On the Probability of a Common Origin for tRNA and 5S rRNA[1]

Mark P. Staves and James C. Lacey, Jr.

Department of Biochemistry, CH19, University of Alabama at Birmingham, Birmingham, Alabama 35294, U.S.A.

Z. Naturforsch. **44c**, 296–306 (1989); received April 5, 1988

tRNA, 5S rRNA, Evolution, Universal Translator, Sequence Similarity

As part of a continuing study on the origin of genetic coding and the process of protein synthesis, we have compared sequences of a large number of transfer RNAs and several 5S ribosomal RNAs using automated routines. Transfer RNAs were found to exhibit a high degree of matching with 5S rRNAs. These matches are considered to be indicative of sequence homology, reflecting common ancestry of the two molecules. Other possible explanations for the matches (convergence, transfection) are discussed and found to be highly implausible. Matches are also found between 5S rRNA and the introns of yeast precursor tRNAs. Many matches extend from before the 3′ end to after the 5′ end of circularized 5S rRNA sequences. Data is presented which indicates a tandemly duplicated or circular molecule could have served as the precursor to both 5S rRNA and tRNA. A derivative of this molecule may have functioned as a universal translator before the evolution of the highly specialized tRNAs.

Introduction

We have a long standing interest in the origin of the genetic code and the process of protein synthesis. Some previous findings include correlations of properties between amino acids and their anticodonic nucleotides [1–2], selective affinities between amino acids and their anticodonic nucleotides [3–5] and selective reactions between amino acids and their anticodonic nucleotides [6–7]. These results indicate an early, more direct involvement of amino acids with the anticodon in a primitive transfer RNA. Numbers of proposals have been made in that regard and include the idea that the first tRNAs (or adaptor molecules) were mononucleotides [8]. Others have suggested larger adaptors. For example, Crick *et al.* proposed a small tRNA with a five-letter anticodon. Hopfield presented a hairpin model in which the amino acid was very near the anticodon for recognition purposes [9]. Kuhn and his coworkers [10] have proposed that early tRNAs were hairpin structures with self associative possibilities allowing them to line up on a template. One attractive line of reasoning is that there might have been an early universal tRNA which preceeded the individualized, specialized tRNAs. Such a molecule would be expected to have features which would allow its binding to a ribosome-like structure and be of sufficient length to contain all possible anticodons in the anticodon loop. The 5S rRNA, which is found in all known ribosomes and contemporary tRNAs may be descendents of such a molecule.

Previous reports of investigations into the evolution of the genetic coding mechanism have described sequence homologies between transfer RNAs (tRNAs) and structural RNAs [11, 12]. The model derived from that research [12] predicts matching base sequences may be found between tRNA and 5S rRNA. In addition, there are also biochemical and biphysical data which suggest an affinity between these two molecules. The lengths of some precursor-tRNAs and 5S rRNAs are approximately the same [13–15] and they share similar nucleotide compositions [16]. Both molecules form highly conserved secondary structures consisting of several loop and stem regions. In eukaryotic cells, the transcription of both 5S rDNA and tDNA is catalyzed by RNA polymerase III. Additionally, in both cases, transcription involves the binding of a regulatory protein to specific intragenic sites [17, 18] representing the DHU and TΨC regions on tRNA and 5S rRNA regions homologous to the DHU loop as we show in this report. This transcription mechanism seems to be unique and suggests a phylogenetic relationship between tRNAs and 5S rRNAs.

Years ago we reported matches between *E. coli* $tRNA^{Tyr}$ precursor and *E. coli* 5S rRNA and suggested an evolutionary relationship between these two molecules [19]. This idea, however, has received little attention until recently. Sequence homologies

[1] This paper is dedicated to the memory of Prof. David P. Bloch, in recognition of his important contributions to cell biology and molecular evolution, and in appreciation of his encouragment, collaboration and friendship.

Reprint requests to Dr. M. P. Staves.

Verlag der Zeitschrift für Naturforschung, D-7400 Tübingen
0341–0382/89/0300–0296 $ 01.30/0

have been used successfully to demonstrate relationships between small subunit ribosomal RNA and tRNA [11, 12] and among tRNAs [20]. Similar methods have been used here to investigate the relationships between tRNAs and 5S rRNAs. The present investigation was undertaken to compare more systematically and thoroughly the sequences of tRNAs and 5S rRNAs as a test of the hypothesis of common ancestry of those molecules.

Methods

Sequences of 123 tRNAs from widely divergent sources (an archaebacterium, *Halobacterium volcanii*, a eubacterium, *E. coli*, bovine mitochondria, *Euglena* chloroplast, yeast and *Drosophilla*) were taken from the literature [21, 22]. The 5S rRNA sequences used are found in the compilation of Erdmann and Wolters [23]. Sequences of 5S rRNAs and tRNAs were compared using the automated routines of Goad and Kanehisa [24] which search for local matching regions. This method compares sequences in a pairwise manner. One sequence is designated as the reference molecule. The computer systematically scans this molecule looking for identical sequences among successive frames of the second molecule. Values for the numbers of expected matches per search, based on coincidence, were calculated using the formulae of Goad and Kanehisa [24] as modified by Bloch *et al.* [12].

Results and Discussion

Table I lists the matches found in searches between *Prochloron* 5S rRNA and all 123 tRNAs listed

Table I. Matching regions between 5S rRNAs of *E. coli*, yeast, *Prochloron* (Prochl), *Anacystis* (Anac), *Tetrahymena* (Tet) and tRNAs from Bovine mitochondria (Bovmt), *E. coli*, *Euglena* chloroplast, *H. volcanii* and *Drosophilla*. The positions of the matches are identified by base numbers. The number of matches, mismatches and gaps as well as the total length of each matching region is indicated. The expected values were calculated from the formulae of Goad and Kanehisa [24] as modified by Bloch *et al.* [11].

