Volume 232, number 1, 115–120 FEB 05836 May 1988

Primary and secondary structure of the 18 S ribosomal RNA of the insect species *Tenebrio molitor*

Lydia Hendriks, Raymond De Baere, Christine Van Broeckhoven and Rupert De Wachter

Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

Received 29 February 1988

The sequence of the 18 S rRNA of *Tenebrio molitor* is reported. A detailed secondary structure model for eukaryotic small subunit rRNAs is proposed. The model comprises 48 universal helices that eukaryotic and prokaryotic small subunit rRNAs have in common, plus a number of helices in areas of variable secondary structure. For the central area of the model, an alternative structure is possible, applicable only to eukaryotic small subunit rRNAs. Possibly, small subunit rRNA switched to this alternative conformation after the eukaryotic branch had been established in evolution. Another possibility is that the two conformers represent a dynamic structural switch functioning during the translational activity of the eukaryotic ribosome.

small ribosomal subunit RNA; 18 S rRNA; Nucleotide sequence; Secondary structure; (*Tenebrio molitor*)

1. INTRODUCTION

Up to now, the sequences of about 100 small subunit ribosomal RNA (srRNA) molecules have been published, 57 of which have been aligned in a sequence compilation [1]. The Metazoa are represented in this collection by eight different species, of which there are six vertebrates and only two invertebrates, viz. *Artemia salina* (phylum Arthropoda, class Crustacea) [2] and *Caenorhabditis elegans* (phylum Nemathelminthes, class Nematoda) [3]. Here, we report the primary structure of the srRNA from another Arthropod, *Tenebrio molitor* (class Insecta).

Proceeding from the primary structure of a ribosomal RNA as determined in different species, a secondary structure can be derived by a comparative approach. This method has proven to be a powerful tool for establishing a secondary structure model for eubacterial srRNA [4,5]. Eukaryotic srRNA molecules are mostly larger than their prokaryotic equivalents and therefore comprise larger and/or extra secondary structural elements. Several authors [1–11] have proposed secondary structure models for srRNAs. Some of these models [5,6,8,10] imply that the molecule is folded essentially in the same way in prokaryotes and eukaryotes, while others [1–3,7,9] assume a different folding in certain areas. In addition, these models must account for variability in primary and secondary structure in several areas of eukaryotic srRNAs that have no equivalent in prokaryotic srRNAs.

The examination of srRNA sequences from additional species belonging to a variety of taxa should be a step towards the elucidation of the secondary structure of the molecule in these variable and eukaryote-specific areas. In addition, these sequences can be used later in attempts to elucidate metazoan phylogeny on a molecular basis.

2. MATERIALS AND METHODS

High-M_r DNA was prepared from frozen larvae of *T. molitor* according to [12]. The DNA was further purified by several subsequent ethanol precipitations. After partial digestion of the

Correspondence address: R. De Wachter, Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

DNA with the restriction enzyme *Eco*RI at an enzyme/substrate ratio of 3 U/μg DNA, the digest was separated by low-melting-point agarose gel electrophoresis. *Eco*RI DNA fragments in the range of 20 kb were isolated from the gel using a phenol extraction protocol [13]. A genomic library of *T. molitor* was constructed by ligating 20 kb *Eco*RI DNA fragments into *Eco*RI-cut Charon 4A arms (Amersham, England). The recombinant DNA was packaged in vitro [14] and used for infection of *Escherichia coli* DP50 Sup F [15]. The original library consisted of 5×10^4 pfu/μg insert DNA.

Charon 4A phages containing the 18 S rDNA of *T. molitor* were selected using the plaque hybridization technique [16]. A 2.3 kb *Eco*RI fragment of a partially digested recombinant Charon 4A phage, comprising the entire 18 S rDNA, was subcloned in the plasmid pUC18 [17]. Competent *E. coli* JM 103 were transformed with 10 ng plasmid DNA [18]. Plasmid DNA was recovered using an alkaline extraction procedure essentially as in [19]. The recombinant pUC18 DNA was then digested with tetramer recognizing restriction enzymes and the fragments were subcloned in the M13mp10 or M13mp11 vector [20]. Sequencing was performed according to the M13-dideoxy method [21].

