Volume 9 Number 21 1981 Nucleic Acids Research

Nucleotide sequence of the Aspergillus nidulans mitochondrial gene coding for the small ribosomal subunit RNA: homology to E. coli 16S rRNA

Heinrich G.Köchel and Hans Küntzel

Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Str. 3, D-3400 Göttingen, GFR

Received 14 September 1981

ABSTRACT

The complete primary structure of the 1437 bp gene coding for mitochondrial 15S rRNA and its flanking regions was determined by Maxam-Gilbert sequencing of cloned HindIII fragment H3 of _A. nidulans_ mtDNA.

The gene product reveals significant homology (59 %) to _E. coli_ 16S rRNA, and the potential secondary structures of both rRNA molecules are very similar, except that the hairpin structures 7, 8 and 30 of the Brimacombe 16S rRNA model are deleted, and that two sequences of 8 and 31 nucleotides are inserted in the mitochondrial species.

INTRODUCTION

The 32 kb circular genome of _Aspergillus nidulans_ (1) is currently being studied by Maxam-Gilbert sequencing (2) of cloned (3) restriction fragments (4). The complete nucleotide sequences of HindIII fragments H3 (4273 bp) and H5 (2210 bp), and partial sequences of Eco RI fragments E3 (6.2 kbp) have revealed the following transcriptional order of genes: URF 4 (unidentified polypeptide reading frame), arginine tRNA, asparagine tRNA, ATPase subunit 6, small ribosomal subunit RNA (S-rRNA), URF 1, cytochrome oxidase subunit 3, tRNA gene cluster I, large ribosomale subunit RNA (L-rRNA) and tRNA gene cluster II (ref. 1,4 and unpublished data).

Here we report the nucleotide sequence of the S-rRNA (15S rRNA) gene and its flanking regions, and we discuss the homology of primary and potential secondary structure between mitochondrial 15S rRNA and _E. coli_ 16S rRNA (5,6).

MATERIALS AND METHODS

Plasmid DNA panH3 (HindIII fragment H3 of mtDNA ligated into pBR322) (3) was isolated as described (4). Restriction fragments were separated by preparative electrophoresis on 5 % polyacrylamide gels, eluted from crushed gel slices (2) and purified by DE 52 chromatography (7).

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

Sequencing was performed according to the method of Maxam and Gilbert (2), as modified by Smith and Calvo (7). DNA fragments were dephosphorylated with calf intestinal phosphatase (Boehringer, grade I, dialyzed prior to use against 100 mM Tris-HCl, pH 7.4, 1 mM $MgCl_2$, 50 % glycerol), and end-labelled with γ-^{32}P-ATP (Amersham, 5 Ci/µmol) using T4 polynucleotide kinase (BioLabs). Labelled fragments were either cleaved by suitable restriction endonucleases (BioLabs), or were denatured in the presence of 25 mM mercuric hydroxyde (7). Chemical cleavage products were separated on 60 x 20 x 0.04 cm acrylamide gels, and autoradiography was performed with Osray T4 films (Agfa) for 1 to 4 days at -20° without intensifying screen.

RESULTS AND DISCUSSION

The upper part of Fig. 1 shows a map of the rRNA-tRNA gene region of A. nidulans mtDNA, covering about 17 % of the genome, and the sequencing strategy is indicated in the lower part.

