J. molec. Evolution 1, 230—240 (1972)

Characterization of the Primary Structural Homology between the 16 S Ribosomal RNAs of *Escherichia coli* and *Bacillus megaterium* by Oligomer Cataloging

Kenneth J. Pechman and Carl R. Woese

Department of Microbiology, University of Illinois, Urbana, Ill., U.S.A.

Received April 23, 1972

Summary. The 16S ribosomal RNAs of two *Procaryotes, Escherichia coli* and *Bacillus megaterium* were characterized by oligomer cataloging (oligomers produced by T1 nuclease digestion), in an attempt to detect their primary structural homology and as an initial step in characterizing this homology. Oligomer sequence coincidence between the two catalogs far in excess of the random expected levels was observed. Statistically significant coincidence was most pronounced for the hexamers and pentamers, suggesting that the overall structure of 16S ribosomal RNA may be such that conservation of large stretches of its primary structure (e.g. over eight nucleotides in length) is not in general essential.

Key words: Escherichia coli — *Bacillus megaterium* — Oligomer Catalog — Molecular Fossil Record — Two Dimensional Electrophoretogram — Sequence Homology — Phylogeny.

Introduction

A basic goal of biology is to account for the evolution of the cell. Emergence of the translation apparatus is the single most important event in this evolution, for capacity to translate is what defines genotype and phenotype. The complexity of translation makes it highly unlikely that its existence can be derived theoretically from "first principles" or arrived at experimentally through chemical, "origin of life" approaches. Therefore, the extent to which science comes to understand the evolution of the translation apparatus, etc., will turn heavily upon two interrelated factors (1) the type of translation mechanism that exists in the cell today and our understanding of it, and (2) the extent to which the "molecular fossil record"[1] of translation's evolution has or has not been obliterated by mutations. If the molecular fossil record still remains reasonably intact, science will eventually be in a position to deduce features of more primitive versions of the translation apparatus than now exist.

1 The "molecular fossil record" is the history of a molecule's evolution that is necessarily written *ab initio* into its primary structure, a record that can be deduced by comparative primary structure studies (Zuckerkandl and Pauling, 1965).

What we have attempted in the present study is to begin defining areas within the primary structure of the ribosomal RNAs that are evolutionarily conserved. The 16S rRNAs from two *Procaryotes, Escherichia coli* and *Bacillus megaterium,* have been chosen for initial study. While these two organisms are not so disparate that their comparison would speak to the farthest possible reaches of evolution, neither are they so closely related that one would expect all homologies encountered to be of trivial significance. For example, it is known from hybridization-competition experiments and other studies that the *Bacillaceae* are less related to the *Enterobacteriaceae* than the latter are to *Achromobacteraceae, Spirillaceae,* or *Pseudomonadaceae* (Pace and Campbell, 1971; Sogin *et al.*, 1972).

The approach used herein is a cataloging of the oligomers produced by (complete) digestion of the rRNAs with T1 nuclease. In view of what is now known from RNA (and protein) "fingerprinting" (Fitch, 1971; DuBuy and Weissman, 1971), it is clear that statistically significant levels of oligomer coincidence (among larger oligomers) in any comparison of 16S rRNAs will for the most part, represent primary structural homology (Sogin *et al.*, 1972). Therefore, in determining common (large) oligomers in these rRNAs one is defining markers that designate the conserved areas of the 16S rRNA molecule and begin to define their primary structure.

Materials and Methods

The two bacterial strains used throughout the present study are *E. coli D-10* (met$^-$, RNAse I$^-$) and *B. megaterium KM* (met$^-$, thy$^-$). Growth media comprised 0.025 M ammonium chloride, 0.32 M glycerol, 2.5×10^{-3} M magnesium sulfate, 3×10^{-4} M calcium chloride, in a 0.02 M TRIS-HCl buffer of pH 7.4, containing 0.1% casamino acids (for *E. coli*) or 0.4% casamino acids (for the *Bacillus*). When cultures had attained an optical density of 8–10 Klett units, carrier free $^{32}PO_4$ was added to a concentration of approximately 0.7 mc/ml (15–25 ml total culture volume). Cells were harvested (by centrifugation) when cultures had reached 55–60 Klett units, in log phase growth.

