Volume 12 Number 7 1984 Nucleic Acids Research

Complete nucleotide sequence of the 23S rRNA gene of the Cyanobacterium, *Anacystis nidulans*

Susan E.Douglas and W.Ford Doolittle

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, B3H 4H7 Canada

Received 30 November 1983; Revised and Accepted 2 March 1984

ABSTRACT

The nucleotide sequence of the Anacystis nidulans 23S rRNA gene, including the 5'- and 3'-flanking regions has been determined. The gene is 2876 nucleotides long and shows higher primary sequence homology to the 23S rRNAs of plastids (84.5%) than to that of E. coli (79%). The predicted rRNA transcript also shares many secondary structural features with those of plastids, reinforcing the endosymbiont hypothesis for the origin of these organelles.

INTRODUCTION

The hypothesis that chloroplasts originated from endosymbiotic bacteria is widely accepted. Overwhelming evidence supporting this idea (1) has come from nucleic acid hybridisation studies of cyanobacterial and Euglena gracilis chloroplast rRNAs (2); from ribosomal reconstitution experiments (3) ; from T_1-oligonucleotide cataloguing of 16S rRNAs of plastids and cyanobacteria (4,5,6) and from comparative primary sequence data of the rRNAs and/or their genes. Most of the latter form of evidence has involved comparisons of chloroplast and E. coli rRNAs. However, a contemporary organism more likely to be related to the ancestor of plastids is the cyanobacterium Anacystis nidulans.

In view of this, we have studied one of the two rRNA gene clusters of A. nidulans, and have shown it to be similar to those of plastids in the complement and order of genes, except that there is no physically distinct 4.5S rRNA (7). The spacer between the 16S and 23S rRNA genes contains genes for $tRNA^{Ile}$ and $tRNA^{Ala}$ (8), as do those of Z. mays (9), N. tabacum (10), and E. gracilis (11) chloroplasts. The A. nidulans tRNAs themselves show higher primary sequence homology to those of plastids than to those of E. coli. The 5S rRNA gene of A.

nidulans also is more homologous to plastid 5S rRNA genes than to the E. coli 5S rRNA gene (7). Sequence data on the 16S rRNA gene of A. nidulans (12) shows the same relationship.

MATERIALS AND METHODS

Materials

Restriction enzymes were purchased from Bethesda Research Laboratories and New England Biolabs and used as suggested by the manufacturer. Phage M13 cloning vectors, primer and host cells were obtained from New England Biolabs. DNA polymerase I (large fragment) was from Boehringer Mannheim. [α-^{32}P]-dATP was synthesized according to the method of Walseth and Johnson (13).

DNA Preparation and Restriction Mapping

The recombinant plasmid containing the rRNA gene cluster of A. nidulans was propagated on E. coli JF1754 and the DNA purified as described previously (8). The restriction map for enzymes Eco RI, Bam HI, Sal I, Pst I, Sma I, Xma III and Hind III is shown in Figure 1 (a). The 23S rRNA gene lies on Sma I fragments of 1.4 and 2.3 kb, Hind III fragments of 0.48, 0.84 and 1.23 kb, and Sma I-Hind III fragments of 0.98 and 1.2 kb. These fragments were purified from low-melting-point-agarose and

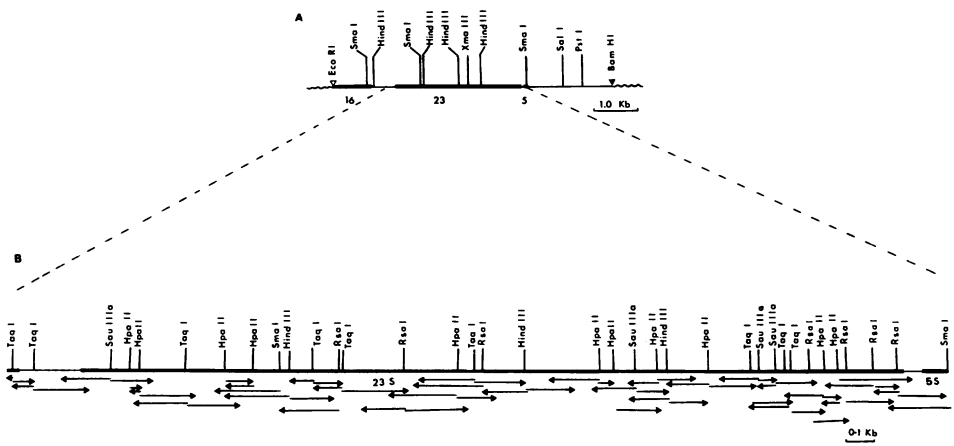


Figure 1. (a). Restriction Map of the 6.3 kb Cloned rRNA Gene Cluster of A. nidulans. The 16S, 23S and 5S rRNA genes are indicated by heavy bars. (b). Strategy for Determining the Sequence of the 23S rRNA Gene and Flanking Regions. Arrows indicate the origin and direction of sequencing.

subcloned into phage M13 using the restriction enzymes Taq I, Sau IIIa, Hpa II and Rsa I, as shown in Figure 1 (b).

