Consensus structure and evolution of 5S rRNA*

Hans Küntzel, Birgit Piechulla and Ulrich Hahn+

Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Hermann-Rein-Str. 3, D-3400 Göttingen, FRG

Received 23 November 1982; Revised and Accepted 5 January 1983

ABSTRACT

A consensus structure model of 5S rRNA presenting all conserved nucleotides in fixed positions has been deduced from the primary and secondary structure of 71 eubacterial, archaebacterial, eukaryotic cytosolic and organellar molecules. Phylogenetically related groups of molecules are characterized by nucleotide deletions in helices III, IV and V, and by potential base pair interactions in helix IV. The group-specific deletions are correlated with the early branching pattern of a dendrogram calculated from nucleotide substitution data: the first major division separates the group of eubacterial and organellar molecules from a second group containing the common ancestors of archaebacterial and eukaryotic/cytosolic molecules. The earliest diverging branch of the eubacterial/organellar group includes molecules from <u>Thermus thermophilus</u>, <u>T. aquaticus</u>, <u>Rhodospirillum rubrum</u>, <u>Paracoccus denitrificans</u> and wheat mitochondria.

INTRODUCTION

The 5S rRNA component of the large ribosomal subunit (1) is a suitable marker molecule to study pre-cambrian phylogeny: a comparison of 5S rRNA sequences from eukaryotes, eubacteria, archaebacteria and organelles (2-14) has revealed a considerable conservation of primary (15,16) and secondary (17-20) structure. Here we present a general consensus structure for all types of 5S rRNAs. Our model is in agreement with previously published secondary structure models for eubacterial (19,20), eukaryotic cytosolic (18,19) and archaebacterial (7) molecules, and in addition presents all conserved nucleotides in fixed positions. The observed group-specific deletions and base pair interactions support the phylogenetic grouping of 5S rRNA molecules based on nucleotide substitution data.

METHODS

55 rRNA was extracted from the snail <u>Arion rufus</u>, purified by preparative gel electrophoresis and sequenced by chemical cleavage methods as described (21). Methods used for computer-aided construction of dendrograms have been reported in detail (16,25).

RESULTS AND DISCUSSION

Fig. 1 shows an alignment of selected 5S rRNA sequences representative for the groups of metazoa [sequence 1], protozoa [19,20], plants [32], archaebacteria [41-46], eubacteria [47,50,56,60,61,67-69], chloroplasts [63] and mitochondria [71].

The alignment is similar, but not identical, to previously published schemes (16, 19) due to the inclusion of more sequence data.

The consensus structure of Fig. 2A is based on the alignment of 71 5S rRNA sequences (2-14) and contains 70 positions where either A,U,G,C or a

