J. Molec. Evolution 3, 63—77 (1974)

The Use of Ribonuclease U_2 in RNA Sequence Determination

Some Corrections in the Catalog of Oligomers Produced by Ribonuclease T_1 Digestion of *Escherichia coli* 16S Ribosomal RNA

T. Uchida*, L. Bonen, H. W. Schaup, B. J. Lewis, L. Zablen, and C. Woese

Department of Microbiology, University of Illinois, Urbana, Illinois 61801

Received June 23, 1973

Abstract. The catalog of oligomers produced by ribonuclease T_1 digestion of *Escherichi coli* 16S ribosomal RNA has been determined by a new method that involves the use of ribonuclease U_2 from *Ustilago sphaerogena*. The sequences for the larger T_1 oligomers (8 or more bases) determined in this way differ in more than 50% of the cases from those reported previously (determined by other methods).

Key words: 16S Ribosomal RNA — Oligonucleotide — Fingerprint — Ribonuclease U_2 — *Escherichia coli*.

Introduction

This laboratory is engaged in a comparative characterization of the primary structures of various ribosomal RNAs—with an immediate goal of establishing a comprehensive and definitive phylogeny for the *Procaryotes*, and an ultimate goal of understanding the evolution of a translation apparatus.

At present the work involves generating catalogs of oligonucleotides (produced by ribonuclease T_1 digestion) for ribosomal RNAs from a variety of organisms—utilizing the two dimensional electrophoretic fingerprinting method developed by Sanger *et al.* (1965). The method used by us to determine sequence for these oligomers differs somewhat from those previously employed, in that sequences are deduced almost exclusively from the partial digestion products resulting from cleavage of a given oligomer (or fragment thereof) by pancreatic ribonuclease and by ribonuclease U_2, from *Ustilago sphaerogena* (Arima *et al.*, 1968; Uchida *et al.*, 1970).

In the course of developing and employing this particular technology, we discovered several discrepancies between T_1 oligomers sequenced by

* Permanent address: Mitsubishi-Kasei Institute of Life Sciences, Machida, Tokyo, Japan.

our methods and those sequenced by methods not utilizing ribonuclease U_2 (Pechman and Woese, 1972). Hence, we have redetermined the sequences of the oligomers produced by ribonuclease T_1 digestion of *Escherichia coli* 16S ribosomal RNA by our methods, and report the resulting catalog herein.

Materials and Methods

Several strains of *Escherichia coli* have been used in this study, of both B and K origin. Differences between their 16S rRNA oligomer catalogs, however, are essentially negligible (involving at most a few oligomers whose occurrence is fractional).

^{32}P labeled RNA was produced by standard methods, involving growth in low phosphate medium, isolation of RNA by phenol extraction, separation of 16S rRNA by polyacrylamide gel electrophoresis, and its purification by passage over CF-11 cellulose columns (Sogin *et al.*, 1971; 1973; Kirby, 1956; Doolittle and Pace, 1971; Franklin, 1966).

Detailed Description of Oligonucleotide Fingerprint Analyses

1. The Primary Pattern

The initial two dimensional electrophoretogram of ^{32}P 16S rRNA is produced in one of two ways—either by the published method of Sanger and coworkers (utilizing a pH 3.5 first dimension on cellulose acetate, followed by transfer to DEAE cellulose paper which is then run in 6.5% formic acid) or by a slight modification of this method (in which the second dimension is run in a "high salt" buffer) (Sanger *et al.*, 1965; Woese and Sogin, manuscript in preparation). The latter provides far better resolution in all isopliths with the exception of certain areas of the G isoplith. This obviates the need to dephosphorylate oligomers prior to electrophoresis, with the attendant disadvantages of such a procedure. Oligomer spots on the primary fingerprint were located by radioautography, cut out, and their ^{32}P content determined by scintillation spectrometry (Sogin *et al.*, 1971). Individual spots were then removed from scintillation vials, washed in three changes of toluene, and dried in preparation for secondary analysis.

For those who are not thoroughly familiar with the ordering of oligomers on a two-dimensional electrophoretogram of a T_1 digest of RNA, a brief explanation should be given (please refer to the accompanying or a comparable figure). The oligomer pattern comprises a series of wedge-shaped "isopliths" (Sanger *et al.*, 1965). Within any isoplith, all oligomers contain the same number of U residues; the fastest moving, G, isoplith contains oligomers devoid of U; the next most rapid isoplith comprises oligomers containing a single U, etc. Within any given isoplith oligomers are arranged in "isomeric", nearly vertical lines; within such a line all oligomers have the *same* number of bases. The number of bases in an oligomer can be determined reliably by this method for any oligomer containing seven bases of less. For larger oligomers, their sizes can generally be determined within *one* base by position in the isoplith. On a given isomeric line the oligomers separate by relative A (*vs.* C) content; the oligomer(s) of lowest (highest) A content travels furthest (least) in the second dimension.

Order of bases within the oligomer also affects its position on a given isomeric line. For example, given two oligomers of identical composition, that with an A in the 5′ position will travel more slowly in the second dimension than those with a pyrimidine in the 5′ position. U residues in the 5′ position cause oligomers (of a given composition) to travel *slightly* more rapidly in both dimensions [see for example, 14a, UCCG, *vs.* 14b, C(UC)G].