DNA Sequencing

Sequencing was done by the dideoxynucleotide chain termination method using the phage M13 system (14). DNA sequencing gels were 33 x 40 x 0.3 cm and 8% in polyacrylamide (19:1, acrylamide:bis) and contained 8 M urea, 50 mM each of tris base and boric acid, 1 mM EDTA (pH 8.3). The gels were run at 1500 V and several loadings were performed to obtain overlapping sequences. Most regions were sequenced several times and in both orientations.

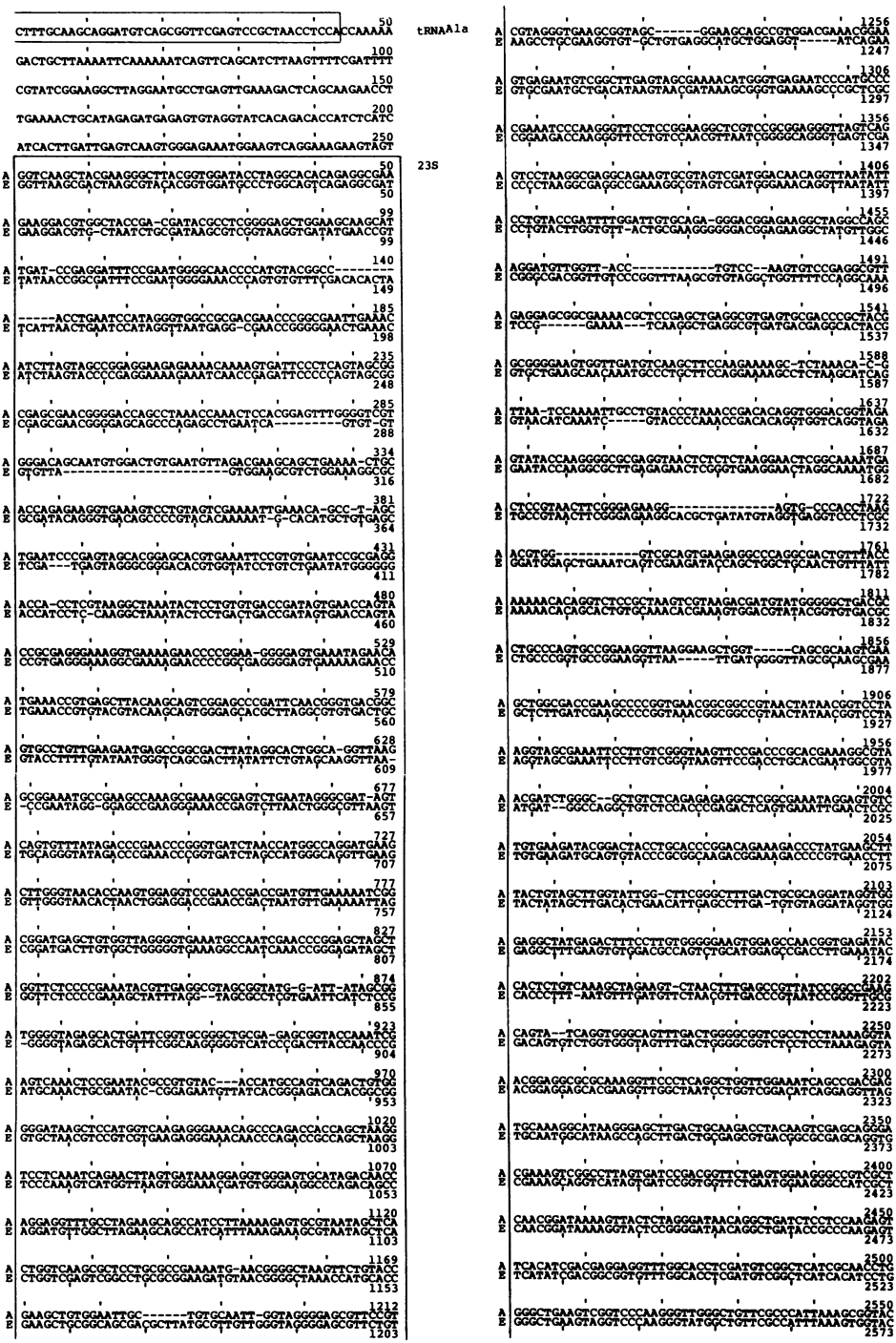
RESULTS AND DISCUSSION

The nucleotide sequence of the 23S rRNA gene and the 5'- and 3'-flanking regions is presented in Figure 2. The termini of the 23S rRNA gene were determined by comparison with the published sequence of the E. coli 23S rRNA (15). The gene is 2876 nucleotides long, slightly shorter than that of E. coli (2904 nucleotides) and longer than that of N. tabacum chloroplast (2804 nucleotides), excluding the 4.5S rRNA region. It is 54.2% G+C.

