The corrected nucleotide sequences of 5 S RNAs from six angiosperms

With some notes on 5 S RNA secondary structure and molecular evolution

Antoon Vandenberghe, Min-Wei Chen, Erna Dams, Raymond de Baere, Els de Roeck, Erik Huysmans and Rupert de Wachter

Laboratory of Molecular Biology, Department of Biochemistry, University of Antwerp, B-2610 Wilrijk, Belgium

Received 10 April 1984

The corrected nucleotide sequences of the cytoplasmic 5 S ribosomal RNAs of 6 angiosperms are reported. The previously reported sequences, reconstructed from oligonucleotide catalogs, had been suspected for some time to contain errors. The corrected seqeuences fit in a universal 5-helix model for 5 S RNA secondary structure. The model involves the existence of a dynamic equilibrium between slightly different base-pairing schemes in two areas of the structure. Reconstruction of a phylogenetic tree from over 200 known 5 S RNA sequences yields a general picture of biological evolution. It confirms the tripartite descent of living species – eubacteria, archaebacteria, eukaryotes – and the endosymbiotic origin of plant mitochondria and chloroplasts.

5 S rRNA *Secondary structure* *Molecular evolution*

1. INTRODUCTION

5 S RNA is an attractive subject for studies in molecular evolution because as a ribosome constituent it occurs universally and is easily isolated, and due to its small size it is accessible to direct RNA-sequencing methods. The potential of 5 S RNA for providing us with an outline of the evolutionary relationships between all living species has been overshadowed for some time by the view that the structures found in eukaryotes and bacteria may not be comparable. At one time it was even suggested [1] that bacterial 5 S RNA might be the equivalent of eukaryotic 5.8 S RNA, rather than 5 S RNA. Although this hypothesis is now abandoned, the view that there is a profound dichotomy in primary, secondary, and tertiary structure between eukaryotic and bacterial 5 S RNA is still widespread [2–5]. A careful alignment of available sequences has led us to the conviction [6], shared by others [7–11], that the secondary structure of 5 S RNA is essentially universal. In retrospect, it appears that the opposite view has been based in part on sequencing errors. One model, specific for eukaryotic 5 S RNA secondary structure [12], has been based on an erroneous sequence [13] for *Chlorella* 5 S RNA, which has been corrected in the meantime [14]. On the basis of our 5 S RNA sequence alignment [6] now extended to 175 structures [15] we have been able to trace and correct two more sequencing errors [16,17] and additional corrections are in progress (unpublished). Here we report the corrections of 6 previously published [18–20] sequences of 5 S RNAs from flowering plants, which have been suspected [21] for some time to be based on a wrong ordering of oligonucleotides. We also comment on equilibria in 5 S RNA secondary structure and present the outline of a phylogenetic tree reconstructed from 213 sequences.

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

2. EXPERIMENTAL

Seeds of rye (*Secale cereale* cv. Somro), dwarf bean (*Phaseolus vulgaris* cv. Limburg), broad bean (*Vicia faba*) and sunflower (*Helianthus annuus* cv. Giganteus) were allowed to germinate on beds of wet vermiculite and shoots were harvested when several centimeters high. Tomatoes (*Lycopersicum esculentum* cv. Marmande) were bought in a greengrocery and duckweed (*Lemna minor*) was collected in the field.

Batches of 100–200 g plant tissue were worked up for the preparation of ribosomes and 5 S rRNA as in [22] for animal tissue. In the case of *S. cereale, P. vulgaris* and *L. esculentum* ribosomes were concentrated by ammonium sulfate precipitation [16] before being purified by ultracentrifugation. On average, 100 g plant tissue yielded 8–35 mg ribosomes, from which 80–270 μg of 5 S RNA was obtained.

For the sequencing of 3′-end-labeled 5 S RNAs by the chemical degradation method [23], electrophoresis on 8, 12 and 20% gels was used to resolve the 5′-terminal, the middle, and the 3′-terminal sequence areas, respectively. The 5′-terminal nucleotide was determined by liquid chromatography of unlabelled 5 S RNA [24] or by enzymatic degradation of 5′-end-labeled 5 S RNA ligated to $(Ap)_4A$ [16].