3. RESULTS AND DISCUSSION

3.1. *Sequencing strategy and primary structure of* T. molitor *18 S rDNA*

The sequencing strategy used to obtain the primary structure of the 18 S rDNA of *T. molitor* is demonstrated in fig.1. The sequence itself together with its 5′- and 3′-flanking regions is shown in fig.2. The sequence of the 18 S rRNA is 1921 nucleotides long, the position of the termini being derived on the basis of homology with known 18 S rRNA sequences [1].

3.2. *Secondary structure model for* T. molitor *18 S rRNA*

Fig.3 shows the secondary structure model derived for the 18 S rRNA of *T. molitor*. Helices are numbered according to the principle set forth previously [2], i.e. in order of their occurrence when the sequence is scanned from the 5′- to 3′-terminus. The same helix number is given to double-stranded segments separated only by internal or bulge loops, in other words the helix number changes at each multibranched loop. Helices occurring in both eukaryotic and prokaryotic srRNAs bear a single number, whereas helices thought to be specific for eukaryotic srRNAs bear a composite number E*a-b*, where *a* is the number of the preceding universal helix and *b* is a serial number. Similarly, numbers of the form P*a-b* are used to designate prokaryote-specific helices.

Renumbering of the helices with respect to previous models [1,2] was necessary because the availability of an increasing number of sequences allows additional secondary structure elements to be discerned. The two helices forming the pseudoknot structure [22] at the 5′-terminus are numbered 1 and 2. Other changes in the numbering system arise from the fact that structures previously considered as bulge loops were proven on a comparative basis to contain short helices. Such is the case for helices 24 and 28. An apparent exception to the numbering rules is made for helices 43 and 44, which are separated by an internal rather than by a multibranched loop. The reason is that in the