5S RNA	tRNA		Position on 5S RNA	tRNA	Matches, Mismatches, Gaps, Length	Expected value (E)
E. coli	*Bovmt*	asp	52–59	50–57	8, 0, 0, 8	0.047
		thr	99–110	38–50	11, 1, 1, 13	0.050
	Yeast	arg 1	43–54	35–45	11, 0, 1, 12	0.027
		lys 1	65–79	8–22	12, 3, 0, 15	0.056
		arg 3	90–103	59–71	12, 1, 1, 14	0.043
	E. coli	thr 1	74–88	33–48	13, 2, 1, 16	0.076
		thr 1	65–79	8–22	12, 3, 0, 15	0.063
		tyr 1	81–92	2–12	11, 0, 1, 12	0.029
		arg 1	62–82	5–25	17, 4, 0, 21	0.0010
		gln 1	82–98	1–17	14, 3, 0, 17	0.0079
		gln 2	82–98	1–17	14, 3, 0, 17	0.0077
		his 1	65–79	9–23	12, 3, 0, 15	0.057
		ile 1	13–25	7–20	12, 1, 1, 14	0.063
		ile 2	11–25	7–19	12, 3, 0, 15	0.078
		val 1	62–77	5–20	13, 3, 0, 16	0.032
		val 1	7–17	1–11	10, 1, 0, 11	0.087
		lys 1	65–79	8–22	12, 3, 0, 15	0.047
		met 1	63–77	6–20	12, 3, 0, 15	0.054
		gly 5	65–79	8–22	12, 3, 0, 15	0.095
	Euch	phe 1	65–79	8–22	12, 3, 0, 15	0.052
		ala 1	65–79	8–22	12, 3, 0, 15	0.068
		gly 1	65–88	8–31	18, 6, 0, 24	0.0010
		gly 1	12–27	19–32	13, 1, 1, 15	0.011
		val 1	65–79	8–22	12, 3, 0, 15	0.057
		ile 1	17–31	63–76	13, 1, 1, 15	0.033
	Dros	ala 1	64–79	7–22	14, 2, 0, 16	0.0025
		met 1	15–29	20–34	13, 2, 0, 15	0.0077
		met 2	15–29	20–34	12, 3, 0, 15	0.088

5S RNA	tRNA		Position on 5S RNA	tRNA	Matches, Mismatches, Gaps, Length	Expected value (E)
Yeast	*Bovmt*	leu 1	52–69	1–17	15, 2, 1, 18	0.0042
		trp 1	63–70	23–30	8, 0, 0, 8	0.097
		asn 1	52–64	26–38	11, 2, 0, 13	0.051
		ser 2	9–17	35–43	9, 0, 0, 9	0.020
	Yeast	gly 2	49–64	7–21	13, 2, 1, 16	0.056
		gly 2	76–88	6–18	11, 2, 0, 13	0.052
		lys 1	50–68	8–25	15, 3, 1, 19	0.022
		thr 1	55–65	13–23	10, 1, 0, 11	0.048
		leu 3	70–84	8–22	12, 3, 0, 15	0.056
		val 3	78–88	8–18	10, 1, 0, 11	0.047
	E. coli	ala 1	50–71	8–30	18, 4, 1, 23	0.0031
		asp 1	49–66	7–24	14, 4, 0, 18	0.016
		thr 1	50–68	8–25	15, 3, 1, 19	0.022
		trp 1	49–68	7–25	16, 3, 1, 20	0.0075
		trp 1	24–36	23–35	12, 0, 1, 13	0.0038
		asn 1	47–63	5–21	14, 3, 0, 17	0.0047
		arg 1	49–67	7–24	16, 2, 1, 19	0.0017
		gln 1	55–65	13–22	10, 0, 1, 11	0.058
		gln 2	55–65	13–22	10, 0, 1, 11	0.058
		his 1	47–69	6–27	18, 4, 1, 23	0.0025
		ile 1	48–68	6–26	17, 4, 0, 21	0.00048
		val 1	49–68	7–26	16, 4, 0, 20	0.0016
		val 2	50–68	8–25	15, 3, 1, 19	0.024
		val 3	50–68	8–26	15, 4, 0, 19	0.0045
		lys 1	50–68	8–25	15, 3, 1, 19	0.021
		met 1	49–68	7–26	16, 4, 0, 20	0.0014
		ala 2	47–71	5–30	20, 5, 1, 26	0.0010
		gly 5	50–74	8–31	19, 5, 1, 25	0.0031
		ser 2	21–34	21–34	12, 2, 0, 14	0.021
	Euch	phe 1	50–68	8–25	15, 3, 1, 19	0.022
		ala	50–71	8–29	18, 3, 2, 23	0.012
		gly 1	50–71	8–28	17, 4, 1, 22	0.0072
		his 1	48–65	6–22	15, 2, 1, 18	0.0049
		ile 1	50–68	8–26	15, 4, 0, 19	0.0041
		tyr 1	70–89	8–26	16, 3, 1, 20	0.0074
		val 1	50–68	8–25	15, 3, 1, 19	0.021
	Hvol	arg 1	68–82	3–17	12, 3, 0, 15	0.053
		arg 1	78–88	8–18	10, 1, 0, 11	0.052
		asp 1	73–88	3–18	14, 2, 0, 16	0.0015
	Dros	ala 1	48–68	6–25	17, 3, 1, 21	0.0024
		ile 1	50–68	8–26	15, 4, 0, 19	0.0043
		leu 1	98–115	26–43	14, 4, 0, 18	0.016
		val 1	77–88	7–18	11, 1, 0, 12	0.013
		val 2	77–88	7–18	12, 0, 0, 12	0.00042
Prochl	*Bovmt*	phe 1	100–114	32–46	12, 3, 0, 15	0.028
		phe 1	41–53	51–62	11, 1, 1, 13	0.074
		gln 1	31–39	53–61	9, 0, 0, 9	0.037
		met 1	37–49	56–68	11, 2, 0, 13	0.046
		cys 1	24–38	41–56	13, 2, 1, 16	0.053
		ser 1	31–43	51–63	12, 1, 0, 13	0.0040
		his 1	51–62	27–38	11, 1, 0, 12	0.011
		leu 2	19–29	32–42	10, 1, 0, 11	0.038
		glu 1	31–41	50–60	10, 1, 0, 11	0.054
	Yeast	ala 1	30–51	48–69	17, 5, 0, 22	0.0017
		ala 1	31–39	54–62	9, 0, 0, 9	0.036
		arg 1	92–102	62–72	10, 1, 0, 11	0.047
		gly 2	31–39	51–59	9, 0, 0, 9	0.033