3. RESULTS

The primary structures of the 6 plant 5 S RNAs that have been reexamined in this study are listed in fig.1. The sequence of *Azotobacter vinelandii* 5 S RNA is added to demonstrate that the number of gaps required for a satisfactory alignment of eukaryotic and bacterial sequences is very limited.

Oligonucleotide catalogs for the six 5 S RNAs listed in fig.1 were originally published by authors in [18,19]. A sequence for *Secale cereale* (rye) 5 S RNA was derived later [20] by partial RNase diges-

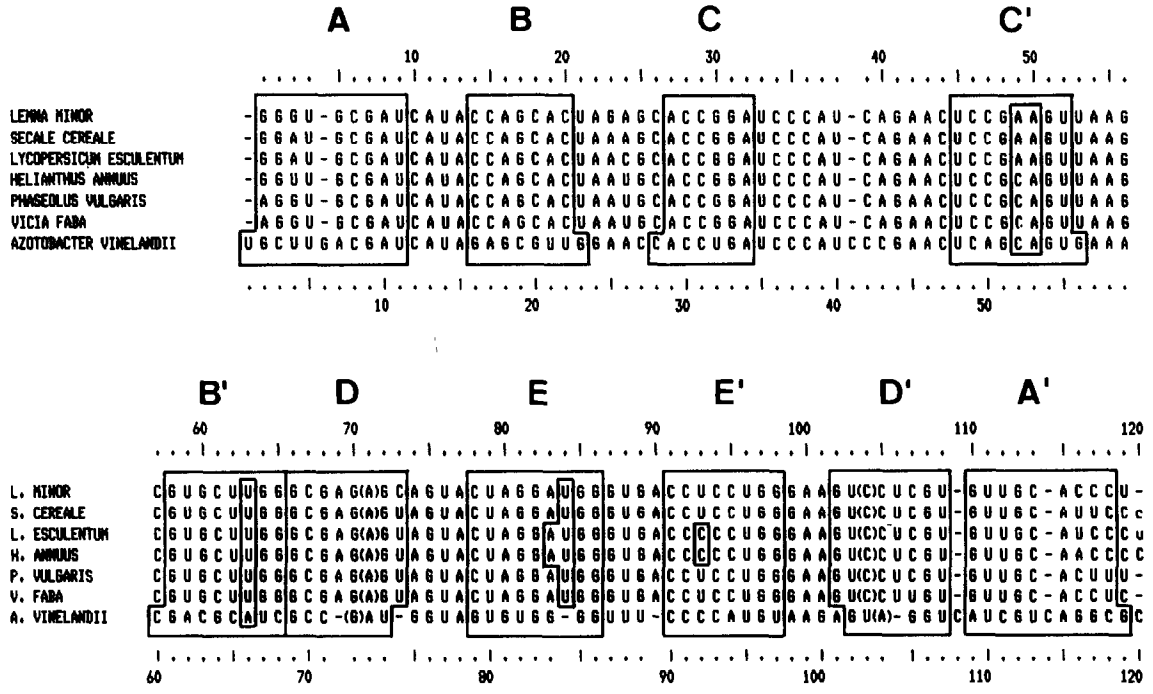


Fig.1. Alignment of the determined plant 5 S RNA sequences. The numbering at the top applies to the 6 plant 5 S RNA sequences, that at the bottom to the sequence of *A. vinelandii* 5 S RNA added for comparison. Boxes labeled A-A′, B-B′, etc., enclose areas involved in base pairing and resulting in helices A, B, etc., in the secondary structure model (fig.3). Bulges are indicated by nested boxes, non-standard base pairs by characters within parentheses. Lower-case characters at the 3′-end of some sequences denote residues present in submolar amounts.