The starting material, cloned HindIII fragment H3, was re-isolated from

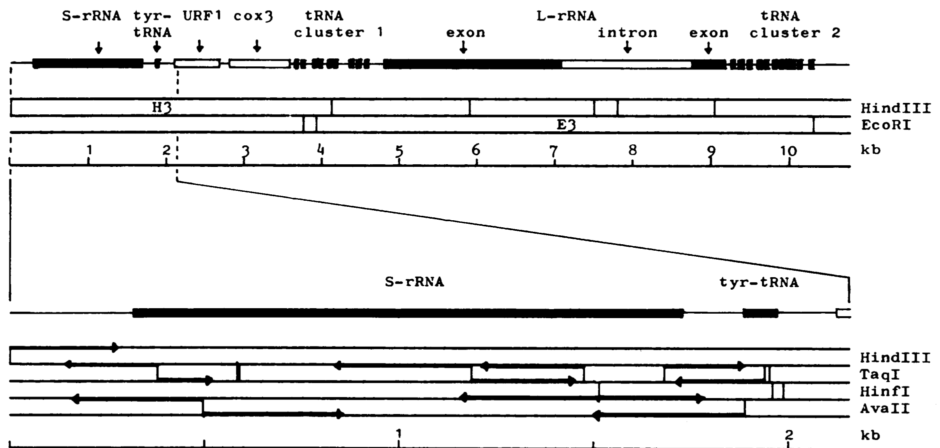

Figure 1. Upper part: map of the rRNA-tRNA gene region of A. nidulans mtDNA. The nucleotide sequence of the tRNA gene clusters I and II and the 5'- and 3'-terminal regions of the large ribosomal subunit RNA (L-rRNA) gene has been reported (4). The map of intron-exon regions of the L-rRNA gene is based on electron microscopy of DNA/RNA hybrids (8) and partial sequence data. The ribosomal spacer region including a tyrosine tRNA gene, an unidentified polypeptide reading frame (URF1) and the gene coding for cytochrome oxidase subunit 3 (cox3) has completely been sequenced (Köchel and Küntzel, unpublished). All genes are on the same DNA strand and are transcribed from left to right. Lower part: Restriction fine map and sequencing strategy of the small ribosomal subunit RNA (S-rRNA) gene.

a HindIII digest of plasmid panH3 (3) and further digested with TaqI, HinfI or AvaII. In some experiments individual subfragments were re-isolated from H3 or plasmid panH3, end-labelled and re-digested. In other experiments the fragments of a whole digest were end-labelled, strand-separated and resolved by gel electrophoresis (2).

The nucleotide sequence starting from the HindIII site at the junction H5/H3 (9) is shown in Fig. 2.