	1	2	· 3	4	5	6	
	-123456789012-3	45678-901234	5678901-234	56789012345	678-9012	34567890	
1	GUCUACGGC-C	AUACC-ACCCUG	AACGCGC-CCC	AUCUCGU-CUG	AUC-UCGG	AAGCUAAG 1	
19	GCUGUCGGC-C	AUACU-AAGGUG	AAAACAC-CGC	AUCCCAU-UCG	AAC-UCCG		
20	GOUGACGGC	AUACC_GUGUCG	AAUGCAC-CGC	AUCUCUU_CUG	ACC-UCCG		
27	GGUUGCGGC-C	AUAUC_UACCAG	AAAGCAC-CGI		AUCAACUG	ILAGITUAAG 27	
22	GAUGCGAU-C	AUACC_ACCACU	AAAGCAC-CGC	AUCCCAU_CAG	AAC-WCCG	AAGIIIIAAG 32	
11		AUACC COUCCO	ANNIACUC CCC	IIIACCCAUCCCC	AAC ACCC	AACAUAAC 21	
1.2		ACACC ACCUCU	GUURCUC-CCC	UNICCOAUCCCG	AAC ACCC	ALCINIALC NO.	
1.1.		AUACO ACCACC	GULACAC-CU	AUCCCADUCCG	AAC NOCA	AAGUUAAG 4)	
44	GGCARCGGU-C	AUAGC-AGCAGG	GAAACAC-CA	AUCCCAUUCCG	AAC-OUGA		
22	GCCCACCCGGU-C	AUAGU-GAGCGG	GUAACAC-CCC	GACUCAUUUCG	AAC-CCGG	AAGUUAAG 45	
40	UAGGUUUGGCGGU-C	AUAGC-GAUGGG	GUAUCAC-CUC	GUCUCGUUUCG	AUC-CCAG	AAGUUAAG 46	
47	UGCCUGGCGGC-C	GUAGC-GCGGUG	GUCCCAC-CUC	ACCCCAUGCCG	AAC-UCAG	AAGUGAAA 47	
50	UGUUCUGUGACGA	GUAGUGGCAUUG	GAA-CAC-CUC	AUCCCAUCCCG	AAC-UCAG	AGGUGAAA 50	
56	UUGGUGGU	AUAGC-AUAGAG	GUCACAC-CUC	UUCCCAUGCCG	AAC-ACAG	AAGUUAAG 56	
58	UCCAGUGUC-U	AUGAC-UUAGAG	GUAACAC-UCO	UUCCCAUUCCG	AAC-AGGC	AGGUUAAG 58	
60	GUUACGGCGGC-U	AUAGC-GUGGGG	GAAACGC-CCC	GCCGUAUAUCG	AAC-CCGG	AAGCUAAG 60	
61	UCCUGGUGUC-U	AUGGC-GGUAUG	GAACCACUCUC	ACCCCAUCCCG	AAC-UCAG	UUGUGAAA 61	
63	-UAUUCUGGUGCUCC	UAGGC-GUAGAG	GAACCAAACCA	AUC-CAUCCCG	AAC-UUGG	UGGUUAAA 63	
67	GUCUGGUGGC-C	AAAGC-ACGAGC	AAAACAC-CCC	AUCCCAUCCCG	AAC-UCGG	CCGUUAAG 67	
68	-UGGCCUGGUGGU-C	AUUGC-GGGCUC	GAAACAC-CCC	AUCCCAUCCCG	AAC-UCGG	CCGUGAAA 68	
69	AA UCCCCCGUGCC-C	AUAGC-GGCGUG	GAACCAC-CCC	UUCCCAUUCCG	AAC-ACGG	AAGUGAAA 60	
71	-AAACCGGGCACUAC	GGUGA-GACGUG	AAAACAC-CCC	AUCCCAUUCCG	ACC-UCGAUAU	AUAUGUGGAA 71	
•	-123456789012-3	45678-901234	5678901-23	56789012345	678-9012	34 567890	
	1	2	3	<u> </u>	5	6	
	•	-	,	-	,	•	
				1	1	1	
	7	. 8	9	1 0	1	1 2	
	7 12345678901234	8 67890-1234	9 5678901234	1 0 567890123456	1 1 -7890123	1 2 4567890123456	
1	7 123456789012345 CAGGGUCGGGCCUGG	8 67890-1234 UUAGU-ACUU	9 5678901234 GGAUGGGAGAG	1 0 567890123456 CGCCUGGGAAU	1 1 <u>-7890123</u> -ACCGGGU	1 2 <u>4 567890123456</u> -GCUGUAGGCUUU	1
1 19	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGGG	8 67890-1234 UUAGU-ACUU	9 5678901234 GGAUGGGAGAG AGGUGGGGGAG	1 0 567890123456 CCGCCUGGGAAU CCGCUUGGGAAG	1 1 <u>-7890123</u> -ACCGGGU -UCCCAGU	1 2 4 567890123456 -GCUGUAGGCUUU -GUCGACAGCCU-	1 19
1 19 20	7 12345678901234 CAGGGUCGGGCCUGG CGCCUUAAGGCUGGG CGGCACAGGCCCGG	8 <u>67 890- 1234</u> 100 AGU-ACUU 100 AGU-ACUA 100 AGU-ACUG	9 5678901234 GGAUGGGAGAG AGGUGGGGGGAG GGGUGGGGGGAG	1 0 567890123456 CCGCCUGGGAAU CCGCUUGGGAAG CCGCCCGGGAAG	1 1 -7890123 -ACCGGGU -UCCCAGU -UCCUUAGGGU	1 2 4 567890123456 -GCUGUAGGCUUU -GUCGACAGCCU- -GCUGUCAGCU	1 19 20
1 19 20 27	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCCUGG CGCCACAGGGCCCGG CUGGUAAGAGCCUGA	8 67890-1234 900-AGU-ACUU 900-AGU-ACUG 400-AGU-ACUG 640-AGU-AGUG	9 5678901234 GGAUGGGAGAG AGGUGGGGGGAG GGGUGGGGGGAG UAGUGGGUGAG	1 0 567890123456 CCGCCUGGGAAU CCGCCUGGGAAG CCGCCCGGGAAG CCAUACGCGAAA	1 -7890123 -ACCGGGU -UCCCAGU -UCCUUAGGGU -CUCAGGU	1 2 4567890123456 -GCUGUAGGCUUU -GUCGACAGCCU- -GCUGUCAGCU -GCUGCAAUCU	1 19 20 27
1 19 20 27 32	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGCUGGC CGGCACAGGGCCGG CUGGUAAGAGCCUGA	8 667890-1234 100AGU-ACUU 100AGU-ACUA 140AGU-ACUG 140AGU-AGUG 160AGU-ACUA	9 5678901234 GGAUGGGAGA AGGUGGGGGA GGGUGGGGGGA UAGUGGGUGA GGAUGGGUGA	1 0 567890123456 CCGCCUGGGAAU CCGCCCGGGAAG CCGCCCGGGAAG CCAUACGCGAAA	1 -7890123 -ACCGGGU -UCCCAGU -UCCUUAGGGU -CUCAGGU -UUCCCGU	1 2 4 567890123456 -GCUGUAGGCUUU -GUGGACAGCCU- -GCUGCAAUCU -GCUGCAAUCC -GUUGCAAUCC-	1 19 20 27 32
1 19 20 27 32 41	7 123456789012345 CAGGGUCGGGCCUGG CGCCACAGGCCCGG CUGGUAAGAGCCUGA CGUGCUGGGCGCAG CCCGCCUGGCGUGCG	8 67890-1234 6UUAGU-ACUU 6UUAGU-ACUA 6AUAGU-ACUG 6GUAGU-ACUA 6GUAGU-ACUA	9 5678901234 GGAUGGGGAGA AGGUGGGGGAG GGGUGGGGGGAG GGAUGGGGGGAG GAGUGGGGGGAG	1 0 567890123456 CCGCCUGGGAAU CCGCUUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG	1 -7890123 -ACCGGGU -UCCCAGU -UCCUUAGGGU -UUCAGGU -UUCCGU -UUCCGU	1 2 4567890123456 -GCUGUAGGCUUU -GUCGACAGCU -GCUGCAAUCU -GUUGCAAUCU -GUUGCAUUCC- CGCCGCCUACU	1 19 20 27 32
1 19 27 32 41 43	7 12345678901234 CAGGGUCGGGCCUGG CGCCUUAAGGCUGGG CGGCACAGGGCCGGC CUGGUAAGACCCUGA CGUGCUUGGGCGAGA CCCGCCUGGGUUCCG ACACCUCGCACCUGGAU	8 <u>67890-1234</u> <u>800AGU-ACUU</u> <u>800AGU-ACUG</u> <u>800AGU-ACUG</u> <u>800AGU-ACUG</u> <u>800AGU-ACUG</u> <u>80CGU-ACUG</u>	9 5678901234 AGGUGGGGGAA GGGUGGGGGAA UAGUGGGGGAA GAGUGGGGGAA AGGUACGCGAA	1 0 567890123456 5CGCCUGGGAAU 5CGCCUGGGAAG 5CAUACGCGGAAA 5CUCCUGGGAAA 5CCUCUGGGAAA	1 1 1 -7890123 -WCCCAGU -UCCCAGU -UCCCGGU -UCCCGGU -UCCCGGU	1 2 4567890123456 -GCUGUAGGCUUU -GCUGCAAGCU -GCUGCAAUCU -GUUGCAUUCC- CGCUGCUACU CGCUGCUAUU-CC-	1 19 20 27 32 41