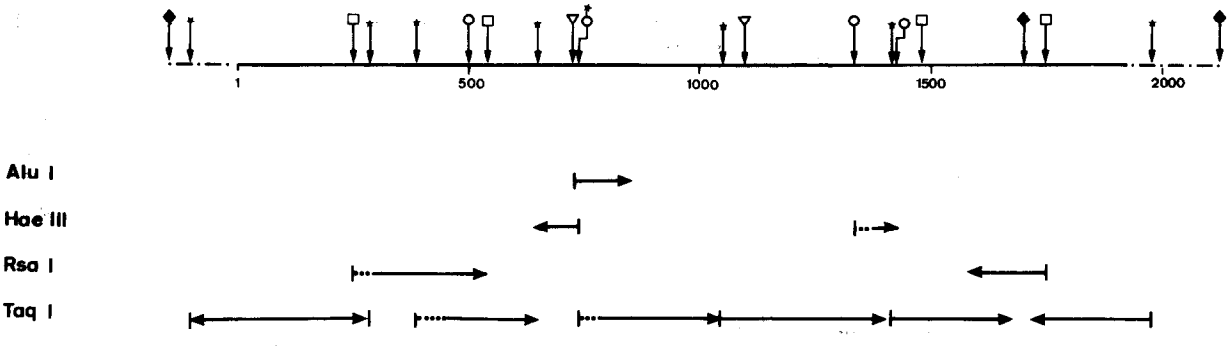

Fig.1. Sequencing strategy of the 18 S rDNA of *T. molitor*. The restriction map covers a DNA fragment containing the complete 18 S rDNA and flanking regions of *T. molitor*, obtained by partial digestion with *Eco*RI. Continuous lines represent the 18 S rDNA and broken lines represent the 5′- and 3′-flanking sequences. Horizontal arrows indicate the direction and distances of sequencing in M13mp10 and M13mp11 subclones. (▽) *Alu*I, (♦) *Eco*RI, (○) *Hae*III, (□) *Rsa*I, (★) *Taq*I. Except for the *Eco*RI sites only those restriction sites are shown which were used for sequencing.

```
                                                  aattc atttggaagt gttcaaattt    -71
ggagtacatg acggatcttt tttatttaat aaataagagg atgcacgtcg tgtctctaaa cgaacgcagt    - 1
TCCCTGGTTG ATCCTGCCAG TAGTCATATG CTTGTCTCAA AGATTAAGCC ATGCATGTCT CAGTACAAGC     70
CGAATTAAGG TGAAACCGCG AAAGGCTCAT TAAATCAGTT ATGGTTCCTT AGATCGTACC CACATTTACT    140
TGGATAACTG TGGTAATTCT AGAGCTAATA CATGCAAACA GAGCTCCAAC CGGAAACGGA AGGAGCGCTT    210
TTATTAGATC AAAACCAATC GGTGGCGGTC TCCGTCATCG TACAACTTGG TGACTCTGAA TAACTTTACG    280
CTGATCGCAC GGTCTTGCAC CGGCGACGCA TCTTTCAAAT GTCTGCCTTA TCAACTGTCG ATGGTAGGTT    350
CTGCGCCTAC CATGGTTGTA ACGGGTAACG GGGAATCAGG GTTCGATTCC GGAGAGGGAG CCTGAGAAAC    420
GGCTACCACA TCCAAGGAAG GCAGCAGGCG CGCAAATTAC CCACTCCCGG CACGGGGAGG TAGTGACGAA    490
AAATAACGAT ACGGGACTCA TCCGAGGCCC CGTAATCGGA ATGAGTACAC TCTAAACCCT TTAACGAGGA    560
TCAATTGGAG GGCAAGTCTG GTGCCAGCAG CCGCGGTAAT TCCAGCTCCA ATAGCGTATA TTAAAGTTGT    630
TGCGGTTAAA AAGCTCGTAG TCGAATCTGT GTCCCGCGCC GCCGGTTCAT CGTTCGCGGT GTTAACTGGC    700
GTGCCGCGGG ACGTCCTGCC GGTGGGCTTA GCTCGTGAGG GCGGCCCAAC TCAATCCCGC CGCGGTGCTC    770
TTCGTTGAGT GTCGAGGTGG GCCGGCACGT TTACTTTGAA CAAATTAGAG TGCTTAAAGC AGGCTAAAAC    840
TTCGCCTGAA TACTGTGTGC ATGGAATAAT GGAATAGGAC CTCGGTTCTA TTTTGTTGGT TTTCGGAATT    910
TTGAGGTAAT GATTAATAGG AACGGATGGG GGCATTCGTA TTGCGACGTT AGAGGTGAAA TTCTTGGATC    980
GTCGCAAGAC GGACAGAAGC GAAAGCATTT GCCAAAAACG CTTTCATTGA TCAAGAACGA AAGTTAGAGG   1050
TTCGAAGGCG ATCAGATACC GCCCTAGTTC TAACCATAAA CGATGCCAGC TAGCGATCCG CCGACGTTCC   1120
TCCGATGACT CGGCGGGCAG CTTCCGGGAA ACCAAAGCTT TTGGGTTCCG GGGGAAGTAT GGTTGCAAAG   1190
CTGAAACTTA AAGGAATTGA CGGAAGGGCA CCACCAGGAG TGGAGCCTGC GGCTTAATTT GACTCAACAC   1260
GGGAAACCTC ACCAGGCCCG GACACCGGAA GGATTGACAG ATTGAGAGCT CTTTCTTGAT TCGGTGGGTG   1330
GTGGTGCATG GCCGTTCTTA GTTGGTGGAG CGATTTGTCT GGTTAATTCC GATAACGAAC GAGACTCTAG   1400
CCTGCTAAAT AGGCGTATTT CGACATCCCA AAGGCCCGCT GGTCCCGGGT TCGCTCGGTG ACCGTGCGCG   1470
GTTTTTACTG TCGGCGTACA AACAATTCTT CTTAGAGGGA CAGGCGGCTT CTAGCCGAAC GAGATTGAGC   1540
AATAACAGGT CTGTGATGCC CTTAGATGTT CTGGGCCGCA CGCGCGCTAC ACTGAAGGAA TCAGCGTGTC   1610
CTCCCTGGCC GAGCGGCCCG GGTAACCCGC TGAACCTCCT TCGTGCTAGG GATTGGGGCT TGCAATTGTT   1680
CCCCATGAAC GAGGAATTCC CAGTAAGCGC GAGTCATAAG CTCGCGTTGA TTACGTCCCT GCCCTTTGTA   1750
CACACCGCCC GTCGCTACTA CCGATTGAAT GATTTAGTGA GGTCTTCGGA CCGGTACGCG GTGGCGTTTC   1820
GGCGTCGCCG ATGTTGCTGG GAAGATGACC AAACTTGATC ATTTAGAGGA AGTAAAAGTC GTAACAAGGT   1890
TTCCGTAGGT GAACCTGCGG AAGGATCATT Aacgtgtctc gcacacacga ataatacaaa cactcacgtc   1960
tgttcgtgcg gtgttgtatg tcgtcgtg                                                 1988
```

