

The structure of 5 S ribosomal RNA in the methanogenic archaeobacteria *Methanobacterium tindarius* and *Methanococcus thermolithotrophicus*

Peter Willekens, Karl O. Stetter*, Antoon Vandenberghe, Erik Huysmans and Rupert De Wachter†

Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium and *Institut für Biochemie, Genetik und Mikrobiologie, Universität Regensburg, Universitätsstrasse 31, D-8400 Regensburg, FRG

Received 29 April 1986

The nucleotide sequences of 5 S rRNA from archaeobacteria *Methanobacterium tindarius* and *Methanococcus thermolithotrophicus* were determined. They can be fitted into a secondary structure model that shows features specific for archaeobacterial 5 S RNAs, as documented in a previous study. Reconstruction of a phylogenetic tree on the basis of sequences from 15 archaeobacterial species presently investigated allows one to discern five branches diverging at deep evolutionary levels within this primary kingdom.

Archaeobacteria 5 S rRNA Nucleotide sequence Secondary structure Molecular evolution

1. INTRODUCTION

The primary structure of 5 S RNA has been determined hitherto for approx. 400 species (238 sequences known at the beginning of 1985 are reviewed in [1]). This extensive set of data has formed the basis for evolutionary studies (e.g. [2–6]). It has also allowed the elaboration of detailed secondary structure models for 5 S RNA on a comparative basis [7–10]. One drawback to the presently available data set is that the distribution of examined species over the three primary kingdoms is rather skewed, since 236 of the reported sequences are from eukaryotic cytoplasmic ribosomes, 138 from eubacteria, 16 from plant organelles, and only 13 [6,11–19] from archaeobacteria. This is a disadvantage for comparative structural studies because it is precisely among the eukaryotes that 5 S RNA secondary structure is most conserved, while it is most variable within the archaeobacterial primary

kingdom [6]. Hence it is the examination of additional archaeobacterial sequences that is most likely to reveal new facts about 5 S RNA secondary structure. Here we report the nucleotide sequence of 5 S RNA from the methanogenic archaeobacteria *Methanobacterium tindarius* [20] and *Methanococcus thermolithotrophicus* [21].

2. MATERIALS AND METHODS

2.1. Culture conditions

M. tindarius DSM 2278 was grown at 37°C for 2 days after 1% inoculation in the mineral base of medium 3 [22] in the presence of 0.5% (v/v) methanol [20] under stirring. Growth was in a 100 l enamel-protected fermentor (Bioengineering, Wald, Switzerland) pressurized with 2 bar N₂/CO₂ (80:20) and yielded 70 g of wet cells. *Mc. thermolithotrophicus* DSM 2095 was grown at 65°C for 7 h after 1% inoculation in the mineral base of medium 3 [22] under stirring. In this case the fermentor was pressurized with 2 bar H₂/CO₂ (80:20) [21] and flushed with 5 l/min of the same gas. Cell yield was 200 g wet wt.

† To whom correspondence should be addressed

2.2. 5 S RNA isolation and sequencing

Batches of 0.5 g freeze-dried cells were used for the preparation [6] of 5 S RNA, yielding about 200 µg material. Nearly the entire sequence of both 5 S RNAs was deduced by 3'-terminal labeling with [5'-³²P]pCp, partial chemical degradation and gel electrophoresis [23]. In the case of *M. tindarius*, the 5'-terminal nucleotide was determined by high-pressure liquid chromatographic identification of the nucleoside bisphosphate present in an alkaline hydrolysate of unlabeled 5 S RNA [24]. In the case of *Mc. thermolithotrophicus*, the 5'-end group and the adjacent nucleotides were identified by 5'-terminal labeling of the dephosphorylated 5 S RNA with [γ-³²P]ATP [25], followed by nuclease degradation and gel elec-

trophoresis [26]. All separations on sequencing gels were performed at 55°C in order to avoid band compression effects.