5S RNA	tRNA		Position on 5S RNA	tRNA	Matches, Mismatches, Gaps, Length	Expected value (E)
Prochl	Yeast	val 2	40–52	47–59	11, 2, 0, 13	0.052
		val 3	41–53	47–59	11, 2, 0, 13	0.053
	E. coli	ala 1	30–45	48–63	13, 3, 0, 16	0.016
		gly 1	106–122	37–53	13, 4, 0, 17	0.042
		gly 1	31–39	52–60	9, 0, 0, 9	0.030
		tyr 1	31–44	63–76	12, 2, 0, 14	0.016
		arg 1	31–45	54–67	13, 1, 1, 15	0.010
		his 1	48–58	28–38	10, 1, 0, 11	0.061
		ile 2	76–95	34–53	16, 4, 0, 20	0.0013
		val 2	33–45	51–63	11, 2, 0, 13	0.051
		phe 1	31–39	54–62	9, 0, 0, 9	0.031
		ala 2	30–45	48–63	13, 3, 0, 16	0.016
		gly 3	31–39	53–61	9, 0, 0, 9	0.033
		ser 2	21–39	66–84	14, 5, 0, 19	0.038
		ser 2	10–21	72–82	11, 0, 1, 12	0.017
	Euch	phe 1	4–17	61–72	12, 0, 2, 14	0.022
		leu 1	1–15	40–53	13, 1, 1, 15	0.019
		tyr 1	31–50	62–80	16, 3, 1, 20	0.010
	Hvol	ser 1	83–97	29–43	13, 2, 0, 15	0.0050
	Dros	ala 1	105–114	8–18	10, 0, 1, 11	0.057
		asp 1	31–45	55–68	13, 1, 1, 15	0.014
		glu 3	70–85	17–33	14, 2, 1, 17	0.020
		glu 3	31–39	53–61	9, 0, 0, 9	0.032
		gly 1	97–117	16–36	17, 3, 2, 22	0.041
		gly 1	31–39	52–60	9, 0, 0, 9	0.032
		his 1	12–28	11–27	13, 4, 0, 17	0.046
		met 1	57–74	56–72	15, 2, 1, 18	0.0042
		met 2	57–74	56–72	15, 2, 1, 18	0.0040
		phe 1	33–45	51–63	11, 2, 0, 13	0.048
		phe 1	102–110	5–13	9, 0, 0, 9	0.028
		val 1	33–45	51–63	11, 2, 0, 13	0.049
		val 1	1–10	3–13	10, 0, 1, 11	0.053
Euch	*Bovmt*	ile 1	79–90	17–29	12, 0, 1, 13	0.014
		asn 1	50–62	34–44	11, 0, 2, 13	0.088
		asp 1	29–37	15–23	9, 0, 0, 9	0.069
		pro 1	65–77	13–24	12, 0, 1, 13	0.0064
	Yeast	asp 1	71–84	6–19	11, 3, 0, 14	0.097
		ser 1	74–81	31–38	8, 0, 0, 8	0.094
		val 1	4–34	7–39	25, 6, 2, 33	0.000077
		arg 3	16–27	46–58	11, 1, 1, 13	0.092
		ser 10	74–81	31–38	8, 0, 0, 8	0.094
	E. coli	leu 1	105–112	42–49	8, 0, 0, 8	0.099
		leu 2	104–115	41–52	11, 1, 0, 12	0.0075
		gly 5	39–50	27–39	11, 1, 1, 13	0.067
	Euch	asp 1	22–31	58–68	10, 0, 1, 11	0.061
		leu 1	11–20	46–55	10, 0, 0, 10	0.0086
	Hvol	val 1	3–15	6–18	11, 2, 0, 13	0.032
	Dros	asp 1	78–85	42–49	8, 0, 0, 8	0.063
Dros	*Bovmt*	met 1	50–57	14–21	8, 0, 0, 8	0.091
		gly 1	28–42	47–60	13, 1, 1, 15	0.0055
		arg 1	75–82	3–10	8, 0, 0, 8	0.055
	Yeast	ala 1	78–90	31–44	12, 1, 1, 14	0.058
		gly 2	57–73	35–52	14, 3, 1, 18	0.081
		leu 1	110–119	2–12	10, 0, 1, 11	0.065
		phe 1	78–94	11–27	13, 4, 0, 17	0.047
		thr 1	26–43	48–65	15, 3, 0, 18	0.0013
		thr 1	91–105	61–75	12, 3, 0, 15	0.046