The 23S rRNAs of A. nidulans and plastids show several additional nucleotides relative to that of E. coli, as shown in Table 1. In C. reinhardii chloroplast, there is a 7S and 3S rRNA 5' to the 23S rRNA gene (16). These small rRNAs are homologous to the 5' terminus of other chloroplast 23S rRNAs and

Table 1. Insertions in 23S rRNAs of A. nidulans and Plastids Relative to E. coli. (A) A. nidulans, (Z) Z. mays, (N) N. tabacum, (C) C. reinhardii. The co-ordinates refer to the location of the additional nucleotides in the A. nidulans sequence relative to the E. coli sequence.

Location	A	Z	N	C
270-278	9	-	-	9
292-313	22	25	26	25
1245-1249	5	5	5	?
1496-1501	6	7	7	?
1601-1605	5	6	6	?
1834-1838	5	5	5	?



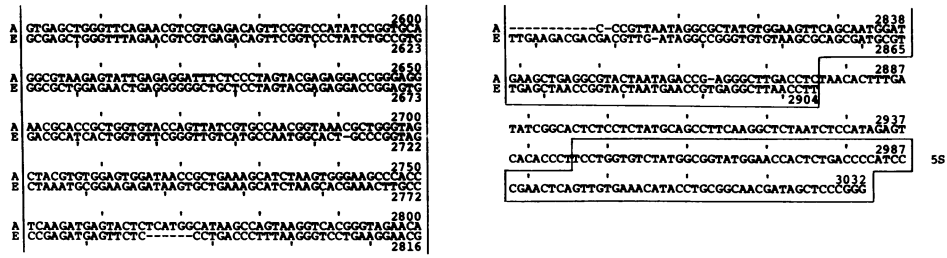


Figure 2. Nucleotide Sequence of the 23S rRNA Gene and Flanking Regions. The 3'-terminus of the tRNA^{Ala}, the 23S rRNA gene and the 5'-terminus of the 5S rRNA gene are boxed. The sequence of the E. coli 23S rRNA gene (E) is given below that of A. nidulans (A).

A. nidulans, as shown in Figure 3. In fact, the insertion of 22 nucleotides in A. nidulans and plastids relative to E. coli (position 292-313) corresponds to the spacer between the 7S and 3S rRNA genes of C. reinhardii. On the other hand, the second spacer, that between the 3S and 23S rRNAs, is present in the E. coli 23S rRNA gene. The deletion of the 7S-3S rRNA gene spacer in E. coli 23S rRNA is therefore puzzling. The region cannot be of any functional importance since it is deleted in E. coli, post-transcriptionally removed in C. reinhardii chloroplast and forms variable secondary structures in A. nidulans and plastids (see Figure 4).

The 3S equivalent is very highly conserved in secondary structure, although primary sequence homology is quite variable (see Table 2). A. nidulans shows by far the highest homology to the 3S region of C. reinhardii chloroplast, and also shares a 9-nucleotide insertion in the 7S region relative to the other organisms (see Figure 3). This may reflect a closer evolutionary relationship between A. nidulans and the plastids of the unicellular green algae, such as C. reinhardii and Euglena gracilis. Both A. nidulans and E. gracilis plastids share the feature of uninterrupted spacer tRNAs (8,17). However, the 23S rRNA gene of C. reinhardii chloroplast is interrupted by a 940 bp intervening sequence near the 3' end (18). A. nidulans 23S rRNA contains no intervening sequence here, but neither do the Z. mays or N. tabacum plastid 23S rRNAs (19,20).

The 3' terminus of the 23S rRNA gene of A. nidulans does

```
E.  GGGAGCAGCCCAGAGCCTGAATCA          GTG TGTGTGTTA                          GTGGAAGC
Z.  ..........T.A.C.G....AAC          G.G.T....G..G.GCAATACAAGCGTTGTGCTGCTA  .GC.....
A.  ....C.....T.A.C.AAACTC..CGGAGTTT G.G.TC...G.AC.GCAATGT  GGACTGTGAATGTTA .AC.....
C.  ....A.....T.A. ...ATG..GCAAGATGTAG.G.TC...G.AGGACAACATAAAAATCGCTATTTTTAATAC.....
N.  ..........T.A.C.G....AAC          .G.T....G.AG.GCAATACAAGCGTCGTGCTGCTA  .GC....A
                  7S                                spacer
```

```
E.  GTCTGGAAAGGCGCGCGATACAGGGTGACAGCCCCGTACACAAAAA TGC ACATGCTGTGAG CTC GA  TGAGTAG
Z.  .GT.. .GT.C...A.CC..G.T..CT.T..T..A...GC.G.T.GCAT. .. ....TAC . ...T..CCC......
A.  AG...A... CT..A.C.G.G.A.....A..T..T...GT.G....T..AA... ..C . .. ..GAATCCC......
C.  AG...A. TCCT..A.C...G.T.....A..T..A...GT .....GAAAATTTA.A.T.TT.T.. AATCCC......
N.  CG.CC...T.CT..A.CC..G.T..C..A..T..A...GC.G...GCAT. ..TA...TAT . ...T..CCC......
                  3S                                spacer
```

