Biochem. J. (1989) 258, 49–56 (Printed in Great Britain)

28 S ribosomal RNA in vertebrates

Locations of large-scale features revealed by electron microscopy in relation to other features of the sequences

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

The 28 S rRNA from several vetebrate species, when examined by electron microscopy, is seen to contain regions of extensive secondary structure, as first reported for HeLa-cell 28 S rRNA by Wellauer & Dawid [(1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 2827–2831]. Here we correlate the locations of these regions, determined from the electron-microscopic data, with the primary structure of 28 S rRNA from human, mouse and *Xenopus laevis* determined by sequence analysis of rDNA. The secondary-structure features observed by electron microscopy correspond closely to phylogenetically variable G+C-rich regions that largely comprise the eukaryotic expansion segments in these three species. In most if not all cases the features can be identified with long G+C-rich helices deduced from sequence data. Correlations are given between the locations of the secondary-structure features and several 'landmark' restriction sites in 28 S rDNA. By correlating the locations of the rRNA methyl groups reported elsewhere [Maden (1988) J. Mol. Biol. **201**, 289–314] with the present findings it is concluded that the rRNA secondary-structure features revealed by electron microscopy are largely or wholly unmethylated.

INTRODUCTION

The 28 S rRNA in vertebrates contains large-scale features of secondary structure that are revealed by electron microscopy (Wellauer & Dawid, 1973, 1974; Wellauer *et al.*, 1974; Schibler *et al.*, 1975). These features were first described in HeLa (human)-cell rRNA; they were observed in 28 S rRNA that was prepared for microscopy from 80% formamide; they occur at characteristic locations and possess characteristic morphologies. The two largest features are a triple loop towards the 5′ end of the molecule and a double loop centred about 36% of the distance from the 3′ end. [The 5′→3′ polarities of the molecules were incorrectly assigned in the above references but were subsequently corrected by Dawid & Wellauer (1976). The corrected polarities are referred to throughout the present paper.] Similar features occur in 28 S rRNA from *Xenopus* (Wellauer & Dawid, 1974), mouse L cells (Wellauer *et al.*, 1974) and other vertebrate species (Schibler *et al.*, 1975). Some of the features, in particular the triple loop and the double loop, show phyletic size increase in the order *Xenopus* < mouse < human (Wellauer *et al.*, 1974). Various lines of evidence indicate that the features are G+C-rich, and suggest that they are based upon (partial or exact) sequence complementarity between adjacent regions in RNA, giving rise to extensive double-helical hairpins (Wellauer & Dawid, 1973, 1974; Schibler *et al.*, 1975).

Nucleotide sequence data from genes encoding the large (23–28 S) rRNA from a wide range of sources have shown that 28 S rRNA in eukaryotes possesses a 'mosaic' arrangement of conserved and variable sequence elements (Gorski *et al.*, 1987). The conserved elements give rise to a core of secondary-structure features whose general architecture is the same across a wide range of prokaryotes and eukaryotes. The variable elements, also called expansion segments (Clark *et al.*, 1984) or D (divergent) regions (Hassouna *et al.*, 1984), are present in eukaryotes but not in prokaryotes. There are about 11 or 12 such segments; some are quite short (about 30 bases), but others are much longer (up to several 100 bases) and show phyletic size increase from yeast to lower and higher vertebrates. In vertebrates the larger variable regions are G+C-rich and are capable of forming extensive but imperfect helical hairpins.

Thus the properties of the larger variable regions, namely phyletic size increase, G+C-richness in vertebrates and the ability to form extensive hairpin helices, appear to correspond to the properties of the large-scale features revealed by electron microscopy. These apparent relationships have been commented upon (Hadjiolov *et al.*, 1984; Michot *et al.*, 1984; Clark *et al.*, 1984; Gorski *et al.*, 1987) but have not been systematically correlated. In the present paper we describe the results of such a correlation for human, mouse and *Xenopus* 28 S rRNA, three sources that have been well characterized by electron microscopy and sequencing. The results show that the electron-microscopic features correspond closely to variable regions identified by sequencing, and generally to long helices deduced from the sequences. Two further correlations are also made: the co-ordinates of some potentially useful restriction sites in rDNA are given in relation to the electron-microscopic features, and comparison of the present findings with those on the locations of the rRNA methyl groups indicates that the electron-microscopic features are largely or wholly unmethylated.