Fig.2. Sequence of the 18 S rDNA of *T. molitor* and its flanking regions. The 95 5′-flanking nucleotides and the 67 3′-flanking nucleotides are indicated by bases in lower-case letters, the 18 S rDNA sequence being printed in upper-case characters. The 5′- and 3′-termini of the 18 S rDNA were derived on the basis of homology with other srRNA sequences.

Rhizopoda species *Acanthamoeba castellanii* [23] and the Sporozoa species *Plasmodium berghei* [24], extra nucleotides are present in the 5′-strand of the internal loop separating these two helices. This extra sequence can be folded into a hairpin structure bearing the number E43-1, which is absent in the remaining known 18 S rRNAs.

The area located between helices 21 and 22 in fig.3 is eukaryote-specific, and very variable in length as well as in secondary structure. Whereas prokaryotic srRNAs comprise a single helix in this area, namely the protein S8-binding helix [5] P21-1, eukaryotes insert at least 200 residues in this area. *A. castellanii* [23] and *Euglena gracilis* [25] even insert 375 and 512 nucleotides, respectively. A partial secondary structure model for this E21 area was proposed by Nelles and co-workers [2]. Recently, other authors [3,7,10,11] advocated various tentative secondary structure models for this area, featuring 5–7 helices. The local structure adopted in fig.3 shows 5 hairpin loops. The partial model proposed in [2] was taken as a starting point and additional helices were constructed. Most eukaryotic srRNAs can be folded into a structure

