doi:10.1016/S0022-2836(02)00515-6 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 320, 455–474

JMB

Sequence and Structural Conservation in RNA Ribose Zippers

Makio Tamura and Stephen R. Holbrook*

Lawrence Berkeley National Laboratory, Structural Biology Department, Physical Biosciences Division
1 Cyclotron Road
132 Melvin Calvin Lab, Bldg 3
Berkeley, CA 94720, USA

The "ribose zipper", an important element of RNA tertiary structure, is characterized by consecutive hydrogen-bonding interactions between ribose 2′-hydroxyls from different regions of an RNA chain or between RNA chains. These tertiary contacts have previously been observed to also involve base–backbone and base–base interactions (A-minor type). We searched for ribose zipper tertiary interactions in the crystal structures of the large ribosomal subunit RNAs of *Haloarcula marismortui* and *Deinococcus radiodurans*, and the small ribosomal subunit RNA of *Thermus thermophilus* and identified a total of 97 ribose zippers. Of these, 20 were found in *T. thermophilus* 16 S rRNA, 44 in *H. marismortui* 23 S rRNA (plus 2 bridging 5 S and 23 S rRNAs) and 30 in *D. radiodurans* 23 S rRNA (plus 1 bridging 5 S and 23 S rRNAs). These were analyzed in terms of sequence conservation, structural conservation and stability, location in secondary structure, and phylogenetic conservation.

Eleven types of ribose zippers were defined based on ribose–base interactions. Of these 11, seven were observed in the ribosomal RNAs. The most common of these is the canonical ribose zipper, originally observed in the P4–P6 group I intron fragment. All ribose zippers were formed by antiparallel chain interactions and only a single example extended beyond two residues, forming an overlapping ribose zipper of three consecutive residues near the small subunit A-site. Almost all ribose zippers link stem (Watson–Crick duplex) or stem-like (base-paired), with loop (external, internal, or junction) chain segments. About two-thirds of the observed ribose zippers interact with ribosomal proteins. Most of these ribosomal proteins bridge the ribose zipper chain segments with basic amino acid residues hydrogen bonding to the RNA backbone. Proteins involved in crucial ribosome function and in early stages of ribosomal assembly also stabilize ribose zipper interactions.

All ribose zippers show strong sequence conservation both within these three ribosomal RNA structures and in a large database of aligned prokaryotic sequences. The physical basis of the sequence conservation is stacked base triples formed between consecutive base-pairs on the stem or stem-like segment with bases (often adenines) from the loop-side segment. These triples have previously been characterized as Type I and Type II A-minor motifs and are stabilized by base–base and base–ribose hydrogen bonds.

The sequence and structure conservation of ribose zippers can be directly used in tertiary structure prediction and may have applications in molecular modeling and design.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: ribosomal RNA; RNA ribose zipper; RNA tertiary interaction; protein–RNA interaction; A-minor motif

*Corresponding author

Introduction

The tertiary interactions responsible for the folding, stability, and maintenance of RNA three-dimensional structure[1] appear to be limited to a

Abbreviations used: RZ, ribose zipper.

E-mail address of the corresponding author: srholbrook@lbl.gov

0022-2836/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved

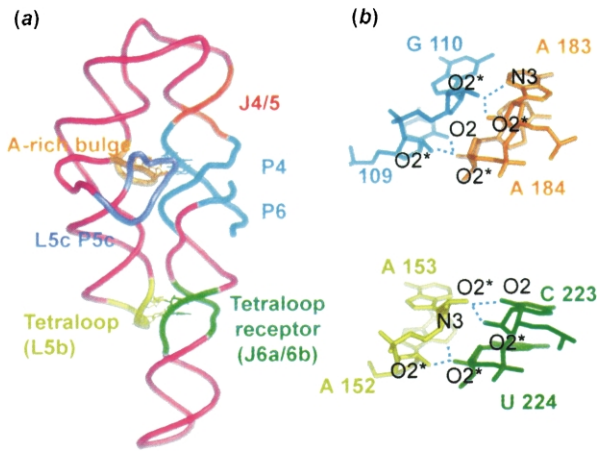


Figure 1. Structure of the P4–P6 group I intron domain and its ribose zippers. (a) There are two ribose zippers found in the group I intron; one ribose zipper mediates the interaction between the A-rich bulge (orange) and the P4 stem (light blue) and another ribose zipper mediates the interaction between the tetraloop (yellow) and the tetraloop receptor (green). (b) In the ribose zippers, there are two residues on each side (109–110, 184–183 and 152–153, 223–224) in which riboses interact by hydrogen bonding (blue broken line) between the 2′-hydroxyl groups (O2′) of the two chain segments in an antiparallel orientation. The 2′-hydroxyl groups of the 3′-ends residues also form minor groove hydrogen bonds to either the N3 atom of a purine (G110, A152) or the O2 atom of a pyrimidine (C109, C223) of the 5′-end residues on the opposite chain segment.