Figure 3. Alignment of 7S-3S rRNA Region of Chlamydomonas reinhardii Chloroplast With Homologous Region of 23S rRNAs. (E) E. coli, (Z) Z. mays, (A) A. nidulans, (C) C. reinhardii, (N) N. tabacum. This region corresponds to nucleotides 246-395 of the A. nidulans sequence.

not contain a sequence homologous to the 23S-4.5S spacer of Z. mays and N. tabacum chloroplasts (7). The region where the 23S-4.5S spacer occurs in plastids is found near a helix which is altered in both plastids and A. nidulans relative to E. coli.

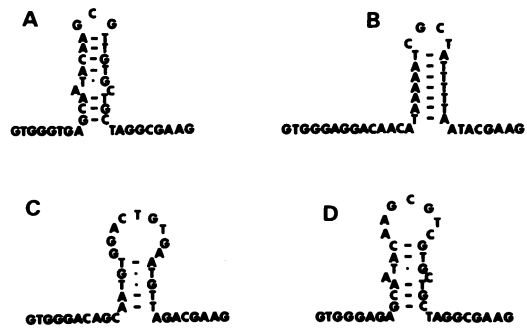


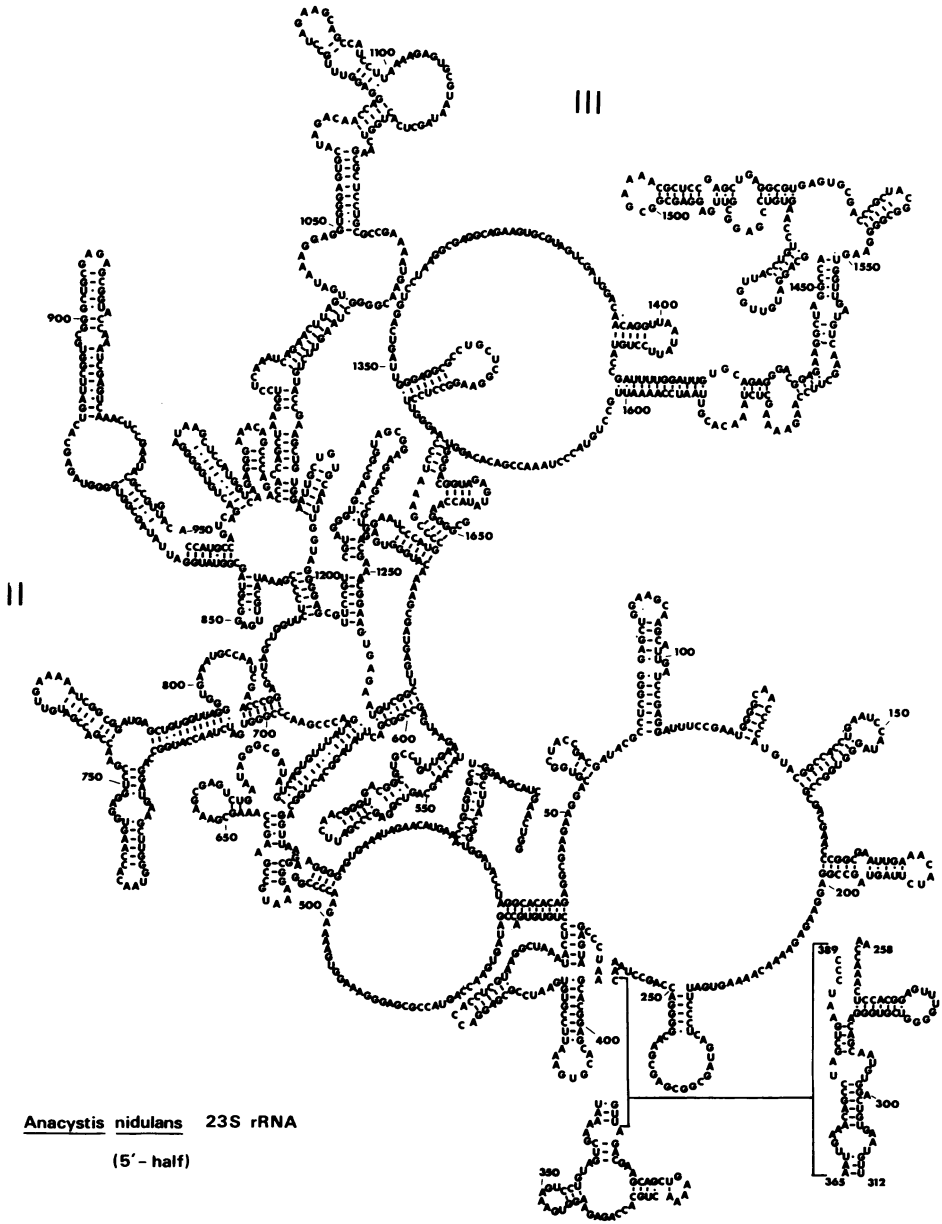
Figure 4. Potential Secondary Structures of the Regions Homologous to the 7S-3S rRNA Spacer of C. reinhardii. (A) Z. mays, (B) C. reinhardii, (C) A. nidulans, (D) N. tabacum.