2.3. Dendrogram construction

A dissimilarity matrix was computed from the alignment shown in fig.1. The dissimilarity D_{AB} between a pair of sequences A and B is defined as

$$D_{AB} = -\frac{3}{4} \ln \left[1 - \frac{4}{3} \left(\frac{S}{I+S} \right) \right] \left(1 - \frac{G}{T} \right) + \frac{G}{T} \quad (1)$$

where I is the number of alignment positions containing identical nucleotides, S is the number of substitutions, G is the number of deletions in sequence A with respect to B or vice versa, and $T = I + S + G$. In contrast to the computation method

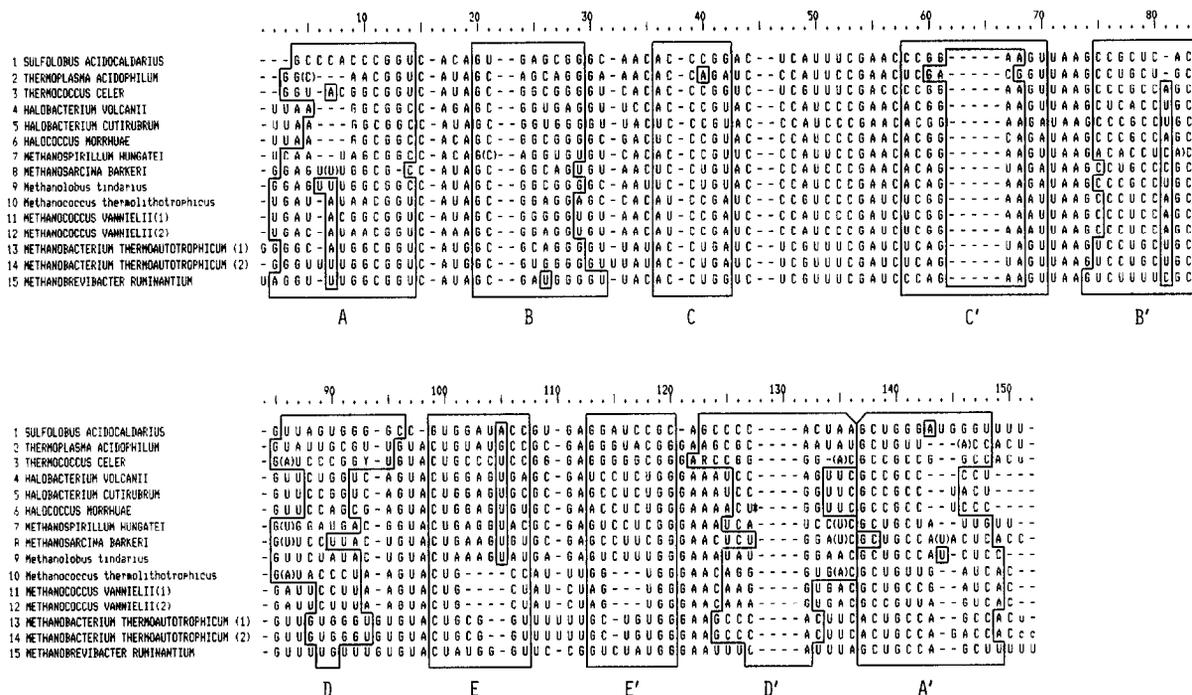


Fig.1. Alignment of archaeobacterial 5 S RNA sequences. Species names printed in capitals correspond to previously published sequences [6,11–19]. Variant sequences in a single species are listed separately if determined after purification of the individual 5 S RNA or its gene. In *Thermococcus celer* 5 S RNA, U·A/C·G heterogeneity at positions 93–123 was found upon sequencing of a mixture. Identical 5 S RNA sequences were found in *S. acidocaldarius* [13] and *S. solfataricus* [16], as well as in *Halobacterium cutirubrum* [14] and *H. halobium* [15]. Length heterogeneity is indicated by terminal residues printed in lower case characters. For the species examined in the present work (nos 9,10) no heterogeneity was detected. Alignment positions that are empty in all listed sequences are required to accommodate residues in 5 S RNAs from other primary kingdoms (see text). The asterisk at position 127 corresponds to a 108-nucleotide insertion in *H. morrhuae* 5 S rRNA [11]. Boxes labeled A-A', B-B', etc. enclose double-stranded areas of the secondary structure model followed in [1].

of D_{AB} followed in previous papers [3,5,6,27], all deletions or insertions in sequence A with respect to B were given the same weight, regardless of their length. Comparison of the sequences in the following imaginary alignment

```

ACUGGU - - - CA
  •   •   •   •
  •   •   •   •
- CAGAUUUUCA
  
```

would give the values $I = 5$, $S = 2$, $G = 2$, $T = 9$. The first term of eqn 1 accounts for substitutions and comprises a correction factor for multiple mutations per site [28]. The second term accounts for deletions or insertions. The standard deviation on the first term of eqn 1, s_D , is given by

$$s_D = \left[\frac{\left(1 - \frac{G}{T}\right)^2 \cdot S \left(1 - \frac{S}{I+S}\right)}{\left[1 - \frac{4}{3} \left(\frac{S}{I+S}\right)\right]^2 (I+S)^2} \right]^{1/2} \quad (2)$$

which is the expression proposed in [28], multiplied by $(1 - G/T)$, the fraction of compared sites comprising nucleotides in both sequences.