small number of general classes including: coaxial helical stacks,[2] kissing hairpins,[3] tetraloop–receptor interactions,[4] A-minor motifs,[5] pseudoknots,[6] loop–loop interactions such as found in tRNA,[2] and ribose zippers.[4] The ribose zipper was first recognized as an intermolecular interaction in hammerhead ribozyme crystals[7] and two intramolecular tertiary interactions in the crystal structure of the P4–P6 domain of the group I intron.[4] One ribose zipper mediates the interaction between an adenosine rich bulge and the P4 stem and the other mediates the interaction between the GAAA tetraloop and tetraloop receptor (Figure 1). In these ribose zippers, two consecutive residues from one chain segment interact in an antiparallel fashion with two consecutive residues from another chain segment distant in sequence but close in three dimensions. The backbones of these chain segments interact through hydrogen bonding between the ribose 2′-hydroxyl groups of the two chain segments (O2′–O2′), and with minor groove atoms of a base on the opposite chain segment (e.g. O2′–N3A). Base triples were also observed in these examples between the residues of the ribose zippers and their Watson–Crick paired partners. Experimental studies in which the ribose 2′-hydroxyls of the ribose zippers are replaced by deoxyribose or 2′-*O*-methyl groups suggested that the backbone–backbone and backbone–base hydrogen bonds contribute only about 2 kcal mol^{-1} total in stabilization energy to the tertiary interaction,[8] whereas stacking of the base triples contributed a significantly greater stabilization energy.[9]

Here, we define a ribose zipper (RZ) as an RNA tertiary interaction, between two distinct chain segments or between two different chains, in which at least two consecutive residues form hydrogen bonds between their ribose 2′-hydroxyl groups bridging the chain segments (see Materials and Methods). In order to understand this motif from a structural, functional, and evolutionary viewpoint, we have searched for RZs in the three-dimensional crystal structures of the large ribosomal subunits of *Haloarcula marismortui* and *Deinococcus radiodurans* and the small ribosomal subunit of *Thermus thermophilus*. We have classified the observed RZs in these rRNAs, characterized their structural and sequence preferences, interactions with ribosomal proteins, and distribution within the ribosomal RNA.

Although ribose zippers were predicted to be pervasive within many biological RNAs to stabilize tertiary folding, it was not evident from these two examples that there may be sequence specificity for this interaction that could be useful for prediction.[1] We find that in both 16 S and 23 S rRNA there is a structurally determined sequence specificity, that there is also a strong specificity for the ribose zipper to link double helical stem regions with loop regions, and that the sequences of the RZs in 16 S rRNA are more highly conserved compared to the average sequence conservation of all residues in prokaryotic 16 S rRNA as found in the Ribosomal Database Project Release 8.1.[10] We also find evidence of covariant conservation of the RZ sequences in 16 S rRNA, suggesting that RZ-mediated tertiary interactions are preserved in evolution. We have also analyzed the interactions between ribose zippers and ribosomal proteins in the small ribosomal subunit of *T. thermophilus*, and the large ribosomal subunit of *H. marismortui*.

Results

We have searched for ribose zippers in the large subunit ribosomal RNAs from *H. marismortui* and *D. radiodurans* and the small subunit ribosomal RNA of *T. thermophilus* using the general condition of hydrogen bonding between at least two consecutive ribose 2′-hydroxyls in different chain segments or different chains (see Materials and Methods).

In the *H. marismortui* large ribosomal subunit, we find 44 RZs in 23 S rRNA and two RZs between 23 S and 5 S rRNA. Likewise, in the *D. radiodurans* large ribosomal subunit, we find 30 RZs in 23 S rRNA, and one RZ between 23 S and 5 S rRNA. Of the 30 RZs in 23 S, 22 are common to *H. marismortui* rRNA as is one of the two interchain RZs between 5 S and 23 S rRNA. In the *T. thermophilus* small ribosomal subunit, we find 20

RZs in the 16 S RNA. Thus, we find a total of 97 RZs in these three ribosomal subunits, 23 of which are common to both large subunit ribosomal RNA structures. These results are detailed in Table 1.