Cells were disrupted by passage through a pressure cell at 15000 psi, after resuspension in a buffer containing 0.14 M TRIS Cl, 0.07 M sodium acetate, 0.003 M EDTA, 0.5% sodium dodecyl sulfate, 1 mg/ml "Macaloid" (Baroid Div., National Lead Co., Huston Texas), and adjusted to pH 7.2 with acetic acid. The resulting lysate was cleared of debris by centrifugation (20000 g for 5–10 min) and the resulting supernatant layered directly onto a so-called "double" polyacrylamide gel (top half 3% polyacrylamide, bottom half 8% polyacrylamide (Sogin, 1971). After electrophoresis for approximately four hours (10 ma per gel, at 50 volts), gels were removed from their plexiglass tubes and the radioactive RNA species located by brief exposure of the whole gel to x-ray film. The individual sections containing the RNAs were thus located and cut out, and the desired rRNA(s) eluted by electrophoresis into an elution chamber (Sogin, 1971).

The resulting RNA was further purified by elution from a CF-11 cellulose (Whatman) column (Pace, personal communication; Doolittle and Pace, 1971). The eluate was centrifuged to remove particulate debris (i.e. cellulose fibers, etc.).

T1 nuclease digests of purified rRNAs were analyzed by the two dimensional electrophoretic procedure of Sanger and coworkers (Sanger *et al.*, 1965; Sogin *et al.*, 1971; Sogin *et al.*, 1972). Secondary characterization of individual spots on the two

dimensional electrophoretograms involved pancreatic nuclease digestion (Sanger *et al.*, 1965) or digestion with U2 nuclease (from *Ustilago sphaerogena*) (Arima *et al.*, 1968; Uchida *et al.*, 1970). The cleavage specificity of U2 is basically for purines, but under certain conditions it will exhibit more complex specificities, revealing the existence of AA stretches, or cleaving at pyrimidine residues (Uchida, personal communication; Uchida and Woese, unpublished; Arima *et al.*, 1968; Uchida *et al.*, 1970). Partial digestion of spots from primary patterns by spleen diesterase was performed by the methods of Sanger and coworkers (1965). These secondary procedures allow one to deduce full sequence for nearly all oligomers up to size six nucleotides, and for many of the larger oligomers as well.

Results

The 16S rRNA molecule comprises approximately 1700 nucleotides (0.5×10^6 daltons) and has a base ratio of G:A:C:U of about 32:27:21:20 (Midgley and McCarthy, 1962). Fig. 1 is a representative two-dimensional electrophoretogram of a T1 nuclease digest of *B. megaterium* 16S rRNA. Tables 1 and 2 show the T1 oligomer catalog for the 16S rRNA from this organism and that from *E. coli*, both determined by the methods described above. Table 3 is the size distribution for each of the T1 oligomer catalogs, and it can be seen that these approximate the distribution one would obtain from a *random* sequence of comparable composition and length.

The degree of oligomer sequence coincidence between the two 16S rRNAs (Tables 1 and 2) is considerably in excess of that expected on a random basis. This can be demonstrated in either of two ways, statistically or experimentally. Table 4 compares (a) the oligomer coincidence levels calculated for two 16S-like RNAs of *random* sequence, (b) the observed coincidence levels, for T1 pentamers and larger, and (c) coincidence levels observed when the two 16S rRNAs in question are compared to "unrelated" RNAs of comparable size. Among the T1 hexamers, two *chance* coincidences would be expected; at least 8 and possibly 10 coincidences are found; i.e. almost half of the (20-odd) hexamers present in the one case are also found in the other. Among T1 heptamers, where less than 0.4 chance coincidences would be expected, three are actually found (Tables 2 and 4). Among T1 oligomers of length eight or larger only one coincidence can be claimed with certainty, a nonamer; however, more detailed characterization of these oligomers could reveal additional ones.