EXPERIMENTAL AND RESULTS

Sequence data: conserved and variable regions

We summarize the relevant sequence data first, since these data delineate the conserved (C) and variable (V) regions of 28 S rRNA, with minor qualifications as

* To whom correspondence should be addressed.

discussed below. The delineations have been progressively refined with successive additions to the vertebrate 28 S rRNA sequence data base and with associated refinements in secondary-structure modelling. The definitions of C and V regions adopted here are derived from those of Gorski *et al.* (1987) for human 28 S rRNA. Their criteria were: C regions differ in length by less than 12% between vertebrate 28 S rRNA and *Escherichia coli* 23 S rRNA; V regions differ in length by more than 20% between vertebrates and *E. coli*. A few of the human 28 S-rRNA V regions first identified in this way have been adjusted by a few nucleotide residues here to give better fit with helical features also depicted by Gorski *et al.* (1987). The resulting identifications (co-ordinates and lengths) of the human V regions are summarized in Table 1. Table 1 also gives the co-ordinates and lengths of the corresponding V regions in the aligned mouse and *X. laevis* sequences. [These sequences are aligned with the human 28 S-rRNA sequence in Fig. 1 of Gorski *et al.* (1987).] The V regions defined in Table 1 comprise some 48% of the human 28 S-rRNA sequence, 44% of the mouse sequence and 36% of the *X. laevis* sequence. Regions V2a, V5, V8 and V11 encompass most of the tracts of phyletic size increase from *Xenopus* to mouse and human 28 S rRNA. Regions V2b and V7 are longer in *X. laevis* than in human 28 S rRNA (as signified by negative numbers in the 'human minus *X. laevis*' column). The sum of the lengths of the C regions is almost constant for the three species, and the small differences may become smaller still when minor uncertainties in the respective published sequences (particularly the *Xenopus* sequence reported by Ware *et al.*, 1983) have been resolved.

In making the above compilation it was necessary for comparative purposes to use uniform definitions of the co-ordinates of the V-region boundaries, by designating a homologous nucleotide in each of the three aligned sequences as the first or last nucleotide residue in a given V region. For cross-reference the previous nomenclatures and co-ordinates of the (generally corresponding) D regions (Hassouna *et al.*, 1984) and expansion segments (Clark *et al.*, 1984) are also given in Table 1. The unresolved differences are insufficient to affect the main conclusions of the present paper, with one relatively minor exception that is discussed below.

Electron microscopy: secondary-structure features

Electron micrographs of 28 S rRNA (and also of 18 S rRNA and ribosomal precursor RNA) were published for HeLa (human) cells, *Xenopus laevis* and mouse as cited above. Each report showed typical micrographs followed by linear tracings of about 40 molecules of each major class (28 S, 32 S etc.) The linear tracings comprise thin and thick sections; the thin sections represent regions where no secondary structure was seen; the thick sections depict secondary structure features in the form of 'out and back' distances along the hairpins, i.e. the lengths of single-stranded RNA regions that make up the hairpins.

The assemblies of tracings of each class of molecule reveal characteristic features but with some degree of variation between individual molecules; thus no single molecule can be chosen as representative, and averaging of the information from many molecules is necessary. The following comments summarize the main features (see also Fig. 1).

Every 28 S rRNA molecule contains a triple loop or related feature near the 5′ end and a double loop centred at about 36% of the distance from the 3′ end (in the corrected 5′→3′ orientation). Other, smaller, features are present in tracings of some molecules but not in others. After treatment of HeLa-cell 28 S rRNA with 3% formaldehyde the triple loop became resolved in many molecules into three separate components. After treatment with 10% formaldehyde many of the secondary-structure features disappeared but the double loop persisted, being resolved into its two components, and implying particular stability of secondary structure in this region.

The measurements on the 28 S rRNA tracings are summarized in Table 2. The features are designated *a* to *g*, where *a* is the feature nearest the 5′ end, *b* is the triple loop and *e* the double loop. For HeLa-cell 28 S rRNA the amount of secondary structure in the tracings was considerable, especially in the vicinity of the triple loop; the measurements were more readily made on 28 S rRNA that had been treated with 3% formaldehyde, in which the secondary structure was largely retained but was more reproducible between tracings than in the non-formaldehyde-treated material. In mouse and *Xenopus* 28 S rRNA the secondary structure was slightly less than in HeLa-cell 28 S rRNA, and these measurements were made from tracings of rRNA that had not been treated with formaldehyde.

Relationship between electron-microscopic features and V regions

Fig. 1 shows that the averaged locations of the electron-microscopic features correspond closely to those of the V regions defined by sequencing. The triple and double loops are represented as their measured lengths with some internal details, as discussed later below. The other features (*a*, *c*, *d*, *f* and *g*) were present in some but not all tracings and were generally short; they are arbitrarily shown as encompassing 100 nucleotide residues each.

Feature *a*. This short feature is present in most molecules and is centred approximately 200 nucleotide residues from the 5′ end. It is slightly to the right of region V1 defined by Gorski *et al.* (1987) and region ES1 defined by Clark *et al.* (1984), but corresponds closely to region D1 defined by Hassouna *et al.* (1984) (see Table 1).

Feature *b*. This feature, which typically appears as a triple loop, corresponds almost exactly to the large variable region V2a. The approximate locations of the tips of the three component loops are shown as triangles, designated *b*1–*b*3. *Xenopus* showed an extra loop in many molecules (feature *b*′); this appears to correspond to region V2b, which is slightly expanded in *X. laevis* relative to mammals (Table 1).