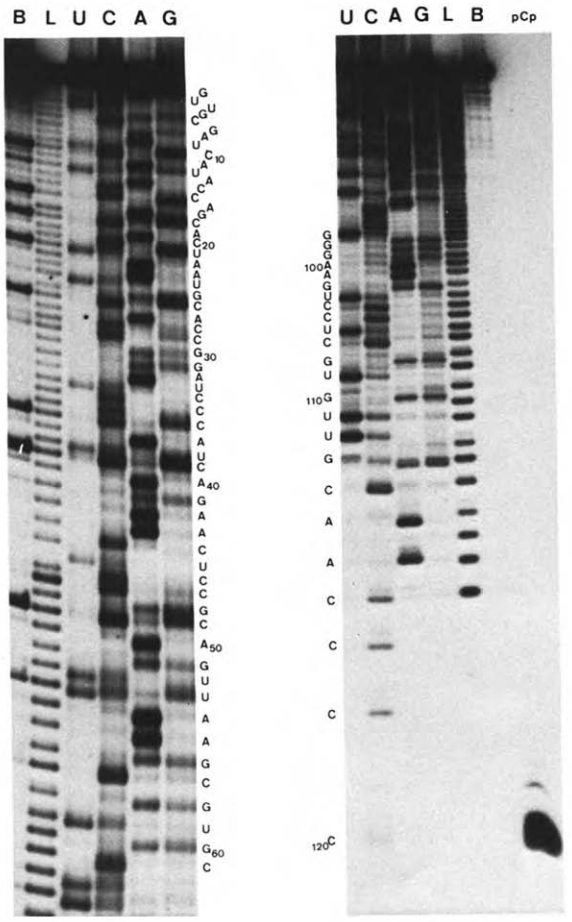


Fig.2. Sequencing gels of *H. annuus* 5 S RNA. The autoradiograms shown are from an 8% (left) and a 20% (right) polyacrylamide gel with the separations of partial base-specific chemical degradation mixtures of 3′-end-labeled RNA [23] indicated by U, C, A, G. An untreated control (B) shows spontaneous hydrolysis at YpA sites, also responsible for spurious bands in the G-lane. Bands in the 'ladder' (L) obtained by acid hydrolysis [25] run slower than bands of the same chain length in the chemical degradation lanes because the latter bear an extra 5′-phosphate.

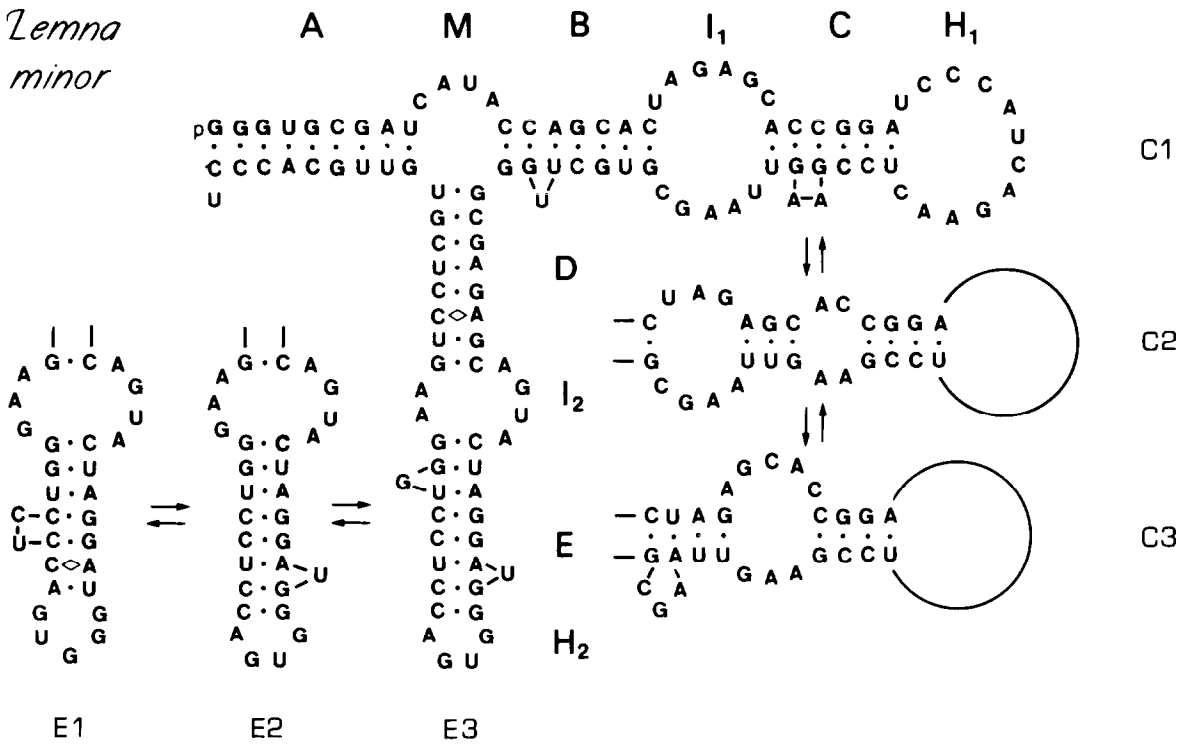
tion, and the oligonucleotide ordering for the 5 remaining species was deduced by analogy. However, when the structure of wheat 5 S RNA was determined by gel-sequencing methods [21] it became apparent that some pancreatic RNase oligonucleotides had probably been assembled in the wrong order in the closely related rye 5 S RNA sequence [20], and that this error had been repeated in the deduction of the other plant sequences. Our results show indeed that the correction of all 6 sequences requires a rearrangement of $ACCAU_{14}$ to $CAUAC_{14}$ and an insertion of C_{34}. The alignment (fig.1) comprises other species-specific corrections, mainly near the termini. An autoradiogram illustrating the determination of the *Helianthus annuus* 5 S RNA sequence is shown in fig.2.