Homology is also demonstrable within the T1 pentamers and tetramers, but is less obvious due to the relatively high levels of coincidence expected on a random basis in these cases (see Table 4). The number of random pentamer coincidences expected is in the range of 10; sixteen are observed for certain, with another four possible (Table 1). In particular the following should be noted: (1) Although 41–48 (i.e. over half) of the possible pentamer sequences are absent from *E. coli* 16S rRNA, no more than 5–9 of these are represented among *B. megaterium*'s 25–26 pentamers. (2) On the one

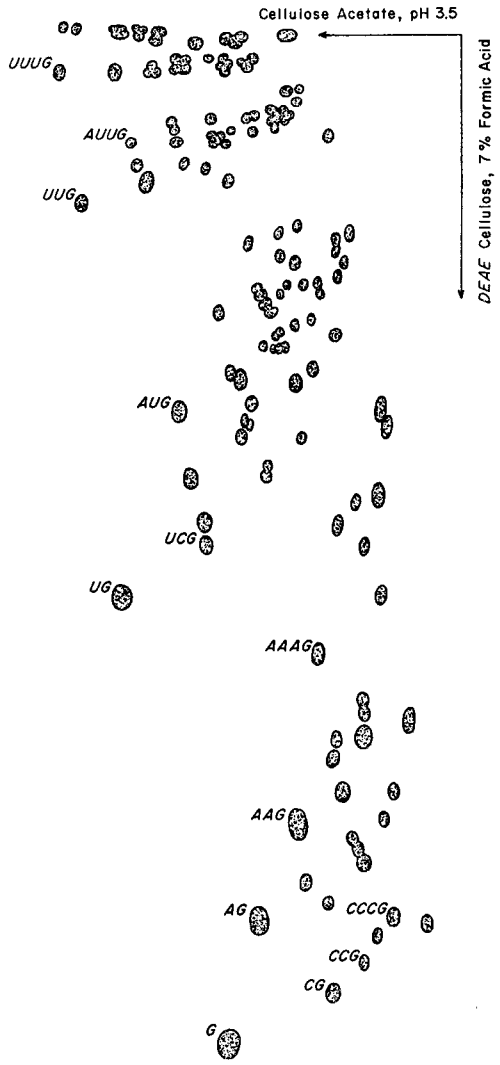

Fig. 1. Two dimensional electrophoretogram of a complete T1 nuclease digest of *Bacillus megaterium* 16S ribosomal RNA. First dimension run on cellulose acetate paper in pH 3.5 acetate buffer containing 7 M urea; second dimension run on DEAE cellulose paper in 7% formic acid (Sanger *et al.*, 1965; Sogin *et al.*, 1971). See text for details

hand, of those tetramer sequences present once in *E. coli*, 55–60% are also present exactly once in *B. megaterium*, but only 15% are present three or more times; yet of the tetramers present three or more times in *E. coli*, *none* are present in single copy and about 60% are represented three or more times in *B. megaterium*.

Table 1. Number of copies of T1 oligomers of $\leqq 5$ nucleotides length present in the 16S rRNAs of *B. megaterium* and *E. coli*. An asterisk preceeding a nucleotide indicates post-transcriptional modification of that nucleotide, as determined by Fellner and Sanger (1968)