5S RNA	tRNA		Position on 5S RNA	tRNA	Matches, Mismatches, Gaps, Length	Expected value (E)
	E. coli	ala 1	78–97	11–30	15, 5, 0, 20	0.021
		trp 1	100–113	36–49	12, 2, 0, 14	0.020
		cys 1	68–77	15–25	10, 0, 1, 11	0.064
		ala 2	78–97	11–30	15, 5, 0, 20	0.017
	Euch	ala 1	27–41	48–63	13, 3, 0, 16	0.018
		tyr 1	79–89	19–28	10, 0, 1, 11	0.064
	Hvol	asp 1	59–69	64–74	10, 1, 0, 11	0.074
		ile 1	34–48	28–42	12, 3, 0, 15	0.076
	Dros	ala 1	75–95	7–28	17, 4, 1, 25	0.011
		gly 1	28–43	53–68	13, 3, 0, 16	0.024
		met 1	31–47	23–39	13, 4, 0, 17	0.049
		met 1	37–51	45–59	12, 3, 0, 15	0.057
		met 2	37–51	45–59	12, 3, 0, 15	0.060
Hvol	*Bovmt*	val 1	105–112	23–30	8, 0, 0, 8	0.066
		leu 1	1–16	2–16	14, 1, 1, 16	0.0018
		met 1	35–42	29–36	8, 0, 0, 8	0.093
		asp 1	51–58	38–45	8, 0, 0, 8	0.067
		glu 1	17–28	48–59	10, 2, 0, 12	0.083
		pro 1	72–87	41–55	13, 2, 1, 16	0.030
	Yeast	asp 1	6–18	37–50	12, 1, 1, 14	0.046
		leu 1	8–19	9–19	11, 0, 1, 12	0.018
		thr 1	2–12	20–29	10, 0, 1, 11	0.053
		tyr 1	110–121	55–68	12, 0, 2, 14	0.029
		val 1	91–107	57–73	14, 3, 0, 17	0.0053
		val 1	69–79	4–16	11, 0, 2, 13	0.086
		arg 3	22–31	51–60	10, 0, 0, 10	0.0080
	E. coli	trp 1	99–122	48–72	20, 4, 1, 25	0.00039
		tyr 1	17–27	1–11	10, 1, 0, 11	0.073
		arg 1	13–26	41–56	14, 0, 2, 126	0.0031
		gln 1	9–23	10–23	13, 1, 1, 15	0.015
		gn 2	9–23	10–23	13, 1, 1, 15	0.015
		his 1	104–121	59–76	14, 4, 0, 18	0.015
		ser 2	8–19	9–19	11, 0, 1, 12	0.026
	Euch	phe 1	98–112	2–17	13, 2, 1, 16	0.055
		arg 1	42–51	36–46	10, 0, 1, 11	0.068
	Hvol	ser 2	7–26	18–36	16, 3, 1, 20	0.013
		tyr 1	57–66	9–19	10, 0, 1, 11	0.088
		val 1	72–84	8–20	12, 0, 2, 14	0.050
	Dros	leu 1	2–14	21–32	12, 0, 1, 13	0.0064
		leu 1	8–19	9–19	11, 0, 1, 12	0.023
		leu 1	68–76	38–46	9, 0, 0, 9	0.041
		met 1	66–76	21–31	10, 1, 0, 11	0.057
Anac	*Bovmt*	asp 1	51–62	10–22	11, 1, 1, 13	0.066
	Yeast	phe 1	47–57	11–21	10, 1, 0, 11	0.049
		val 1	48–23	10–25	14, 1, 2, 17	0.04
		val 3	16–26	63–72	10, 0, 1, 11	0.064
	E. coli	gly 1	72–84	22–35	12, 1, 1, 14	0.044
		thr 1	97–109	1–13	11, 2, 0, 13	0.053
		trp 1	91–103	27–39	12, 2, 0, 13	0.062
		arg 1	38–54	2–18	13, 4, 0, 17	0.056
		gln 1	15–27	15–27	11, 2, 0, 13	0.065
		gln 2	15–27	15–27	11, 2, 0, 13	0.063
		met 1	69–82	1–13	12, 1, 1, 14	0.040
		met in	83–97	15–29	12, 3, 0, 15	0.077
		phe 1	82–100	2–21	16, 3, 1, 20	0.0088
		leu 3	79–89	75–84	10, 0, 1, 11	0.080
	Euch	leu 1	1–11	41–50	10, 0, 1, 11	0.051

5S RNA	tRNA		Position on 5S RNA	tRNA	Matches, Mismatches, Gaps, Length	Expected value (E)
	Hvol	gln 1	109–119	45–54	10, 0, 1, 11	0.067
		lys 1	90–106	3–19	13, 4, 0, 17	0.057
		ser 1	61–75	70–84	12, 3, 0, 15	0.094
		ser 2	83–92	2–12	10, 0, 1, 11	0.088
		val 1	30–43	62–74	12, 1, 1, 14	0.046
	Dros	arg 1	10–24	15–19	12, 3, 0, 15	0.050
		leu 1	26–35	72–82	10, 0, 1, 11	0.074
		lys 1	82–100	2–21	16, 3, 1, 20	0.0086
		lys 2	82–100	2–21	16, 3, 1, 20	0.0070
		phe 1	47–57	11–21	10, 1, 0, 11	0.046
		phe 1	101–109	5–13	9, 0, 0, 9	0.029
		pro 1	74–88	38–52	13, 2, 0, 15	0.0042
		pro 1	8–20	10–20	11, 0, 2, 13	0.072
Tet	*Bovmt*	ile 1	10–26	39–56	14, 3, 1, 18	0.050
		thr 1	53–68	54–69	12, 4, 0, 16	0.092
	Yeast	phe 1	98–115	42–59	15, 2, 2, 19	0.082
		phe 1	78–94	11–27	13, 4, 0, 17	0.047
		val 1	104–115	50–60	11, 0, 1, 12	0.015
		val 2	99–114	43–59	15, 0, 3, 18	0.0061
		val 2	31–44	46–61	14, 0, 2, 16	0.0020
		val 3	102–115	46–59	13, 0, 2, 15	0.010
	E. coli	ala 1	78–97	11–30	15, 5, 0, 20	0.014
		cys 1	38–48	52–61	10, 0, 1, 11	0.057
		his 1	29–39	57–67	10, 1, 0, 11	0.049
		val 2	69–90	1–23	17, 5, 1, 23	0.027
		val 2	76–89	34–46	12, 1, 1, 14	0.042
		phe 1	100–112	36–48	11, 2, 0, 13	0.058
		ala 2	78–97	11–30	15, 5, 0, 20	0.013
		leu 2	80–90	4–14	10, 1, 0, 11	0.063
	Euch	arg 1	40–52	36–48	11, 2, 0, 13	0.057
		glu 1	37–47	53–63	10, 1, 0, 11	0.046
		val 1	38–48	54–63	10, 0, 1, 11	0.058
	Hvol	arg 1	38–48	54–65	11, 0, 1, 12	0.018
		ile 1	38–48	55–64	10, 0, 1, 11	0.066
		leu 1	38–48	66–75	10, 0, 1, 11	0.078
	Dros	asp 1	51–69	19–35	15, 2, 2, 19	0.052
		glu 3	99–115	42–58	13, 4, 0, 17	0.040
		lys 2	100–113	57–69	12, 1, 1, 14	0.040
		thr 1	51–61	19–30	11, 0, 1, 12	0.018
		val 2	38–48	54–63	10, 0, 1, 11	0.055

in Methods. Forty-five matches ($E<0.10$) were found and these have been plotted in Fig. 1a. Of interest is the large number of matches corresponding to the TΨC arm of the tRNAs and the C1 stem and H1 loop [23] of the 5S rRNA. The spreading of the cluster of matches at the TΨC arm is a result of differences in the length of variable loops among tRNAs [21].

