J Mol Evol 17, 52–55 (1981)

Journal of Molecular Evolution

The Nucleotide Sequence of *Beneckea harveyi* 5S rRNA

Kenneth R. Luehrsen and George E. Fox

University of Houston, Central Campus, Department of Biophysical Sciences, Houston, Texas 77004, USA

Summary. The complete sequence of the 5S rRNA from the bioluminescent bacterium, *Beneckea harveyi* has been determined to be

p U G C U U G G C G $\overset{10}{A}$ C C A U A G C G $\overset{20}{A}$ U U-
G G A C C C A $\overset{30}{C}$ C U G A (U) C U U C $\overset{40}{C}$ A U U C C-
G A A C $\overset{50}{U}$ C A G A A G U G $\overset{60}{A}$ A C G A A U U A-
G $\overset{70}{C}$ G C C G A U G G U $\overset{80}{A}$ G U G U G G G G C $\overset{90}{U}$ U-
C C C C A U G U $\overset{100}{G}$ A G A G U A G G A $\overset{110}{C}$ A U C G-
C C A G G $\overset{120}{C}$ U (U)$_{OH}$.

Two sites of sensitivity to ribonuclease T_2 cleavage were identified; at A_{41} and either A_{54} or A_{55}. Comparison with existing sequence information from *Escherichia coli* and *Photobacterium phosphoreum* clarifies the amount of diversity among the bioluminescent bacteria and provides further insight into their phylogenetic position. Sequence heterogeneities were encountered and the importance of these in interpreting 5S rRNA data is discussed.

Key words: 5S rRNA – *Beneckea harveyi* – Procaryotic phylogeny – Sequence heterogeneity

Introduction

Like tRNA, 5S rRNA is ubiquitous in occurrence and of relatively small size. By contrast, post-transcriptionally modified nucleotides are rarely encountered and most procaryotic organisms have a characteristic 5S rRNA species, though sequence heterogeneity is observed. These favorable considerations imply that 5S rRNA may afford unique opportunities for the study of molecular evolution. Currently this is manifested in comparisons of 5S rRNA primary structure (Klotz et al. 1979; Hori and Osawa 1979) which allow determination of phylogenetic relationships. In the longer view, since structural variability is also known to occur in 5S rRNA (Fox and Woese 1975, Bellemare et al. 1973; Wrede and Erdmann 1973) it will be very important to conduct comparative studies of structural features in 5S rRNA. Such studies may provide insight into the mechanisms by which a macromolecule can evolve from one structural domain to another. In the present report we expand the existing comparative data base by reporting the primary sequence of 5S rRNA isolated from the free living bioluminescent marine bacterium, *Beneckea harveyi* strain 392 (formerly *Photobacterium fischeri* strain MAV (Reichelt and Baumann 1973) and apparently soon to be redesignated as *Vibrio harveyi* (Baumann, personal communication)).

Materials and Methods

Uniformly ^{32}P-labeled or non-labeled 5S rRNA were isolated from *Beneckea harveyi* strain 392 (kindly provided by Shiao-Chun Tu of the University of Houston) as described previously (Woese et al. 1975). Primary digests of uniformly labeled RNA were made with ribonuclease T_1, and pancreatic ribonuclease. The digestion products were separated by two-dimensional paper electrophoresis and located by autoradiography (Sanger et al. 1965; Woese et al. 1976). Subsequently each oligonucleotide product on the primary "fingerprint" was sequenced by established secondary and tertiary procedures (Woese et al. 1976; Uchida et al. 1974). The 3'-terminal oligonucleotides were located and sequenced in the following novel way (Luehrsen, Hecht and Fox, unpublished results). T_4 RNA ligase was used to attach [5' ^{32}P] pGp to the 3' end of unlabeled 5S rRNA. The 3'-terminal fragment(s) generated by subsequent complete

0022–2844/81/0017/0052/$ 01.00

ribonuclease T_1 digestion have a normal G terminus and are labeled. The 3'-oligonucleotide(s) can then be identified by the usual secondary procedures when utilized in conjunction with appropriate uniformly labeled marker oligonucleotides.

Following these preliminary characterizations sequencing studies were performed using 5'-labeled RNA. T_4 polynucleotide kinase was used to end label either dephosphorylated whole 5S rRNA or the large fragment produced by ribonuclease T_2 digestion of whole 5S rRNA (Meyhack et al. 1977; Jordan 1971). Sequencing ladders were generated by the enzymatic method (Donis-Keller et al. 1977). In addition to the usual enzymes two new enzymatic activities were employed, *Neurospora crassa* endonuclease (Krupp and Gross 1979) and Phy-M from *Physarum polycephalum* (Donis-Keller 1980). These activities facilitate the distinction between cytosine and uracil. Digestions were done as described (Donis-Keller et al. 1977) on aliquots of 5S rRNA, layered on to separate lanes of thin (0.5 mm) polyacrylamide gels (12% or 20%) and electrophoresed (1.6 kv or 2.0 kv) for varying lengths of time.