1 190 272 4 3 4 5 4 5 4 5 4 5 4 5 4 5 4 5 4 5 4 5	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGGCACAGGGCCUGG CUGGUAGAGGCCUGG CGUGCUUGGGCGAGA CCCGCCUGCGUGCGUCCC ACACCUCACGUGGAU CCUGCU-CCUUGGAU	8 <u>167890-1234</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u> <u>100-AGU-ACUA</u>	9 5678901234 GGAUGGGAGAA AGGUGGGGGAA UAGUGGGGGAA GGAUGGGUGA GAGUGCGCGAA AGGUACGCGAA	1 0 557890123456 CCGCCUGGGAAU CCGCCUGGGAAG CCUCCGGGAAG CCUCCUGGGAAA CCUCCUGGGAAA SUCCUCGGGAAA	1 1 -7890123 -ACCGGGU -UCCULAGGGU -UCCCAGU -UCCCGGU -UCCGGUU -UCCGGUU -UCCUCCU	1 2 4567890123456 -GCUGUAGGCUUU -GUGGACAGCCU- -GCUGCAAUCU- -GUUGCAAUCC- CGCUGCUAUUCCC- CGCCGCCUACU- CGCUGCUAUUGUU UGCUGUUAUGUU	190272 1927231
1907213444	7 123456789012345 CAGGGUCGGGCCUGG CGCCLUAAGGCUGGG CGGCACAGGGCCCG CUGGUAAGAGCCUGA CGUGCUUGGCGGACA CCUGCUUGGCGUCGC ACACCUCACGUGGAC CCUGCU-GCGUAUUC CC-GCUCACGUGGAC	8 <u>100</u> AGU-ACUU 100 AGU-ACUU 100 AGU-ACUU 100 AGU-ACUG 100 AGU-ACUG 100 AGU-ACUG 100 AGU-ACUG 100 AGU-ACUG 100 AGU-ACUG 100 CGU	9 5678901234 GGAUGGGAGAA AGGUGGGGGAA UAGUGGGGGAA GAGUGGGGGAA GAGUGCGCGAA AGGUACGCGAA UAUGCCGCGAA	1 0 557890123456 CCGCCUGGGAAU CCGCCUGGGAAG CCGCCCGGGAAG CCUCCGGGAAA CCUCCUGGGAAA SIGCUCCGGAAA SIGCUCCGGAAA	1 -7890123 -ACCGGGU -UCCCAGU -UCCUUAGGU -UCCAGGU -UCCCGGU -UCCCGGU -UCAUCCU -CGCAQUA	1 2 4567890123456 -GCUGUAGGCUUU -GUGGACAGCCU- -GCUGCAGCU- -GCUGCAGUU- -GUUGCAUUCC- CGCCGCCUACU- CGCUGCUACUAUU UGCUGUUACCACU	192721 192734344
1907213456	7 123456789012345 CAGGGUCGGGCCUGG CGCCULAGGCGGCCUGG CUGGUAGAGCCUGA CCGGCUGGGUGGA CCCGCCUGCGUUCCG ACACCUCACGUGGAU CCCGCCUGCGUUCCG CCGCCUGCGUUCGG CCGCCUGCGUUAGG	8 <u>67890_1234</u> <u>800_AGUACUA</u> <u>800_AGUACUA</u> <u>800_AGUACUA</u> <u>600_AGUACUA</u> <u>600_AGUACUA</u> <u>600_AGUACUA</u> <u>600_AGUACUA</u> <u>600_AGUACUA</u> <u>600_AGUACUA</u> <u>600_AGUACUA</u>	9 5678901234 GGAUGGGAGAA GGGUGGGGGAA UAGUGGGUGA GGAUGGGUGA GAGUGCGCGAA UAUGCCGCGAA GAUACCGUGAA	1 0 557890123456 CCGCCUGGGAAU CCGCCUGGGAAG CCGCCCGGGAAG CCCUCGGGAAG CCUCCUGGGAAG SCCUCCUGGGAAA SCCUCCGGAAG SCAUACGGAAG SCAUACGGAAG	1 -7890123 -ACCGGGU -UCCUUAGGU -UCCUUAGGU -UCCAGGU -UUCCCGU -UCCCCGU -UCCCCGU -UCCACUA -CCCACUA	1 250223456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAAUCU- -GCUGCAUUCC- -GUUGCAUUCCC- CGCCGCCUACU- CGCUGCUAUUGUU UGCUGUACACU AGCUGGCAUGGGU	1907213454 1907213454
19072134444 19072134567	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGCCUUAAGGCCUGG CGUGCUUGGGCGAGA CCUGCUGGGCGAGA CCCGCCUGCGUGGUUCG CC-GCUCACGUGGAUUCG CC-GCUCACGUUAG UCUUUCGCGUUUUG	8 <u>67 - 890 - 1234</u> <u>800 - AGU - ACUU</u> <u>800 - AGU - ACUU</u> <u>800 - AGU - ACUG</u> <u>800 - CAGU - ACUG</u> <u>800 - ACUA</u> <u>800 - ACUA - ACUA <u>800 - ACUA - ACUA</u> <u>800 - ACUA - ACUA - ACUA <u>800 - ACUA - ACUA</u></u></u>	9 5678901234 GGAUGGGGAGA AGGUGGGGGAA UAGUGGGGGAA UAGUGGGGGAA GAGUGCGCGAA UAUGCCGCGAA UAUGCCGCGAA GAUACCGUGAA UGGGUUCCG-	1 0 557890123456 CCGCCUQGGAAU CCGCUUGGGAAQ CCUCUGGGAAA CCUCCUCGGAAA SCCUCCUCGGAAA SCCUCCGGAAA SCCUCCGGAAA SCGUACCGCAAC	1 -7890123 -ACCGGUU -UCCCAGU -UCCULAGGU -UUCAGGU -UUCCCGU -UCCAGUA -UCCUCCU -CGCACUA -UCCACUA	1 2 4567890123456 -GCUGUAGGCUUU -GUCGACAGCU- -GCUGUCAGCU- -GUUGCAUUC- -GUUGCAUUCC- CGCUGCUAUUGUU UGCUGUUACCACU AGCUGGAUGGGU AGCUGGAGUUGU	19072134567
190721344444	7 123456789012345 CAGGGUCGGGCCUGG CGCCULAAGGCUGG CGCCULAGGCCCGG CUGGUAGAGCCUGA CCGCCUGGGCUGG CCGCCUGGCUUCCG ACACCUCACGUGGAU CCUGCU-GCGUAUUC CC-GCUCACGUUAG CCCGCUACGCUUAG CCCCGUAGCGC-CCA	8 <u>67 - 890 - 1234</u> <u>100 - AGU - A</u>	9 5678901234 GGAUGGGGGA GGGUGGGGGA UAGUGGGGGGA UAUGGGUGGG	1 0 567890123456 CCGCCUGGGAAG CCGCCUGGGAAG CCGCCCGGGAAG CCUCCUGGGAAA SUCCUCGGGAAA SUCCUCGGGAAA SUCCUCGGGAAA SUCCUCGGGAAG GGUA-CCGCAGC CCCAUGCGA-G	1 -7890123 -ACCCAGU -UCCCAGU -UCCCGGU -UUCCCGU -UCCCGGU -UCCGGUU -UCAUCCU -CGCACUA -CCCACUA -UUCAUUU -AGUAGCG	1 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU -GCUGCAUCC -GCUGCAUUCC CGUGCAUUCC CGCUGCUACU CGCUGCUACU CGCUGCUUCCAU AGCUGGAUGGGU AGCUGGAGGCUUU ACUGGCAGCCUU	190721345670
1907213456706	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGGCAAGGGCCUGG CUGGUAGAGCCUGA CCCGCCUGCUUCGGACACCUGACUCGGAC CCCGCCUGCGUCGGAC CCCGCCUGCGUCGGAC CCCGCCUGCGUCGGAC CCGCGUCAGCGCCCGA CGACGGCGCCCGAC CGAUGGAUCGCCCCGA	8 <u>67 - 890 - 1234</u> <u>800 - AGU - A</u>	9 5678901234 GGAUGGGGGAA GGGUGGGGGGAA GGGUGGGGGGAA GGAUGGGUGAA GGAUGGGUGAA GAUACCGUGA GAUACCGUGAA UGGGUUCCG-GUUU- UGG-GGUUU- UGG-GGUUU-	1 0 567890123456 CCGCCUGGGAAG CCGCCUGGGAAG CCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAA SUCCUCGGGAAG SGUACCGGAAG SGUACCGCAGGAG SGUACCGCA-G CCCCAUGUCA-A	1 -7890123 -ACCGGGU -UCCUAGGGU -UCCUAGGGU -UUCCGGU -UUCCGGU -UCCGGUU -UCCGGU -UCCAUA -CCCAUA -UUCAUA -UUCAUA	1 2 4567890123456 -GCUGUAGGUUU -GUGACAGCU- -GCUGUCAGCU- -GUUGCAUUC- -GUUGCAUUCC- CGCUGCUAUUGU UGCUGUUACCACU AGCUGGAUGGU AGCUGGAUGGU AGCUGCAGGCUUU -ACCAJUGAGCAU	19072134445670