The spot designation convention we employ is based upon isoplith pattern; (Sogin *et al.*, 1971); see Table 1 or Fig. 1 for examples. The initial number refers to the number of U residues in the oligomer, the second to the total number of bases. The lower case letter which follows distinguishes among oligomers having the same first two numbers; the number of A residues tends to increase in alphabetical order.

2. Secondary Analysis

In general each oligomer spot (whose sequence was not determined by position on the primary pattern) was divided into several portions, one being digested with pancreatic ribonuclease, the other with ribonuclease U_2.

Pancreatic ribonuclease (RNAse A, Worthington Biochemicals) was used at a concentration of 1 mg/ml (in water). Sufficient enzyme solution was applied to each spot so that the paper appeared thoroughly moistened (glistening). Incubation time was not critical, and varied from several hours at 37 °C to overnight at room temperature on occasion. For oligomers containing larger relative amounts of U, it was necessary to use the enzyme at 5 mg/ml in order to obtain complete digestion in a reasonable length of time. To prevent drying during digestion spots were sealed between layers of Parafilm.

Ribonuclease U_2 for secondary analysis of oligomers containing one or more A residues was applied in a 0.1 M sodium acetate buffer, pH 5.5, containing 0.0015 M EDTA, at a concentration of 5 or 10 units/ml (Uchida *et al.*, 1970). The paper spot was "saturated" with enzyme solution, as just described. The treated spots, sealed in Parafilm, were incubated at 37 °C for 4–6 hrs.

For digestion of oligomers not containing A residues, the ribonuclease U_2 digestion procedures was changed to the extent that the enzyme concentration was increased 3–4 fold and the pH of the acetate buffer reduced to 4.5. [This will be referred to as the "overcutting", as opposed to the "normal" mode (above) for ribonuclease U_2 digestion.]

Each spot receiving a secondary digestion procedure was then inserted, by maceration, into a sheet of DEAE cellulose paper—about twenty spots plus suitable marker spots per 18 inch width of paper. When individual spots were large, they were folded before being macerated into the DEAE sheet.

Pancreatic ribonuclease secondary digestions were electrophoresed in one of two solvent systems—6.5 % formic acid, or 5 % acetic acid containing 0.5 % pyridine (PA). The following "marker" spots were generally used to identify the pancreatic nuclease digestion products: U, AU, AAU, AAAU, C, AC, AAC, AAAC, G, AG, AAG, and AAAG.

Ribonuclease U_2 secondary digestions were generally run in 6.5 % formic acid. However, for oligomers high in U content the above "high salt" buffer was also used, in that 6.5 % formic acid does not always provide definitive resolution of spots containing various numbers of C residues when the sum of U and G residues is three or greater. Marker oligomers appropriate for this system were CA, CCA, UA, UUA, (CU)A, CG, CCG, UG, CUG, UCG, UUG, etc. [CA, UA, etc. are, of course, generated by ribonuclease U_2 cleavage of simple oligomers such as CAG and UAG].

When used under the (normal) conditions described, U_2 enzyme in addition to producing fragments of the form $(C_xU_y)A$ and $(C_xU_y)G$, also produces composite fragments. For example, an oligomer such as CCUAACACAUG in addition to yielding the expected fragments $(C_2U)A$, CA, and UG (resulting from cleavage on the 3′ side of A residues), yields high levels of $(C_2U)AA$, and detectable levels of CAUG, CACAUG, etc. Under these conditions the frequency of cleavage at bases other than A is negligible—which is definitely not the case when the enzyme is used under the above "overcutting" conditions.

3. Tertiary Analysis

Oligomer spots generated by secondary analysis whose sequences are not unequivocally determined by position relative to given marker spots, require one or more tertiary procedures: either rerunning in a different solvent, or further ribonuclease digestion, etc. Electrophoretic separations in 6.5 % formic acid are primarily a function of the U + G content of an oligomer (and secondarily a function of its size and A *vs.* C content). Hence, certain oligomers of distinctly different size, but of "equivalent" U + G content can be confused in this solvent. However, these cases are always resolvable upon reelectrophoresis in the above "PA" buffer system—which separates

primarily as a function of oligomer *size*. Thus, for example, CA and CCA, or $(C_2U)G$ and UUA not clearly resolved in 6.5% formic acid, manifest electrophoretic mobilities that differ almost two-fold in the PA solvent. CCA and CAA are also clearly resolvable in the latter solvent.

Compositional isomers, such as UCA and CUA, that require further characterization in order to determine sequence are digested (as described above) with U_2 enzyme under "overcutting" conditions. For oligomers of the form $(C_xU_y)A_z$, tertiary electrophoresis is performed in the PA solvent. In this system removal of a C residue from such an oligomer—e.g., $(C_2U_2)A \rightarrow (CU_2)A$—increases electrophoretic mobility two-fold, whereas removal of a U residue increases mobility about 2.8 fold. In other words, although electrophoretic mobilities in this system are a strong function of oligomer size, one can also detect compositional changes within defined limits.

For oligomers of the form $(C_xU_y)G$ ("overcut" with ribonuclease U_2) tertiary electrophoresis is done either in 6.5% formic acid ($y = 1$), or in the above "high salt" buffer ($y \geqq 2$). The former resolves oligomers of the form C_xG, and UG, CUG, UCG, and larger members in the $(C_xU)G$ series, (by composition), but does not resolve well those oligomers containing two U residues and variable amounts of C. These latter are resolved in the "high salt" buffer, however; removal of a C residue [i.e., $(C_xU_2)G \rightarrow (C_{x-1}U_2)G$] increases electrophoretic mobility 10–20% in general.