The fact that so many matches were associated with the highly conserved TΨC arm of the tRNAs may be interpreted to indicate that this large number reflects only the conserved nature of tRNA rather than a common origin of the two classes of RNAs. To investigate this question, the *Prochloron* 5S rRNA sequence was randomized, maintaining the same base frequencies, and compared to the same set of tRNAs. The resulting matches are shown plotted in Fig. 1b. Two important points are apparent. Fewer matching regions are found in comparisons with the randomized molecule than with the native RNA. In fact, only 13 matches (the number expected by chance) were found. This finding underscores the

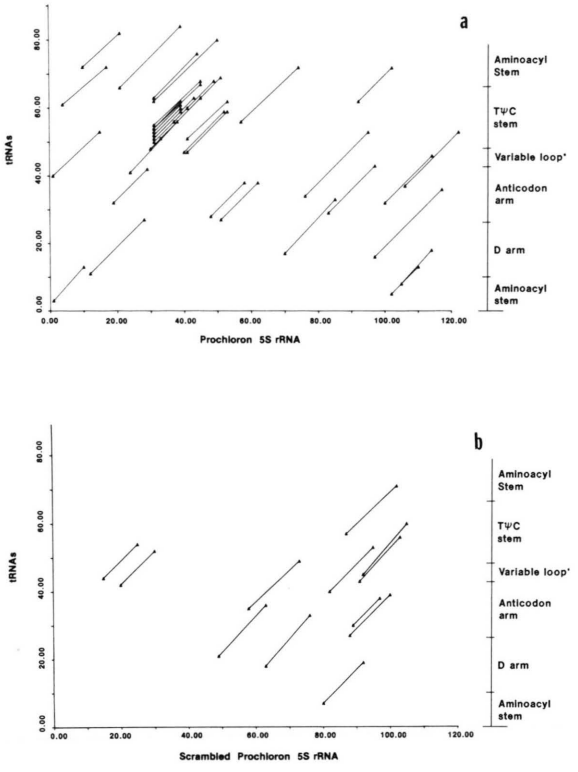

Fig. 1. Two plots showing the positions of matches between the a) *Prochloron* 5S rRNA and b) scrambled *Prochloron* 5S rRNA sequences (abscissa) and 123 tRNA sequences taken from *E. coli, H. volcanii, Euglena* chloroplast, bovine mitochondria, yeast and *Drosophilla* (ordinate). Small departures from the 45° line reflect different numbers of gaps or insertions in the two sequences. The spreading of matches corresponding to the TΨC arm is a result of the different lengths of the variable loop.

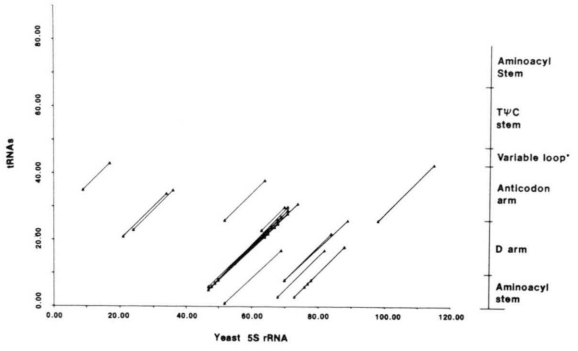

Fig. 2. A plot showing the positions of matches between the yeast 5S rRNA sequence (abscissa) and 123 tRNA sequences taken from *E. coli, H. volcanii, Euglena* chloroplast, bovine mitochondria, yeast and *Drosophilla* (ordinate). Small departures from the 45° line reflect different numbers of gaps or insertions in the two sequences.

fact that the sequence of the native *Prochloron* shows many more matches with tRNAs than can be attributed to chance. Secondly, using the randomized sequence the matches which are found do not fall into any obvious pattern nor are they clustered in any way. Thus, the clustering shown in Fig. 1a is not an inevitable result of the conserved nature of tRNA.

A list of the 44 matching regions found in comparisons between the 5S rRNA of yeast and all tRNAs is shown in Table I. These matches are plotted in Fig. 2. Approximately two-thirds of the matches are localized on the DHU arm of the tRNAs and extend from the H1 loop to the I1 loop of the 5S rRNA. The matches form a narrow band on the plot since the DHU arm is 5′ to the variable loop.

While searches between tRNAs and the 5S rRNA sequences of *Prochloron* and yeast yield matches with a very conservative distribution, this is not always the case. A composite plot of eight 5S rRNA sequences compared to all 123 tRNAs is shown in Fig. 3. The 5S sequences were taken from *E. coli, Halobacterium volcanii, Anacystis nidulans, Prochloron, Euglena gracilis* chloroplast, *Tetrahymena,* yeast and *Drosophilla.* The 239 matches illustrated on the plot are listed in Table I. A large number of the matches clearly correspond to the conserved DHU and TΨC arms of tRNAs. The sequences involved in these matches seem to be highly conserved not only in tRNA but also in 5S rRNA. In addition to the conserved regions, any portion of the tRNA or rRNA may be involved in a match. However, specific sites on the tRNAs seem to be prohibited from matching with specific sites on the 5S rRNA. This is manifested as a "zone of prohibition" and is a striking feature of the plot shown in Fig. 3. Approximately 24% of the plot shown in Fig. 4 is covered by the zone of prohibition, yet only *ca.* 3.8% of the matches fall within this zone. This zone of prohibition is *ca.* 20 bases wide and extends diagonally for the length of the tRNAs and from approximately base 10 to base 90 on the 5S rRNAs. Interestingly, while matches involving tRNAs from bovine mitochondria make up less than 13% of the total number of matches shown in Fig. 4, they represent more than 50% of the matches which are found within the zone of prohibition. If one considers only nonmitochondrial tRNA,