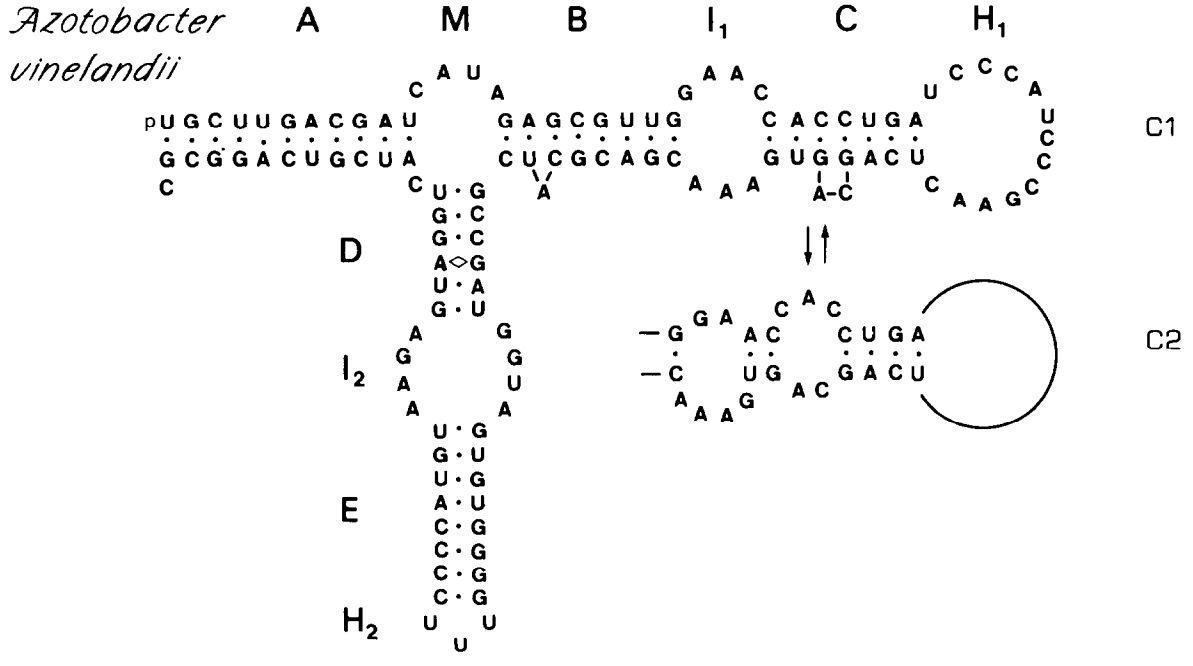
4. DISCUSSION

4.1. *Secondary structure of 5 S RNA*

In fig.3 the duckweed cytoplasmic 5 S RNA sequence is fitted in the universal secondary structure model that we previously proposed [6] and is compared with the corresponding model for 5 S RNA of the bacterium *A. vinelandii* [16]. In spite of the striking similarity between the two models, some authors [3,5] still question the existence of helix D in bacterial 5 S RNAs. The arguments in favor of the existence of helix D, discussed extensively in [6,16], are that it is evident on a comparative basis, that it improves the estimated stability of the secondary structures, and that it is compatible with the result of experiments designed to probe secondary structure. Moreover, so-called compensating substitutions are observed in helix D when the structures of cyanobacterial or mycoplasmal 5 S RNAs are compared with those of other bacteria [15]. Specific differences between the eukaryotic and eubacterial 5 S RNA base-pairing schemes do exist, however. Features distinguishing the eubacterial model from its eukaryotic counterpart are a larger size of loop H_1, a symmetrical I_2-loop, the absence of bulges in helix E and the presence of an unpaired base between helices D and A. More detailed taxon-specific differences within the eukaryotic and eubacterial primary kingdoms have been discussed [9,10,25].

In all hitherto sequenced 5 S RNAs, two alternative base-pairing schemes can be considered in area I_1-C of the model [6,16,17,22,25,26]. These alternative structures are designated as C1 and C2 in fig.3. We have observed that 3 alternative shapes are often possible in area C, and that 2–5 alternative shapes for helix E can consistently be considered in eukaryotes (unpublished). The different shapes of helix E distinguish themselves by the number and position of the bulges protruding