On a few occasions we have also used spleen diesterase digestion to generate fragments (Sanger *et al.*, 1965).

Results and Discussion

As can be seen above ("Materials and Methods" and references cited therein) the utility of ribonuclease U_2 lies in the fact that its mode of cleavage can be altered rather dramatically by altering pH, etc. Under the above "normal" conditions it cleaves T_1 oligomers not only at A residues but also leaves partially intact AA ... stretches, if these are present—an example being the production of UCA and UCAA from the oligomer UAUCAAUG. And, the enzyme also produces appreciable levels of larger, partial digestion products under these conditions. Under what we call "overcutting" conditions (above), pyrimidine stretches are also cleaved, permitting sequencing of the initial fragments of T_1 oligomers produced by ribonuclease U_2 cleavage under "normal" conditions.

Fig. 1 is a two dimensional electrophoretogram of a ribonuclease T_1 digest of *E. coli* B *236* 16S rRNA, done by the method of Sanger and coworkers as modified slightly by Woese and Sogin to improve resolution in the second (DEAE) dimension (Sanger *et al.*, 1965; Woese and Sogin, manuscript in preparation). Spot numbering is according to the convention of Sogin *et al.* (1971) explained above.

The analysis of the various spots in the electrophoretogram, both as regards number of oligomers and their compositions (determined by the 2^d and 3^d procedures described above) are given in the accompanying table. Where points of difference exist between sequences determined by us and those reported by other workers, an explanation for our conclusion is given.

What is immediately apparent from Table 1 is that very few differences in sequences determined by different laboratories are found when oligomer