Results

The 5S rRNA sequence which has been established for *Beneckea harveyi* is shown in Fig. 1. Sufficient sequencing ladders were generated to encompass the entire length of the molecule and to insure the reproducibility of the result. A representative sequencing gel is illustrated in Fig. 2. The complete sequence of each oligonucleotide product generated by complete digestion with ribonuclease T_1 or pancreatic ribonuclease was also determined. Both of these catalogs were identical to those determined earlier (L. Zablen, unpublished and cited in Woese et al. 1975) and entirely consistent with the sequence as presented, except at two sites where heterogeneity is indicated by the catalog data. The large ribonuclease oligomer AUCUUCCAUUCCG was found to have an alternative form, ACUUCCAUUCCG, which differs by a single deletion of a uracil residue. Although these oligomers were present in approximately equal mole ratios in the catalogs, the shorter ACUUCCAUUCCG sequence was never observed in the gel ladders. This was fortunate as the presence of such length heterogeneity would be expected to make the gel ladders uninterpretable. No definitive explanation can be made for why this problem was not encountered except that the material excised from the 5'-labeled preparative gel may have been enriched for the longer form since the slowest migrating material was utilized. The cataloging studies also verified the presence of two alternative 3' terminii, ... CU_{OH} and .... CUU_{OH} in *B. harveyi* 5S rRNA.

Ribonuclease T_1 and pancreatic ribonuclease catalogs were also determined for four major fragments produced by partial digestion with ribonuclease T_2. These studies verified the sequence heterogeneities discussed above and were entirely consistent with the placement of the T_1 and pancreatic oligomers deduced from the sequencing gels. The two sites of T_2 cleavage were also established, being after A_{41} and after either A_{54} or A_{55}.

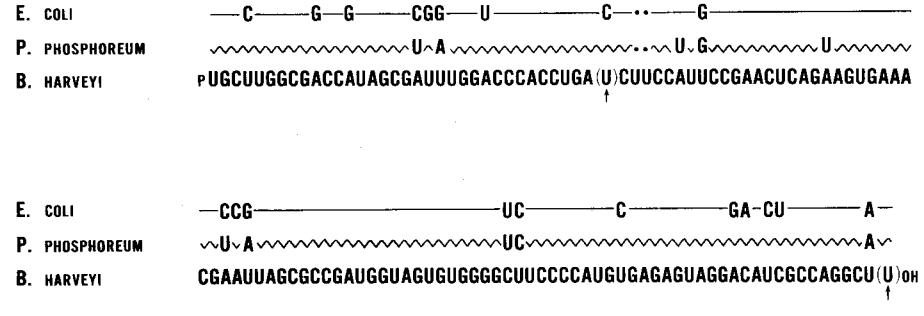

Fig. 1. Nucleotide sequence of *Beneckea harveyi* 5S rRNA aligned with *Photobacterium phosphoreum* and *Escherichia coli*. Dots in the *P. phosphoreum* and *E. coli* sequences represent deletions relative to *B. harveyi*. Arrows indicate the sites of sequence heterogeneity in *Beneckea harveyi* which are described in the text