19072134567068	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGCCUUGAGCCUGG CGGCACAGGGCCGG CCGGCUGGCUGGGCGGA CCCGCCUGCGUGGCGA CCCGCCUGCGUGGU CC-GCUCACGUGGU UCUUUCGCGUUUU CGCCGUCGC-CCA CGAUGCAUCCC-CCA CGAUGCAUCCC-CCA	8 657890_1234 800AGU_ACUU 900AGU_ACUU 900AGU_ACUG 900_AGU_ACUG 900_AGU_ACUG 900_ACUG 900_ACUG 900_ACUG 900_ACUG 900_BUC 900_ACUG 900_BUC 900_ACUG 900_ACUG 900_ACUA 900_AC	9 5678901234 GGAUGGGGGA UAGUGGGGGA UAGUGGGGGA UAGUGGGGGA GGAUGGGGGA UAUGCGGGA UAUGCGGGA UGGGGUUCCG- UGG-GGUUC- ACUGAU	1 0 0 2567890123456 CCGCCUGGGAAU CCGCCUGGGAAG CCUCUGGGAAA CCUCCUGGGAAA CCUCCUGGGAAA CCUCUGGGAAA CCUCUGGGAAA CCUCUGGGAAA CCUCUCGGAAA CCUCUAGGAAU CCCCAUGCGA-G CCCAUGCGA-G	1 -7890123 -ACCGGGU -UCCCAGU -UCCUAGGU -UCCAGGU -UCCAGGU -UCCAGGU -UCAUCCU -CGCAUU -CCCAUU -CCCAUU -AGUAGGG -AAUAGCA	1 2 4567890123456 -GCUGUAGGCUUU -GUCGACAGCU- -GCUGCAGCU- -GCUGCAAUCU- -GUUGCAUUCC- CGCUGCUACU- CGCUGCUACU- CGCUGCUACU- AGCUGGAUGGU AGCUGCAGUUU AACUGCAGGCAU -ACCAUGAGCAU	19072134567060 19072134567060
19072134567068	7 123456789012345 CAGGGUCGGGCCUGG CGCCULAAGGCUGG CGCCULAAUGCCUGG CUGGUAGGGCCUGA CCCGCCUGCGUUCG ACACCUCACGUGGAU CCCGCCUGCGUUCG CCGCCUACGUUGG CCGCGUAGGCC-CGA CGAUGCAUCGC-CGA CGAUGCAUCGC-CGA CUCUAUGGC-UGG	8 <u>67_890-1234</u> <u>800-AGU-ACUA</u> <u>800-AGU-ACUA</u> <u>800-AGU-ACUA</u> <u>800-AGU-ACUA</u> <u>800-AGU-ACUA</u> <u>800-AGU-ACUA</u> <u>800-AGU-ACUA</u> <u>800-GGC-GGU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u> <u>8000GU-ACUA</u>	9 5678901234 GGAUGGGGGAA GGGUGGGGGAA UAGUGGGGGAA GAUGGGGGGAA GAUGCGCGGAA GAUACCGGGA GAUACCGGGA GAUACCGGGA UGGGUUCC9- UGG-GGUUC- UGG-GGUUC- UGG-GGUUC- UGG-GGUUC- CAG-GGGAAG	1 0 567890123456 CCCCUGGGAAG CCCCUGGGAAG CCCCUCGGGAAG CCUCUGGGAAG CCUCUGGGAAA SUCCUCGGGAAG SCCUCUGGGAAA SUCCCACGGAAG SCCUCUGGGAAG SCCCAUGUGA-G CCCCAUGUGA-G CCCUGUGGA-G	1 -7890123 -ACCCAGU -UCCCAGU -UCCCGGU -UCCCGGU -UCCGGUU -UCCGGUU -UCCAUAU -CCCAUAU -UUCACUA -UUCACUA -GAUCUCG -GAUCUCG -AAUAGCA	1 2 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAAUCU- -GCUGCAUUCC- CGCUGCAUUCC- CGCCGCCUACU- CGCUGGAUUGCU AGCUGGAUGGGU AGCUGCAGGUUU AACUGCAGGAU -ACCAJAGAGCAU AGCUGCAGUU- CGACGCUGGGU-	19072134567068
190721345670680	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGCCUUGAGGCGGGC CUGGUAGGGCGAGA CCCGCCUGCGUUGCG CCCGCCUGCGUUGC CCCGCCUCACGUGGU CCUCCACGUGCGCCCGA CGAUGCAUCGCCCGA CUCUAAUACGGC-CGA CUCUAAUGGCC-CGA	8 <u>67 - 890 - 1234</u> <u>100 - AGU - ACUA</u> <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA <u>100 - AGUA</u> <u>100 - AGUA <u>100 -</u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u>	9 5678901234 GGAUGGGGGA GGGUGGGGGA UAGUGGGGGA GGAUGGGGGA UAGCGGGA GAUGCGGGA UAUGCCGGA GAUACCGGA GAUACCGGA UGG-GGUUL-C UGG-GGUUL-C CAG-GGGAG UAA-CCGGGAG	1 0 567890123456 CCCCCUGGGAAG CCCCCUGGGAAG CCCCUCGGGAAG CCUCCUGGGAAG SCCUCCUGGGAAG SCCUCCUGGGAAG SGCUCCUGGGAAG SGCUCCUGGGAAG CCCCAUGCGA-G CCCCAUGCGA-G CCCCUGGGA-G CCUUCUGGA-G	1 -7890123 -ACCGGGU -UCCCGGU -UCCULAGGGU -UUCCCGU -UUCCCGU -UCCGGUU -UCCACUA -UUCAUUU -GGUAUUU -AGUCUCG -AAUCUCG -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GUGACAGCU- -GCUGUCAGCU- -GUUGCAUUC- -GUUGCAUUCC- CGCUGCUAUUGUU UGCUGUUACCACU UGCUGUUACCACU UGCUGCUAGGAUGGGU AGCUGCGAGCUUU -ACCAUAGAGCAU -ACCAUAGAGCAU -GGCCGCGUGA-	1 90721 3456 70680 1 22344456 70680
1907213456706801	7 123456780012345 CAGGGUCGGGCCUGG CGCCUIAAGGCUGG CGCCUIAAGGCUGG CUGGUAGAGCCUGA CCGCCUGGGUGGCCGA ACACCUCACGUGGAT CCUGCU-GCGUAUU CCUGUU-GCGUUAU CCUGUUACGCUUA CGCGUAGGCCCGA CUCUAUUACGC-UGA CCCCAUACGCC-CA	8 <u>67 - 890 - 1234</u> <u>100 - AGU - A</u>	9 5678901234/ GGAUGGGGGA GGGUGGGGGA GGGUGGGGGA GGGUGGGGGGA GAUGGGUGCGGA AGGUACGCGGA MAGGUACGCGA UGGGUUCCG- UGG-GGUCCG- UGG-GGUUCG UGGGUCCG- UGG-GGUUCG UGG-GGUUCG CAC-GGGAG UAA-CCGGGA	1 0 567890123456 CCGCCUGGGAAG CCGCCUGGGAAG CCGCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG SCUCCUGGGAAG SCUCCUCGGAAG SCUCCUCGGAAG SCUCCUCGGAAG SCUCCUCGGAAG CCCAUGUCA-A CC-CUGUGGA-G SCUUGUGGA-G SCUUGUGGCUA-A	1 -7890123 -ACCGGGU -UCCCAGU -UCCCGGU -UCCCGGU -UCCGGUU -UCAUCCU -UCAUCCU -UCAACGA -GAUAGGU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAUCC- -GCUGCAUUCC- CGCUGCUACU- CGCUGCUUUCC- CGCUGCUUUCCACU AGCUGCAGGCUUU AGCUGCAGGCUUU AACUGCCAGGCU -ACCAUAGACAU AGCUGCAGUU- CGACGCCGUGA- CGACGCCAGGU-	1 90721 3456706801 122344456706801
19072134567068013	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGGCLAGGGCCUGG CUGGUAGGCCUGA CCCGCCUGCGUUCCG ACACCUCACGUGGAU CCCGCCUGCGUAGGCCCGA CGCCUACGCUAGGCCCGA CGCCUAGGCCCCGA CGCCUAGGCCCCGA CUCUAUGCGCCCGA CUCUAUGCGCCCGA CAUACUGCGCCCGA	8 <u>67 - 890 - 1234</u> <u>8</u> UU - AGU - AGU - AGU <u>6</u> UU - AGU - AGU <u>6</u> UU - AGU - AGU <u>6</u> UU - AGU - AGU <u>6</u> U - AGU - AGU <u>6</u> GU - AGU <u>6</u> GC - UGU - AGU <u>6</u> GG - GGC - GGG <u>6</u> GU - AGU <u>6</u> GU - AGU <u>6</u> UU-AGU <u>6</u> U - AGU <u>6</u> U - AGU	9 5678901234 GGAUGGGGGAA GGGUGGGGGAA GGGUGGGGGGAA GAUGGGGGGAA GAUACCGGGA GAUACCGGGA GAUACCGGGAU UGG-GGUUCG- 	1 0 567890123456 CCCCUGGGAAG CCCCUGGGAAG CCCCUCGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCCAUGUCA-A SUGAGA-A SUGAGA-G CCCGUUGCGA-G CCCUCUGGGA-G CCCUCCGGAA-A	1 -7890123 -ACCGGGU -UCCUUAGGGU -UCCUAGGU -UCCCGU -UCCGGUU -UCCGGUU -UCCAUA -CCCAUA -CCCAUA -UUCAUU -UUCAUU -AUUAGGU -AGUAGGU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAAUCU- -GCUGCAUUCC- CGCUGCUAUUCC- CGCUGCUAUUCU- CGCUGGUAUUGUU AGCUGGAUUGGU AGCUGGAGGUUJU AGCUGCAGGCUUJU -ACCAUAGAGCAU AGCUGCCAGGUU- CGACGCCUGGU- CGACGCCUGGU- CGACGCCAGGUC- CGACGCCAGGAU-	1 90 7 2 1 3 4 56 706 80 1 3.