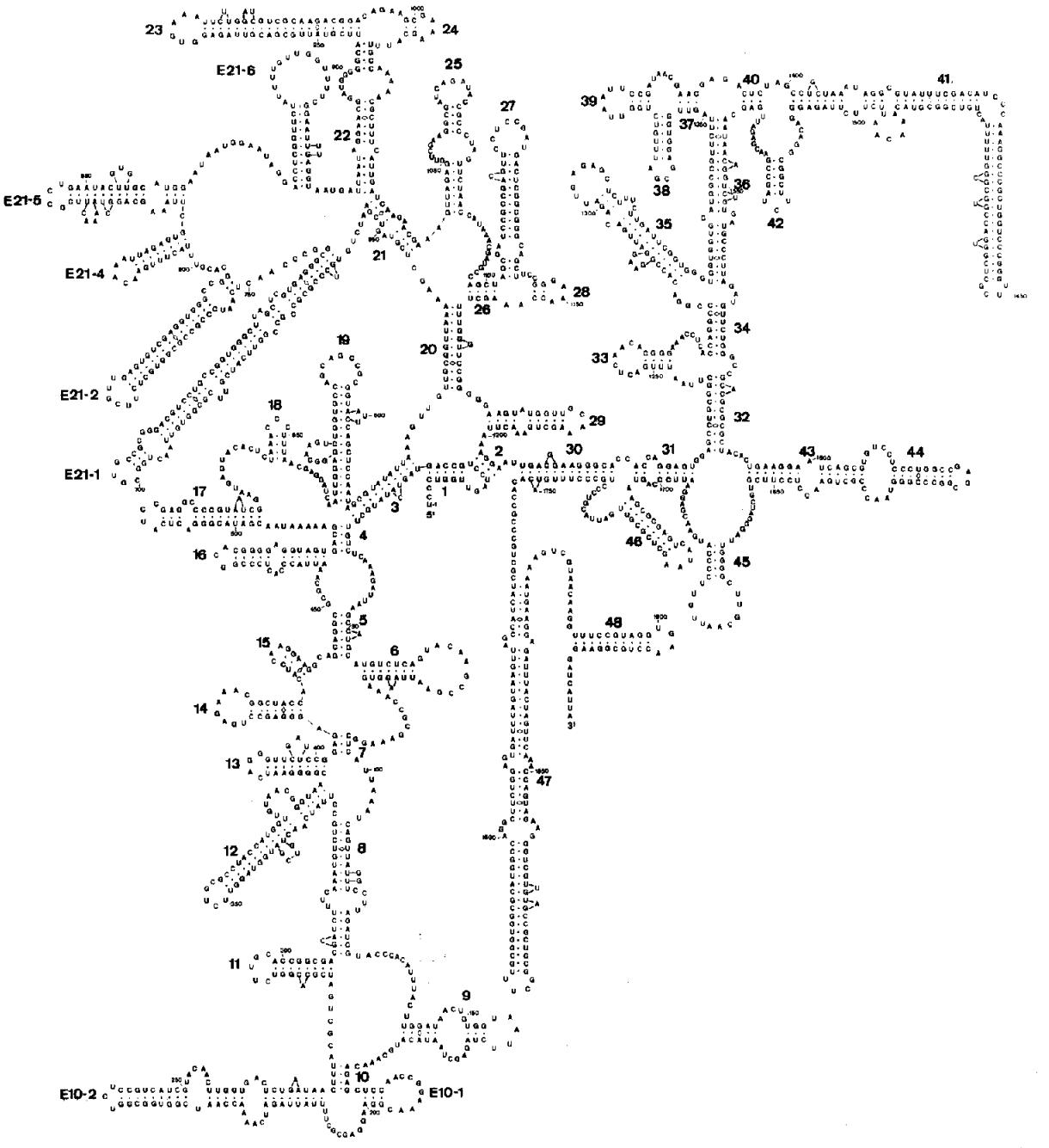

Fig.3. Secondary structure model of the 18 S rRNA of *T. molitor*. Helices are numbered as described in the text. E preceding a helix number refers to a helix present only in eukaryotic srRNAs. The base pairs G·C, A·U and G·U are indicated by dots, non-standard base pairs presumably intercalated in the helices are indicated by a lozenge. Every 50th nucleotide of the sequence is numbered.

containing helices E21-1, E21-2, E21-4, E21-5 and E21-6 drawn in fig.3. Certain species contain additional insertions which may form extra hairpin loops not shown in fig.3: E21-3 (between E21-2 and E21-4) can be formed in *Euplotes aediculates* [26], *A. castellanii* [23] and *E. gracilis* [25] which

possess an insertion here; E21-7 (between E21-6 and 22) can be formed in *E. gracilis* [25], *Crithidia fasciculata* [27] and *Trypanosoma brucei* [25]. In contrast, the Microsporidia species *Vairimorpha necatrix* [28] has a large deletion wiping out the entire area between helices 21 and 22. Among the hairpins proposed in fig.3, E21-1, E21-2 and E21-6 can be proven by compensating substitutions. This is not possible for hairpin E21-4 because it is too conserved in sequence, or for E21-5 which has a very irregular structure and whose existence therefore seems most dubious. The secondary structure models proposed by others are not entirely applicable to every eukaryotic srRNA molecule. Our model comes closest to that proposed by Ellis and co-workers [3] except that hairpin E21-5 replaces two hairpins which are postulated in their model but cannot be constructed for every known sequence.