Oligomer	*B. megaterium*	*E. coli*	Oligomer	*B. megaterium*	*E. coli*	Oligomer	*B. megaterium*	*E. coli*
G	170	170	*CC*CG	1	1	ACUAG	0	0
			CC*GCG	1	1	AUCAG	1	1
CG	33	42	*CAACG	2–1	1	AAUCG	1	1
AG	42	47				AACUG	1	3–4
UG	45	45	CCCCG	0	0	UAAAG	1	3
			CCACG	1	1	AAUAG	0	0
CCG	7	6	CCCAG	1	1	AUAAG	0	0
CAG	6	9	CACCG	0	1	AAAUG	1	2
ACG	7	10	ACCCG	0	0			
AAG	12	18	CAACG	1–2	1	UCUCG	0	
UCG	8	11	CACAG	0	0	UUCCG	0	2
CUG	10	11	CCAAG	0	0	UCCUG	0	
UAG	8	14	ACACG	1	2	CUUCG	0	
AUG	14	13	ACCAG	0	1	CCUUG	1	1
UUG	7–8	9	AACCG	0	1	CUCUG		
			CAAAG	0	0	UUACG	0	0
CCCG	2	1	AACAG	0	0	UACUG	0	0
CACG	0–1	2	ACAAG	0	0	UAUCG	1	0
CCAG	2	2	AAACG	0	1	CAUUG	0	0
ACCG	1	1	AAAAG	1	0	CUAUG	0	0
CAAG	3	7				UCAUG	0	0
AACG	1	2	UCCCG	1	1–2	CUUAG	0	0
ACAG	2	2	CUCCG	0	0	UCUAG	0	0
AAAG	4–5	1	CCUCG	0	0	UUCAG	0	0
UCCG	0	2	CCCUG	0	0	ACUUG	0	2–1
CCUG	2	3	UACCG	0	0	AUCUG	0	1–2
CUCG	2	2	UCACG	0	0	AUUCG	0	0
UACG	1–2	1	CUACG	0	0	UAAUG	0	1
UCAG	3	3	CACUG	0	0–1	UAUAG	0	0
CUAG	1	1	CAUCG	0	0	UUAAG	1	2
CAUG	2	3	CCAUG	0	1	AUAUG	0	0
ACUG	4	4	CCUAG	0	2	AUUAG	2	2
AUCG	1	0	CUCAG	1		AAUUG	1	1
UAAG	2	0	UCCAG	0	0			
AAUG	1	2	ACCUG	1	0	UUUCG	1	
AUAG	2	0	AUCCG	0	0	UCUUG	1	1–2
UUCG	2	3	ACUCG	0	0	UUCUG	0	
UCUG	2		UAACG	1	1	CUUUG	0	2–1
CCUG	2	3	UCAAG	1	1	UUAUG	0	0
UAUG	1	1	CUAAG	1		UAUUG	0	0
UUAG	3	0	CAUAG	1	0	UUUAG	0	0
AUUG	3	3	UACAG	0	0	AUUUG	0	0
UUUG	3	3	CAAUG	0	0			
			ACAUG	0	0	UUUUG	0	0
			AUACG	0	0			

Table 2. Occurrence of T1 oligomers of $\geqq 6$ nucleotides length in the 16S rRNAs of *B. megaterium* and *E. coli*. For oligomers in parentheses the ordering is determined only to the extent indicated. Oligomers containing post-transcriptionally modified nucleotides are indicated as in Table 1