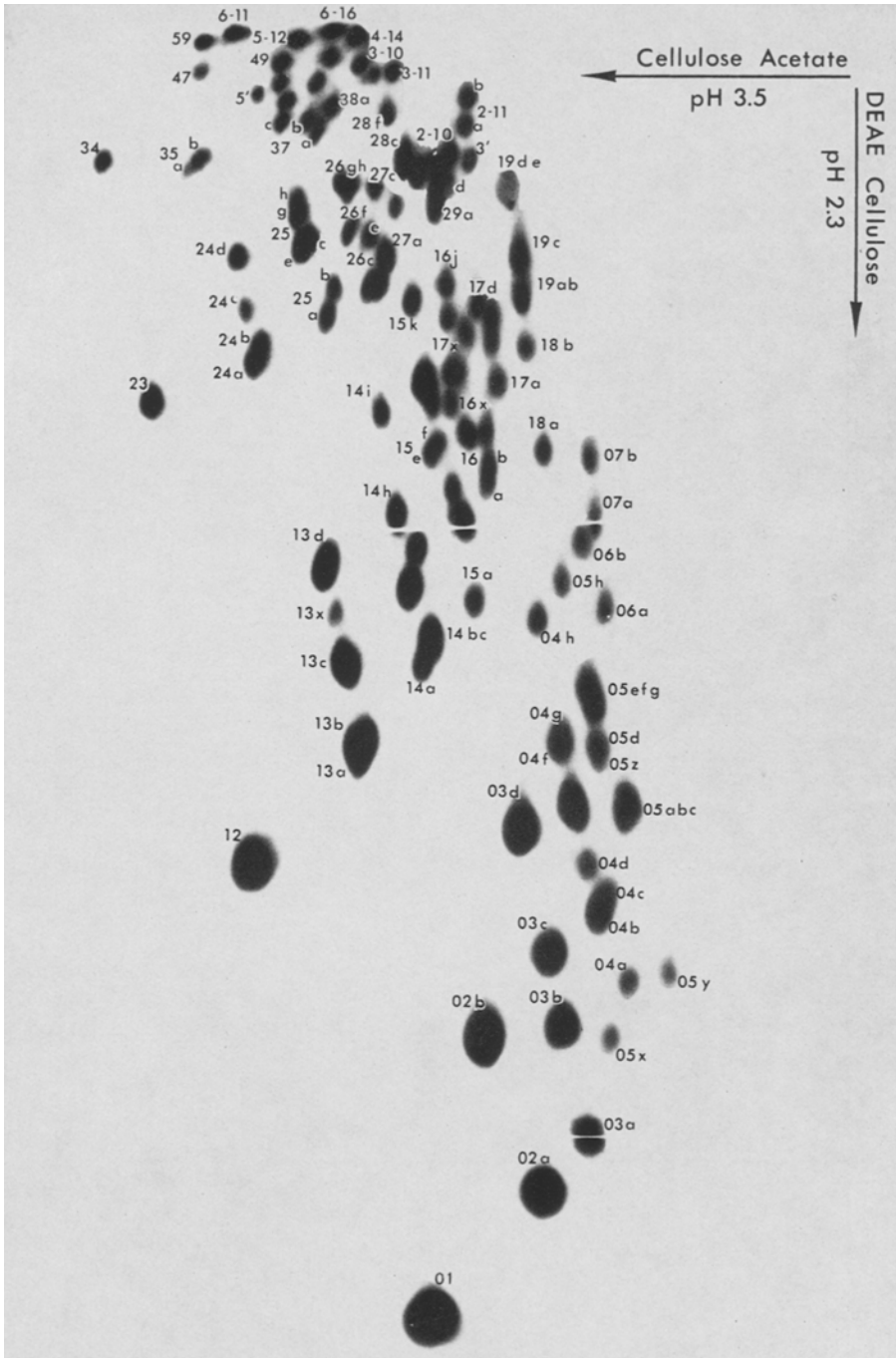

Fig. 1. Two dimensional electrophoretogram of a ribonuclease T_1 digest of *E. coli* 16S ribosomal RNA. *E. coli* 16S rRNA labelled with ^{32}P and isolated as described has been digested with ribonuclease T_1 and the resulting products separated electrophoretically, according to the basic method of Sanger and coworkers (1965). In this case the second (i.e., DEAE paper) dimension has been run in the modified "high salt" solvent of Woese and Sogin (see above). The oligomer spot naming convention has been explained above (the 5′ and 3′ terminal oligomers are designated as such)

Tabelle 1. Oligonucleotide catalog for *E. coli* 16S ribosomal RNA

Composition	A	B	C	D	Comments
			Dimers		
CG	42	(02a)	27	(108)	Repeated determinations give 42 ± 10 %
AG	47	(02b)	42	(107)	
UG	45	(12)	42	(91)	
			Trimers		
CCG	6	(03a)	5	(109)	
CAG	9	(03b)	7	(97)	
ACG	10	(03c)	11	(96)	
AAG	18	(03d)	25	(95)	
UCG	10	(13a)	10	(90)	
CUG	11	(13b)	10	(89)	
UAG	14	(13c)	11	(87)	
AUG*	1	(13x)	1	—	See footnote [a]
AUG	13	(13d)	16	(82)	
UUG	9	(23)	12	(54)	
			Tetramers		
CCCG	1	(04a)	$1^1/_2$	(110)	
CACG	2	(04b)		(101b)	CA; CG
CCAG	3	(04c)	$5^1/_2$	(101a)	
ACCG	1	(04d)		(100)	CCG
CAAG	7	(04e)	6	(99)	
ACAG	2	(04f)	2	(98b)	
AACG	2	(04g)		(98a)	
AAAG	1	(04h)	3	(94)	
UCCG	2	(14a)	$1^1/_2$	(88)	CCG; CG

CUCG	2	(14b)	0	—	CUG; UCG; UG; CG
CCUG	3	(14c)	4	(86)	
UACG	1	(14d)	1	(83)	
UCAG	1–3	(14e)	2	(84)	
CUAG	3–1	(14f)	2	(81)	
CAUG	2	(14g)	2	(79)	
ACUG	4	(14h)	4	(78)	
AAUG	2	(14i)	4	(73)	Repeated determinations give 2 copies
UCUG	3	(24a)	4	(53)	CUG; UG
CUUG	$2^1/_2$	(24b)	3	(51)	UUG; UG
UAUG	1	(24c)	$1^1/_2$	(46)	
AUUG	$2^1/_2$	(24d)	$2^1/_2$	(44)	
UUUG	3	(34)	3	(20)	
			Pentamers		
$\overset{*}{C}CGCG$	1	(05x)	1	(111)	n.d.
$\overset{*}{C}\overset{*}{C}CCG$	1	(05y)	1	(101c)	n.d.
$\overset{*}{C}AACG$	1	(05z)	1	(105)	See foortnote [b]
CCACG	1	(05a)	1	(106a)	CCA; CA; CG; CCG
CACCG	1	(05b)			
ACCCG	0	—	1	(106b)	No CCCG in this region
CCCAG	1	(05c)	1	(102)	$\overline{C}_3$, $\overline{AG}$; CCCA; G
CAACG	1	(05d)	1	(104d)	$\overline{AAC}$; $\overline{C}$; $\overline{G}$; CA; CAA; CG
ACCAG	1	(05e)	1	(104c)	$(\overline{AC})_3$; $\overline{AAC}$; $\overline{C}_2$; $\overline{AG}$; $\overline{G}_2$; CCA; CA; CCG; CG
ACACG	1–2	(05f)	1	(104a)	
AACCG	1	(05g)	1	(103b)	
CACAG	0	—	1	(104b)	See footnote [c]
AAACG	1	(05ih)	1	(92)	$\overline{A_3C}$; $\overline{G}$; CG
UCCCG	$1^1/_2$	(15a)	2	(85)	CCCG; CCG; CG

Table 1 (continued)