Two alternative structures have been proposed for the central area enclosed by helices 3, 21, 25 and 29. Several authors [5,6,8,10] adopt a similar topology for the secondary structure of eukaryotic and prokaryotic srRNAs in this area. Others [1–3,7,9] advocate a eukaryote-specific topology for this region. However, as discussed in [2], eukaryotic srRNAs can also be folded in the alternative topology applicable to prokaryotes. In fig.3 the 18 S RNA of *T. molitor* is fitted into the latter 'universal' secondary structure. However, the potential for formation of an alternative structure in this area, possible only in eukaryotic srRNAs, is demonstrated in fig.4. In this structure, helices 20, 26 and 28 of the universal model are dissolved in favor of the formation of two alternative helices, labeled E20 and E28 in fig.4.

If the model of fig.4b correctly describes the local secondary structure in eukaryotic srRNAs, then this would mean that the molecule underwent a structural switch after the eukaryotic branch had

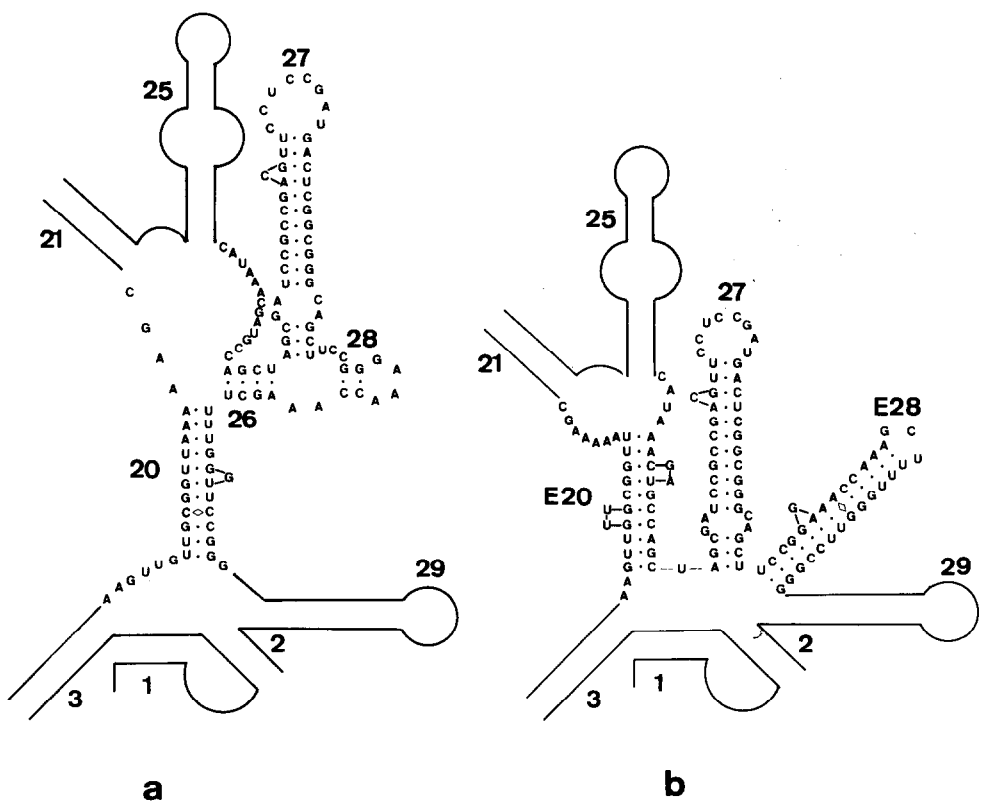

Fig.4. Alternative folding patterns for the central area of *T. molitor* 18 S rRNA. Topology a is as in fig.3. In topology b, possible in eukaryotic srRNAs, universal helices 20, 26 and 28 are dissolved in favor of eukaryotic-specific helices E20 and E28.