Lemna minor
A
M
B
I_1
C
H_1
C1
C2
C3
D
I_2
E
H_2
E1
E2
E3


Azotobacter vinelandii
A
M
B
I_1
C
H_1
C1
C2
D
I_2
E
H_2

←

Fig.3. Secondary structure models for *L. minor* and *A. vinelandii* 5 S RNAs. Helices are labeled A–E (cf. fig.1); loops are labeled M (multi-branched), I (internal) and H (hairpin). The base pairs G·C, A·U and G·U are indicated by a dot, isolated non-standard base pairs within a helix by a lozenge. Of the equilibrium secondary structures in areas C and E, C1 and C2 are universal and C3 is found frequently. In addition, eukaryotic 5 S RNAs have up to 5 equilibrium structures in area E, depending on the considered taxon. Forms E1–E3 depicted here are possible in all land plants.

from the helix. In land plant cytoplasmic 5 S RNAs, shapes E1–E3 (fig.3) are always possible. In our view, the alternative base-pairing schemes for areas C and E correspond with alternative structures, in dynamic equilibrium with each other in biologically active 5 S RNA. This may endow the 5 S RNA molecule with a flexibility needed for its (unknown) function within the ribosome.

Apart from those listed in fig.1, the cytoplasmic 5 S RNA sequences of the following plants have been determined: wheat [21], spinach [27], lupin [28] and flax [29]. All can be fitted in exactly the same model as the plant sequence in fig.3, except for one of the two lupin 5 S RNA sequences deduced from the gene structures. In the structure derived from clone pAR1 base-pairs essential for the formation of helices C–E are missing. A possible explanation is that clone pAR1 contained a 5 S pseudogene and clone pAR4 a functional 5 S RNA gene.

4.2. *Molecular evolution*

An evolutionary tree based on 5 S RNA was reconstructed from an alignment of 213 sequences, among which 175 are listed in a recent collection [15] and the remainder are based on