Some generalizations can be made about these RZs: (1) in both small and large subunit ribosomal RNAs, the orientation of the two chain segments linked by the RZ is always antiparallel (no parallel RZs are observed); (2) only one ribose zipper was found to extend for more than two consecutive residues. The single exception of three consecutive riboses bridging two chain segments (discussed below) may be considered as two overlapping RZs. These findings add length and orientation constraints to our understanding of RZs.

Beyond these generalities based on the RZ backbone–backbone interactions (O2′–O2′), we are also able to classify the observed examples into a limited number of types according to their base–backbone interactions.

Types of ribose zippers

We define 11 possible types of ribose zippers according to their base–backbone hydrogen-bonding pattern as shown in Figure 2(a). Only seven of these types are observed in our database. All of these have two consecutive O2′–O2′ hydrogen bonds between the chain segments and the orientation of the chain segments of all observed RZs is antiparallel. These types are explained below.

1. *Canonical RZ*. Two base–backbone hydrogen bonds between the N3 atom of a purine or the O2 atom of a pyrimidine at the 5′ end on one side and the O2′ of the 3′ end on the other side. This is the same hydrogen-bonding pattern for ribose–ribose and base–ribose interactions found in the group I intron.[4]
2. *Cis RZ*. The base–ribose interactions are between the 5′ base and 3′ ribose and between the 3′ base and 5′ ribose: consequently the two bases and the two riboses of ribose–base hydrogen bond pairs are on the same side of ribose zipper.
3. *Reverse RZ*. Two base–backbone hydrogen bonds between a purine N3 or a pyrimidine O2 at the 3′ end on one side and the O2′ hydroxyl of the 5′ end residue on the other side. The orientations of the residues forming base–ribose interactions are opposite to a canonical RZ. *No examples are observed*.
4. *Single RZ*. Only one hydrogen bond between the N3 atom of a purine or the O2 atom of a pyrimidine at the 5′ end and the 2′ hydroxyl group of the 3′ end of the other side is present.
5. *Reverse single RZ*. Only one ribose–base hydrogen bond between the N3 atom of a purine or the O2 atom of a pyrimidine at the 3′ end and the 2′ hydroxyl group of the 5′ end of the other side is present. *No examples are observed*.
6. *Naked RZ*. Both ribose–base hydrogen bonds are missing (no ribose–base hydrogen bond).
7. *Pseudo canonical RZ*. The hydrogen bond pattern is similar to that of the canonical ribose zipper, except that at least one ribose–base hydrogen bond is not between canonical atoms (e.g. the H-bond is between the O4′ of the ribose and the N2 of a purine base). *No examples are observed*.
8. *Pseudo cis RZ*. The hydrogen bond pattern is similar to that of the cis ribose zipper, except that at least one base–ribose hydrogen bond is not between canonical atoms.
9. *Pseudo reverse RZ*. The hydrogen bond pattern is similar to that of reverse ribose zipper, except that at least one ribose–base hydrogen bond is not between canonical atoms. *No examples are observed*.
10. *Pseudo single RZ*. Only one hydrogen bond between base atom at the 5′ end and ribose atom at the 3′ end of another side, but this base–ribose hydrogen bond is not between canonical atoms.
11. *Pseudo reverse single RZ*. Only one hydrogen bond between base atom at the 3′ end and ribose atom at the 5′ end of another side, but this base–ribose hydrogen bond is not between canonical atoms. *No examples are seen*.

In *T. thermophilus* 16 S rRNA we find five types of RZs: the canonical RZ, the cis RZ, the pseudo cis RZ, the single RZ, and the naked RZ. In *H. marismortui* 23 S rRNA there are four types of RZs: the canonical RZ, the single RZ, the reverse single base RZ, and the naked RZ. Finally, in *D. radiodurans* 23 S rRNA we find five types of RZs: the canonical RZ, the pseudo cis RZ, the single RZ, the pseudo single RZ, and the naked RZ. The other RZ types defined above are not observed in the ribosomal RNAs. Molecular model building indicates that all of the proposed types of ribose zippers are sterically feasible, even those we do not observe. Their absence may be coincidental or reflect small energy differences of preferred folding pathways.