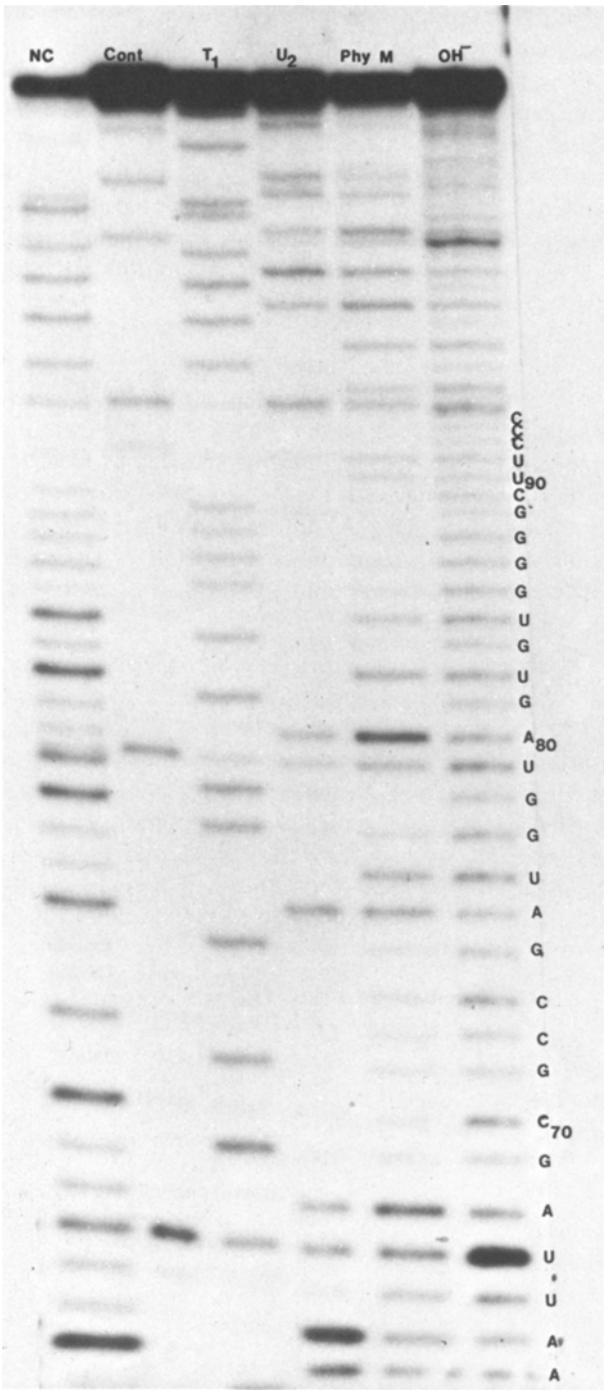

Fig. 2. Autoradiogramm of the large ribonuclease T_2 fragment of *B. harveyi* 5S rRNA which has been 5' end labeled and sequenced enzymatically on a thin (0.5 mm) polyacrylamide gel (12% acrylamide, 0.6% bisacrylamide). The lanes indicated are: NC – *Neurospora crassa* endonuclease, CONT – control, T_1 – ribonuclease T_1, U_2 – *Ustilago* ribonuclease Phy – Phy-M nuclease from *Physarum polycephalum*

Discussion

Figure 1 indicates sites of base variation relative to the *Escherichia coli* (Brownlee et al. 1968) and *Photobacterium phosphoreum* (Woese et al. 1975) 5S rRNA sequences. As Fig. 3 shows, the *B. harveyi* sequence can be arranged into the same four helix secondary structure as *E. coli* and a variety of other procaryotic 5S rRNAs (Fox and Woese 1975).

In *B. harveyi* 5S rRNA the characteristic procaryotic hairpin, helix IV, can be extended from five to eight pairs by inclusion of three consecutive G-U pairs (positions 81–83 with 97–99). *E. coli* and *P. phosphoreum* 5S rRNAs can similarly form such an extended helix, though not all procaryotic 5S rRNAs can (Fox and Woese 1975). Although it is not obvious that this extended structure would be thermodynamically favorable, recent chemical modification data (Noller and Garrett 1979) suggests that at least in *E. coli* (where one of the three pairs is of the G-C type) it is. Helix II is a site of considerable sequence variation relative to *E. coli* and *P. phosphoreum*. In all three 5S rRNAs the pairing is preserved, which is as good evidence as any that this helix is in fact actually present in a biochemically significant tertiary stucture. One unusual feature of *B. harveyi* 5S rRNA is the presence of sequence heterogeneity and a larger than normal number of bases in the -CGAAC- containing loop defined by helix III. Typically procaryotic 5S rRNAs contain thirteen bases in this region. *B. harveyi* has either fourteen or fifteen bases depending upon which sequence variant is considered.

The sequence heterogeneity encountered in *B. harveyi* 5S rRNA may represent two different phenomena. The variability at position 35 is almost certainly due to the existence of multiple cistrons of differing sequence. Variability in length at the 3' end, however, is commonly encountered in 5S rRNA (Luehrsen, unpublished observations) and may be an artifact of RNA processing. If this is the case there could be as many as four 5S rRNA species varying in length from 120 to 122 nucleotides present in the *B. harveyi* cytoplasm. If the variation in length at the 3' end is due to cistron variability there still would be at least two 5S rRNA species present and it would not be obvious which 3' terminus went with which variant at position 35. It also should be noted that there is neither any assurance that all sites of sequence variability have been located nor any evidence that all of the 5S rRNA species are biochemically active.