The S-rRNA coding region is operationally defined by primary and secondary structure homology of its transcription product with E. coli 16S rRNA (5,6), but the exact start and end of the mature 15S rRNA gene remains to be determined by comparison with terminal rRNA sequences. The putative gene

```
                                                                                                   AAGCTTT ATTAAAAAAT
                                                                                                   HindIII

AAGGAATTTA ACAGAAACTA AAAATTATTA AAACGTTATA ATAACGTTTT TTATATAATT GCGAACATTA CTCCTATATT TAGTATTAAT TTATAATTAT

AAATGTTTTT AACATAAACG GGCTTAATTA TATCTTTATA CTAATATGAT AAGTAAAAAA TTTTTTTTTT TCAAAAAGTG TAAAAAAAAA TATAATCTTG

CACTATATAT TATAAAAAAA CAACATAAAA TATAAGATAA AAAAGTATTT ATCTATACAA TAAAGGTATA GAGTAATATC CCCATTTCCC TATACTATGA

        10         20         30         40         50         60         70         80         90        100
AAAATTTTGA GTTTGATGAT GGCTCTGATT GAACGCTGTC CAAGTACTTG ACACATGCTA ATCGAACGAC TAATTAGTTA TTAAACTAAG TAGTAGTGGT
                                                                   TaqI
       110        120        130        140        150        160        170        180        190        200
GTACAGGTGA GTAAAAGATA ATTTGGCTAC CTTAAAGTAA GGGCAAAAAT CCCTTATAAA AAGAAAGGAA AAAGAAGGTC CGCTTTAAGA TGTTAATTAT
                                                                                AvaII
       210        220        230        240        250        260        270        280        290        300
TATCACGGTG AGTAGTAGTA AAGGTAATGA CTTTACTAGC TAAATCCGTA GTCGTGACTG AGAGGTCGAT CGACCACATT GGGTCTGAAA AAACCCCAAT
                                                                        TaqI TaqI
       310        320        330        340        350        360        370        380        390        400
GCGTTTTAGT ACAGCAGTGA GGAATATTGG TCAATGGCCG AAAGGCGAAC CAGTAACTTG GAAGAATGAA AGTGTATTAT AATAATACAA TAACGATTAT
                                       HaeIII
       410        420        430        440        450        460        470        480        490        500
ATCGTATAAA ATTCTAAATA GAATAATGAT AATGACAATT TTCTATTTAT AAGTCTTGAC CAAACTACGT GCCAGCAGTC GCGGTAATAC GTAGAAGACT

       510        520        530        540        550        560        570        580        590        600
AGTGTTAATC ATCTTTATTA GGTTTAAAGG GTACCTAGAC GGTAAATTAA ACTCTAAATG AGTACTTATT TACTAGAGTT TTATGTAAGA AGGAAGAAGC
                                                                                                        AluI
       610        620        630        640        650        660        670        680        690        700
TCTGGAGTAG TGATATAATA CGTATATACC AGAGAGACTG GTAACGGCGA AGGCATCCTT CTATGTAAAA ACTGACGTTG AGGGACGAAG GCTTGGGTAG

       710        720        730        740        750        760        770        780        790        800
CAAGAAGGAT TAGATACCCT AGTAGTCCAA GCAGACAATG ATGAATGTCA TAGACTAGAA AAAGTCGTTT AGACTATAAA TTTAGTCTAT AAATGAAAGT

       810        820        830        840        850        860        870        880        890        900
GTAAGCATTC CACCTCAAGA GTAATATGGC AACATATAAA CTGAAATCAT TAGACCGTTT CTGAGACCAG TAGTGAAGTA TGTTATTTAA TTCGATGATC
                                                                                                  TaqI MboI
       910        920        930        940        950        960        970        980        990       1000
CGCGAAAAAC CTTACCACAG TTTGAATATT TTACAAGCGC TGCACGGCTG TTTTTAGTTA ATGTCGTGAG ATCTGGTTAA GTCCTTTAAT TAACGAAAAC
                                       HhaI
      1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
CCTCACTTTA TTTGCATTTA TAAAGTTGAT CGCCTTTATA TTGGTTAGAT AATAGGGATT AAGACAAGTC ATCATGGCCT TAATACTGTG GGCTATAGAC
                                                                                HaeIII
      1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
GTGCCACATA TGCCTTTACA AAGGGATGCG ATTTTGTGAA ATTGAGCTAA TCCCCCAAAA AAGGATATAA TATGGATTGT AGTCTGTAAC TCGACTACAT
                                                Alu I                                              TaqI
      1210       1220       1230       1240       1250       1260       1270       1280       1290       1300
GAATAAGGAA TTACTAGTAA TCGTGAATCA CCATCGTCAC GGTGAATTAA AAATCAGCTG GGTACTAACC ACTCGTCAGG CGCTGAAAGA AGTATGTGCA
                           HinfI                            AluI                  HhaI
      1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
AGAAGTTTGA TTTACTTATA TTTATAATAT ATATAATCAG TTATATATAT TTATAAGTTA AATTTTCGCA TGCATGACTT TGATTGGTGT TAAGTCGAAA
                                                                                                    TaqI
      1410       1420       1430   1437
TATGGTTCGT GTAATGGAAA TTGCACGGGA TGAATTAACC ACTTAGCAAT AAATAAATGC ATATATATAC CAAATTTTAT ATACTTTTGT ATATAAAAGA

TGGAACAAAT ATATATATAG CATTATCTTT AAATAAAGGT AAATTGAAAA AATTTTTTCA ATCTGGTTCA AATCCAGGAA AAGATAAAAT TTTAAGGAAG
```

Figure 2. Nucleotide sequence of the non-transcribed strand of the S-rRNA gene and flanking regions. The putative start and end of the mature rRNA coding region (based on the alignment of Fig. 4) and the start of the tyrosine tRNA gene at nucleotide 1595 are marked.

length of 1437 bp and the position of the gene within HindIII fragment H3 (the termini separated by 217 and 2629 bp, respectively, from the two HindIII sites) is in excellent agreement with a previously published map based on physical mapping of rRNA/DNA hybrids (8).

The S-rRNA gene is separated from flanking genes coding for ATPase subunit 6 and tyrosine tRNA (sequences to be published elsewhere) by AT-rich spacer regions of 347 and 157 bp lengths, respectively, which contain stop codons in all three frames and several inverted repeats (Fig. 3).

The hairpin structures shown in Fig. 3 could possibly be formed in a precursor transcript of the S-rRNA gene and may serve as signals for processing endonucleases.

Fig. 4 shows a sequence alignment of E. coli 16S rRNA (5) and A. nidulans mt 15S rRNA (inferred from the sequence of Fig. 2). The two molecules were superimposed to maximal sequence homology (59 % of all base vs. base positions) without considering secondary structure homologies in the first instance, and the mt 15S rRNA was than back-folded in analogy to the 16S rRNA model of Brimacombe (6). The striking coincidence of primary and secondary structure homology seen in Fig. 4 further supports the previously noted phylogenetic conservation of rRNA structure (10).