	B. megaterium	*E. coli*
6-mer		
C_3ACG	1	0
CACAAG	1	1
A_3CCG	1	0
CA_4G	0	1
A_2CAAG	1	0
A_3CAG	0	1/2
$[C_2, U]ACG$	1	1–2
C_2AUCG	0	1
AC_2UCG	0	1
CUA_2CG	1	1
C_2UAAG	1	0
UA_2CCG	1	1
UA_3CG	1	1
A_2CCUG	0	1
A_2UACG	1	1
A_3CUG	1	1
*A*ACCUG	1	1
U_2C_3G	1	1
$C[C_2, U_2]G$	0	4
[C, AU]CUG	1	1
UA_2UCG	1	1
[U, C, AU]AG	1	0
AUCAUG	0	1
UA_2CUG	1	0
AUACUG	0	1
A_2CUUG	1	0
$[C_2, U_3]G$	2	0
UCAUUG	1	0
7-mer		
CA_3CAG	1	1
$ACCA_3G$	0	1
$CACUC_2G$	1	0
CA_2CUCG	1	1
CA_2CCUG	1	0
$CAUA_2CG$	0	1
$UACA_3G$	1	0
A_2UAC_2G	0	1
$CACA_2UG$	0	1
UA_2CACG	1	0
$*UA_2CA_2G$	1	1
U_2AC_3G	0	1
UA_2UACG	1	1
$[AU, C_2, U]AG$	1	0
CAU_2AAG	1	0
$AC[C_2, U_2]G$	0	1
$[C_3, U_3]G$	0	1
CU_2AUCG	1	0
$CUACU_2G$	1	0
[UUA, UCA]G	1	0
CAU_3AG	1	0
[AAU, AU]UG	0	1
U_2A_2UCG	0	1
$[C_2U]AUUG$	0	1
U_4AAG	1	0
$[U_4, C]AG$	0	1
8-mer		
A_2CAC_2AG	1	0
$[A_3C, AC]CG$	1	0
C_6UG	0	1
CCACACUG	1[a]	1
$[A_2C, C, U]CCG$	1	0
AUA_3CUG	0	1
2U oligos	4	—
uncharacterized	—	4
$A_2UUACUG$	0	1
[U, C]AUCAUG	0	1
$A\begin{bmatrix}U\\A\end{bmatrix}U\begin{bmatrix}A\\U\end{bmatrix}AUUG$	1	0
9-mer and larger		
UACACACCG	1	1
A_3UC_4G	0	1
$[C_2, AU, AC]ACG$	0	1
$A_2C[C, U]A_3G$	0	1
$[A_2C, U, C_2, AC]G$	1	0
$[U, C_5, AC]G$	1	0
$[A_3CU, C]A_3G$	1	0
$[A_2C, (AC)_2, U]G$	1	0
$[U_2, C_4]ACG$	?	1
$[(AC)_2, C_2, U_2]AG$	0	1
$[A_2C, AC, AU, U, C_2]G$	0	1
$[AC, AU, C_3, U]A_3G$	0	1
$[A_2CU, C, U, AC]AG$	1	0
$[A_2C, U_2, C_2, AC]AG$	2	0
2U oligos	5	—
uncharacterized	—	6

Table 2 (Continued)

	B. megaterium	*E. coli*		*B. megaterium*	*E. coli*
$[U_3, C_4]AG$	1	0	U_3A_2UUCG	?	1
$[U_3, C_5]G$	0	1	UUA_2UACU_3G	0	1
$[A_2C, AU, U, AC]UG$	0	1	$[CCA, CUUA]CU_3G$	0	1
$A_2C_2U_2AC_2UG$	0	1	$[A_2U, AU, U_3]G$	1	0
$[U, A_2C]AUUCG$	1	0	$[AU, U_4, C]G$	1	0
$[C, U]AAUCCUG$	1	0			
$[U_2, C_4]AUCG$	1	0	$[A_2U, U_5]AAG$	1	0
$[U_2, C_2, AU. AC]UG$	1	0			
$[A_2U, AU, U, AC, C_2]G$	1	0	$[(AU)_2U_3, C_2]UG$	1	0
3U oligos	1	—	$[A_2U, (AU)_2, C_8, U_2]AG$	1	0
uncharacterized	—	1	$[C_xU_2A, U_2AA][C_yU_3]G$	0	1

[a] This ordering and several more are possible.

Table 3. Random expected and observed size distributions for oligomers of the T1 type present in 16S rRNAs. In calculation G content is taken to be 32% and the length of the RNA 1700 nucleotides. The random level of occurrence of an oligomer of size n is given by the expression $N_n = (0.32)^2\ (0.68)^{n-1} \cdot 1700$

Oligomer size	Random Expected	*E. coli*	*B. megaterium*
1	174	170	170
2	118	134	120
3	80	101	80
4	55	52	54
5	37	45	29
6	25	22	21
7	17	15	15
8	12	9	9
$\geqq 9$	25	~ 25	~ 30

Discussion

That primary structural homology between *E. coli* and *B. megaterium* 16S rRNAs exists is known from nucleic acid hybridization studies (Moore and McCarthy, 1967; Pace and Campbell, 1971). There can be little doubt then, that the statistically significant degree of oligomer coincidence observed in the present studies reflect such an homology. [However, it should be noted that the primary structural constraints reported herein do not of themselves *prove* the existence of this homology, merely make it highly likely.] While nucleic acid hybridization measurements limit one to stating whether or not detectable homology exists, the present approach begins to define the nature of the homology in molecular terms—and so to define both what is primitive and what is of structural and/or functional importance in ribosomal RNA.