Starting from the dissimilarity matrix, sequences were clustered into a tree by the weighted pairwise grouping method [29].

3. RESULTS AND DISCUSSION

3.1. Primary structure of the examined 5 S RNA

Fig.1 shows an alignment of 5 S RNA sequences from 15 archaeobacterial species that have been hitherto published, including the two reported here. The alignment is comparable with the 1985 edition of the 5 S RNA sequence compilation [1] except that the total length has been increased from 147 to 152 positions. Of the 5 extra positions, 32 and 110 are required to accommodate nucleotides occurring in archaeobacterial sequences. Position 16 is required to accommodate nucleotides occurring in certain *Mycobacterium* 5 S RNA sequences (unpublished). In addition, areas 1–8 and 143–152 each contain an extra position with respect to [1] in order to optimize the homology among archaeobacterial sequences. Alignment in these terminal areas should be considered as tentative.

3.2. Secondary structure model

The boxes superimposed on the alignment in fig.1 enclose complementary areas that can be distinguished in all 5 S RNAs and that lead to a 5-helix model [7–10] followed in the 5 S RNA sequence compilation [1]. Characteristic variations on this basic design have been shown [6] to exist in 5 S RNAs of the three primary kingdoms. The secondary structure models presented in fig.2 incorporate the extra base pairing potential brought to light by the latter study, hence they do not coincide exactly with the delimitation of complementary areas indicated in fig.1. Also shown in fig.2 are potential secondary structure switches that may endow 5 S RNAs with flexibility according to a hypothesis set forth elsewhere [30].

The following characteristics are shared by the two models shown in fig.2. Helix A contains an irregularity consisting of an asymmetric internal loop (fig.2a) or a bulge (fig.2b), with the extra base occurring in the 5'-proximal strand. Such an irregularity is found in most archaeobacterial 5 S RNAs but is not strictly characteristic for this primary kingdom, since it also occurs in certain eubacterial 5 S RNAs such as those of Actinomycetes and relatives (unpublished). Helix B bears two bulges on opposite strands. This characteristic is shared by 5 S RNA of most archaeobacteria and eubacteria [6]. The internal loop I_2 separating helices D and E is asymmetrical and its strands show partial complementarity which, if it gave rise to base pairing, could convert this loop into a two-bulge system. The latter characteristic is shared by most archaeobacterial and eukaryotic 5 S RNAs but not by eubacterial 5 S RNAs which have a symmetrical I_2 loop [6,30].

In nearly all 5 S RNAs, area I_1 -C of the model can be fitted into a structure showing a bulge on the 3'-proximal strand of helix C, but also into one or more alternative structures [8,30]. In *M. tindarius* 5 S RNA (fig.2a) the bulge could turn into an asymmetrical internal loop, or it could switch to the opposite strand. In *Mc. thermolithotrophicus* 5 S RNA (fig.2b), an alternative structure at the I_1 -C boundary is also possible, although in this case there is less opportunity for base pairing at the expense of loop I_1 . This sequence shows the most obvious potential for additional base pairing in loop H_1 , which could be reduced in size to 3 nucleotides in favour of a longer C-helix. The ex-