Sequences are not only conserved in some of the apparently singel-stran-

```
                                                                T T
                                                               T   T
                                                                C·G
                                                                A·T
                                                                T·A
                                                                A·T
                                                                T·A
                                                                A·T
                                                                T·A
                                                                T·A
                                                                T·A
                                                              AAT·AGATGG
                                                                        A
                                                              CCA·TAAACA
                                                                T·A
       A A A                                                    A·T
       A   G                                                    T·A
        A·T        C                  AA GT                     A·T
        C·G      A   T                A   A                     T·A
        T T      C   A                 A·T                      A·T
        T·A       G·T                  A·T                      T·A
        T·A      TT·A                  A·T                      A·TA
        T·A      C   T                 T·A                      C·G
        T·A       CT·A                 A·T                      G·C
        T·A        A·T                 G·C                      T·A
        T·A        A·T                 A·T                      A·T
  A     T·A        T·A                AT·A               B     AA·T
        T·A        A·T                 A·T                     AT·A
AAAAAATT·A_____T·AAAAAAACAACATAAAAT·ACAATAAA          AATAAA·TCTTTAAA
```

Figure 3. Inverted repeats in nucleotide sequences flanking the S-rRNA gene. A, upstream sequence, nucleotides -146 to -37. B, downstream sequence, nucleotides 1448 to 1533.

