if the secondary structure model (Ref. 8, Figure 5) is correct.

Modified nucleotides. Um14 is incompletely methylated in vertebrate 5.8 S RNA, as discussed elsewhere[13,15]. The extent of methylation is approximately 0.2 molar in HeLa, 0.4 molar in Xenopus and within the range 0.7 - 0.9 molar in chick (Ref. 13 and this work; the precise value obtained depends on the analytical procedure used). Otherwise the modification pattern of Xenopus and chick 5.8 S RNAs is the same as that of HeLa cell 5.8 S RNA. In contrast to Ref. 8, however, we find Ψ57 to be fully modified in HeLa cells.

DISCUSSION

The sequence differences reported here comprise five transitions, one transversion (C152 HeLa → G153 Xenopus) and the presence of an extra C residue next to C 111 in the Xenopus sequence. Interestingly, the yeast 5.8 S sequence[7] contains an extra A at a corresponding position.

It is of interest to examine these differences in the light of the proposed secondary structure[8]. Substitution of G for A at position 68 in Xenopus results in replacement of an A-U base pair by a G-U pair in the middle of a relatively long stem (Figure 5). The stem contains one G-U pair in HeLa and chick, and is probably rendered somewhat unstable by the relatively long hairpin loop[8,16]. The additional G-U

pair in Xenopus would be expected to result in an overall stability of +2 Kcal/mole for this arm, as compared with +3 Kcal/mole in the mammalian sequence[8,16]. The extra C residue in Xenopus, C 112, is accommodated in a short single-stranded region of the secondary structure model; nothing can be inferred about possible tertiary interactions in this region. As already discussed, U138 (HeLa)→ C139 (Xenopus) results in replacement of a possible U-G pair by a C-G pair, with yet further stabilization of this already very stable GC-rich stem (+12 Kcal/mole instead of +10 Kcal/mole)[16].

The sequences near the 5' and 3' termini are depicted in the secondary structure model as interacting to form an imperfectly paired stem. C152 (HeLa) → G153 (Xenopus) would permit formation of a G-U pair, thus slightly stabilizing the presumed stem. U146 (HeLa) → C146 (chick) leaves an unpaired pyrimidine in either case. In chick (as well as in turtle[12]), G2 → A2 would be expected to destabilize the stem interaction near the termini. A direct chemical analysis of secondary structure in this region is in progress (Kelly, Goddard, Khan and Maden, unpublished observations).

The 3' terminal U in HeLa 5.8 S RNA is known to be in a single stranded region[11] and this is presumably also the case for the corresponding C in Xenopus. G2 → A2 and U159 → C160 are both extremely close to the actual termini of mature 5.8 S RNA. This implies that conservation of primary sequence within these regions is not critically important for correct excision of the adjacent transcribed spacer during ribosome maturation.

The chick and turtle[12] 5.8 S sequences are ∿ 99% homologous to that of mammals, the Xenopus laevis sequence is 97% homologous and the yeast sequence is 75% homologous[7]. Since the 5.8 S sequence appears to have remained very stable during vertebrate evolution it will be of interest to determine to what extent the 18 S and 28 S sequences have been conserved. Yeast 5.8 S RNA differs from that of vertebrates in the terminal sequences (near the excision sites from the transcribed spacer, mentioned above), in the numbers of nucleotides in single-stranded regions in the proposed secondary structure[8], and in the content of modified nucleotides. It is possible that sequence analysis of 5.8 S RNAs from other distantly related eukaryotes, such as Drosophila and higher plants, will be of value for extending our insight into the structure and evolution of this accessible region of the ribosomal transcription unit.

ACKNOWLEDGEMENTS

We thank Dr. P. Ford for discussion and Mrs. Mary Minto for skilled technical assistance. This work was supported by a grant from the Medical Research Council.

REFERENCES

1. Pene, J.J., Knight, E. and Darnell, J.E. (1968). J.Mol. Biol. 33, 609-623.

2. King,H.W.S. and Gould, H. (1970). J. Mol. Biol. 51, 687-702.

3. Speirs, J. and Birnstiel, M.L. (1974). J. Mol. Biol. 87, 237-256.

4. Walker, T.A. and Pace, N.R. (1977). Nucleic Acids Research 4, 595-601.

5. Maden, B.E.H. and Robertson, J.S. (1974). J. Mol. Biol. 87, 227-235.

6. Nazar, R.N., Owens, T.W., Sitz, T.O. and Busch, H. (1975). J. Biol. Chem. 250, 2475-2481.

7. Rubin, G.M. (1973). J. Biol. Chem. 248, 3860-3875.

8. Nazar, R.N., Sitz, T.O. and Busch, H. (1975). J. Biol. Chem. 250, 8591-8597.

9. Nazar, R.N., Sitz, T.O. and Busch, H. (1976). Biochemistry 15, 505-508.

10. Hampe, A., Eladari, M. and Galibert, F. (1976). Biochimie 58, 943-951.

11. Khan, M.S.N., and Maden, B.E.H. (1976). FEBS Lett. 72, 105-110.

12. Nazar, R.N. and Roy, K.L. (1976). FEBS Lett. 72, 111-116.

13. Khan, M.S.N. and Maden, B.E.H. (1976). J. Mol. Biol. 101, 235-254.

14. Brownlee, G.G. (1972) in: Determination of sequences in RNA. North-Holland/Elsevier, Amsterdam.

15. Nazar, R.N., Sitz, T.O. and Busch, H. (1975). FEBS Lett. 59, 83-87.

16. Tinoco, I., Uhlenbecke, O.C. and Levene, M.D. (1971). Nature 230, 362-367.