190721345670680137	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGCCUUGAGGCGGGC CUGGUAGGCGGGG CCCGCCUGCGUGGC CCCGCCUGCGUGGC CCCGCCUGCGUAGG CCCGCCUCAGUUGG CGCCGUAGGGC-CGA CUCUAAUGCGC-CGA CUCUAAUGCGC-CGA CUCUAAUGCGC-CGA CUCUAAUGCGC-CGA	8 <u>67 - 890 - 1234</u> <u>100 - AGU - ACUA</u> <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - AGUA</u> <u>100 - ACUA</u> <u>100 - CGU - ACUA</u>	9 5678901234 GGAUGGGGGA GGGUGGGGGGA UAGUGGGGGA (UAGGGGGGA (UAGGGGGA (UAGGGGGA (UAGGGGAUCGGA (UGGGGUUCG- UGG-GGUU- (CAG-GGGAGG UAA-CCGGGA UAA-CCGGGA (UA-CGGGAGG (UA-CGGGAGG (UA-CGGGAGG (UA-CGGGAGG (UA-CGGGAGG (UA-CGGGAGG (UA-CGGGAAG (UA-CGGAAG (UA-CGAAG (UA-CGAAG (UA-CGGAAG (UA-CGGAAG (UA-CGGAAG (UA-CCAAG (UA-CCAAG (UA-CCAAG (UA-CCAAG (UA-CCAAG (UA-CCAAG (UA-CAAG	1 0 567890123456 CCCCCUGGGAAU CCCCUGGGAAU CCCCUGGGAAG CCUCCUGGGAAA CCUCCUGGGAAA GCUCCUGGGAAA GCUCCUGGGAAG CCCCAUGCGA-G CCCCAUGUGA-A SC-CUGUGGA-G CCCGUGGGA-G CCGUCGCUA-A AC-CGUCGGA-G	1 -7890123 -ACCGGGU -UCCCAGU -UCCAGGU -UCCAGGU -UCCAGGU -UCCAUU -UCCAUU -UCCAUU -UCCAUU -GAUCUCG -AAUAGGU -AGUAGGU -AUAGCU -AGUAGCU	1 2 4567890123456 -GCUGUAGGUUU -GUGACAGCU- -GCUGUCAGCU- -GUUGCAUUC- -GUUGCAUUCC- CGCUGCUAUUGUU UGCUGUUACCACU AGCUGGAUGGU AGCUGGAUGGUUU AACUGCCAGGCU -ACCAUAGACGAU -ACCAUAGACGU -CGACGCCGUGA CGACGCCGUGA CGGCGCCCGUGA CGGCGCCAGAUC- CGCCGCCAGAUC-	190721345670680137
1907213456706801378	7 123456780012345 CAGGGUCGGGCCUGG CGCCULAAGGCUGG CGCCULAAGGCUGG CUGGUAGGGCCUGA CCCGCCUGCGUCGA ACACCUCACGUGGA CCCGCCUGCGUCGA CCUCGCUAGGGCCCA CGCUGAGGCCCCA CGCUGAGGCCCCA CUCUAUGGGCCCA CAUACUGCGCCCA CAUACUGCGCCCA	8 <u>67890_1234</u> <u>600_AGU_AGUA</u> <u>600_AGU_AGUA</u> <u>600_AGU_AGUA</u> <u>600_AGU_AGUA</u> <u>600_AGU_AGUA</u> <u>600_AGU_AGUA</u> <u>600_AGU_AGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u> <u>600_AGUAAGUA</u>	9 5678901234 GGAUGGGGGA GGGUGGGGGA GGGUGGGGGA GAUGGGGGGA GAUGCGCGGA GAUACCGGGA GAUACCGGGA UGGGGUUCCG UGG-GGUUCCG UGG-GGUUCC CAC-GGGAG UAA-CCGGGA CCC-GGGAG UAA-CCGGGA CCC-GGGAG CG-UCUUAA CG-UCUUAA	1 0 267890123456 CCGCUGGGAAG CCGCUGGGAAG CCGCUGGGAAG CCUCUGGGAAG CCUCUGGGAAG SCCUCUGGGAAG SCCUCUGGGAAG SCUCUGGGAAG SCCUCUGGGAAG SCCCUGUGGA-G SCUCUGCGAAC SC-CUGUGGA-G SCUCUGCGGAAC SA-CGUGGGAAG	1 -7890123 -ACCGGGU -UCCCAGU -UCCCGGU -UCCCGGU -UCCGGUU -UCCGGUU -UCCAUAU -UCCAUAU -UCCAUAU -UCCAUAU -UCCACUA -UUCACUA -GAUAGGU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAAUCU- -GCUGCAUUCC- CGCUGCAUUCU- CGCUGCAUUCU- CGCUGGAUGGGU AGCUGGAUGGGU AGCUGCAGGUU -ACCAUAGAGCAU AGCUGCCAGGU- CGACGCUGGU- CGACGCUGGU- CGACGCCAGGUC- CGCCGCCAGGUC- CGCCGCCAGGUC-	1 90 7 2 1 3 4 5 6 7 0 6 8 0 1 3 7 8 1 2 2 3 4 4 4 4 4 5 5 5 6 6 6 6 6 8
19072134567068013789	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGGCAAGGGCCUGG CUGGUAGGCCUGG CUGGUCGGCGCAG CCCGCCUGCGUUCG ACACCUCACGUGGUUCG CC-GCUCACGUGGUUCG CC-GCUCACGUGGUUCG CC-GCUCACGUGGC-CGA CGAUGCAUCGC-CCA CGAUGCAUCGC-CCA CCUCUAAUACGG-CCA CACUACUGCGC-CA CACUACUGCGC-CA GACCCUGCGC-CA GACCCUCCGC-CA	8 <u>67 - 890 - 1234</u> <u>100 - AGU - GGU - AGU - AGU - GGU - AGU - AGU - GGU - GU - GGU - GU - GGU - GU - GGU - GU - GU - GU - GU - GGU - GU </u>	9 5678901234 GGAUGGGGGA GGGUGGGGGGA GGGUGGGGGGA GGUGGGGGGA GAUGCGGGGA GAUGCCGGGA GAUGCCGGGA GAUACCGUGA UGG-GGUUCG- UGG-GGUUCG- CAC-GGGAG CCC-GGGUAG CG-UCAAAG CG-UCAAAG GC-GGAGGG	1 0 567890123456 CCGCCUGGGAAG CCGCCUGGGAAG CCGCCUGGGAAG CCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAG SCCCAUGUCA-A SUGAGA-A SUC-CUGGGA-G CCCGUGGGA-G CCCUGGGA-A AA-CGUGGGA-G CC-CUGGGA-G CC-CUGGGA-G	1 -7890123 -ACCGGGU -UCCULAGGGU -UCCULAGGGU -UUCCGGU -UUCGGUU -UCCGGUU -UCCGGUU -UCCAUA -UUCAUUU -GCCAUA -UUCAUUU -GGUAGGU -AGUAGCU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GUGACAGCU- -GCUGUCAGCU- -GCUGCAUCC- CGCUGCAUUCC- CGCUGCUAUUGU AGCUGGAUGGU AGCUGCAGGUUU AGCUGCAGGCUUU -ACCAUAGAGCAU -ACCAUAGAGCAU -ACCAUGGGU- CGACGCUGGGU- CGACGCUGGGU- CGACGCAGGCU CGCCGCGAGACU- CGCCGCCAGACC- CGCCGCCAGACC- CGCCGCCAGGCU	1 90 7 2 1 3 4 5 6 7 0 6 8 0 1 3 7 8 9 1 2 2 3 4 4 4 4 4 5 5 5 6 6 6 6 6 6 9