Table 4. Expected *vs* observed oligomer coincidence levels

Column A—Size of T1 oligomer.

Column B—Number of oligomers of given size found in given RNA.

Column C—Expected level of oligomer sequence coincidence for two unrelated (random sequence) RNAs of the size and composition of 16S rRNA.

Column D—Observed level of oligomer sequence coincidence for *B. megaterium* and *E. coli* 16S rRNAs. The number in parenthesis indicates the probability that a level of coincidence at least this high will be found on a random basis (Molina, 1942).

Column E—Observed level of oligomer coincidence between *B. megaterium* or *E. coli* 16S rRNA and either of two "unrelated" RNAs of comparable size: (1) the 18S rRNA from yeast, and (2) the 14S rRNA from the 50S ribosomal subunit of *Rhodopseudomonas spheroides* (Sogin, unpublished; Zablen, unpublished; Marrs and Kaplan, 1970).

A	B				C	D		E	
Size	Number found				Random	Observed		Coincidence unrelated	
	unrelated		*B. megaterium*	*E. coli*					
	(1)	(2)						(1)	(2)
5	24	28	26[a]	42[a]	~11	16–20[a]	(~3%)	6–10	11–12
6	21	18	20[b]	21[b]	2	8–10[b]	(≦0.1%)	2–3	3–4
7	11	11	14[c]	14[c]	<0.4	3[c]	(0.1%)	0–1	0–1
8			9	9	<0.1	0–1			
≧9			~30	~25	<0.03	≧1	(<3%)		

[a] Excludes all 3 post-transcriptionally modified oligomers, which are conserved.
[b] Excludes the 1 post-transcriptionally modified oligomer, which is conserved.
[c] Excludes the 1 post-transcriptionally modified oligomer, which is conserved.

As might be expected, a number of the conserved oligomers contain posttranscriptionally modified nucleotides. Of the oligomers of this type reported in the *E. coli* 16S rRNA (approximately seven) we have found at least five to be present in *B. megaterium* (Tables 1 and 2), and have yet to demonstrate any to be absent (Fellner and Sanger, 1968; Fellner *et al.*, 1970).

If we assume that the probability of mutational replacement of a nucleotide is roughly constant over most of the rRNA sequence and that T1 hexamers are representative of the rRNA primary structure, then it follows that *E. coli* and *B. megaterium* 16S rRNAs have about 90 percent of their total residues in common—i.e. $(0.9)^7 = 0.5$ [2]. However, these assumptions are questionable, the first in view of the "conserved" *vs* the "mutable" regions found in protein and 5S rRNA sequences (Fitch and Margoliash, 1966; DuBuy and Weissman, 1971), the second in view of the fact that the degree of conservation seen for T1 hexamers does not hold for heptamers

2 A T1 hexamer is defined by a stretch of seven contiguous nucleotides.

and larger (see Tables 2 and 4). In any case, it appears safe to conclude that a reasonably large portion of the 16S rRNA sequence is highly conserved between these two organisms, and perhaps that the structure of the 16S rRNA is such that long stretches of nucleotides (say, over eight residues length) are generally not required to be conserved.

Along these lines it should be noted that the sequence of the 3′ terminal 40 nucleotides of the *E. coli* 16S rRNA is now known (Hayes *et al.*, 1971). It comprises for the most part four large and very characteristic T1 oligomers, *UAACAAG, *A*ACCUG, CUAACG, and the 3′ end T1 oligomer itself, AU[AC, U_3, C_4]A_{OH}. The first three oligomers are all found in *B. megaterium* 16S rRNA, as is a 3′ end oligomer very like that found in *E. coli*. [Both 3′ end T1 oligomers yield AU, AC, and multiple C's and U's when digested with pancreatic nuclease; however, the *B. megaterium* version contains two more U's and less C than its *E. coli* counterpart. Both yield (UC)A and a large 3′OH fragment (different in the two cases) when digested with U2 nuclease. As the reader can calculate, it is highly improbable that these two 3′ end T1 oligomers are not closely related.]