Feature *c*. This is present in most molecules of HeLa-cell and mouse 28 S rRNA but is not evident in *X. laevis*. It corresponds in location to variable region V5, which is expanded in mammals but not in *Xenopus* (Table 1).

Feature *d*. This occurs in most molecules of all three species. It appears to correspond to regions V6 and V7 in *Xenopus* but to be shifted slightly to the right of regions

Table 1. Summary of variable (V) regions in human, mouse and *X. laevis* 28 S rRNA

The data are from the following sources. The human 28 S rRNA sequence with comparative sequence alignments and definitions of C and V regions in the three aligned sequences are from Gorski *et al.* (1987), with minor modifications as detailed in a few instances below. The mouse 28 S rRNA sequence was determined by Hassouna *et al.* (1984), who also defined the original D regions listed towards the right of the Table. The *X. laevis* 28 S rRNA sequence was determined by Ware *et al.* (1983), and the expansion segments (ES in right-hand column) were defined by Clark *et al.* (1984). The boundaries of human region V1, given by Gorski *et al.* as 116–164, have been adjusted here to give a better fit with their helix V1 and also with the corresponding feature defined by Clark *et al.* (1984). Human region V2 was given as a single large region in Fig. 1 of Gorski *et al.* (1987), but is interspersed with some conserved helices in their secondary-structure model. The subdivision into regions V2a and V2b shown here reflects this pattern and is in approximate agreement with the previous mouse and *Xenopus* designation. Human regions V3, V5 and V9 are represented by much shorter tracts in *Xenopus*, which were not originally recognized as expansion segments. The original designation of region D6 in mouse was somewhat longer at the 5′ end than the designation of the corresponding region V5 given here by alignment with human region V5. The original region D11 defined by Hassouna *et al.* (1984) is not listed here as it is significantly expanded only in *Physarum* and not in vertebrates. The numbering schemes may require minor revision when a few uncertainties in the sequences have been resolved and a systematic method is agreed for numbering sections within some variable regions that show intraspecies heterogeneity among genes (Gonzalez *et al.*, 1985; Maden *et al.*, 1987).

	Human		Mouse		*X. laevis*				Previous nomenclature and co-ordinates			
									Mouse		*X. laevis*	
Variable region	Co-ordinates	Length	Co-ordinates	Length	Co-ordinates	Length	Human minus mouse	Human minus *X. laevis*				
V1	115		115		115				D1	122	ES1	113
	153	39	156	42	153	39	−3	0		277		156
V2a	455		464		459				D2	436	ES2	483
	1290	836	1123	660	949	491	176	345		1124		976
V2b	1335		1168		994				D3	1166	ES3	1052
	1480	146	1313	146	1164	171	0	−25		1315		1162
V3	1674		1507		1359				D4	1507	–	
	1700	27	1522	16	1366	8	11	19		1525		
V4	1790		1612		1457				D5	1606	ES4	1451
	1819	30	1642	31	1489	33	−1	−3		1635		1477
V5	2064		1886		1737				D6	1879	–	
	2243	180	2021	136	1760	24	44	156		2032		
V6	2424		2201		1940				D7a	2207	ES5	1963
	2488	65	2265	65	2004	65	0	0		2263		2000
V7	2530		2307		2046				D7b	2302	ES6	2045
	2553	24	2327	21	2105	60	3	−36		2342		2126
V8	2877		2652		2436				D8	2648	ES7	2458
	3586	710	3272	621	2779	344	89	366		3259		2754
V9	3943		3629		3136				D9	3629	–	
	4005	63	3686	58	3162	27	5	36		3658		
V10	4048		3729		3205				D10	3727	ES8	3214
	4113	66	3804	76	3275	71	−10	−5		3789		3278
V11	4696		4387		3864				D12	4379	ES9	3871
	4908	213	4595	209	4001	138	4	75		4619		4004
Total V		2399		2081		1471	318	928				
Total sequence		5025		4712		4110	313	915				
Total C		2626		2631		2639	−5	−13				
$\frac{V}{Total}$ (%)		48		44		36						

Table 2. Co-ordinates of electron-microscopic features in human, mouse and *X. laevis* 28 S rRNA

The human measurements are from Fig. 5(*a*) of Wellauer & Dawid (1973). The mouse measurements are from Fig. 3 of Wellauer *et al.* (1974). The *X. laevis* measurements are from Fig. 3 of Wellauer & Dawid (1974). The 28 S sections of the Figures were enlarged to 150 mm for making the measurements. For the smaller features the distance of the centre point of each feature along each molecule was measured. For the triple loop (feature *b*) the measurements were of the start and end of the complete feature and the approximate midpoints of the individual loop tracings (*b*1, *b*2 and *b*3) corresponding to the tips of the loops. For the double loop (feature *e*) the measurements were of the start and end and the junction between the two loops, delineated by a vertical bar in the original tracings. Features from the different molecules were grouped according to their locations; their average distances from the 5′ end, ±the standard deviations, were converted into numbers of nucleotide residues on the basis of complete sequence lengths of 5025 nucleotide residues (human), 4712 nucleotide residues (mouse) and 4110 nucleotide residues (*X. laevis*). The numbers in parentheses in the Table are the numers of molecules in which the given feature was present (or sufficiently clearly recognizable) in the tracings. The total numbers of molecules available for measurement were as follows: Human 28 S rRNA treated with 3% formaldehyde, 40 molecules; mouse 28 S rRNA, no formaldehyde treatment, 44 molecules; *X. laevis* 28 S rRNA, no formaldehyde treatment, 40 molecules. For further details see the text.