1 90 7 2 1 3 4 5 6 7 0 6 8 0 1 3 7 8 9 1 1 90 7 2 1 3 4 5 6 7 0 6 8 0 1 3 7 8 9 1	7 123456780012345 CAGGGUCGGGCCUGG CGCCULAAGGCUGG CGCCULAAGGCUGG CUGGULAGAGCCUGA CCUGCUCAGGUGAU CCUGCU-GCGUAUUCC ACACCUCACGUGAU CCUGCU-GCGUAUUCC CGCGUACGCUCACGUUAG CCUCUAUGCGUUAC CCUCUAUGCGUUAC CCCCAUACGGC-CCA CUCUAUGCGC-CCA CCCAUACUGCGC-CA CAUACUGCGC-CA CAUACUGCGC-CA GCCCUACGCC-CA GCCCCAGCCCCCA	8 <u>67 - 890 - 1234</u> <u>100 - AGU - A</u>	9 5678901234/ GGAUGGGGGA GGGUGGGGGA GGGUGGGGGA GGGUGGGGGA GAUGGGUGCGGA AGGUACGCGGA AGGUACGCGGA UAGGCUCCGG- UGG-GGUUC- UGG-GGUUC- UGG-GGUUC- CAC-GGAUAC CCC-GGGUAC CCC-GGGUAC CCC-GGGUAC CGUCAAAA CGUCUUAA CGUCUUAA	1 0 567890123456 CCGCCUGGGAAG CCGCCUGGGAAG CCGCCUGGGAAG CCCCCGGGAAG CCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGGGAAG SCUCCUGCGAAC SCUCUGCGGAAC SCUCGGGCCAAA CCCUGCGGAAC CCGUGCGAAC CCCCUGCGGAAC CCCCUGCGGAAC CCCCUGCGGAAC CCCCUGCGCAAC	1 -7890123 -ACCCAGU -UCCCAGU -UCCCAGU -UCCCGGU -UCCCGGU -UCCCGGU -UCCAGU -UCCACUA -CCCACUA -GGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GCUGUAGGUU- -GCUGCAGCU- -GCUGCAUUCC- -GUUGCAUUCC- CGCUGCUACU- CGCUGCUUUCCAC CGCUGCUUUACCACU AGCUGCAGGCUUU AGCUGCAGGCUU AGCUGCAGGCU- CGACGCCAGUU- CGACGCCAGGU- CGACGCCAGGU- CGCCGCCAGAC- CGCCGCCAGAC- CGCCGCCAGACC- CGCCGCCAGACC- CGCCGCCAGACC-	19072134444555666666671
1 90 7 2 1 3 4 5 6 7 0 6 8 0 1 3 7 8 9 1 90 7 2 1 3 4 5 6 7 0 6 8 0 1 3 7 8 9 1	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGCCUGG CGGCLAAGGGCCCGG CUGGUAGAGCCUGA CCCGCCUGCGUUCCG ACACCUCACGUGGAU CCUGCU-GCGUUAGG CC-GGUCACGUUAG CCCGCUAGGCC-CCA CGAUGCAUCGC-CCA CUCUAAUGUGC-UGA CCCCAUAGGGC-CCA CUCUAAUGUGC-CA CACUACUGCGG-CA CACUACUGCGC-CA CACUACUGCGC-CA CACUACUGCGC-CA CACUACUGCGC-CA CACUACUGCGCC-CA CACUACUGCGCC-CA CACUACUGCGCC-CA CACUACUGCGCC-CA CACUACUGCGCC-CA CACUACUGCGCC-CA	8 <u>67-890-1234</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUA</u> <u>600-ACUACUACUA</u> <u>600-ACUACUACUA</u> <u>600-ACUACUACUA</u> <u>600-ACUACUACUACUACUA</u> <u>600-ACUACUACUACUACUA</u> <u>600-ACUACUACUACUACUACUACUA</u> <u>600-ACUACUACUACUA</u> <u>600-ACUACUACUA</u> <u>600-ACUACUACUACUA</u> <u>600-ACUACUACUACUA</u> <u>600-ACUACUACUACUACUA</u> <u>600-ACUACUACUACUA</u> <u>600-ACUACUACUACUACUA</u> <u>600-ACUACUACUACUACUACUACUA</u>	9 5678901234 GGAUGGGGGAA GGGUGGGGGAA GGGUGGGGGGAA GAUGGGGGGAA GAUGCGGGGAA GAUACCGGGA GAUACCGGGA UGG-GGUUCU- UGG-GGUUCU- UGG-GGUUCU- CAG-GGAAG UAA-CCGGAA CCC-GGGAAG UAA-CCGGAAG UAA-CGGGAAG UAA-CGGGAAG UAA-CGGGAG CGUCUAAAA GCGGACGAA 	1 0 567890123456 CCCCUGGGAAG CCCCUGGGAAG CCCUCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGGGAAG CCUCCUGCGAAG CCCCUUGCAAG CCCCUUGCAAG CCCUUGCGAAG CCCUCCGGAACA CCCUCCGGAACA CCCUCCGGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA CCCUCCGAACA	1 -7890123 -ACCGGGU -UCCUUAGGGU -UCCUAGGU -UCCGGUU -UCCGGUU -UCCGGUU -UCCGUUC -UCCAUA -UCCAUA -UUCAUUU -GGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAAUCU- -GCUGCAUUCC- CGCUGCUAUUC- CGCUGCUAUUCC- CGCCGCCUAUU- CGCUGGAUGGGU AGCUGCAGGCUU AGCUGCAGGCUU AGCUGCCAGGCU -ACCAUAGAGCAU AGCUGCCAGGUU- CGACGCUGGGU- CGACGCCUGGU- CGCCGCCAGAU- CGCCGCAGGUC- CGCCGCCAGGUC- CGCCGCCAGGCC CGCCGCCAGGCC CGCCGCCAGGCC CGCCGCCAGGCC CGCCGCCAGGCC CGCCGCCGCAGAAU	1907213444455566666667
19072134444555566666667	7 123456789012345 CAGGGUCGGGCCUGG CGCCUUAAGGCUGG CGGCAAGGGCCUGG CUGGUAGGGCGAGA CCCGCCUGCGUUGCG CCCGCUGCGUUGCG CC-GCUCACGUGGAU CCC-GCUCACGUGGC CGAUGCAUCGC-CGA CUCUAAUACGC-CCA CUCUAUUACGG-UGA CUCUAUUACGC-CGA CCCCAUAGCGC-CGA CACUACUGCGC-CAA CGCCCUGCGC-CAA GAGCCCUGCGC-CAA GAGCCCUGCGC-CAA GAGCCCUGCGC-CAA CGCCCCAGCCC-CAA CGCCCCAGCCC-CAA	8 <u>67 - 890 - 1234</u> <u>100 - ACU-ACUA</u> <u>100 - ACU-ACUA</u> <u>100 - AGU-ACUA</u> <u>100 - AGU-ACUA</u> <u>100 - AGU-ACUA</u> <u>100 - AGU-ACUA</u> <u>100 - CGU-ACUG</u> <u>100 - CGU-ACUG <u>100 - CGU-ACUG</u> <u>100 - CGU-ACUG <u>100 - C</u></u></u>	9 5678901234 GGAUGGGGGAA GGGUGGGGGGAA GGGUGGGGGGAA GGAUGGGGGGAA GGAUGGGGGAA (JAUGGGGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	1 0 567890123456 CCCCUGGGAAG CCCCUGGGAAG CCCCUCGGGAAG CCUCCUGGGAAG SCCUCCUGGGAAG SCCUCCUGGGAAG SGCUCCUGGGAAG SGCUCCUGGGAG CCCCAUGUCA-A SUGCCACA CCCCAUGCA-G CCCCUGGGA-G CCCGUGGGA-G CCCGCUGGGA-G CC-CUCUGGA-G CC-CUCUGGA-G CC-CUCUGGA-G CC-CCUCGGA-G CC-CCUCGGA-G C-CCUCGCA-G C-CCUCGGA-G C-CCUCGGA-G C-CCUCGGA-G C-CCUCGGA-G C-CCUCGGA-G C-CCUCGGA-G C-CCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGGA-G CCCCUCGCGA-G CCCCUCGCGA-G CCCCUCGCGA-G CCCCUCGCA-G CCCCUCCCCA-G CCCCUCCCCUCGA-G CCCCUCCCCUCCCA-G CCCCUCCCCCUCGCA-G CCCCUCCCCA-G CCCCCUCCCCA-G CCCCCCUCCCCCUCCCA-G CCCCCUCCCCCUCCCA-G CCCCCCUCCCCCCCUCCCA-G CCCCCCUCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	1 -7890123 -ACCGGGU -UCCUAGGU -UCCUAGGU -UCCCCGU -UCCCCGU -UCCCCGU -UCCACUA -UCCACUA -UCCACUA -UUCAUUU -GAUCUCG -AAUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU -AGUAGGU	1 2 4567890123456 -GCUGUAGGUUU -GCUGUCAGCU- -GCUGCAAUCU- -GUUGCAUUCC- CGCUGCUAUUCC- CGCUGCUAUUGUU UGCUGUUACCACU UGCUGUUACCACU AGCUGGAUGGGU AGCUGCAGGCUUU -ACCAUAGAGCAU -ACCAUAGAGCAU -ACCAUAGAGCAU -CGCCGCCGGGU- CGCCGCCGGGU- CGCCGCCGGGCL CGCCGCCAGAUC- CGCCGCCAGACC- CGCCGCCAGACC- CGCCGCCGGGGAA 4567890123456	190721345670680137891