Location and distribution of ribose zippers

Figure 3(a)–(d) shows the location of the RZs in the ribosomal RNAs of *T. thermophilus* and *H. marismortui*. The ribose zippers found in the rRNAs of the large ribosomal subunit of *D. radiodurans* are discussed and compared separately. There are no RZs mediating stem–stem interactions. Out of the 20 RZs in 16 S rRNA and 46 RZs in 23 S (and 5 S) rRNA, 15 (75.0%) and 33 (71.7%), respectively, are found in stem–loop interactions. The secondary structure of each ribosomal RNA corresponds to the secondary structure diagram from the Comparative RNA Web Site.[11] A summary of the number of interacting secondary structural elements mediated by RZs in the small

Table 1. List of ribose zippers in the *T. thermophilus* small ribosomal subunit RNA and the *H. marismortui* large ribosomal subunit RNA

	Residue[b]							Base pair[f]					
Entry[a]	s5′	s3′	15′	13′	SL[c]	Domain[d]	Type[e]	S5′	S3′	15′	13′	Protein interaction[g]	Notice[h]
Small subunit ribosomal RNA (16 S)													
Canonical RZ													
1S	**A16**	U17	G1079	A1080	E–E	I-III	AU/GA	A919	A918		G1077	S5	
2S	C221	U222	A195	A196	S–J	I	CU/AA	G142	A141	U180	A179	S20	
3S	C334	C335	A1433	A1434	S–I	I–IV	CC/AA	G319	G318		G1467		
4S	C400	C401	A621	A622	S–E	I–II	CC/AA	G42	G41		C618	S4	
5S	G542	C543	A509	A510	S–I	I	GC/AA	C503	G502				
6S	C882	C883	A573	A574	S–J	II	CC/AA	G568	G567			S12	
7S	C1524	G1525	A766	A767	I–J	II–IV	CG/AA	G1511	U1510	U813		S11	
8S	A1375	U1376	A938	G939	J–J	III	AU/AG	A1346	U1345		C1344	S7 (S9)	
9S	C984	C985	A958	A959	S–J	III	CC/AA	G1221	G1220			S19	
10S	C1217	C1218	A1015	A1016	S–E	III	CC/AA	G988	G987		G1013	S14 (S19)	
11S	C1096	C1097	A1169	A1170	S–E	III	CC/AA	G1089	G1088		G1166		
12S	C1325	C1326	A1268	A1269	S–E	III	CC/AA	G1312	G1311		G1266	THX (19S)	
cis RZ													
c1S	C1369	G1370	A1250	A1251	S–I	III	CG/AA	G1353	C1352			S9, S14 (THX)	
c2S	C1352	G1353	A1287	A1288	S–I	III	CG/AA	G1370	C1369		C1249	THX (S9, S14)	
c3S	C1404	G1405	A1518	A1519	S–E	IV	CG/AA	G1497	C1496				
Single RZ													
s1S	G66	C67	A171	A172	S–I	I	GC/AA	C103	G102	C150	A149	(S20)	Type A
s2S	G785	G786	A696	U697	S–E	II	GG/AU	C797	C796	G691	G690	(S11)	Type A
s3S	U1510	G1511	U813	A814	S–J	II–IV	UG/UA	G1525	C1524		A766	(S11)	Type A
S4S	G1405	U1406	G1517	A1518	I–E	IV	GU/GA	C1496	U1495				Type B
Pseudo cis RZ													
pc1S	G112	G113	A353	G354	J–J	I	GG/AG	A315	G314	G113	C58		
Large subunit ribosomal RNA (23 S)													
Canonical RZ													
1L	U26	U27	A1318	G1319	I–J	I–II	UU/AG	A516	U517	G1339	U1338		
2L	A152	C153	A439	C440	J–J	I	AC/AC G185		G184		G41	L15e, L37e	
3L	C769	C770	A160	A161	S–I	I	CC/AA	G892	G891	U176	A174	L15e (L37e, L4)	
4L	A204	U205	A189	G190	J–J	I	AU/AG	A436	A437		A435	L15e	
5L	C208	G209	A665	A666	S–I	I–II	CG/AA	G231	C230		G680	L4	
6L	C376	C377	A242	A243	S–J	I	CC/AA	G274	G273	G269		L15e, L7ae	
7L	U325	G326	A305	A306	J–J	I	UG/AA	A340	C330	G345	G345	L4, L24	
8L	C637	C638	A520	A521	S–I	I–II	CC/AA	G1364	G1363	G23		L32e (L4, L22)	
9L	C1334	C1335	A551	A552	S–I	II	CC/AA	G1323	G1322			L32e	
10L	C1263	U1264	A565	A566	S–I	II	CU/AA		A1092		G592	L30	
11L	C783	A784	A1458	A1459	S–I	II–III	CA/AA	G863	U862		G1484	L2 (L37e)	

(*continued*)