Feature	Human	Mouse	*X. laevis*	V correlation
a	201±40 (30)	220±34 (42)	214±44 (38)	V1/D1
b start	436±97 (39)	471±57 (44)	411±60 (38)	V2a
*b*1	536±60 (34)	587±57 (44)	501±60 (36)	V2a
*b*2	837±64 (38)	826±53 (44)	712±58 (39)	V2a
*b*3	1122±67 (32)	1068±57 (44)	917±66 (36)	V2a
b end	1206±114 (39)	1194±60 (44)	1027±77 (38)	V2a
b′	–	–	1134±68 (32)	V2b
c	2164±67 (29)	1979±79 (40)	–	V5
d	2633±97 (19)	2387±80 (24)	2048±181 (28)	V7?
e start	2848±94 (40)	2670±63 (44)	2383±71 (40)	V8
e junction	3216±87 (39)	3031±64 (44)	2641±66 (40)	V8
e end	3585±97 (40)	3280±69 (44)	2781±69 (40)	V8
f	4060±167 (6)	3748±134 (40)	3222±82 (14)	V10
g	4737±40 (14)	4483±64 (38)	3915±36 (22)	V11

V6 and 7 in mammals. Region V7 is expanded in *Xenopus* relative to mammals (Table 1).

Feature *e*. This is the double loop that is such a prominent feature in micrographs of vertebrate 28 S rRNA. It corresponds closely to the large variable region V8 in all three species. It is approximately symmetrical in HeLa-cell 28 S rRNA, slightly asymmetrical in mouse, and smaller and asymmetrical in *X. laevis*. The black triangle indicates the junction between the two loops and hence the degree of asymmetry between the loops.

Feature *f*. This feature is evident in some molecules and corresponds approximately to variable regions V9 and V10.

Feature *g*. This feature is evident near the 3′ end of many or most molecules and coincides closely with variable region V11.

Molecular basis of the electron-microscopic features

Can the electron-microscopic features be correlated with helical elements deduced from sequencing? Detailed discussion is rendered somewhat complicated by differences between secondary-structure models as well as by the large sizes of the molecules. Following a comment on the secondary structure models we discuss in outline the relationship between electron-microscopic morphology and sequence-deduced helices.

The first vertebrate 28 S rRNA secondary-structure model to be published was for mouse (Michot *et al.*, 1984), followed shortly by *Xenopus* (Clark *et al.*, 1984) and, later, human (Gorski *et al.*, 1987). All three models are in fairly good agreement in the internal architecture of several conserved regions, but the mouse and human models differ in some large-scale respects from the *Xenopus* model, notably in the positioning of some long-range interactions. The mouse and human models are in approximate agreement in most of the long-range interactions in the conserved regions, particularly in the domain boundary regions, but there are several internal differences between the arrangements of V (or D) regions in the mouse and human models, even where the sequences are fairly highly conserved among vertebrates. It seems likely that tracts of V regions that are highly conserved among vertebrates do in fact conform to a common secondary structure in the different species. We therefore refer to the models eponymously (Michot, Clark and Gorski) rather than by species.

A survey of the models suggests that the Michot model can be related straightforwardly to the electron-microscopic features. The Gorski model can also be related to several electron-microscopic features, with differences from the Michot model as outlined below. The Clark model is less easy to relate in detail to the major electron-microscopic features. Table 3 shows the co-ordinates (first and last nucleotide residues) of helical features in the Michot model whose locations correspond approximately to the electron-microscopic features. The nucleotides are listed for the mouse sequence and for the homologous points in the aligned human and *X. laevis* sequences.