Composition	A	B	C	D	Comments
			Pentamers (continued)		
UCCAG	1	(15b)	1	(77b)	UCCA; CUCA; $\overline{AG}$, etc.
CUCAG	2	(15c)	2	(77a)	
CCAUG	1	(15d)	1	(76a)	$\overline{C}_2$; $\overline{AU}$; $\overline{G}$; CCA; UG
UAACG	1	(15e)	2	(74)	$\overline{U}$; $\overline{AAC}$; $\overline{G}$; UA; UAA; CG
UCAAG	1	(15f)	1	(72)	$\overline{AAG}$; UCAA, etc.
AUCAG	1	(15g)	0	—	$\overline{AU}$; $\overline{C}$; $\overline{AG}$: UCA; c.f. (70b) below
AAUCG	1	(15h)	1	(70a)	$\overline{AAU}$; $\overline{C}$; $\overline{G}$; UCG
AACUG	2–3	(15i)	3	(68)	$\overline{AAC}$; $\overline{U}$; $\overline{G}$; CUG
UAAAG	3–2	(15j)	$2^1/_2$	(66)	$\overline{AAAG}$; UA; UAA; UAAA
AAAUG	2	(15k)	2	(60)	$\overline{AAAU}$; UG
UCUCG	1–2	(25a)	$1^1/_2$	(52)	UCUCG; CUCG
CCUUG	1	(25b)	1	(47a)	CCUUG
ACUUG	1	(25c)	1	(42b)	$\overline{AC}$; $\overline{AU}$; $\overline{C}$; $\overline{U}$; $\overline{G}$; UCUG; CUUG, etc.
AUCUG	1	(25d)	1	(42a)	
UUAAG	3	(25e)	4	(39a)	$\overline{AAU}$; $\overline{AAG}$; UUA; UUAA; UA; UAA; UG
UAAUG	1	(25f)		(39b)	
AUAUG	—	—	2	(38)	$\overline{AU}$; $\overline{U}$; $\overline{AG}$; UUA; No UA; No UG; No $\overline{G}$
AUUAG	2	(25g)	—	—	
AAUUG	1	(25h)	1	(35)	$\overline{AAU}$; UUG, etc.
UCUUG	2–1	(35a)	$2^1/_2$	(18)	UCUUG
CUUUG	1–2	(35b)	0	—	CUUUG (sequence probable but not certain)
			Hexamers		
CACAAG	1	(06a)	1	(103a)	$\overline{C}$; $\overline{AC}$; $\overline{AAG}$; CA; CAA
CAAAAG	1	(06b)	1	(93b)	$\overline{C}$; $\overline{AAAAG}$; CA; CAA; etc.

UCCACG	1	(16a)	1	(76b)	$\overline{U}$; $\overline{C}$; $\overline{AC}$; $\overline{G}$; $\underline{CG}$; $\underline{UCCA}$
CCAUCG	1	(16b)	1	(75a)	$\overline{C}$; $\overline{AU}$; $\overline{G}$; $\underline{CCA}$; $\underline{UCG}$
ACCUCG	1	(16c)	—	—	$\overline{AC}$, etc; $\underline{CCUCG}$; No oligomer corresponding to (75b) found
ACUCCCG	—	—	1	(75b)	
U[AACC]G	1	(16d)	1	(69b)	$\overline{AAC}$; $\overline{U}$; $\overline{C}$; $\overline{G}$; $\underline{UA}$; $\underline{UAA}$; $\underline{CCG}$; [u.F.]
CUAACG	1	(16e)	1	(69a)	$\overline{AAC}$, etc., $\underline{CUAA}$; $\underline{CG}$
AACCUG	1/2	(16f)	1/2	(67)	$\overline{AAC}$, etc.; $\underline{CCUG}$
AUCCAG	0	—	1	(70b)	No evidence for; c.f. (15g.)
UAACAG	1	(16g)	1	(63b)	$\overline{AAC}$; $\overline{AG}$, etc.; $\underline{UA}$; $\underline{UAA}$; $\underline{CA}$
UAAACG	1	(16h)	1	(63a)	$\overline{AAAC}$, etc.; $\underline{UA}$; $\underline{UAA}$; $\underline{UAAA}$; $\underline{CG}$
AAUACG	1	(16i)	1	(62)	$\overline{AAU}$; $\overline{AC}$; $\overline{G}$; $\underline{UA}$; $\underline{CG}$
AAACUG	1	(16j)	1	(57a)	$\overline{AAAC}$; $\overline{U}$; $\overline{G}$; $\underline{CUG}$; $\overline{AC}\,\overline{UG}$
UUCCCG	1	(26a)	1	(47b)	$(\underline{UC_3})G$; $\underline{CCCG}$; $\underline{CCG}$
CUUCCG	1	(26b)	1	(48)	$\underline{UUCCG}$; $\underline{UCCG}$; $\underline{CUUCG}$; $\underline{UUCG}$
CCUUCG	2	(26c)	2	(49)	
CCCUUG	1	(26d)	1	(45)	$\underline{CCCUUG}$
CAUCUG	1	(26e)	1	(41)	$\overline{AU}$, etc.; $\underline{CA}$; $\underline{UCUG}$
UAAUCG	1	(26f)	1	(40)	$\overline{AAU}$, etc.; $\underline{UA}$; $\underline{UAA}$; $\underline{UCG}$
AUCAUG	1	(26g)	1	(34)	$\overline{AU}$, etc.; $\underline{UG}$; $\underline{UCA}$; $\underline{UCAUG}$
AUACUG	1	(26h)	1	(33)	$\overline{AU}$; $\overline{AC}$, etc.; $\underline{UA}$; $\underline{CUG}$
$\overset{**}{A}$ACCUG	1	(16x)	1	(71)	See footnote [d]
				Heptamers	
CAAACAG	1	(07a)	1	(93c)	$\overline{AAAC}$; $\overline{C}$; $\overline{AG}$; $\underline{CA}$; $\underline{CAA}$; $\overline{AC}\,\overline{AG}$
[ACC]AAAG	1	(07b)	1	(93d)	$\overline{AC}$; $\overline{AAAG}$; $\overline{C}$; $\underline{CCA}$; $\underline{CCAA}$; [u.F.]
CAUCACG	0	—	1	(64)	Not found in our preparations
C[AACUC]G	1	(17a)	1	(65)	$\overline{AAC}$, etc.; $\underline{CAA}$; $\underline{CUCG}$; [u.F.]