Table 2. Percent Sequence Homology of the 3S rRNA Gene of C. reinhardii and its Homologs. (C) C. reinhardii, (E) E. coli, (Z) Z. mays, (A) A. nidulans, (N) N. tabacum.

	E	A	Z	C	N
E. coli		63	58	58	56
A. nidulans			67	85	71
Z. mays chloroplast				71	79
C. reinhardii chloroplast					75

Other helices which are altered (or even missing) in A. nidulans and Z. mays chloroplast relative to E. coli occur at positions 131 (where the helix is absent), 257-312/365-389 (where the equivalent of the 7S-3S and 3S-23S spacers of C. reinhardii chloroplast are found) and 1721 (where the helix is shortened) of the A. nidulans sequence (see Figure 5). For the most part, the secondary structures of the rest of the 23S rRNA molecules of E. coli, A. nidulans and chloroplasts are superimposable. The potential helices are of three types: highly conserved (involving only Watson-Crick base pairs), irregular (containing G:U pairs) and highly irregular (containing non-Watson-Crick base pairs, especially G:A pairs and bulged residues). Some helices are highly conserved phylogenetically, while others are extremely variable but maintain secondary structural features.

In the following calculations, each potential helix in the A. nidulans 23S rRNA molecule was tabulated according to whether it was highly conserved, variable or highly irregular with respect to the 23S rRNAs of E. coli and Z. mays chloroplast (data not shown). Of the highly conserved helices (greater than 85% homology among the three organisms), A. nidulans 23S rRNA shares approximately equal homology with those of E. coli and Z. mays chloroplast. Of the variable helices, 80% show more similarity to those of the Z. mays chloroplast 23S rRNA than to to those of the E. coli 23S rRNA. Of the highly irregular helices, 67% show more similarity to those of the Z. mays chloroplast 23S rRNA than to those of the E. coli 23S rRNA. Of the 13 positions where bulged residues occur in A. nidulans, nine are common to all three organisms. Of the remaining four,


III
II
I
Anacystis nidulans 23S rRNA
(5′- half)