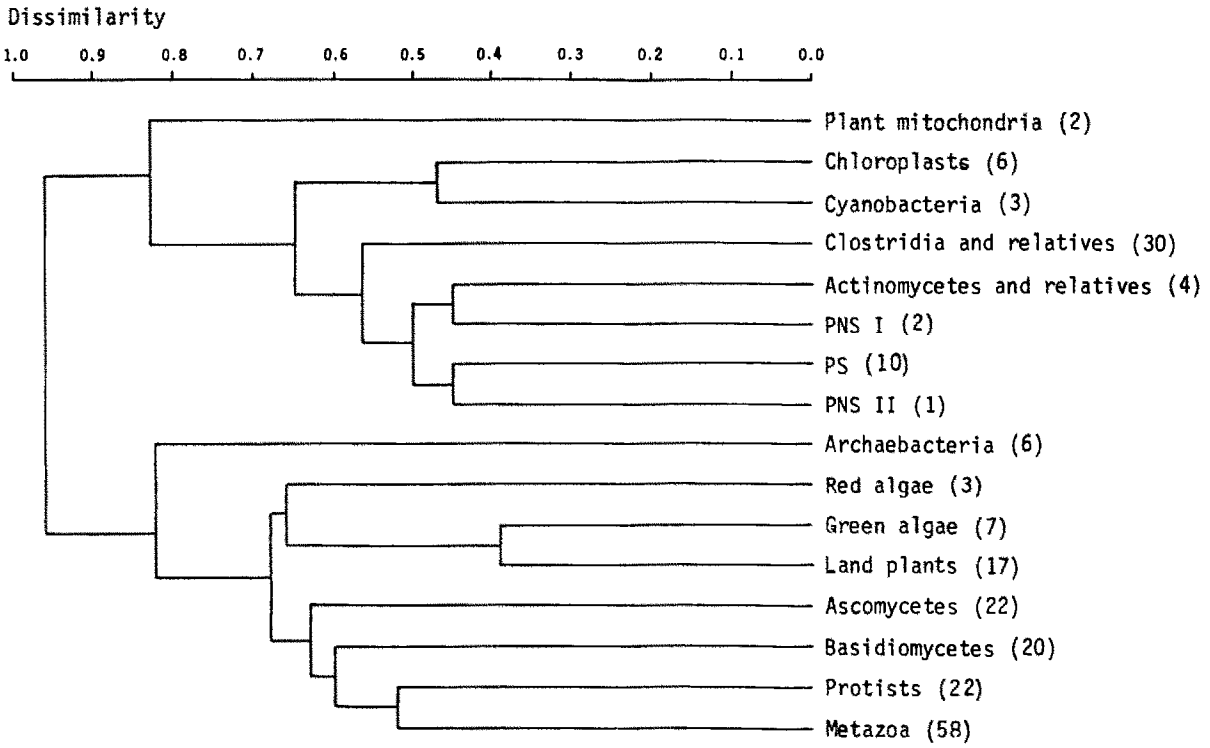


Fig.4. Outline of evolution as reconstructed from 5 S RNA sequences. The construction of the phenogram and the meaning of the dissimilarity scale have been described in [26]. The complete tree comprises 213 sequences but the ramifications within each of the major taxa indicated here are simplified to a single branch. Each taxon name is followed by the number of sequences comprised. The abbreviations PS (purple sulfur bacteria and relatives) PNS I and II (groups I and II of the purple non-sulfur bacteria and relatives) refer to the classification in [30].

oligonucleotide catalogs, recently published, or unpublished results from our laboratory. The alignment of the sequences was as in [15]. Fig.4 shows a simplified version of the complete phenogram (unpublished) constructed by weighted pair grouping starting from a difference matrix as in [26]. The ramifications within each of the major taxa mentioned have been omitted and reduced to a single branch.

The primary kingdoms eubacteria, eukaryotes and archaebacteria form 3 major branches. The archaebacterial branch, which is not further elaborated because it comprises merely 6 sequences, is more closely related to the eukaryotes than to the eubacteria. Among the eukaryotes, there is an early divergence into a branch leading to the red algae, green algae and land plants, and one containing the fungi, protists and metazoa. The latter branch first sees a divergence of ascomycetes, followed by basidiomycetes, and finally bifurcates into protists and metazoa. Species found among the protists include 12 protozoa, 5 mastigomycetous fungi, 2 slime molds, 2 brown algae and 1 cryptophytan species.

For the clusters within the eubacterial primary kingdom the nomenclature in [30] is followed. The most ancient taxon represented by 5 S RNA sequences are the cyanobacteria. Next, the Gram-positive clostridia, bacilli and lactic acid bacteria branch off. The remaining branches consist of Gram-negative purple sulfur and purple non-sulfur bacteria and relatives, and one branch leading to the Gram-positive actinomycetes. The position of the latter cluster is rather surprising and may be due to a distortion resulting from a shortage of relevant sequences. Chloroplast 5 S RNA sequences are all included in the cyanobacterial lineage. A much looser relationship with the eubacteria is seen for the only two plant mitochondrial 5 S RNA sequences analyzed so far.