Figure 1. Alignment of 21 selected 5S rRNA sequences. The nucleotide numbers refer to the consensus structure of Fig. 2, the sequence numbers refer to the list of organisms shown in Fig. 4.

purime or pyrimidime nucleotide is conserved in more than 80 % of all sequences. The consensus sequence is folded according to the five helix model (18,19), and potential base pairs shown in straight lines are conserved in all molecules, with few exceptions. Some alternative or additional base pair interactions are possible, although in a smaller number of molecules: the base pairs 55/29 and 56/28 could be formed in 48 sequences (19), the base pairs 12/113 and 13/112 in six eubacterial molecules (20), and helix III could be extended by two base pairs (36/46 and 37/45) in 60 molecules (19). A tertiary interaction between nucleotides 42-45 and 80-77 observed in <u>E.coli</u> 55 rRNA (22) is possible in other eubacteria except cyanobacteria.

Fig. 2B shows a mollusc 5S rRNA (from the snail Arion rufus) written in



Figure 2 A: Consensus structure of 5S rRNA, based on 72 sequenced molecules (2-14). Nucleotides conserved in more than 80 % of all analyzed sequences are indicated (R, purine nucleotide, Y, pyrimidine nucleotide). Group-specific deletions are marked by shaded circles, group-specific insertions by arrows. Potential hydrogen bonds between the base pairs A/U, G/C and G/U conserved in at least 95 % of all molecules are shown in straight lines, alternative base pair interactions in helices IV and V in broken lines. B: Nucleotide sequence and potential secondary structure of 5S rRNA from the snail Arion rufus. The sequence was determined by chemical cleavage methods as described (21). Nucleotides 42 and 114 of the consensus sequence (shaded circles) are deleted in all eukaryotic (cytoplasmic) 5S rRNA molecules.

the consensus configuration. The nucleotide sequence was determined independently by us and by Fang et al. (14).

The group-specific patterns of nucleotide deletions and helical interactions are shown in Fig. 3. Nucleotides 42 and 114 of the consensus sequence are deleted in all eukaryotic cytosolic species, nucleotides 73, 77, 88 and 106 are deleted in all eubacterial and organellar sequences, and none of these nucleotides are deleted in the six archaebacterial sequences. Instead, some individual deletions (nucleotides 67, 94 and 99) are found in the archaebacterial sequences 44, 45 and 46 (sequence numbers refer to Fig. 4). An addi-





Figure 3. Deletions, insertions and potential base-pairing interactions in 55 rRNAs. Residues marked by black circles are deleted in all members of a given group, those marked by shaded circles are deleted in some organisms. Inserted nucleotides are indicated by arrows. See also the alignment of Fig. 1. The Halococcus molecule contains a long insert of 108 nucleotides at 109/110 (2).



Figure 4. Phylogenetic tree based on the nucleotide substitution analysis of 5S rRNAs. Nodal points are expressed as the number of nucleotide substitutions between pairs of sequences or protosequences (16,24,25), and the early branching pattern is correlated with a phylogenetic interpretation of group-specific deletions (see Fig. 3). All sequences except 15 (14), 18 (3), 23 (4), 24 (5), 31 (6), 43 (7), 46 (7), 57 (8), 59 (9), 60 (10), 62 (11), 67 (11), 68 (12) and 69 (13) are compiled in ref. 2.

tional deletion of nucleotide 97 is a common feature of four eubacterial sequences from <u>T</u>. <u>thermophilus</u> (13), <u>T</u>. <u>aquaticus</u> (2), <u>R</u>. <u>rubrum</u> (12) and <u>P</u>. <u>denitrificans</u> (11). In two molecules (from wheat mitochondria (2) and <u>Mycoplasma</u> (2)) helix IV is further reduced by deletions of 7 to 9 nucleotides including positions 88 and 97. However, this could be a convergence phenomenon, because the two sequences belong to different affinity groups (see below).