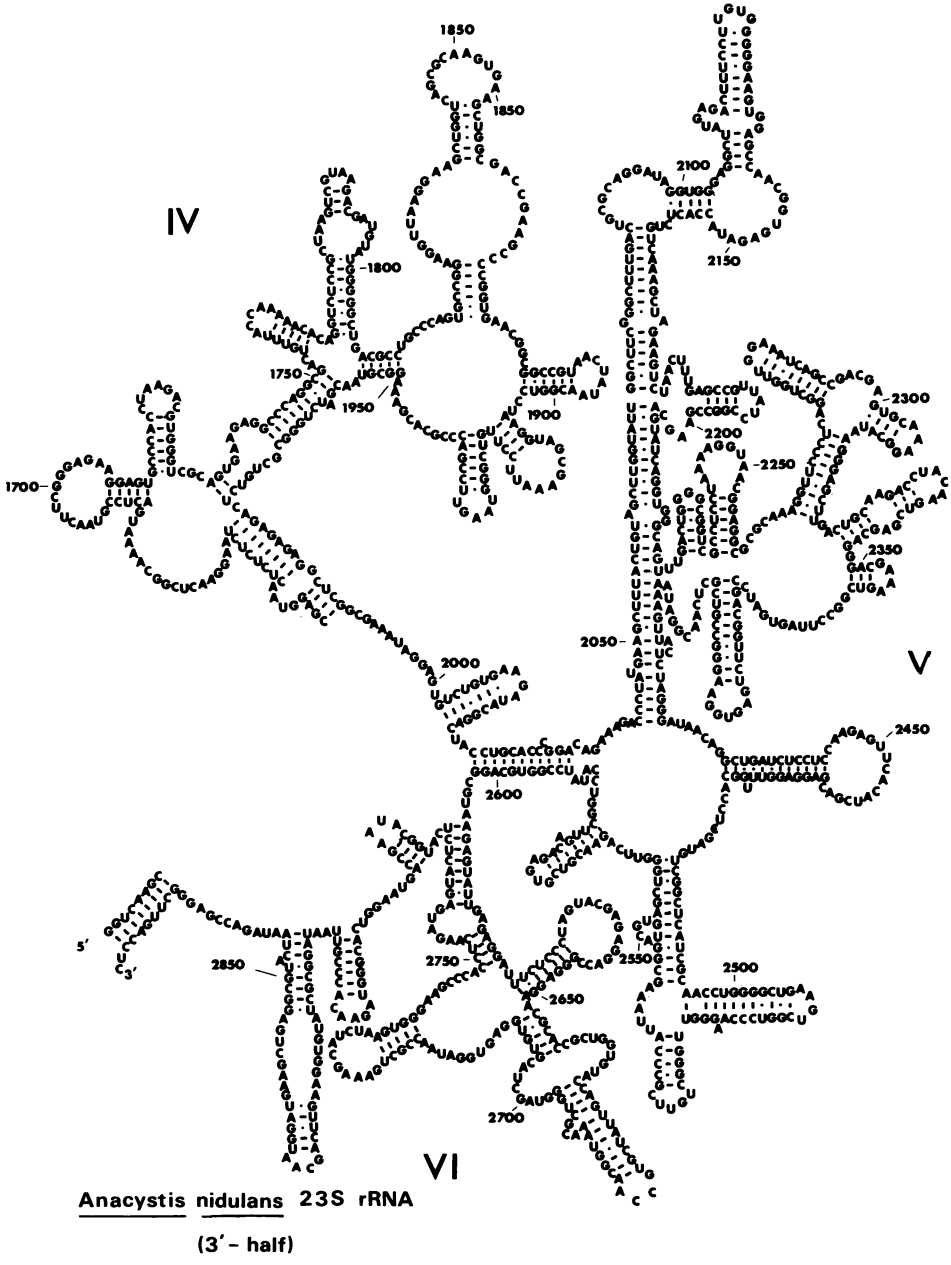


Figure 5. Potential Secondary Structure of the 23S rRNA of A. nidulans (adapted from the model of Noller et al [21]). The region between nucleotides 258 and 389 is highly variable in different organisms and does not conform to Noller's model. An alternative secondary structure is shown in brackets.

one is shared with Z. mays chloroplast, one with E. coli and two are different in all 3 organisms. Those nucleotides found in single-stranded regions of the molecule tend to be highly conserved, although this is not as pronounced as in the case of the 16S rRNA (21).

Certain regions of the 23S rRNA molecule are postulated to interact with proteins, tRNAs or other rRNAs. Nucleotides 72-83 of the E. coli 5S rRNA are complementary to nucleotides 143-154 of the E. coli 23S rRNA (15) and an interaction between these two regions has been proposed. There is no complementarity between maize 23S and 5S rDNA in this region. In fact, part of this region has been deleted in the 23S rRNA genes of both A. nidulans and Z. mays chloroplast. A region complementary to nucleotides 68-74 of maize 5S rRNA has been found between nucleotides 1846 and 1855 of maize 23S rRNA (19). Comparable sequence homology between this region of A. nidulans 5S rRNA and nucleotides 1933-1938 is also present (see Figure 6). The region of the 23S rRNA molecule of E. coli corresponding to that delimited by nucleotides 2220-2347 in the A. nidulans 23S rRNA molecule, is also believed to interact with 5S rRNA and proteins L5, L18 and L25 (22). This region shows very similar secondary structure in A. nidulans and plastids and is also the region which is deleted in mammalian mitochondrial large subunit rRNA. Mammalian mitochondrial ribosomes lack a 5S rRNA.

Several interactions have also been postulated to occur between 16S and 23S rRNAs in both E. coli and Z. mays chloroplast (19). These interactions are supported by chemical modification studies and similar interactions could occur in A. nidulans (see Figure 6). Compensatory base changes have sometimes occurred in the A. nidulans sequence which maintain the interactions. In addition, some nucleotide substitutions strengthen these interactions. For example, an A to C change at position 2438 of the A. nidulans 23S rRNA molecule relative to that of E. coli allows additional base-pairing to occur with the 16S rRNA molecule.