Electron-microscopic feature *a*, as noted above, maps just to the right of the narrowly defined region V1 but corresponds well to the larger region D1 defined by Hassouna *et al.* (1984). This whole D1 region is G+C-

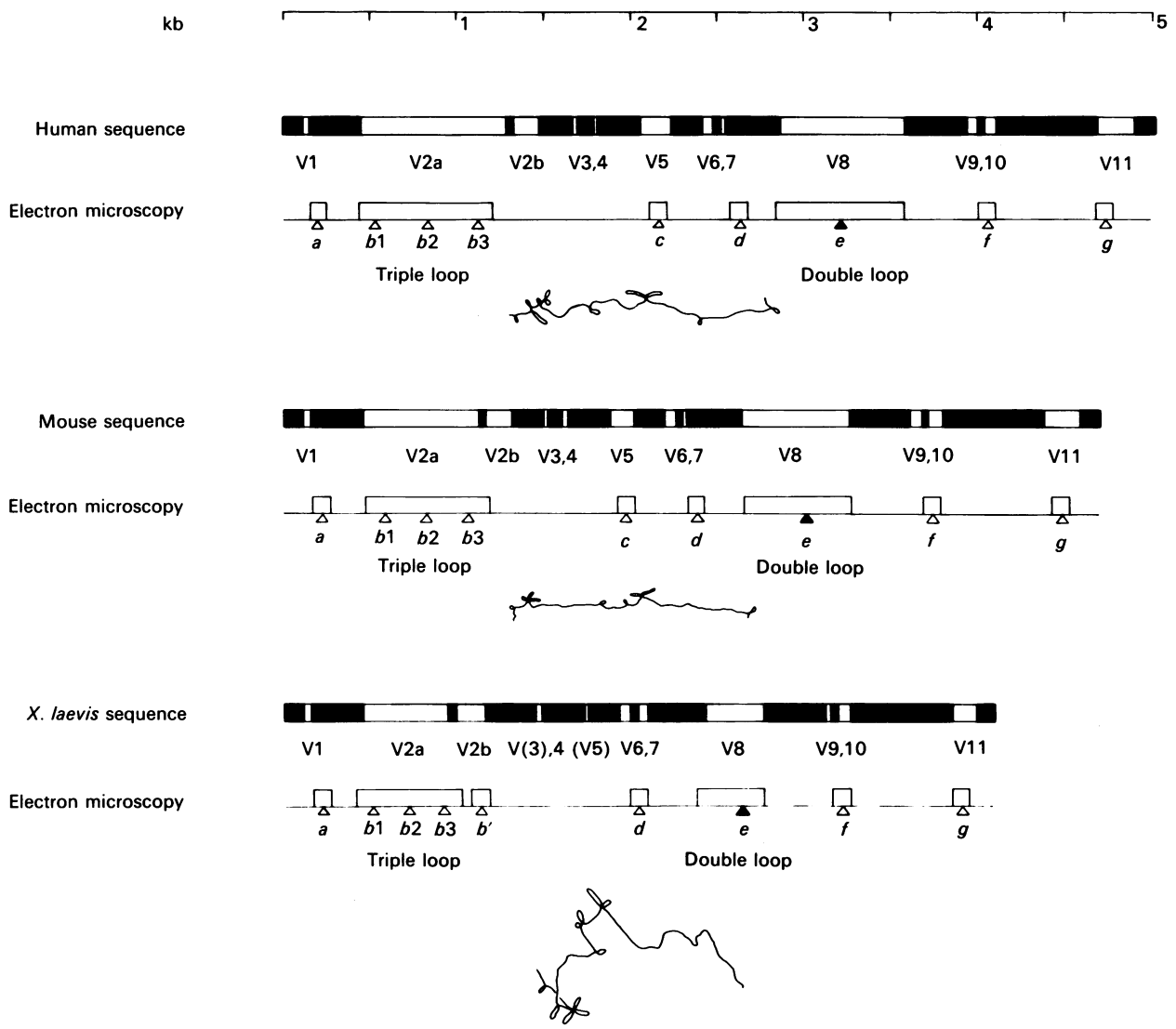

Fig. 1. Distribution of conserved and variable regions and of electron-microscopic features in human, mouse and *X. laevis* 28 S rRNA

The variable regions deduced from sequencing are shown as white tracts, V1 etc., in the upper of each pair of lines; the conserved regions are in black. The data are from Table 1. The electron-microscopic features are shown as boxes corresponding to their (mean) locations and approximate lengths (see Table 2 and the text). The white triangles define the centre points of the smaller features *a*, *c*, *d*, *f* and *g* and the approximate locations of the tips of the three loops in feature *b*. The black triangle indicates the junction between the two loops in feature *e*. Underneath each linear electron-microscopic diagram is shown a representative 28 S rRNA electron-microscopic tracing on a smaller scale. These tracings are redrawn in the correct 5′→3′ orientation from the 28 S segments of 45 S (or 40 S) ribosomal precursor RNA shown in Wellauer & Dawid (1973) (human), Wellauer *et al.* (1974) (mouse) and Wellauer & Dawid (1974) (*X. laevis*).

rich (80%). The tract indicated in Table 3 is arranged in the Michot model as a single long helix with some bulges and a small branch at the top. The arrangement is fairly similar, but not identical, in the Gorski model.

Feature *b*, the triple-loop region, is well explained in the Michot model by three long G+C-rich helical systems (Table 3), which comprise most of their region D2 (V2a in Fig. 1). The first two helices as listed in Table 3 are unbranched but contain some bulges; the third helix is branched near the base, with one long arm and a cluster of shorter branches. The Gorski arrangement of region V2 also contains three quite long helices; however, the tracts that comprise the proximal regions of the long helices in the Michot model are arranged instead in a nest of smaller helices in the Gorski model.