Table 1 (continued)

Composition	A	B	C	D	Comments
			Heptamers (continued)		
CAUAACG	1	(17b)	1/2	(59)	$\overline{AU}$; $\overline{AAC}$, etc.; CA; UA; UAA; CG
CACAAUG	1	(17c)	1	(56a)	$\overline{AC}$; $\overline{AAU}$, etc.; CA; CAA; UG
AAUACCG	1	(17d)	1	(58)	$\overline{AAU}$; $\overline{AC}$, etc.; UA; CCG
UUACCCG	1	(27a)	1	(43)	$\overline{AC}$, etc.; UUA; CCCG
ACCUUCG	1	(27b)	1	(36)	$\overline{AC}$, etc.; CCUUCG
UAAUACG	1	(27c)	1	(27)	$\overline{AAU}$, $\overline{AC}$, etc.; UA; UAA; CG
CCUCUUG	1	(37a)	1	(17)	CCUCUUG; {(U$_3$C)G identified by mobility in PA / solvent
CUCAUUG	1	(37b)	1	(16)	$\overline{AU}$, etc.; CUCA; UUG
UUAAUCG	1	(37c)	1	(15)	$\overline{AAU}$, etc., UUA; UUAA; UCG
AUAUUCG	—	—	1	(13)	$\overline{AU}$, etc.; UUCA; UG; no UA or (U$_2$C)G found
AUUCAUG	1	(37d)	—	—	
AAUAUUG	1	(37e)	1	(11)	$\overline{AAU}$, $\overline{AU}$, etc.; UA; UUG; no UAA
UUUUCAG	1	(47)	1	(6)	UUUUCA; $\overline{CAG}$[†]
*UAACAAG	1	(17x)	1	(57b)	$\overline{AAG}$; *$\overline{UAAC}$; CAA. See footnote [e]
			Octamers		
CCCCCCUG	1	(18a)	1	(93a)	CUG; (C$_2$U)G; (C$_3$U)G; etc.; 1° position demands ≧8 bases
CACCACUG	—	—	1	(61)	$\overline{AC}$; $\overline{C}$; $\overline{U}$; $\overline{G}$; CA; CCA; CUG; CACUG
CCACACUG	1	(18b)	—	—	
AUACCCUG	1	(28a)	—	—	$\overline{AU}$; $\overline{AC}$, etc.; CCCUG; $\overline{UACCCUG}$; No CCG
AUACUCCG	—	—	1	(28)	
CUACAAUG	1	(28b)	1	(24a)	$\overline{AAU}$; $\overline{AC}$, etc.; CUA; $\overline{CAAUG}$; CAA; UG

AAUUCCAG	1	(28c)	1	(24b)	$\overline{AAU}$; $\overline{AG}$, etc.; UUCCA; $\overline{CCAG}$†; UUCCAG
ACUCCAUG	1	(28d)	1	(29a)	$\overline{AC}$; $\overline{AU}$; $\overline{AG}$; $\overline{G}$, etc.; UG; CUCCA; UCCCUA
AUCCCUAG	1	(28e)	—	—	
AU(C_2U)AG	—	—	1	(29b)	1° position alone demands > 7 bases
AUAAACUG	1	(28f)	—	—	$\overline{AU}$; $\overline{AAAC}$, etc.; UA; UAA; UAAA;
AUUAAACG	—	—	1	(21)	CUG; no UUA or CG found
UCAUCAUG	1	(38a)	—	—	$\overline{AU}$, etc.; UG; UCA; UCAUG; no UCG found
UCAUCAUCG	—	—	1	(14)	
AAUUACUG	1	(38b)	1	(12)	$\overline{AAU}$; $\overline{AC}$, etc; UUA; CUG

Nonamers

UACACACCG	1	(19a)	—	—	
U[$(AC)_3$, C_2]G	—	—	1	(56b)	$(\overline{AC})_3$; $\overline{C}$; $\overline{U}$; $\overline{G}$; UA; CA; CCG; CACACCG
[CA_2CA_2CUA]CG	1	(19b)	—	—	
C[$(AC)_3$, U]G	—	—	1	(55b)	CUA; CG; sequence not certain
AAAUCCCCG	1	(19c)	1	(55a)	$\overline{AAAU}$, etc.; UCCCCG
AACUCAAAG	1	(19d	1	(50a)	$\overline{AAC}$; $\overline{AAAG}$, etc.; CUCAAA; etc.
CAUACAAAG	1	(19e)	1	(50b)	$\overline{AU}$, $\overline{AC}$; $\overline{AAAG}$; UA; CA; No UAA, etc.
CCCCUUACG	1	(29a)	—	—	CCCCUUA; CG; $\overline{AC}$, etc.
CCCCCUUCACG	—	—	1	(37)	1° position demands < 10 bases
CUAACUCCG	1	(29b)	1	(32)	$\overline{AAC}$, etc.; CUAA; CUCCG
CUCAACCUG	1	(29c)	—	—	$\overline{AAC}$, etc.; CUCAA; CCUG; No CCCUG
CCUAACCCUG	—	—	1	(30a)	1° position strongly suggests 9 bases
CUCACCUAG	1	(29d)	—	—	C[UC]A → CUA and UCA; CCUAG
CUCACCUACAG	—	—	1	(31)	1° position suggests 9–10, not 11 bases