It is important to explain the existence of sequence homology between these two 16S rRNA species. If it reflects the fact that certain portions of their common ancestral primary structure are locked into the present sequences due to stringent constraints imposed by structural and/or functional considerations, then the conservation becomes highly significant. However, were the frequency of occurrence of mutations in rRNA cistrons to be sufficiently low for some reason, then the bulk of the observed conservation could merely reflect the fact that mutations had not occurred in those regions in either organism, and conservation would be of trivial significance. Given that mutation rates in *Procaryotes* are relatively high (Drake, 1970), that the *Bacilli* and the *Enterobacteriaceae* do not appear to be closely related (Wittmann *et al.*, 1970), and that *Procaryotic* lineages seem to date much further back in time than Metazoan ones (Margulis, 1972), it seems highly unlikely that the latter, trivial explanation is correct. However, the matter need not be settled by argument; it is ammenable to experiment.

Were the trivial explanation correct, a comparison of 16S rRNAs from three or more organisms of comparable relatedness would show that the set of oligomers conserved for one pair of them would tend to be unlike the sets conserved for any other pairs. [The situation is analogous to starting with n identical sequences of playing cards and then altering randomly a certain fraction, x, of the cards in each sequence. When the altered sequences are compared, the probability that given stretches of the original sequence will be conserved over the full set of n sequences drops rapidly as a function of increasing x when n is large.] Partial characterization of the 16S rRNA from *Alcaligenes faecalis* show that five of the

eight T1 hexamers and all three of the T1 heptamers definitely common to *E. coli* and *B. megaterium* are common to all three organisms (Pechman and Arlow, unpublished). [Also of the two remaining hexamers *possibly* common to *E. coli* and *B. megaterium*, both are, in the same sense, possibly common to *A. faecalis* as well.] Consequently there can be little doubt that the oligomer conservation reported herein results from constraints on molecular primary structure, not merely low mutation rates in the rRNA cistrons. It is also clear from such a triple comparison that the oligomer coincidences in *E. coli* and *B. megaterium must* represent primary structural homology (a point discussed above). The alternative, that large unique oligomers of identical sequence in all cases occupy *different* (nonhomologous) places in different 16S primary structures, becomes totally absurd as the number of organisms in which they are found increases.

In a molecule as large as the 16S rRNA, all residues are clearly not equivalent in their importance to molecular function. There must reasonably be a class of residues for which the replacement of any residue by another would have negligible effect on the overall molecular function. This class corresponds to what Kimura and others call "neutral" replacements. Such residues would be replaced with a frequency proportional to the mutation rate of the corresponding cistrons (Kimura, 1968). They are "short half-life" residues, and can be used to measure phylogenetic "distance" for the more closely related phylogenetic groupings—i.e. within *Species* and *Tribes*.

Many if not most of the residues in an rRNA undoubtedly *are* subject to functional constraints, etc. While altering any one of these would be deleterious to molecular function, it is conceivable that in some instances a more or less simultaneous alteration of several residues would be compensatory, and so leave molecular function unchanged (e.g. a G ... C pair altered to an A ... U pair). Thus, one can consider the rRNAs to contain various types of "replacement units" (each comprising two or more residues). Overall molecular function would remain undisturbed when the composition of such a unit is changed, *provided* that certain critical overall features of that unit remained invariant (i.e. multiple, "compensating" changes occur "simultaneously" within the unit). Acceptable alterations of such units would then occur as higher powers of the mutation frequency, giving such units rather long "half-lives". (See also Fitch, 1971.) Such replacement units can be used to measure phylogenetic relatedness among the more distantly related organisms—on the *Family, Order*, etc. level. In the limit, certain of these units could be sufficiently complex that their "half-lives" would exceed the evolutionary time available. Indeed in certain cases a unit could be uniquely indispensable, in that only one composition of the unit would produce the required function (in a given environment of the whole molecule).