Electron-microscopic feature *c*, as noted above, corresponds to region V5 (D6), which is expanded in mammals but not in *Xenopus*. The region is depicted as a helical system with bulges and a branch in the Michot and Gorski models; the details differ, partly reflecting species differences between the sequences. [This region also shows intraspecies microheterogeneity in man (Gonzalez *et al.*, 1985; Maden *et al.*, 1987).]

Electron-microscopic feature *d* is not well explained by region V6 or V7 (D7a or D7b) in the secondary-structure models. However, a tract just to the right of these regions

Table 3. Summary of co-ordinates of suggested helices corresponding to electron-microscopic features of human, mouse and *X. laevis* 28 S rRNA

The helical features are as depicted in the mouse 28 S rRNA model proposed by Michot *et al.* (1984). The co-ordinates are for the first and last nucleotide residues in each helical element for which a correlation with an electron-microscopic feature is suggested. The actual bases are given for ease of cross-reference with the original model. The corresponding co-ordinates and bases are given for the human and *X. laevis* sequences, for which, in general, similar helical elements can be constructed. However, region *c* in *Xenopus* (in parentheses) is short and does not correspond to an electron-microscopic feature. The region shown for helix *g* in *Xenopus* (also in parentheses) is by homology with the mammalian sequences. However, this tract on its own is too short to generate an electron-microscopic feature in *Xenopus* and must be augmented by adjacent material to account for the observed feature. For further details see the text.

Electron-microscopic feature	Human		Mouse		*X. laevis*	
	Co-ordinates	Length	Co-ordinates	Length	Co-ordinates	Length
a	G157 C266	110	G160 C275	116	G157 U272	116
*b*1 (triple loop)	C463 G674	212	C472 G688	217	C467 G588?	111?
*b*2 (triple loop)	C701 G939	239	C719 G851	133	C614 G725	112
*b*3 (triple loop)	G945 C1269	325	G857 C1102	246	G731 C928	198
c	C2056 G2254	199	C1879 G2032	154	(C1729) (G1772)	(44)
d	U2569 G2730	162	U2343 G2505	163	U2121 G2288	168
*e*1 (double loop)	G2889 C3249	361	G2664 C3023	361	G2448 C2647	200
*e*2 (double loop)	G3253 C3551	299	G3027 C3237	211	G2651 C2744	94
f	C4046 G4128	83	C3727 G3819	93	C3203 G3290	88
g	G4724 U4824	101	G4416 U4510	95	(G3893) (C3920)	(38)

is arranged as a long helical system in the Michot model (Table 3) and, with some variations, in the Gorski model. This tract, like the extended D1 region mentioned above, is G+C-rich and differs at many points between vertebrates and yeast.

Feature *e*, the double loop, is represented in the Michot model as two long highly G+C-rich helices (Table 3) that spring directly from a supporting helix, the whole structure resembling a 'T' with very long horizontal arms. In the Gorski model the proximal part of the left arm of the Michot model is arranged instead into a nest of smaller helices but the distal part of this long arm is retained, comprising some 90 base-pairs interspersed with bulges. There are also some differences between the two models in the basal region supporting the right long arm; these differences arise from a different base-pairing arrangement in the supporting vertical limb of the 'T'. In the Gorski model the 3′ strand of the limb is 'slipped' upwards so as to feed some 17 extra bases back into the right arm of the double-loop system with respect to the Michot model. It is not easy to choose between some of the details of the two models. However, for all three species the two helices in the Michot arrangement are in good accord with the electron-microscopic morphology: the human helices account for an almost symmetrical double loop, the mouse helices for a slightly asymmetric double loop and the *Xenopus* helices for a smaller and asymmetric double loop (see the respective sequence lengths in Table 3). When *Xenopus* is fitted to the Gorski model much of the material is taken up instead into the nest of small loops, and too little remains to construct a convincing double loop. This consideration would appear to favour the main features of the Michot model in this region.

The overall validity of this general correlation in the double-loop region is further supported as follows. The junction between the two long helices in the Michot model coincides with a conserved *Pvu*II site (CAGCTG) in rDNA. The co-ordinates of this site in the human, mouse and *Xenopus* sequences are in remarkable agreement with the co-ordinates of the junction of the two arms of the double loop as measured from the electron-microscopic tracings. This and other useful correlations between the electron-microscopic features and restriction sites in rDNA are summarized in Table 4, below.

Feature *f* can best be correlated with region V10 (D10), which is represented as a branched G+C-rich helical arrangement in both the Michot and the Gorski models.