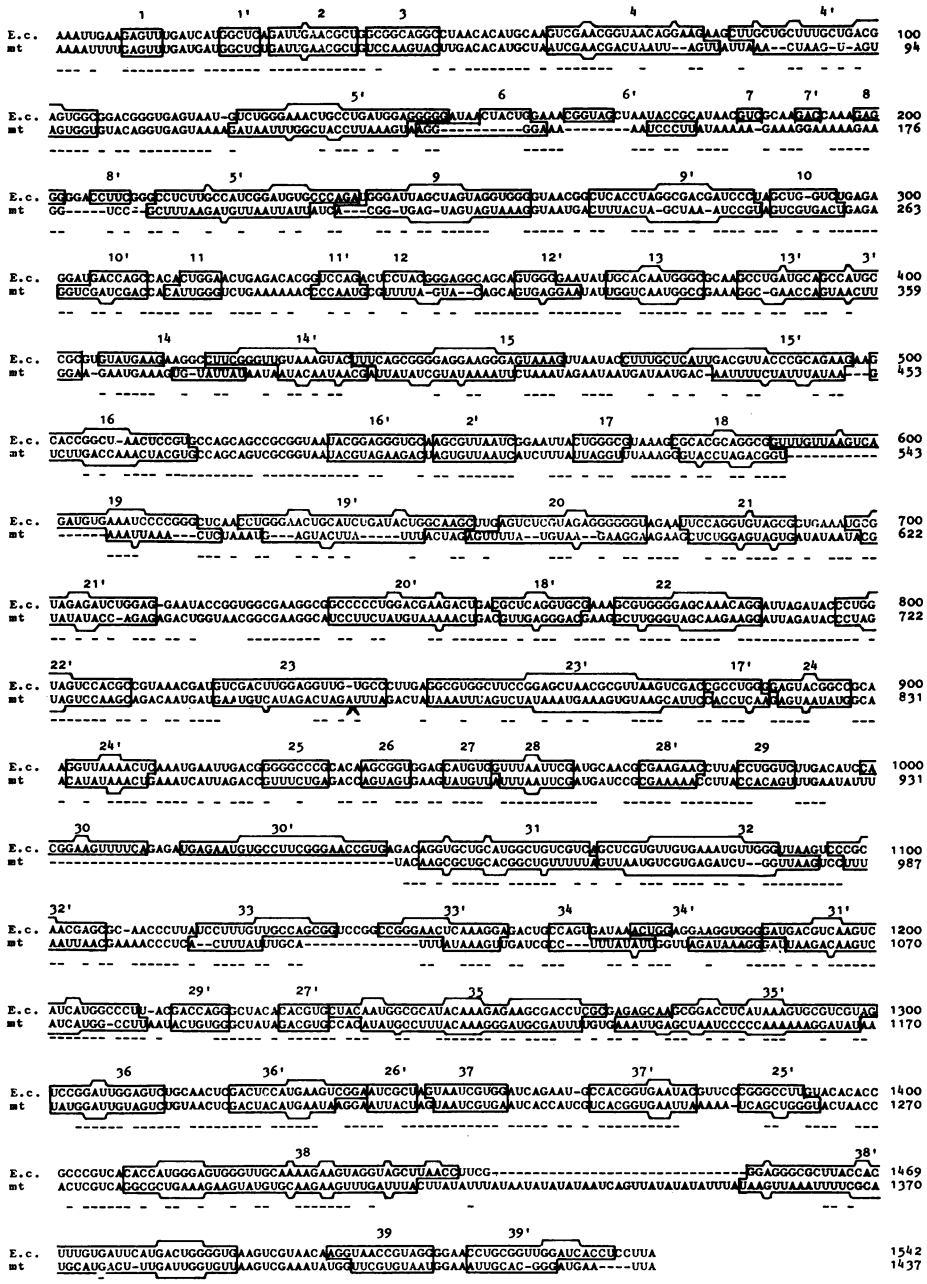

Figure 4. Sequence alignment of E. coli 16S rRNA (5) and A. nidulans mt 15S rRNA. The boxed regions are base-paired according to the secondary structure model of Brimacombe (6).

ded regions (e.g., hairpin loops 1, 16, 22, and stem-connecting loops 4/5 and 38/39), but also in some base-paired regions (e.g., stems 1, 2, 28 and 37) whereas other double-stranded regions are conserved in their secondary, but not in their primary structure (e.g., stems 3, 23 and 38).

The general architecture of the two molecules, as characterized by spacing and lengths of double-stranded regions, and by the position and size