The region corresponding to nucleotides 1962-1978 of A. nidulans 23S rRNA in E. coli has been postulated to interact with $tRNA_f^{Met}$. Very strong base-pairing could occur in E. coli

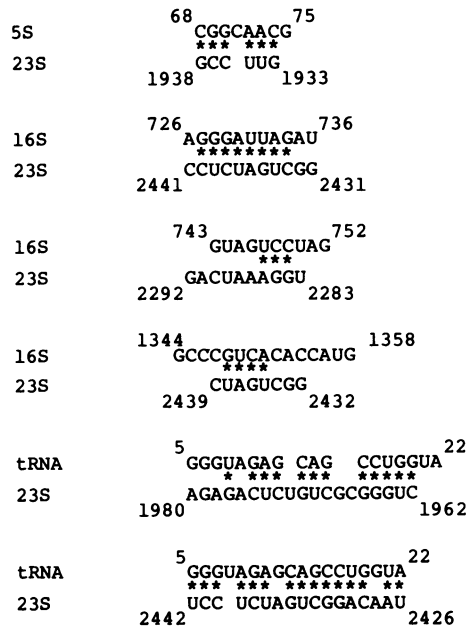


Figure 6. Potential Interactions Between 23S rRNA and 16S rRNA, 5S rRNA and $tRNA_f^{Met}$ in A. nidulans. Asterisks indicate nucleotides which could partake in base pairing.

(17 bp with a single-base bulge), but this is much weaker in both A. nidulans and Z. mays chloroplast (12 bp with 3 insertions and 1 mismatch, and 11 bp with 3 insertions and 2 mismatches, respectively). Much stronger base-pairing could occur in A. nidulans between $tRNA_f^{Met}$ and 23S rRNA between nucleotides 2426-2442. Strong base-pairing could also occur here between the 23S rRNAs and $tRNA_f^{Met}$s of both E. coli and Z. mays chloroplast. However, this is not substantiated by chemical modification or cross-linking studies and is therefore speculative.

A. nidulans and chloroplast 23S rRNAs can form very similar secondary structures in the regions believed to interact with proteins L24 and L1 in E. coli (22) as shown in Figure 7.

The region between nucleotides 2339 and 2562 is believed to be responsible for peptidyl transferase function in E. coli 23S rRNA. Very similar secondary structures can be formed by the large subunit rRNAs of A. nidulans, chloroplasts and yeast mitochondria. Chloramphenicol mutants map in this region in

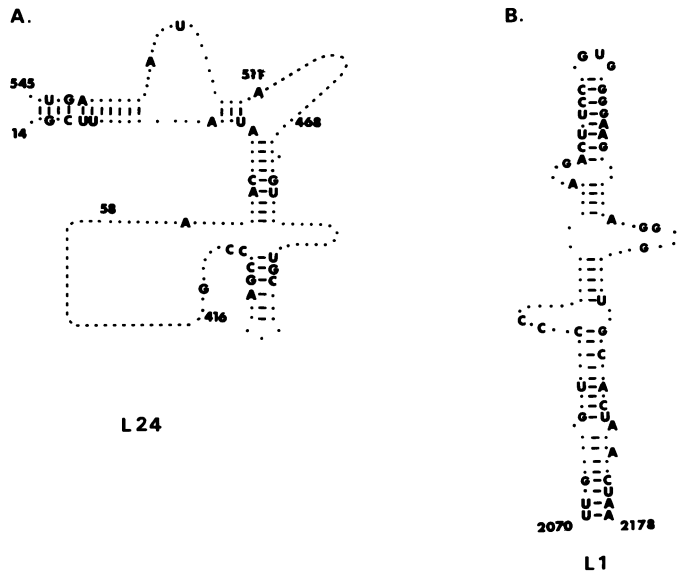


Figure 7. Potential Secondary Structure of the Regions of the 23S rRNA of A. nidulans Proposed to Interact with Proteins L1 and L24 (22). Dots represent identical nucleotides to those found in E. coli 23S rRNA. Letters indicate nucleotide differences.

yeast mitochondrial 23S rRNA (23) and have been found to contain single-base substitutions at positions corresponding to nucleotides 2424 and 2479 of the A. nidulans sequence. These positions are conserved in A. nidulans, E. coli and Z. mays chloroplast 23S rRNAs and lie in long stretches of identical nucleotides.

The 23S rRNAs of some plastids contain "hidden breaks" which may result in the formation of the 4.5S rRNA of higher plants and the 3S and 7S rRNAs of C. reinhardii. Broad bean chloroplast 23S rRNA has been shown to be unstable at high temperatures and to dissociate into three fragments (24). This may be a common feature of chloroplast 23S rRNAs, and the sequences in these organelles and in A. nidulans corresponding to the 7S-3S and 3S-23S rRNA spacers of C. reinhardii may also be removed. It is of interest that the 23S rRNA of A. nidulans also contains a hidden break which results in its dissociation into two fragments (25). Since the fragments formed are 0.17 and 0.88 daltons, the break is likely to occur after the first domain or within domain V. The exact location is under investigation. Cyanobacteria are

Table 3. Percent Sequence Homology of 23S rRNA Genes. The 23S rRNA genes from E. coli (E), A. nidulans (A), N. tabacum chloroplast (N) and Z. mays chloroplast (Z) were compared. The spacer between the 23S and 4.5S rRNAs of plastids was not included in the comparison.