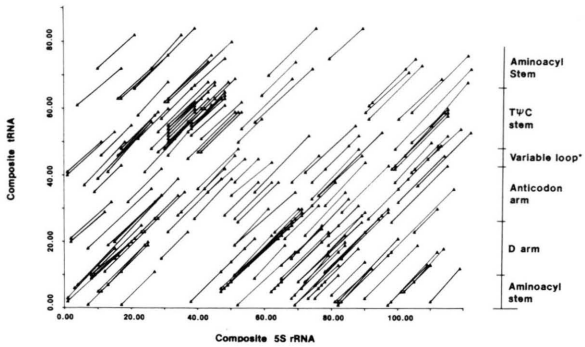

Fig. 3. A plot showing the positions of matches between the 5S rRNA sequences of *E. coli, H. volcanii, Euglena* chloroplast, *Anacystis, Prochloron, Tetrahymena,* yeast (abscissa) and 123 tRNA sequences taken from *E. coli, H. volcanii, Euglena* chloroplast, bovine mitochondria, yeast and *Drosophilla* (ordinate). Small departures from the 45° line reflect different numbers of gaps or insertions in the two sequences. The "zone of prohibition" is clearly visible extending diagonally for the length of the tRNAs and from approximately base 10 to base 90 on the 5S rRNAs.

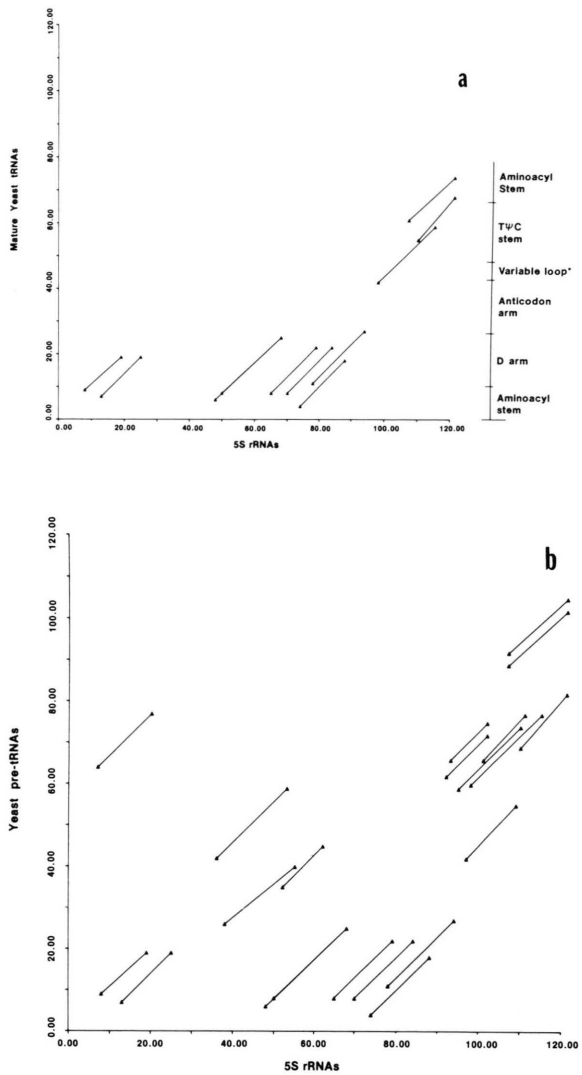

Fig. 4. Plots showing the number and positions of matches between the 5S rRNA sequences of *E. coli, H. volcanii, Tetrahymena* and yeast (abscissa) and a) ten mature yeast tRNAs (ordinate). The corresponding precursor (intron containing) yeast tRNAs are plotted on the ordinate of b. Twice as many matches are found in searches with precursor tRNA sequences as compared to mature sequences.

the occurrence of matches in the zone of prohibition is reduced to 1.9%.

Convergence as an explanation for the matches?

When two structures share similar features, this similarity is generally the result of one of three phenomena: chance; convergence (usually through functional requirements); or common origin. In the present case, chance may be eliminated as a possibility as the number of matches between 5S rRNA and tRNA exceeds levels attributable to chance alone (> 99% confidence). When the similarities are found at the molecular level, as in this case, another explanation is possible. However, transposition by genetic processes and interspecific exchange seem to be an improbable cause for the matches. Many tRNAs which are found to participate in interspecies matches do not exhibit corresponding matches with their own 5S rRNAs. The *E. coli* $tRNA^{Val}$, for example, shows a match with yeast 5S rRNA but not with *E. coli* 5S rRNA. When corresponding RNAs are found to take part in both inter- and intraspecific matches, the former is as likely as not to be of a higher quality than the latter. Interspecific matches that are not mirrored by intraspecific matches among corresponding RNAs suggest common origins, since the physically separate existence of the two molecules exhibiting the match would preclude common functional interactions from shaping the sequences during their evolution. The possibility of forming and maintaining interspecies matches of a higher quality than those between the putative parental species seems remote. Such a series of events might entail transfection [25, 26] of tRNA genes which mimic, by

chance, 5S rRNA genes in the new host. Alternatively, transfection of a gene which exhibited a match in the donor species may occur. This element must then be maintained in the new host and subsequently lost by the donor species. The widespread occurrence of interspecific matches make recent recombination events an unlikely cause. Additionally, it would require that the matches between species would be subject to a more rigorous selection pressure for their maintenance than the matches within the species of origin. Such an invocation of selection would be outside of any meaningful biological context.

One method which may be used to determine if matches reflect convergence between 5S rRNA and tRNA involves comparing the frequencies of intra- and interspecific matches. If convergence were the cause of the similarities between 5S rRNA and tRNAs, one would expect to find more matches in intraspecific searches than interspecific comparisons. Table II shows the frequences of matching regions found between 5S rRNA and tRNA sequences of *E. coli, H. volcanii, Euglena* chloroplast, yeast and *Drosophilla.* The table indicates that there is no general trend for rRNAs from a given organism having a higher frequency of matches with tRNAs from the same source. The mean frequency for intraspecific matches is not significantly higher than that for interspecific matches. In fact, the highest frequency of matches (61.3%) occurs between yeast 5S rRNA sequences and tRNAs from *E. coli.* Additionally, the frequency of matching between 5S rRNA and tRNAs from *Euglena* chloroplast is among the lowest in the table. Since interspecific matches occur with the same frequency as intraspecific matches, convergent evolution may be eliminated as a possible explanation for the matches between the tRNAs and 5S rRNAs. No known evolutionary mechanism can effect a convergence of molecules in different cellular environments with different functions.