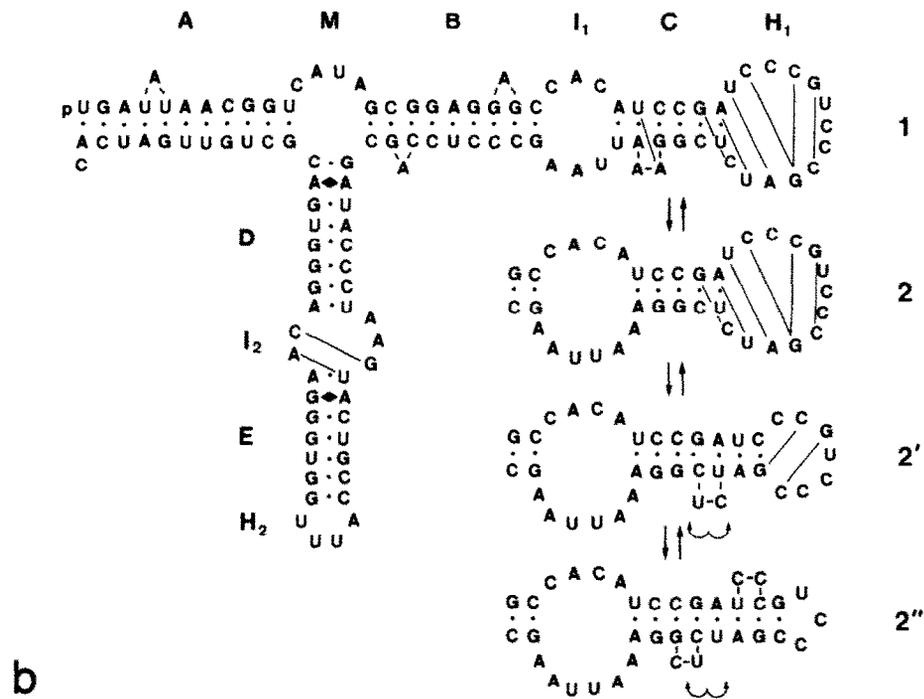
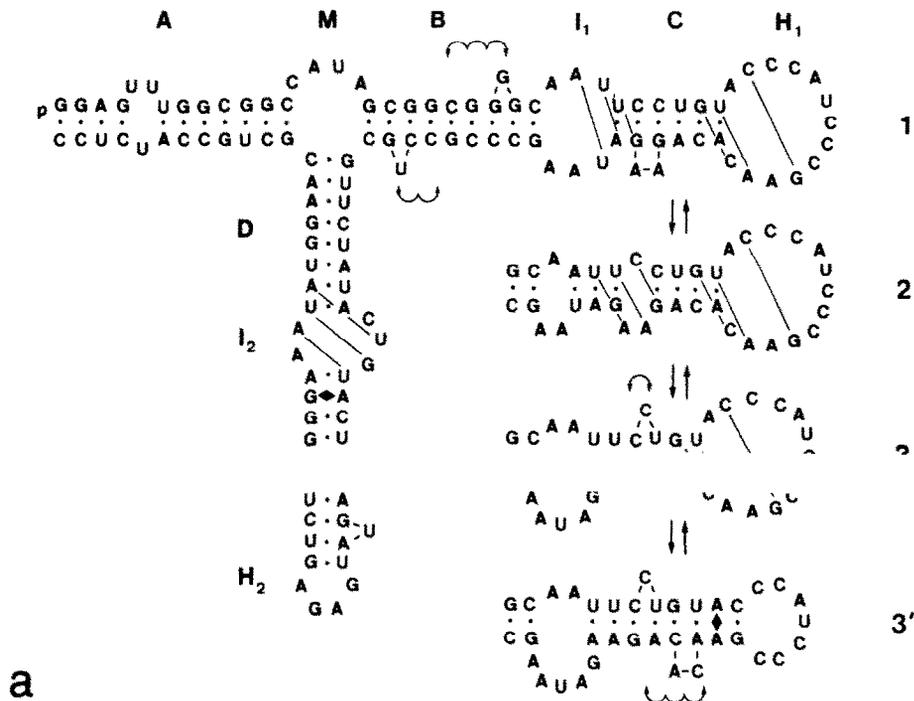


Fig.2. Secondary structure model for *Methanobolus tindarius* (a) and *Methanococcus thermolithotrophicus* (b) 5 S rRNAs. With respect to the base pairing scheme superimposed on the alignment (fig.1), helices B and E are extended at the expense of the adjacent loops I₁ and I₂. Odd base pairs that may be present within a helix are symbolized by a losenge. Bulges that may 'migrate' along a helix [30] are indicated by curled arrows pointing at the positions that can be occupied. Alternative base pairs that may form at certain loop-helix boundaries are indicated by lines connecting involved bases. Some of the possible results are drawn for area I₁-C-H₁. Structures 1,2 (and 3 in case a) would result from alternative pairing at the I₁-C boundary. Additional base pairing at the C-H₁ boundary would convert structure 2 to 2' and 3 to 3'. Even more base pairing in loop H₁ could transform structure 2' into 2'' in case b. The set of combinations shown is not exhaustive.

istence of base pairing in loop H₁ has been postulated on the basis of a cross-linking study [31] but the proposed structure cannot be extended to all known 5 S RNA sequences [30] unless one invokes the existence of adjacent odd base pairs (pairs other than G·C, A·U and G·U).