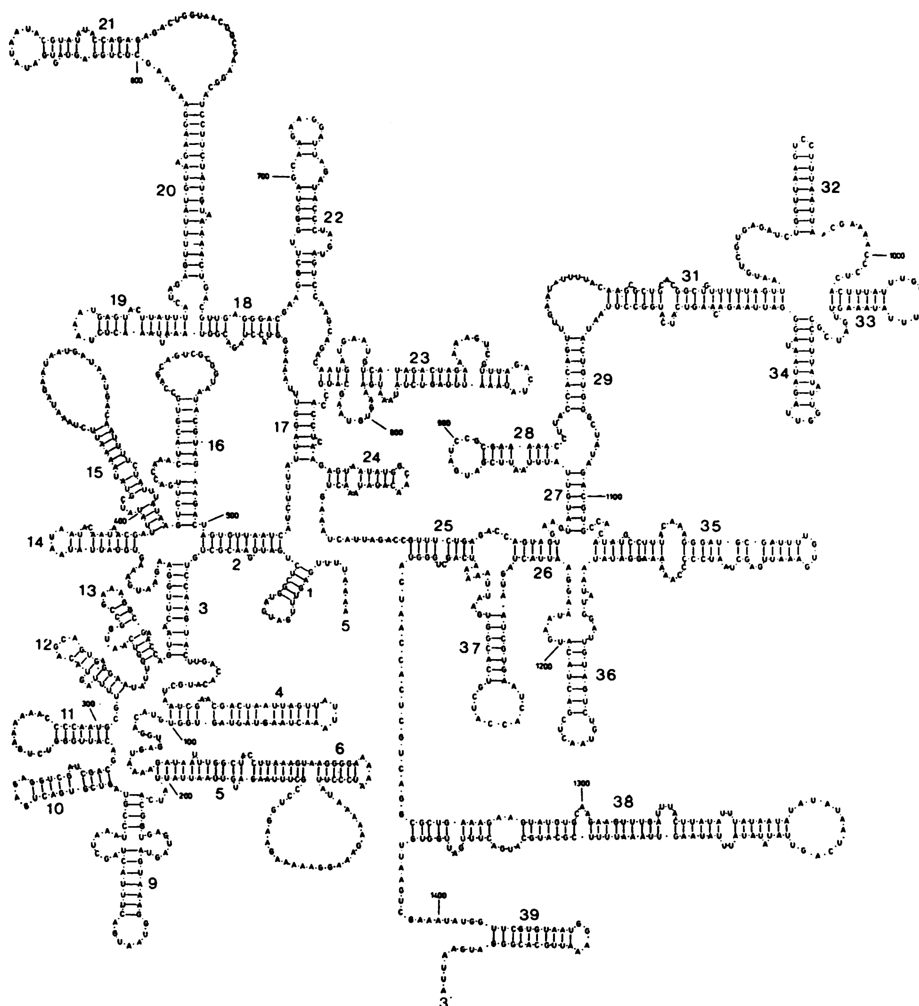

Figure 5. Secondary structure model of mitochondrial 15S rRNA. The molecule was folded in analogy to E. coli 16S rRNA (6). The numbers refer to the base-paired regions of Fig. 4.

of stem-interrupting loops, is largely conserved, except that the hairpin structures 7, 8 and 30 of the bacterial molecule are deleted from the mitochondrial species. Furthermore, hairpin loop 38 is enlarged in mt 15S rRNA by an insert of 31 nucleotides (mainly A and U residues), and another eight nucleotides (760 to 767, omitted from Fig. 4) are looping out from stem 23 of the mitochondrial molecule at the indicated position.

Fig. 5 shows the secondary structure model of A. nidulans mt 15S rRNA based on the alignment of Fig. 4. The hairpin structures 32 and 38 are presented in an alternative configuration.

The 3'-terminal sequence -CCUCCU- of bacterial and chloroplast 16S rRNA (5,11) interacting with mRNA during polypeptide chain initiation (12) is absent from A. nidulans mt 15S rRNA as well as from all other nuclear (13,14) and mitochondrial (15-17) S-rRNAs, indicating that a bacteria-like mechanism of mRNA recognition does not operate in mitochondria. On the other hand, the mitochondrial molecule is significantly more related to bacterial than to nuclear-coded S-rRNAs, both in primary and secondary structure (18), and a phylogenetic tree analysis of all published S-rRNA gene sequences (18,19) supports the endosymbiotic eubacterial origin of fungal and animal mitochondria.

REFERENCES

1. Küntzel, H., Köchel, H.G., Lazarus, C.M. and Lünsdorf, H. (1981) in Mitochondrial Genes, Slonimski, P., Borst, P. and Attardi, G., eds., Cold Spring Harbor Laboratory, in press.
2. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology, Grossmann, L. and Moldave, K., eds. Vol. 65, pp. 499-595, Academic Press, New York.
3. Bartnik, E., Biderman, A.W., Hahn, U., Küntzel, H. and Stepien, P.P. (1981) Molec. Gen. Genet. 182, 332-335.
4. Köchel, H.G., Lazarus, C.M., Basak, N. and Küntzel, H. (1981) Cell 23, 625-633.
5. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801-4805.
6. Brimacombe, R. (1980) Biochem. Int. 1, 162-171.
7. Smith, R.D. and Calvo, J.M. (1980) Nucl. Acids Res. 8, 2255-2274.
8. Lazarus, C.M., Lünsdorf, H., Hahn, U., Stepien, P.P. and Küntzel, H. (1980) Molec. Gen. Genet. 177, 389-397.
9. Stepien, P.P., Bernard, U., Cooke, H. and Küntzel, H. (1978) Nucl. Acids Res. 5, 317-330.
10. Zwieb, C., Glotz, C. and Brimacombe, R. (1981) Nucl. Acids Res. 9, 3621-3640.
11. Schwarz, Z. and Kössel, H. (1980) Nature 283, 739-742.
12. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71,1342-1346.
13. Robtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skryabin, K.G. and Bayev, A.A. (1980) Nucl. Acids Res. 8, 5779-5794.
14. Salim, M. and Maden, B.E.H. (1981) Nature 291, 205-208.
15. Sor, F. and Fukuhara, H. (1980) C.R. Acad. Sc. Paris 291, 933-936.
16. Eperon, I.C., Anderson, S. and Nierlich, D.P. (1980) Nature 286, 460-467.
17. Van Etten, R.A., Walberg, M.W. and Clayton, D.A. (1980) Cell 22, 157-170.

18. Küntzel, H. and Köchel, H.G. (1981) Nature, in press.
19. Küntzel, H. (1981) Int. J. Microbiol. Hyg., in press.