Feature *g* is well explained in mammals by a relatively long helix in the Michot model (Table 3). Much of this helix is also incorporated into the Gorski model, but the

Table 4. Restriction sites in rDNA and electron-microscopic features of human, mouse and *X. laevis* 28 S rRNA

These hexanucleotide restriction sites are potentially useful for studies on the V regions and associated electron-microscopic features. They are the only sites for these enzymes in the respective 28 S rDNA sequences. Each site is designated by its first nucleotide residue. Note (a) This *Spl*I site is on the 5′ boundary of the putative helix corresponding to electron-microscopic feature *a* (Table 3) and is also the nearest generally useful hexanucleotide restriction site to the 5′ end of the 28 S-rDNA sequence. Note (b) This *Bam*HI site is in region V2b (D3); region V2a with flanking sequences can be excised by an *Spl*I/*Bam*HI double digest. Note (c) These two sites are highly conserved, being present in 28 S rDNA from *Saccharomyces* to man. They encompass region V5 and allow excision and preliminary sizing. Note (d) The first *Sst*I site is not evident in the *X. laevis* published sequence owing to an extra, lower-case, 't'; however, the site is experimentally demonstrable. The two *Sst*I sites encompass region V8 (see the text). Note (e) The second *Bam*HI site, although in a structurally conserved region of the sequence, is absent from *Xenopus* owing to a local base change. The site comes shortly before region V8. Note (f) The first *Pvu*II site is conserved in the rDNA of all three species; the second site occurs in *Xenopus* but not in mammals owing to a base change. The first site is at the junction of the two loops in the double-loop region of the Michot secondary-structure model, and correlates as shown in the Table with the electron-microscopic double-loop junction. Note (g) The *Eco*RI site to the left of region V11 is highly conserved. The second *Nco*I site, also to the left of region V11, and the second *Spl*I site, to the right of region V11, occur in mammals but not in *X. laevis*.

Restriction enzyme	Restriction site			Notes
	Human	Mouse	*X. laevis*	
*Spl*I	152	155	152	(a)
*Bam*HI	1405	1237	1063	(b)
*Nco*I	1955	1778	1628	(c)
*Bgl*II	2324	2101	1840	(c)
*Sst*I	2701	2476	2258	(d)
(*Bam*HI)	2839	2614	–	(e)
*Pvu*II	3246	3020	2644	(f)
Double-loop junction	3216±87	3031±64	2641±66	(f)
(*Pvu*II)	–	–	2760	(f)
*Sst*I	4198	3889	3360	(d)
*Eco*RI	4438	4129	3605	(g)
(*Nco*I)	4542	4233	–	(g)
(*Spl*I)	4959	4647	–	(g)

basal arrangement differs. The region comprising this helix is expanded in mammals relative to *Xenopus*, and flanking material in the V11 (D12, ES9) region in *Xenopus* must be invoked to contribute to the observable electron-microscopic feature.

In summary, convincing explanations can be offered for most of the electron-microscopic features. and particularly the triple and double loops, on the basis of G+C-rich helical elements in the Michot model, and perhaps slightly less satisfactorily in the Gorski model. It is interesting that the distal regions of the longest helices are pyrimidine-rich in the ascending (outwards) strand, purine-rich in the descending (return) strand, and are largely made up of simple sequence tracts (see the Discussion section).

Restriction sites and the electron-microscopic features

In addition to the conserved *Pvu*II site, mentioned above, several other restriction sites that are useful in relation to the electron-microscopic features are listed in Table 4. The sites have been experimentally demonstrated in human and *Xenopus* 28 S rDNA (Maden *et al.*, 1987; B. E. H. Maden, unpublished work). The two *Sst*I sites are particularly useful. They encompass regions V8–V10, and, since most of the phyletic size variation in this tract occurs in region V8 (the electron-microscopic double loop), sizing of the *Sst*I fragment affords a rapid preliminary screening method for detecting size differences in this region. We have detected a marginal size difference of 10–20 bp between this region of *X. laevis* rDNA and that of *X. borealis* rDNA in this way (B. E. H. Maden, Y. Akhtar & N. Ramasinghe, unpublished work).

rRNA methylation is not associated with the electron-microscopic features

A relatively high content of methyl groups, particularly 2′-*O*-methyl groups, is a characteristic of eukaryotic rRNA. Are the methyl groups associated with eukaryotic specific tracts in rRNA and perhaps with the electron-microscopic features? Data on the methyl group locations indicate that this is not so. Some 58 methyl groups have been located in human and *Xenopus* 28 S rRNA and all of these are in the conserved core of the molecule (Maden, 1988). Indirect evidence indicates that the remaining few (about 10–12) unlocated methyl groups are also in the conserved core (also summarized by Maden, 1988). Similarly the 18 S rRNA methyl groups are concentrated in conserved core regions (Salim & Maden, 1981; Maden, 1986). The molecular basis of recognition for methylation during rRNA maturation appears to be intimately associated with features in the core structure and not with the eukaryotic V regions or their associated electron-microscopic features.