A decision as to what class or classes of replacement units the conserved sequences found between *E. coli* and *B. megaterium* represent must await a wider search of the *Procaryote* spectrum (and a *Eucaryote* or two). Hopefully, some of the present cases are sufficiently conserved (ancient) that they carry information about the evolution of the ribosome.

In closing we feel it necessary to point out that many of the T1 oligomer sequences reported by us for *E. coli* 16S rRNA are in disagreement with those reported by Fellner *et al.* (1970). We will not discuss these discrepancies in the present context, but will say here only that they do not reflect strain differences for the most part.

Acknowledgements. The work reported herein was supported by NSF grant GB 16298 to C.W. We are grateful to Prof. F. Egami and Dr. T. Uchida for their generous gift of ribonuclease U2, and to Prof. J. T. Wachsman for supplying us with *B. megaterium* strain KM (met⁻, thy⁻).

References

Arima, T., Uchida, T., Egami, F.: Biochem. J. **106**, 601 (1968).
Doolittle, W. F., Pace, N. R.: Proc. nat. Acad. Sci. (Wash.) **68**, 1786 (1971).
Drake, J.: The molecular basis of mutation. San Francisco: Holden-Day 1970.
DuBuy, B., Weisman, S. M.: J. biol. Chem. **246**, 747 (1971).
Fellner, P., Ehresmann, C., Ebel, J. P.: Nature (Lond.) **217**, 26 (1970).
Fellner, P., Sanger, F.: Nature (Lond.) **219**, 236 (1968).
Fitch, W. M.: J. molec. Evolution **1**, 84 (1971).
Fitch, W. M., Margoliash, E.: Science **155**, 279 (1966).
Hayes, F., Hayes, D., Fellner, P., Ehresmann, C.: Nature (New Biol.) **232**, 54 (1971).
Kimura, M.: Nature (Lond.) **217**, 624 (1968).
Margulis, L., in: Exobiology, p. 342 (Ponnamperuma, C., ed.). Amsterdam and London: North-Holland Pub. Co. 1972.
Marrs, B., Kaplan, S.: J. molec. Biol. **49**, 297 (1970).
Midgley, J. E. M., McCarthy, B. J.: Biochim. biophys. Acta (Amst.) **61**, 696 (1962).
Molina, E. C.: Poisson's exponential binomial limit. Princeton: D. van Nostrand Co. 1942.
Moore, R. L., McCarthy: J. Bact. **94**, 1066 (1967).
Pace, B., Campbell, L. L.: J. Bact. **107**, 543 (1971).
Sanger, F., Brownlee, G. G., Barrell, B. G.: J. molec. Biol. **13**, 373 (1965).
Sogin, M., Pace, B., Pace, N., Woese, C.: Nature (Lond.) **232**, 48 (1971).
Sogin, S.: Ph. D. thesis, University of Illinois, 1971.
Sogin, S., Sogin, M., Woese, C.: J. molec. Evolution **1**, 173 (1972).
Uchida, T., Arima, T., Egami, F.: J. Biochem. **67**, 91 (1970).
Wittmann, H. G., Stoffler, G., Kaltschmidt, E., Rudolff, V., Janda, H. G., Dzionara, M., Donner, D., Nierhaus, K., Ceck, M., Hindennach, I., Wittmann, B.: FEBS Symp. **21**, 33 (1970).
Zuckerkandl, E., Pauling, L., in: Evolving genes and proteins, p. 97 (Bryson, V., Vogel, H., eds.). London and New York: Academic Press 1965.

C. R. Woese
K. J. Pechman
Department of Microbiology
University of Illinois
131 Burrill Hall
Urbana, Illinois 61801, U.S.A.