The most plausible interpretation of the matches is that the matches are true homologies, reflecting common origins. This is the explanation that is generally accepted when matches are found among molecules that are known to be similar. The sequences are probably conserved because they became locked into essential functions. The functional constraints on the primary structure have resulted in selection against the maintenance for mutational changes.

Matches with the excised portions of yeast tRNAs

Many tDNA transcripts require significant processing before yielding a mature tRNA. This processing often involves the excision of an intervening sequence. The best studied example of processing of this sort involves the tRNAs of yeast [14, 15]. A mutant yeast strain allows the unprocessed tRNA precursors to accumulate [27], and thus to be identified and sequenced [28].

The sequences of yeast precursor-tRNAs reported by Ogden *et al.* [14] and Lee and Knapp [15] provide an opportunity to test the common origin hypothesis. The intervening sequences (introns) range in length from 14 to 60 bases. If tRNA and 5S rRNA share a common ancestor, traces of this common origin should be found not only in the mature tRNAs but also in those portions of the precursor-tRNAs which are later excised. In comparisons between the mature yeast tRNA sequences and five 5S rRNAs, 12 matches were found (Fig. 4a). However, when the

Table II. The frequency of inter- and intraspecific matches between tRNAs and 5S rRNAs from *Euglena* chloroplast, *E. coli, H. volcanii* and yeast. The first number gives the frequency, as a percentage, of matches between the indicated RNAs, while the numbers enclosed in brackets show the actual number of matches. The mean frequency of intraspecific matches (26.9 ± 20.3) is not significantly different from that of interspecific matches (24.3 ± 11) at the 95% confidence level.

rRNA/tRNA	*Euglena* chlor.	*E. coli*	*H. volcanii*	Yeast
Euglena chlor.	15.4 [2]	9.7 [3]	7.7 [1]	20.0 [5]
E. coli	30.8 [4]	45.2 [14]	7.7 [1]	12.0 [3]
H. volcanii	15.4 [2]	22.6 [7]	23.1 [3]	28.0 [7]
Yeast	58.3 [7]	61.3 [19]	23.1 [3]	24.0 [6]

precursors of these tRNAs were compared with the 5S rRNA sequences, 23 matches were found (Fig. 4b). Table III lists matches observed with both the precursor and mature yeast sequences. It is clear from these comparisons that matches occur with the introns of the pre-tRNAs with at least the same frequency as with the mature tRNAs. These results support the common origin hypothesis.

The nature of the common ancestor of tRNA and 5S rRNA

Mullins *et al.* [19] suggested that the precursor to 5S rRNA was the result of the tandem duplication. They further suggested that some tRNAs may have been derived from this same precursor molecule. Others [29, 30] have suggested that both tRNA and 5S rRNA arose from tRNAs lacking the DHU arm. An example of such a molecule from contemporary organisms is $tRNA^{Ser}$ from mammalian mitochondria. Wolters and Erdmann [30] suggest that these tRNAs represent degenerate forms in which back mutations have created molecules of suboptimal function. It is unclear how a suboptimal adaptor molecule could be selected for. Additionally, it seems unlikely that the degenerate tRNAs represent the ancestral molecule as this would require that all tRNAs from other sources have converged to the common *ca.* 90-base long $tRNA^{Ser}$. It seems more plausible to consider the degenerate (*ca.* 60-base long) tRNAs derived, and characteristic of mitochondrial tRNAs, rather than analogs of ancient molecules. In addition, the large number of matches corresponding to the DHU arms of tRNAs shown in Fig. 2 and 3 would seem to argue against the 5S rRNA precursor having been a DHU arm-less tRNA.

Table III. A comparison of the matches found between 5S rRNAs from yeast, *E. coli*, *H. volcanii* and *Tetrahymena* with mature yeast tRNAs and their unprocessed, intron-containing, precursors.

Mature tRNA		Unprocessed tRNA	
5S RNA	tRNA	5S RNA	tRNA
Yeast	ile 1	Yeast	ile 1
	leu 3		leu 2
	lys 1		leu 3
	pro 1		lys 1
E. coli	ile 1		lys 1
	lys 1		pro 1
Hvol.	leu 1		pro 4
	pro 1	*E. coli*	ile 1
	tyr 1		lys 1
	tyr 2		pro 1
Tet.	phe 1		pro 4
	phe 1	*Hvol.*	leu 3
			pro 1
			pro 4
			trp 1
			tyr 1
		Tet.	ile 1
			leu 3
			lys 1
			phe 1
			phe 1
			phe 2
			trp 1

We propose that tRNAs arose from a molecule that was also the precursor to 5S rRNA or an ancient 5S rRNA itself. Transfer RNAs (and 5S rRNA?) were copied (transcribed?) from this molecule. Transcription could begin at many different sites, but was prohibited from beginning at the site corresponding to bases 10–30 on present-day 5S rRNA. This accounts for the "zone of prohibition" evident in Fig. 3. The explanation for this prohibition is unknown. Perhaps it corresponds to a binding site with another nucleic acid or a protein. This may have effectively protected the site from being a site of transcription initiations. However, transcription begun before the putative binding site might have caused the binding protein or nucleic acid to become dislodged, allowing transcription to continue.

The molecule we propose to have given rise to these two classes of RNAs may itself have been derived from the tandem duplication of a shorter molecule [19]. Alternatively, the precursor molecule may have been a circularized RNA composed of approximately 120 bases. In either case, some tRNAs should have been formed from transcripts that bridge the region corresponding to the 3′ and 5′ ends of contemporary 5S rRNAs. This should be reflected by matches between tRNAs and 5S rRNAs extending from before base 120 to after base 1 on a circularized 5S sequence. To investigate this possibility, the 5S rRNA sequences used in the comparisons shown in Fig. 3 were cut after base 46 and were spliced front to back. The resulting sequences were compared with the same tRNAs as described for the plot shown in Fig. 3. If no matches were found extending across the spliced region, fewer matches should be found than the 239 listed in Table I. Fig. 5 shows a plot of the matches found in comparisons of "spliced" 5S rRNAs and tRNAs. Although many matches shown