Table 1 (continued)

Composition	A	B	C	D	Comments
			Nonamers (continued)		
ACUCCUACG	1	(29e)	1	(30b)	$\overline{AC}$, etc.; CUCCUA; CG; sequence probable but not certain
CUAAUACCG	1	(29f)	—	—	$\overline{AAU}$; $\overline{AC}$, etc.; CUAA; UA; CCG; No CG found
CUAAUACG	—	—	1	(25)	
UACUUUCAG	1	(49)	—	—	$\overline{AC}$; $\overline{AG}$. etc.; UA; CUUUCA; $\overline{CAG}$†
UACUUCUAG	—	—	1	(7)	
UUUAAUUCG	1	(59)	—	—	$\overline{AAU}$, etc.; UUUA; UUCG; No UA(A) or $(U_4C)G$
UAAUCUUUG	—	—	1	(3)	
			Decamers and larger		
*UCACACCAUG	1	(2, 10)	1	(26)	$\overline{AC}$; $\overline{AU}$, etc.; UG; CA; CCA; UCA; CCAUG; CACCAUG; see footnote [f]
CCUAACACAUG	1	(2, 11, a)	—	(22)	$\overline{AAC}$; $\overline{AC}$; $\overline{AU}$, etc.; CCUAA; UG; CA; $\overline{CAUG}$; $\overline{CACAUG}$; No CAACAUG found
C(U, C)ACAACAUG	—	—	1	—	
ACCUCAUAAAG	1	(2, 11, b)	—	—	$\overline{AC}$; $\overline{AU}$; $\overline{AAAG}$, etc.; CCUCA; UAA, etc.; $C\overline{C}(\overline{UC})AUAAAG$; No $(C_4U)A$ found
ACCCUCAUAAAG	—	—	1	(19)	
AUAACUACUG	1	(3, 10)	—	—	$\overline{AU}$; $\overline{AAC}$; $\overline{AC}$, etc.; UA; UAA; CUA; CUG; No CA or $(CU_2)G$ found
AUAACACUUG	—	—	1	(9)	
AACCUUACCUG	1	(3, 11)	1	(10a)	$\overline{AAC}$; $\overline{AC}$, etc.; CCUUA; CCUG
CUUAACCUUCG	1	(4, 11)	—	—	$\overline{AAC}$, etc.; CUUAA; CCUUCG; No CCUUG found [i.e., UUG]
CUUAACCUUG	—	—	1	(8)	
UUAAAACUCAAAUG	1	(4, 14)	—	—	$\overline{AAAU}$; $\overline{AAAAC}$, etc.; UUA; UUAA, etc.; CUCA; CUCAAA, etc.; UG; No UCG or CUA; etc. found
UUAAAACUAAAUCG	—	—	1	(5)	
CUUACCACUUUG	1	(5, 12)	1	(4)	$\overline{AC}$, etc.; CUUA; CCA; CUUUG; CCACUUUG

UUAAUACUUUG	1	(6, 11)	—	—	$\overline{AAU}$; $\overline{AC}$; etc., $\underline{UUA}$; $\underline{UUAA}$; $\underline{UA}$; $\underline{CUUUG}$;
UCUAAUACUUUG	—	—	1	(1)	$\underline{No}$ $(U_2C)AA$ found
CAACCCUUAUCCUUUG	1	(6, 16)	—	—	$\overline{AAC}$; $\overline{AU}$, etc., $\underline{CAA}$; $\underline{CCCUUA}$; $\underline{UCCUUUG}$;
CAACCCUUUUU(AU, C_2)G	—	—	1	(2)	$\underline{No}$ $(C_xU_5)A$ or $(U, C_y)G$ found
pAAAUUG	1		1	(23)	$\overline{pAAAU}$; $\overline{U}$; $\overline{G}$; $\underline{pA}$; $\underline{AA}$; $\underline{UUG}$
AUCACCU(U, CC)U(A_{OH})	1		1	(10b)	$\overline{AU}$; $\overline{AC}$; etc.; $\underline{UCA}$; $(\underline{C_4U_3})(A_{OH})$; CCU†; CCUU†; 3′ portion of sequence somewhat tentative; A_{OH} not determined by us

Oligomer sequence and frequency of occurrence as determined by us and by Fellner *et al.* (1972) are given for all oligomers found in a ribonuclease T_1 digest of *E. coli* 16S rRNA (dimers and larger). Columns A and B give, respectively, the number of copies we find and our oligomer code designation (for cross-reference to Fig. 1); columns C and D do the same for the oligomers reported by Fellner *et al.* (1972).

The "Comments" column lists (where not completely self-evident) the secondary and tertiary products from which oligomer sequence was deduced (not all of which are necessarily given). Also given is our reason for accepting our version of a sequence when it differs from that reported by Fellner *et al.* (1972).