	E	A	N	Z
E. coli		79	67	71
A. nidulans			85	84
N. tabacum				92

one of the few groups of prokaryotes known to contain hidden breaks in their 23S rRNA, another feature shared by chloroplast 23S rRNAs.

The 3'-flanking region of the 23S rRNA gene does not contain any typical eubacterial transcription stop or start signals (26) and transcription probably proceeds through the spacer and 5S rRNA gene, as is usual in eubacteria and some, if not all, plastids (27). Potential transcription termination signals have been located downstream from the 5S rRNA gene (7). The 5'- and 3'-flanking regions of the 23S rRNA gene can partake in base pairing. This occurs in E. coli and probably serves as a recognition signal for processing by RNase III (28). Base-pairing also occurs between the regions flanking the 23S rRNA (and 4.5S rRNA in those organisms which contain it) in chloroplasts (20).

In conclusion, from primary sequence studies (see Table 3), secondary structural features and physical characteristics of the large subunit rRNA of A. nidulans, it is clear that it shows more homology to chloroplast 23S rRNAs than to E. coli 23S rRNA, thus giving further support to the claim that plastids arose from cyanobacterium-like ancestors.

Acknowledgements

This work was supported by a grant from the Medical Research Council to W.F.D. and a post-graduate scholarship from the Natural Science and Engineering Research Council to S.E.D.

REFERENCES

1. Gray, M. W. and Doolittle, W. F. (1982) Microbiol. Revs. 46, 1-42.
2. Piggott, G. H. and Carr, N. G. (1972) Science 177, 575-582.
3. Gray, G. E. and Herson, D. E. (1976) Arch. Microbiol. 109, 95-99.
4. Bonen, L. and Doolittle, W. F. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 2310-2314.
5. Bonen, L. and Doolittle, W. F. (1976) Nature 261, 669-673.
6. Zablen, L. B., Kissil, M. S., Woese, C. R. and Buetow, D. E. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 2418-2422.
7. Douglas, S. E. and Doolittle, W. F. (1983) FEBS Lett., submitted.
8. Williamson, S. E. and Doolittle, W. F. (1983) Nucleic Acids Res. 11, 225-235.
9. Koch, W., Edwards, K. and Kossel, H. (1981) Cell 25, 203-213.
10. Takaiwa, F. and Sugiura, M. (1982) Nucleic Acids Res. 10, 2665-2676.
11. Graf, L., Kossel, H. and Stutz, E. (1980) Nature 186, 908-910.
12. Tomioka, N. and Sugiura, M. (1983) Mol. Gen. Genet. 184, 359-363.
13. Walseth, T. F. and Johnson, R. A. (1979) Biochim. Biophys. Acta. 562, 11-31.
14. Messing, J., Crea, R. and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321.
15. Brosius, J., Dull, T. J. and Noller, H. F. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 201-204.
16. Rochaix, J. D. and Darlix, J. L. (1982) J. Mol. Biol. 159, 383-395.
17. Orozco, E. M. Jr., Rushow, K. E., Dodd, J. R. and Hallick, R. B. (1980) J. Biol. Chem. 255, 10997-11003.
18. Rochaix, J. D. and Malnoe, P. (1978) Cell 15,661-670.
19. Edwards, K. and Kossel, H. (1981) Nucleic Acids Res. 9, 2853-2869.
20. Takaiwa, F. and Sugiura, M. (1982) Eur. J. Biochem. 124, 13-19.
21. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F. and Herr, W. (1981) Nucleic Acids Res. 9, 6167-6189.
22. Branlant, C., Krol, A., Srwidada, J. and Brimacombe, R. (1976) Eur. J. Biochem. 70, 483-492.
23. Dujon, B. (1980) Cell 20, 185-197.
24. Leaver, C. J. (1973) Biochem. J. 135, 237-240.
25. Doolittle, W. F. (1973) J. Bacteriol. 113, 1256-1263.
26. Gilbert, W. (1976) in RNA Polymerase, Losick, R. and Chamberlin, M. Eds., p193, Cold Spring Harbor Laboratory, New York.
27. Surzycki, S. J. and Rochaix, J. D. (1971) J. Mol. Biol. 62, 89-109.
28. Bram, R. J., Young, R. A. and Steitz, J. A. (1980) Cell 19, 393-401.
29. Kumano, M., Tomioka, N. and Sugiura, M. (1983) Gene 24, 219-225.

Note added in proof

While this manuscript was being reviewed, the nucleotide sequence of the 23S rRNA gene from the other rRNA gene cluster of A. nidulans was published (29). That sequence differs from that presented here in eight positions, mostlyin the 3'-terminal 600 nucleotides. These positions have been determined absolutely unambiguously by us by sequencing in both directions. This may reflect true differences between the two 23S rRNA gene copies.