The most conserved secondary interactions are in helices I, II and III: note the looped out base 67 of helix II (potentially base-paired only in the mitochondrial sequence) and the two looped out bases 53 and 54 of helix III (helix I contains a looped out base only in the two archaebacterial sequences 45 and 46).

Helices IV and V are more variable in their base pairing and deletion patterns: in helix IV nucleotides 80 and 82-86 can form base pairs either with 103-97 (configuration 1) or with 104-98 (configuration 2). The helix IV configurations shown in Fig. 3 are energetically (23) favourable over the alternative configurations, and a certain phylogenetic group specificity of these structural modulations cannot be overlooked: configuration 1 dominates in most eubacteria (j,n) and archaebacteria (f,g,i), in all plants and algae (d,e) and fungi (c, with the exception of <u>Neurospora</u> (2) due to the deletion of nucleotide 84). In all six protozoa both configurations are equivalent, and configuration 2 dominates in all 18 metazoan molecules. Helix IV is interrupted by a looped out base (87 or 88) in plants, algae, fungi and in all archaebacteria (one exception). In all protozoa configuration 1 produces a looped out base 87, and configuration 2 results in a mismatch at either 84/101, 85/100 or 86/99. In all metazoan molecules the bases 87/98 are mismatched (A/C).

A phylogenetic tree analysis of 46 sequences of 5S rRNAs has previously been described (16). The tree of Fig. 4 is constructed by the same method and is based on a larger number of sequences.

The nematode <u>C. elegans</u> (3) diverges from the metazoan branch prior to the separation between invertebrates (including the snail sequence of Fig. 2B) and vertebrates. Three base pairs in helix II and III (21/64, 30/56 and 35/49) are either C/G, G/C and A/U, respectively, in all 12 vertebrates, or G/C, A/U and U/A in the five invertebrate sequences 13-17. The nematode sequence is invertebrate-like in two positions and vertebrate-like in one position. The protozoan group includes <u>Euglena gracilis</u> (4) and the slime mold <u>Dictyostelium</u> <u>discoideum</u> (5) due to sequence affinities and the group-specific ambiguity of helix IV configurations. The lower fungus Phycomyces blakesleeanus (6) appears to diverge from the fungal branch, together with the fission yeast <u>S</u>. <u>pombe</u> (2), prior to the separation between filamentous fungi and budding yeasts.

The subtree of Gram-positive bacteria includes two actinomycetes (<u>Strepto-myces griseus</u> (9) and <u>Micrococcus lysodeikticus</u> (10)), and the two mycoplasmas <u>M. capricolum</u> (2) and <u>Spiroplasma</u> (8). The two related eubacteria <u>P. denitri-ficans</u> and <u>R. rubrum</u> appear to share a common ancestor with the two <u>Thermus</u> species, and the wheat mitochondrial sequence has a higher affinity to the protosequence of this early diverging group (53 substitutions) than to the protosequences of other eubacterial groups (between 61 and 66 substitutions). <u>P. denitrificans</u> and <u>R. rubrum</u> have earlier been recognized as possible relatives to proto-mitochondria on the basis of cytochrome c homologies (26).

The six archaebacterial sequences form an extremely heterogenous group (7) but appear to exhibit a common root, as suggested by substitution analysis and by the group-specific presence of the six nucleotides 42,73,77,88,106 and 114. The archaebacterial protosequence is significantly more related to the eukaryotic-cytosolic protosequence (52 substitutions) than to the eubacterial/ organellar protosequence (65 substitutions), and a common ancestor of archaebacterial and eukaryotic molecules is also suggested by five eukaryotic features (presence of nucleotides 73,77,88,106, and looped out base 87) versus two prokaryotic features (presence of nucleotides 42 and 114).

ACKNOWLEDGEMENT

We thank Dr. R.T. Walker for communicating to us the <u>Spiroplasma</u> 5S rRNA sequence prior to publication and M. Heidrich for helpful advice.

*Dedicated to Friedrich Cramer on the occasion of his sixtieth birthday.

+ Present address: Institut für Kristallographie, Freie Universität, 1000 Berlin, 33, FRG.

REFERENCES

- 1. Erdmann, V.A. (1976) Progr. Nucl. Acid Res. Mol. Biol. 18, 45-90.
- 2. Erdmann, V.A. (1982) Nucl. Acids Res. 10, r93-r115.
- Butler, M.H., Wall, S.M., Luehrsen, K.R., Fox, G.E. and Hecht, R.M. (1981) J. Mol. Evol. <u>18</u>, 18-23.
- 4. Hori, H., Osawa, S. and Iwabuchi, M. (1980) Nucl. Acids Res. 8, 5535-5539.
- Delihas, N., Anderson, J., Andresini, W., Kaufman, L. and Lyman, H. (1981) Nucl. Acids Res. 9, 6627-6633.
- Andersen, J., Andresini, W. and Delihas, N. (1982) J. Biol. Chem. <u>257</u>, 9114-9118.
- 7. Fox, G.E., Luehrsen, K.R. and Woese, C.R. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 330-345.
- Walker, R.T., Chelton, E.T.J., Kilpatrick, M.W., Rogers, M.J. and Simmons, J. (1982) Nucl. Acids Res. <u>10</u>, 6363-6367.
- 9. Simoncsits, A. (1980) Nucl. Acids Res. 8, 4111-4124.

- Hori, H., Osawa, S., Murao, K. and Ishikura, H. (1980) Nucl. Acids Res. <u>8</u>, 5423-5426.
- MacKay, R.M., Salgado, D., Bonen, L., Stackebrandt, E. and Doolittle, W.F. (1982) Nucl. Acids Res. 10, 2963-2970.
- 12. Newhouse, N., Nicoghosian, K. and Cedergren, R.J. (1981) Can. J. Biochem. 59, 921-932.
- Kumagai, I., Digweed, M., Erdmann, V., Watanabe, K. and Oshima, T. (1981) Nucl. Acids Res. <u>9</u>, 5159-5162.
- Fang, B., DeBaere, R., Vandenberghe, A. and DeWachter, R. (1982) Nucl. Acids Res. 10, 4679-4685.
- 15. Hori, H. and Osawa, S. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 381-385.
- 16. Küntzel, H., Heidrich, M. and Piechulla, B. (1981) Nucl. Acids Res. 9, 1451-1461.
- 17. Fox, G.E. and Woese, C.R. (1975) Nature 256, 505-507.
- Luehrsen, K.R. and Fox, G.E. (1981) Proc. Natl. Acad. Sci. U.S.A. <u>78</u>, 2150-2154.
- 19. DeWachter, R., Chen, M. and Vandenberghe, A. (1982) Biochimie <u>64</u>, 311-329.
- Pieler, T., Kumagai, I. and Erdmann, V.A. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 69-78.
- Piechulla, B., Hahn, U., McLaughlin, L. and Küntzel, H. (1981) Nucl. Acids Res. 9, 1445-1450.
- Pieler, T. and Erdmann, V.A. (1982) Proc. Natl. Acad. Sci. U.S.A. <u>79</u>, 4599-4603.
- 23. Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
- 24. Küntzel, H. and Köchel, H.G. (1981) Nature 293, 751-755.
- 25. Küntzel, H. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 31-39.
- 26. Schwartz, R.M. and Dayhoff, M.O. (1978) Science 199, 359-403.