Heavy underlining indicates fragment produced by ribonuclease U_2 secondary digestion, heavy overlining those produced by pancreatic ribonuclease secondary digestion (see Materials and Methods). Light underlining (overlining) indicates a tertiary fragment produced by ribonuclease U_2 (pancreatic ribonuclease) digestion of the secondary fragment in question.

n.d. = Sequence not totally determined by us (see Fellner and Sanger, 1968).
† = Fragment generated by a technique to be published later.
[u.F.] = Sequence in brackets not determined by Fellner *et al.* (1972a).
[u.W.] = Sequence in brackets not determined by ourselves.
* = Over a base indicates post-transcriptional modification.

[a] "$\overset{*}{G}$" runs as monomer *very* near G in "PA" solvent and between G and AC in 6.5% formic acid; $U\overset{*}{G}$ produced by ribonuclease U_2 has mobility slightly greater than normal UG.

[b] C modified at 5 position; 5 H^3C loses H^3 (Woese *et al.*, unpublished).

[c] 2° U_2 digestion yields no CA:G = 2:1 (m.r.), and 2° pancreatic digestion yields no AG in this region of 1° pattern.

[d] $\overset{*}{A}\overset{*}{A}C$ has slightly greater mobility than AAC in 6.5% formic acid and runs with AG in a PA solvent system.

[e] $\overset{*}{U}AAC$ not cleaved by pancreatic ribonuclease at normal concentrations, but is by 5-fold excess; the resulting $\overset{*}{U}$ moves faster than normal U in all solvent systems. $\overset{*}{U}A$ also moves faster than normal UA.

[f] In unpublished double label experiments yielding 5-^{3}H labeled U and C, the $\overset{*}{C}A$ resulting from tertiary cleavage of $(U\overset{*}{C})A$ contains ^{32}P but no ^3H; therefore, C is modified on the 5 position (Woese *et al.*, unpublished).

length is trivial—i.e., ≤ 7 bases. For octamers and larger, however, over half of the sequences disagree. Almost without exception the discrepancies do not involve the fragments generated by *pancreatic* ribonuclease digestion [a rare exception being AUUAG, our sequence, *vs.* AUAUG, the sequence of Fellner *et al.* (1972a)]. Rather, they involve the order in which the pancreatic ribonuclease fragments are placed within the oligomer. In our case this ordering in done through the use of overlapping pancreatic and U_2 ribonuclease fragments, while Fellner *et al.* use venom diesterase products and other methods (Fellner *et al.*, 1972a).

Several sources of systematic error are possible in any of these sequencing methods. For one, the amount of radioactivity remaining by the time tertiary cleavage products are obtained generally precludes any further, "ultimate" analysis; i.e., tertiary cleavage products generally have to be identified by position only on a one dimensional electrophoretogram, in the presence of known, marker, oligomers. The large variety of tertiary fragments produced can in *some* cases lead to misidentification. (To a lesser extent such ambiguities exist for secondary cleavage products, but these we always resolve by an appropriate tertiary procedure.)

A second source of error results from oligomers coincident (or overlapping) on the primary pattern. Determining which secondary fragments are associated with which oligomer is not always certain in these instances.

A third source of error is possible endonuclease activity in exonuclease preparations, which could lead to misinterpretation of digestion products (unless these are identified unequivocally by some further procedure).

Needless to say, corrections in many of the T_1 oligomer primary structures will alter the projected secondary structures for some of the "loops" reported in the 16S ribosomal RNA. It is pointless to go into this at this time. [The interested reader might, for example, note the changes that occur in fragment "P" of Fellner *et al.* (1972b). P appears to be a long coaxial largely base paired helix. Our corrections to the sequences of T_1 oligomers numbers 1, 4, 21, and 37 change the projected base pairing sufficiently to make postulation of a single coaxial helix, as opposed to two noncoaxial helical segments, unlikely.]

What the present study has made clear is that ribonuclease U_2 is a powerful oligomer sequencing tool, one that can reveal systematic errors inherent in certain other sequencing approaches.

Acknowledgments. We are most grateful to Professor Fujio Egami for making possible these studies, through his generous gift of ribonuclease U_2 and arranging for the stay of one of us (T. U.) in the United States.

This work was supported by a National Science Foundation grant (NSF GB 16298) awarded to C. W.

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)
Fellner, P., Ehresmann, C., Ebel, J. P.: Biochimie **54**, 880 (1972a)
Fellner, P., Ehresmann, C., Stiegler, P., Ebel, J. P.: Nature, New Biol. (Lond.) **239**, 1 (1972b)
Fellner, P., Sanger, F.: Nature (Lond.) **219**, 236 (1968)
Franklin, R. M.: Proc. Nat. Acad. Sci. (Wash.) **55**, 1504 (1966)
Kirby, K. S.: Biochem. J. **64**, 405 (1956)
Pechman, K. J., Woese, C.: J. Molec. Evolution **1**, 230 (1972)
Sanger, F., Brownlee, G. G., Barrell, B. G.: J. Molec. Biol. **13**, 373 (1965)
Sogin, M., Pace, B., Pace, N. R., Woese, C.: Nature (Lond.) **232**, 48 (1971)
Sogin, M., Woese, C., Pace, B., Pace, N. R.: J. Molec. Evolution **2**, 167 (1973)
Uchida, T., Arima, T., Egami, F.: J. Biochem. (Tokyo) **67**, 91 (1970)

Dr. C. R. Woese
Department of Microbiology
University of Illinois
Urbana, Illinois 61801, USA