been established in evolution. It is also possible, however, that the molecule, after acquiring a sequence that allows two different foldings, retained this property because it allows a structural switch from which it gains some functional benefit. A search for compensating substitutions in the conformation-specific helices should in principle provide evidence as to which of the possible structures exists, or whether both coexist, in eukaryotic srRNAs. Our preliminary evidence points to the presence of compensating substitutions in helices 26 and 28 of the universal model (fig.4a), but also in helix E28 of the eukaryote-specific model (fig.4b), a finding that would support the coexistence of both structures in eukaryotic srRNAs.

Acknowledgement: Our research was supported in part by an FKFO grant to R.D.W.

REFERENCES

[1] Huysmans, E. and De Wachter, R. (1986) Nucleic Acids Res. 14, r73–118.
[2] Nelles, L., Fang, B.-L., Volckaert, G., Vandenberghe, A. and De Wachter, R. (1984) Nucleic Acids Res. 12, 8749–8768.
[3] Ellis, R.E., Sulston, J.E. and Coulson, A.R. (1986) Nucleic Acids Res. 14, 2345–2364.
[4] Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) Microbiol. Rev. 47, 621–669.
[5] Gutell, R.R., Weiser, B., Woese, C.R. and Noller, H.F. (1985) Prog. Nucleic Acid Res. Mol. Biol. 32, 155–217.
[6] Mankin, A.S., Kopylov, A.M. and Bogdanov, A.A. (1981) FEBS Lett. 134, 11–14.
[7] Atmadja, J., Brimacombe, R. and Maden, B.E. (1984) Nucleic Acids Res. 12, 2649–2667.
[8] Chan, Y.-L., Gutell, R., Noller, H.F. and Wool, I.G. (1984) J. Biol. Chem. 259, 224–230.
[9] Olsen, G.J., McCarroll, R. and Sogin, M.L. (1983) Nucleic Acids Res. 11, 8037–8049.
[10] Gonzalez, I.K. and Schmickel, R.D. (1986) Am. J. Hum. Genet. 38, 419–427.
[11] Herzog, M. and Maroteaux, L. (1986) Proc. Natl. Acad. Sci. USA 83, 8644–8648.
[12] Karn, J., Brenner, S. and Barnett, L. (1983) Methods Enzymol. 101, 3–19.
[13] Wieslander, L. (1979) Anal. Biochem. 98, 305–309.
[14] Hohn, B. (1979) Methods Enzymol. 68, 299–309.
[15] Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.-A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) Science 1986, 161–169.
[16] Benton, W.D. and Davis, R.W. (1977) Science 1986, 180–182.
[17] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[18] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[19] Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513–1523.
[20] Messing, J. (1983) Methods Enzymol. 101, 20–78.
[21] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[22] Pleij, C.W.A., Rietveld, K. and Bosch, L. (1985) Nucleic Acids Res. 13, 1717–1731.
[23] Gunderson, J.H. and Sogin, M.L. (1986) Gene 44, 63–70.
[24] Gunderson, J.H., McCutchan, T.F. and Sogin, M.L. (1986) J. Protozool. 33, 525–529.
[25] Sogin, M.L., Elwood, H.J. and Gunderson, J.H. (1986) Proc. Natl. Acad. Sci. USA 83, 1383–1387.
[26] Sogin, M.L., Swanton, M.T., Gunderson, J.H. and Elwood, H.J. (1986) J. Protozool. 33, 26–29.
[27] Schnare, M.N., Collings, J.C. and Gray, M.W. (1986) Curr. Genet. 10, 405–410.
[28] Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck, B.A. and Woese, C.R. (1987) Nature 326, 411–414.