3.3. Evolution of archaeobacteria as derived from 5 S RNA structure

Fig.3 shows a dendrogram constructed as described in section 2. Since the alignment (fig.1) is considered to be arbitrary in the terminal areas, only positions 9–142 were taken into account for D_{AB} computation. The tree shows an early divergence of *Sulfolobus*. Four other branches diverge at nearby D_{AB} levels ranging from 0.64 to 0.54. One branch leads to *Thermoplasma*, the other three lead to the methanogen orders [22] Methanococcales, Methanobacteriales and Methanomicrobiales. The halobacteria form an

offshoot of the Methanomicrobiales branch. *Thermococcus celer* shows a high degree of similarity to the Methanococcales. This is surprising since this species was assigned to the Thermoproteales, an order thought to be phylogenetically closer to the Sulfolobales than to any of the other archaeobacteria [32]. In fig.3 *Thermococcus* is proposed to belong to a tentative order named Thermococcales. A previous tree constructed on the basis of 10 archaeobacterial species [6] also showed *Sulfolobus* as the earliest diverging species, but a different order of divergence for the remaining four branches. The overlapping standard deviations on the estimated branching points indicate that the determined branching order is actually not statistically significant. Also, the branching order changes when adjacent insertions or deletions are counted as separate mutations (see section 2) or when terminal alignment positions are included in D_{AB} computation. We conclude that the presently

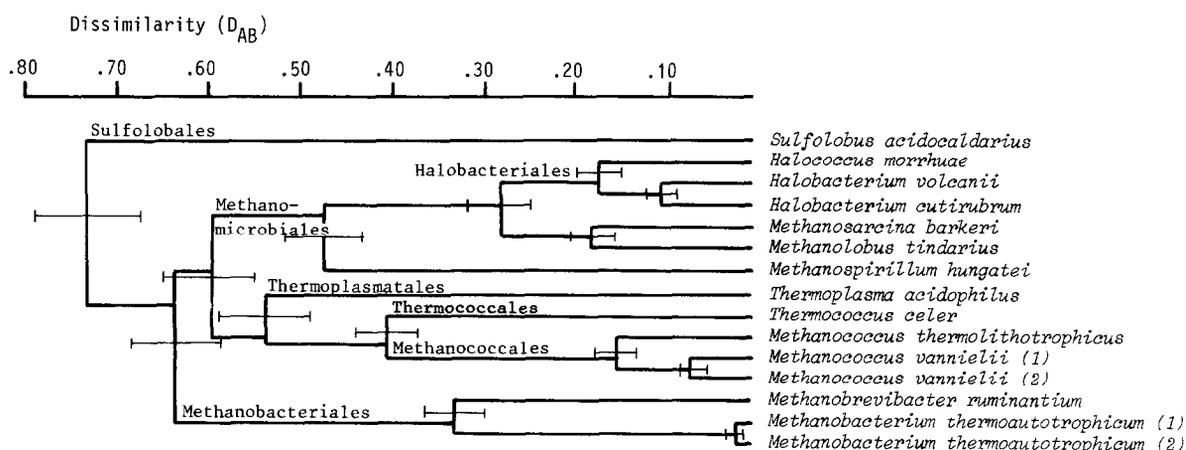


Fig.3. Tree constructed from archaeobacterial 5 S rRNA sequences. The tree was constructed as described in section 2, from positions 9–142 of the sequence alignment (fig.1), the extremities of the alignment being considered unreliable at present. The standard deviation on each D_{AB} value calculated according to eqn 2 is indicated by a bar.

available set of sequences and the clustering method employed do not allow one to resolve the branching order of *Thermoplasma*, Methanomicrobiales, Methanobacteriales and Methanococcales.

ACKNOWLEDGEMENTS

P.W. holds an IWONL scholarship and E.H. is an FKFO research assistant. Our research was subsidized in part by the Fonds voor Kollektief Fundamenteel Onderzoek and by the Deutsche Forschungsgemeinschaft (Schwerpunkt: Methanogene Bakterien).