Intercomparisons of the sequences of *B. harveyi, P. phosphoreum* and *E. coli* can be made according to the alignment shown in Fig. 1. Treating insertion/deletion events as a single change, the following binary differ-

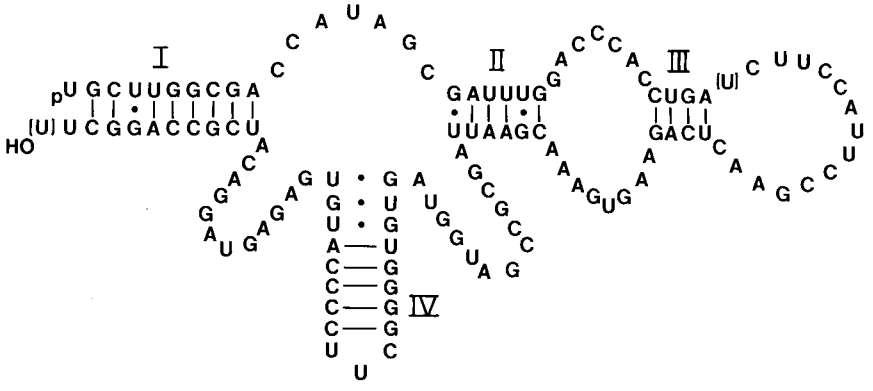

Fig. 3. Schematic drawing of the *Beneckea harveyi* 5S rRNA sequence arranged in the secondary structural scheme proposed by Fox and Woese, 1975. The sites of sequence heterogeneity are indicated in parenthesis

ences are found: *E. coli – B. harveyi* – 22, *E. coli – P. phosphoreum* – 18, and *B. harveyi – P. phosphoreum* – 12. As observed earlier for the *P. phosphoreum* case (Woese et al. 1975), the vast majority of these changes involve a G or C in *E. coli* being replaced by an A or U in the marine strains. Likewise most of the variation between *E. coli* and the two marine strains is located in either the helical regions or the single stranded hypervariable regions of the *E. coli* sequence near the molecular stalk, helix I.

With only three species under consideration, well resolved differences and no significant alignment problems, it is easy to deduce a phylogenetic tree. *B. harveyi* and *P. phosphoreum* are thus found to be specifically related and to share a common ancestor which diverged from an ancestor of *E. coli* at a somewhat earlier time. This is as expected and quite consistent with earlier hybridization studies (Reichelt et al. 1976; Baumann and Baumann 1976). It should be remarked, however, that the widespread occurrence of sequence heterogeneity in 5S rRNA is a factor which should not be ignored when constructing phylogenies based on 5S rRNA sequences. For example, since there is heterogeneity in the 5S rRNA genes of *E. coli* as well as those of *B. harveyi*, the 22 differences counted in the alignment of Fig. 1 at best represents an average. Although not a factor here, in instances such as *Bacillus licheniformis* 5S rRNA where substantial heterogeneity exists (Raué et al. 1976) this may be a difficulty.

Acknowledgements: We are grateful to Dr. H. Donis-Keller for her helpful advice and generous gift of Phy-M nuclease. Dr. N.R. Pace kindly provided us with polynucleotide kinase and Dr. C.R. Woese provided us access to unpublished data. Technical assistance was provided by Mark Aguilar. This work was performed under National Aeronautics and Space Administration grant NSG-7440 to G.E.F.

References

Baumann L, Baumann P (1976) Microbios Letters 3:11–20
Bellemare G, Vigne R, Jordan B (1973) Biochimie 55:29–35
Brownlee GG, Sanger F, Barrell BG (1968) J Mol Biol 34:379–412
Donis-Keller H (1980) Nucleic Acids 8:3133–3142
Donis-Keller H, Maxam AM, Gilbert W (1977) Nucleic Acids 4: 2527–2538
Fox GE, Woese CR (1975) Nature 505–507
Hori H, Osawa S (1979) Proc Natl Acad Sci USA 76:381–385
Jordan BR (1971) J Mol Biol 55:423–439
Klotz LC, Komar N, Blanken RL, Mitchell RM (1979) Proc Natl Acad Sci USA 76:4516–4520
Krupp G, Gross HJ (1979) Nucleic Acids 6:3481–3490
Meyhack B, Pace B, Pace NR (1977) Biochemistry 16:5009–5015
Noller HF, Garrett RA (1979) J Mol Biol 132:621–636
Raué HA, Heerschap A, Planta RJ (1976) Eur J Biochem 68: 169–176
Reichelt JL, Baumann P (1973) Arch Microbiol 94:283–330
Reichelt JL, Baumann P, Baumann L (1976) Arch Microbiol 110:101–120
Sanger F, Brownlee GG, Barrell BG (1965) J Mol Biol 13: 373–398
Uchida T, Bonen L, Schaup H, Lewis BJ, Zablen L, Woese C (1974) J Mol Evol 3:63–77
Woese C, Sogin M, Stahl D, Lewis BJ, Bonen L (1976) J Mol Evol 7:197–213
Woese CR, Pribula CD, Fox GE, Zablen L (1975) J Mol Evol 5:35–46
Wrede P, Erdmann VA (1973) FEBS Letters 33:315–319

Received June 5, 1980; Revised August 8, 1980