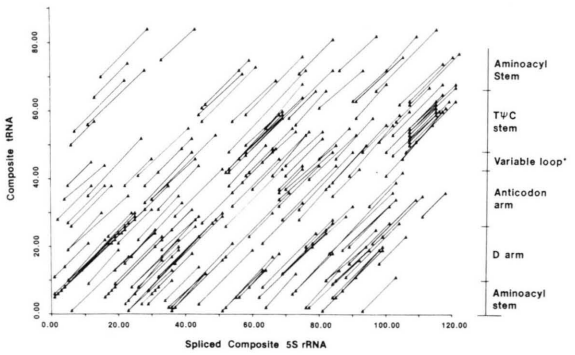

Fig. 5. The 5S rRNA sequences described in Fig. 3 were cut after base 46 and rejoined so that the new sequence consisted of based 47−120+1−46. These spliced 5S sequences were compared with the same tRNAs referred to in Fig. 3. The positions of the resulting matches are shown here. The 5S sequences are plotted on the abscissa and the tRNAs are plotted on the ordinate.

in Fig. 4 were lost as a result of cutting the 5S rRNA sequence, a large number of new matches were created by joining the 3′ and 5′ ends of the sequence. In fact 246 matches are shown in Fig. 5 indicating that more matches were created than lost by the splicing. This result supports the hypothesis that the precursor to the 5S rRNA and tRNA was a tandemly duplicated, or circular molecule.

5S rRNA and the universal translator

A general trend pervading biological evolution is that from generalized to specialized [31]. We believe that we can find traces of this evolutionary trend in the translational apparatus of contemporary organisms. Before the appearance of highly specific tRNAs for translating the information stored in nucleotide sequences to protein, there may have existed an RNA capable of functioning as a general translation molecule. This "universal translator" was able to read all the codons and attach all the amino acids to form a peptide chain. We propose that an ancestor of 5S rRNA may have functioned in this capacity [32]. Additionally, this proto-5S rRNA may have served as a template for proto-tRNAs. Alternate splicing could yield a collection of proto-tRNAs with unique anticodons, and result in extensive sequence homology in other portions of the molecules, as seen in contemporary tRNAs [20]. Further discussion of the "universal translator" hypothesis will be presented in a separate paper.

[1] A. L. Weber and J. C. Lacey, Jr., J. Mol. Evol. **11,** 199−210 (1978).
[2] J. C. Lacey, Jr. and D. W. Mullins, Jr. Orig. Life **13,** 3−42 (1983).
[3] J. C. Lacey, Jr., D. P. Stephens, and S. W. Fox, Biosystems **11,** 9−17 (1979).
[4] M. A. Khaled, D. W. Mullins, Jr., and J. C. Lacey, Jr., J. Mol. Evol. **20,** 66−70 (1978).
[5] M. A. Khaled, D. W. Mullins, Jr., and J. C. Lacey, Jr., Orig. Life **13,** 87−96 (1983).
[6] J. C. Lacey, Jr. and D. W. Mullins, Jr., J. Theo. Biol. **115,** 595−601 (1985).
[7] J. C. Lacey, Jr., L. M. Hall, and D. W. Mullins, Jr., Orig. Life **16,** 69−79 (1985).
[8] J. C. Lacey, Jr., A. L. Weber, and W. E. White, Jr., Orig. Life **6,** 273−284 (1975).
[9] J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **75,** 4334−4338 (1978).
[10] U. Lehmann and H. Kuhn, Orig. Life **14,** 497−504 (1984).
[11] D. P. Bloch, B. McArthur, R. B. Widdowson, D. Spector, R. C. Guimaraes, and J. Smith, J. Mol. Evol. **19,** 420−428 (1983).
[12] D. P. Bloch, B. McArthur, and S. Mirrop, Biosystems **17,** 209−225 (1985).
[13] S. Altman and J. D. Smith, Nature New Biol. **233,** 35−37 (1971).
[14] R. C. Ogden, M.-C. Lee, and G. Knapp, Nuc. Acids Res. **12,** 9367−9382 (1984).
[15] M.-C. Lee and G. Knapp, J. Biol. Chem. **260,** 3108−3115 (1985).
[16] A. S. Spirin and L. P. Gavrilova, The Ribosome, Springer Verlag, New York 1969.
[17] S. Sakonju, D. F. Bogenhagen, and D. D. Brown, Cell **19,** 13−25 (1980).
[18] D. F. Bogenhagen, S. Sakonju, and D. D. Brown, Cell **19,** 27−35 (1980).
[19] D. W. Mullins, J. C. Lacey, and R. A. Hearn, Nature New Biol. **424,** 80−82 (1973).
[20] M. P. Staves, D. P. Bloch, and J. C. Lacey, Z. Naturforsch. **42c,** 129−133 (1987).
[21] M. Sprinzl, J. Moll, F. Meissner, and T. Hartmann, Nucl. Acids Res. **13,** r1−r49 (1985).
[22] M. Sprinzl, T. Vorderwülbecke, and T. Hartmann, Nucl. Acids Res. **13,** r51−r104 (1985).
[23] V. A. Erdmann and J. Wolters, Nucl. Acids Res. **14,** r1−r59 (1986).
[24] W. B. Goad and M. I. Kanehisa, Nucleic Acids Res. **10,** 247−263 (1982).
[25] C. D. Wilde, C. E. Crowther, T. Cripe, M. G.-S. Lee, and N. Cowan, J. Nature **297,** 83−84 (1982).
[26] L. B. Bernstein, S. M. Mount, and A. M. Weiner, Cell **32,** 461−472 (1983).
[27] A. K. Hopper, F. Banks, and V. Evangelidis, Cell **14,** 211−219 (1978).
[28] G. Kanpp, J. S. Beckmann, P. F. Johnson, S. A. Fuhrman, and J. Abelson, Cell **14,** 221−236 (1978).
[29] M. Ujhelyi, Z. Naturforsch. **38c,** 501−504 (1983).
[30] J. Wolters and V. A. Erdmann, Endocyt. C. Res. **3,** 157−166 (1986).
[31] D. P. Bloch, Origins of Life, in the press.
[32] J. C. Lacey, Jr. and M. P. Staves, in preparation.