DISCUSSION

The main conclusions of this paper are, first, that the electron-microscopic features of vertebrate 28 S rRNA correlate closely with V regions defined by sequencing, and, secondly, that the electron-microscopic morphology arises from long G+C-rich helical arms that can be deduced from sequence data. The correlations are particularly clear in the triple-loop and double-loop regions, V2a and V8.

Several tracts of repetitive simple sequences occur in the larger V regions, namely V2a, V5 and V8. Intraspecies heterogeneity has been found in human regions V5 and V8 within or in the vicinity of simple sequence tracts in the helical arms (Gonzalez *et al.*, 1985; Maden *et al.*, 1987). Typically the heterogeneities affect the precise numbers of repeats of short motifs of a few nucleotides. The heterogeneities are thought to arise by unequal recombination (Gorski *et al.*, 1987; Maden *et al.*, 1987), as first postulated in general terms by Smith (1976), and they are likely to be foci of particularly rapid evolutionary change. Thus the phylogenetic variation in large-scale features that were first detected by electron microscopy arises from rapid evolution in simple sequence 'hot spots' in the long helical arms.

The transcribed spacers in vertebrates are also G+C-rich with extensive secondary structure, as revealed by electron microscopy (Wellauer & Dawid, 1973, 1974; Wellauer *et al.*, 1974; Schibler *et al.*, 1975). Qualitatively similar evolutionary change to that in the 28 S rRNA 'hot spots' occurs extensively in the transcribed spacers, which in *Xenopus* show substantial intraspecies microheterogeneity (Stewart *et al.*, 1983) and major interspecies divergence that appears to have arisen by repeated insertions and deletions (Furlong & Maden, 1983; Furlong *et al.*, 1983). Presumably the detailed sequences, both in the most variable parts of 28 S rRNA and in large tracts of the transcribed spacers, are not critical to the respective functions.

(No studies have yet been reported on whether the V regions in 28 S rRNA of mouse and *Xenopus* harbour microheterogeneity. In the light of the human findings some degree of microheterogeneity may be expected.)

It is interesting to recall that the electron-microscopic features and the rRNA methylation patterns (Maden *et al.*, 1972; Maden & Salim, 1974) both played key roles in the early description of the vertebrate rRNA maturation pathway. However, the recent work based on the complete rRNA sequences shows that the electron-microscopic features and rRNA methylation relate to very different structural regions and molecular processes. The electron-microscopic features comprise rapidly evolving eukaryotic expansion segments (and transcribed spacers) that provide models for studying mechanisms of evolutionary change. The rRNA methyl groups occur in conserved core regions that are likely to be intimately involved in the working machinery of the ribosome.

We thank Elaine Mayers for help with the measurements, Julie McGreavey for typing the manuscript and Pauline Dickinson for help with drawing Fig. 1. This work was supported by the Wellcome Trust.

Received 23 June 1988/22 August 1988; accepted 31 August 1988

REFERENCES

Clark, C. G., Tague, B. W., Ware, V. C. & Gerbi, S. A. (1984) Nucleic Acids Res. **12**, 6197–6220
Dawid, I. B. & Wellauer, P. K. (1976) Cell **8**, 443–448
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., 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
Hadjiolov, A. A., Georgiev, O. I., Nasikov, V. V. & Yavachev, L. P. (1984) Nucleic Acids Res. **12**, 3677–3693
Hassouna, N., Michot, B. & Bachellerie, J. P. (1984) Nucleic Acids Res. **12**, 3563–3583
Maden, B. E. H. (1986) J. Mol. Biol. **189**, 681–699
Maden, B. E. H. (1988) J. Mol. Biol. **201**, 289–314
Maden, B. E. H. & Salim, M. (1974) J. Mol. Biol. **88**, 133–164
Maden, B. E. H., Salim, M. & Summers, D. F. (1972) Nature (London) New Biol. **237**, 5–9
Maden, B. E. H., Dent, C. L., Farrell, T. E., Garde, J., McCallum, F. S. & Wakeman, J. A. (1987) Biochem. J. **246**, 519–527
Michot, B., Hassouna, N. & Bachellerie, J. P. (1984) Nucleic Acids Res. **12**, 4259–4279
Salim, M. & Maden, B. E. H. (1981) Nature (London) **291**, 205–208
Schibler, U., Wyler, T. & Hagenbuchle, O. (1975) J. Mol. Biol. **94**, 503–517
Smith, G. A. (1976) Science **191**, 528–535
Stewart, M. A., Hall, L. M. C. & Maden, B. E. H. (1983) Nucleic Acids Res. **11**, 629–646
Ware, V. C., Tague, B. W., Clark, G. C., Gourse, R. L., Brand, R. C. & Gerbi, S. A. (1983) Nucleic Acids Res. **11**, 7795–7817
Wellauer, P. K. & Dawid, I. B. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 2827–2831
Wellauer, P. K. & Dawid, I. B. (1974) J. Mol. Biol. **89**, 379–395
Wellauer, P. K., Dawid, I. B., Kelley, D. E. & Perry, R. P. (1974) J. Mol. Biol. **89**, 397–407