REFERENCES

- [1] Erdmann, V.A., Wolters, J., Huysmans, E. and De Wachter, R. (1985) *Nucleic Acids Res.* 13, r105-r153.
- [2] Hori, H., Lim, B.-L. and Osawa, S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 820-823.
- [3] Vandenberghe, A., Wassink, A., Raeymaekers, P., De Baere, R., Huysmans, E. and De Wachter, R. (1985) *Eur. J. Biochem.* 149, 537-542.
- [4] Lane, D.J., Stahl, D.A., Olsen, G.J., Heller, D.J. and Pace, N.R. (1985) *J. Bacteriol.* 163, 76-81.
- [5] De Wachter, R., Huysmans, E. and Vandenberghe, A. (1985) in: *Evolution of Prokaryotes* (Schleifer, K.H. and Stackebrandt, E. eds) pp.115-141, Academic Press, London.
- [6] Willekens, P., Huysmans, E., Vandenberghe, A. and De Wachter, R. (1986) *Syst. Appl. Microbiol.* 7, 151-159.
- [7] Nishikawa, K. and Takemura, S. (1978) *J. Biochem.* 84, 259-266.
- [8] De Wachter, R., Chen, M.-W. and Vandenberghe, A. (1982) *Biochimie* 64, 311-329.
- [9] Böhm, S., Fabian, H. and Welfle, H. (1982) *Acta Biol. Med. Germ.* 41, 1-16.
- [10] Delihias, N. and Anderson, J. (1982) *Nucleic Acids Res.* 10, 7323-7344.
- [11] Luehrsen, K.R., Nicholson, D.E., Eubanks, D.C. and Fox, G.E. (1981) *Nature* 293, 755-756.
- [12] Luehrsen, K.R., Fox, G.E., Kilpatrick, M.W., Walker, R.T., Domdey, H., Krupp, G. and Gross, H.J. (1981) *Nucleic Acids Res.* 9, 965-970.
- [13] Stahl, D.A., Luehrsen, K.R., Woese, C.R. and Pace, N.R. (1981) *Nucleic Acids Res.* 9, 6129-6137.
- [14] Fox, G.E., Luehrsen, K.R. and Woese, C.R. (1982) *Zbl. Bakt. Hyg. I. Abt. Orig. C3*, 330-345.
- [15] Mankin, A.S., Kagramanova, V.K., Belova, E.N., Teterina, N.L., Baratova, L.A. and Bogdanov, A.A. (1982) *Biochem. Int.* 5, 719-726.
- [16] Dams, E., Londei, P., Cammarano, P., Vandenberghe, A. and De Wachter, R. (1983) *Nucleic Acids Res.* 11, 4667-4676.
- [17] Wich, G., Jarsch, M. and Böck, A. (1984) *Mol. Gen. Genet.* 196, 146-151.
- [18] Jarsch, M. and Böck, A. (1985) *Mol. Gen. Genet.* 200, 305-312.
- [19] McDougall, J. and Nazar, R.N. (1986) *Biochem. Biophys. Res. Commun.* 134, 1167-1174.
- [20] König, H. and Stetter, K.O. (1982) *Zbl. Bakt. Hyg. I. Abt. Orig. C3*, 478-490.
- [21] Huber, H., Thomm, M., König, H., Thies, G. and Stetter, K.O. (1982) *Arch. Microbiol.* 132, 47-50.
- [22] Balch, W.E., Fox, G.E., Magrum, L.J., Woese, C.R. and Wolfe, R.S. (1979) *Microbiol. Rev.* 43, 260-296.
- [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] Ursi, D., Vandenberghe, A. and De Wachter, R. (1982) *Nucleic Acids Res.* 10, 3517-3530.
- [26] Dams, E., Vandenberghe, A. and De Wachter, R. (1983) *Nucleic Acids Res.* 11, 1245-1252.
- [27] Chen, M.W., Anné, J., Volckaert, G., Huysmans, E., Vandenberghe, A. and De Wachter, R. (1984) *Nucleic Acids Res.* 12, 4881-4892.
- [28] Kimura, M. and Ohta, T. (1972) *J. Mol. Evol.* 2, 87-90.
- [29] Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy*, Freeman, San Francisco.
- [30] De Wachter, R., Chen, M.-W. and Vandenberghe, A. (1984) *Eur. J. Biochem.* 143, 175-182.
- [31] Thompson, J.F., Wegnez, M.R. and Hearst, J.E. (1981) *J. Mol. Biol.* 147, 417-436.
- [32] Zillig, W., Schnabel, R., Tu, S. and Stetter, K.O. (1982) *Naturwissenschaften* 69, 197-204.