ACKNOWLEDGEMENTS

Peter Adriaenssens contributed to the sequencing of the plant 5 S RNAs. Our research was aided by an NFWO grant to R.D.W. and A.V. and by an IWONL scholarship to E.D.

REFERENCES

[1] Erdmann, V.A. (1976) Prog. Nucleic Acid Res. Mol. Biol. 18, 45–90.
[2] Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci. USA 76, 381–385.
[3] Douthwaite, S. and Garrett, R.A. (1981) Biochemistry 20, 7301–7307.
[4] Garrett, R.A. and Olesen, S.O. (1982) Biochemistry 21, 4823–4830.
[5] Pieler, T. and Erdmann, V.A. (1982) Proc. Natl. Acad. Sci. USA 79, 4599–4603.
[6] De Wachter, R., Chen, M.-W. and Vandenberghe, A. (1982) Biochimie 64, 311–329.
[7] Nishikawa, H. and Takemura, S. (1974) FEBS Lett. 40, 106–109.
[8] Böhm, S., Fabian, H. and Welfle, H. (1982) Acta Biol. Med. Germ. 41, 1–16.
[9] Delihas, N. and Andersen, J. (1982) Nucleic Acids Res. 10, 7323–7344.
[10] Neimark, H., Andersen, J. and Delihas, N. (1983) Nucleic Acids Res. 11, 7569–7577.
[11] Küntzel, H., Piechulla, B. and Hahn, U. (1983) Nucleic Acids Res. 11, 893–900.
[12] Vigne, R. and Jordan, B.R. (1977) J. Mol. Evol. 10, 77–86.
[13] Jordan, B.R., Galling, G. and Jourdan, R. (1974) J. Mol. Biol. 87, 205–225.
[14] Luehrsen, K.R. and Fox, G.E. (1981) Proc. Natl. Acad. Sci. USA 78, 2150–2154.
[15] Erdmann, V.A., Wolters, J., Huysmans, E., Vandenberghe, A. and De Wachter, R. (1984) Nucleic Acids Res. 12, in press.
[16] Dams, E., Vandenberghe, A. and De Wachter, R. (1983) Nucleic Acids Res. 11, 1245–1252.
[17] Dams, E., Londei, P., Cammarano, P., Vandenberghe, A. and De Wachter, R. (1983) Nucleic Acids Res. 11, 4667–4676.
[18] Payne, P.I., Corry, M.J. and Dyer, T.A. (1973) Biochem. J. 135, 845–851.
[19] Dyer, T.A. and Bowman, C.M. (1976) in: Genetics and Biogenesis of Chloroplasts and Mitochondria (Bucher, T. et al. eds) pp.645–651, Elsevier, Amsterdam, New York.
[20] Payne, P.I. and Dyer, T.A. (1976) Eur. J. Biochem. 71, 33–38.
[21] McKay, R.M., Spencer, D.F., Doolittle, W.F. and Gray, M.W. (1980) Eur. J. Biochem. 112, 561–576.
[22] Fang, B.-L., De Baere, R., Vandenberghe, A. and De Wachter, R. (1982) Nucleic Acids Res. 10, 4679–4685.
[23] Peattie, D.A. (1979) Proc. Natl. Acad. Sci. USA 76, 1760–1764.

[24] Vandenberghe, A. and De Wachter, R. (1982) J. Liq. Chromatogr. 5, 2079–2084.
[25] Dams, E., Vandenberghe, A. and De Wachter, R. (1982) Nucleic Acids Res. 10, 5297–5302.
[26] Huysmans, E., Dams, E., Vandenberghe, A. and De Wachter, R. (1983) Nucleic Acids Res. 11, 2871–2880.
[27] Delihas, N., Andersen, J., Sprouse, H.M., Kashdan, M. and Dudock, B. (1981) J. Biol. Chem. 256, 7515–7517.
[28] Rafalski, J.A., Wiewiorowski, M. and Söll, D. (1982) Nucleic Acids Res. 10, 7635–7642.
[29] Goldsbrough, P.B., Ellis, T.H.N. and Lomonossoff, G.P. (1982) Nucleic Acids Res. 10, 4501–4514.
[30] Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980) Science 209, 457–463.