Table 1 Continued

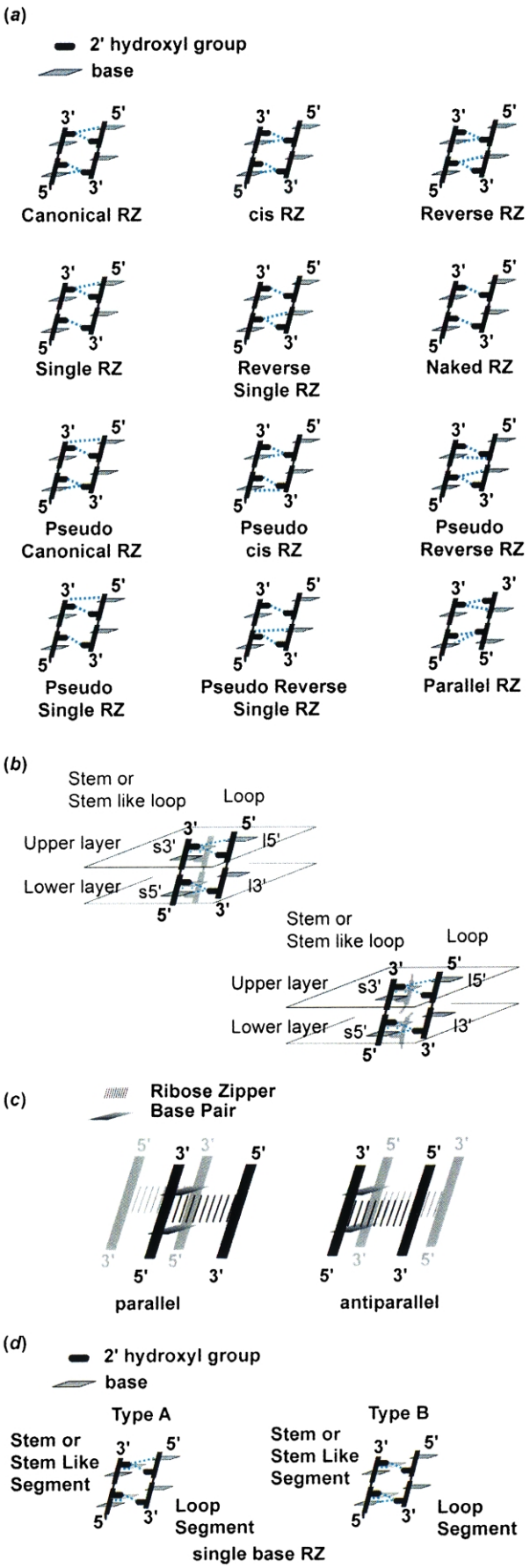
Entry[a]	Residue[b] s5′	s3′	15′	13′	SL[c]	Domain[d]	Type[e]	Base pair[f] S5′	S3′	15′	13′	Protein interaction[g]	Notice[h]
12L	A790	A791	C1708	G1709	E–J	II–III	AA/CG	G824	U823			L19e	
13L	C905	C906	A1329	A1330	S–E	II	CC/AA	G1300	G1299		G1327	L32e, L15	
14L	A1294	G1295	A1040	U1041	J–J	II	AG/AU	G911	C910	C930		L15	
15L	C1068	C1069	A1081	A1082	S–E	II	CC/AA	G1046	G1045	U626		L30	
16L	C2077	U2078	A1078	A1079	S–E	II–IV	CU/AA	G2068	A2067				
17L	C1084	C1085	A1097	A1098	S–I	II	CC/AA	G1075	G1074		G1258	L30 (L32e)	
18L	C1256	C1257	A1106	A1107	S–I	II	CC/AA	G1100	G1099	C1105		L13	
19L	C1208	C1209	A1188	A1189	S–E	II	CC/AA	G1159	G1158				
20L	C1467	G1468	A1865	A1866	S–E	III	GC/AA	G1475, A1476	C1474		G1863	L15e (L2)	
21L	C1513	C1514	A1492	A1493	J–J	III	CC/AA	G1672	G1449			(L23)	
22L	C1553	U1554	A1631	A1632	S–E	III	CU/AA	G1568	A1567		G1629		
23L	C1643	C1644	A1615	A1616	S–I	III	CC/AA	G1542	G1541		C1579	(L19e)	
24L	C1853	C1854	A1858	A1859	S–I	IV	CC/AA	G1878	G1877	U1871	C1870	L2	
25L	C2594	U2595	C1993	A1994	S–E	IV–V	CU/CA	G2584	A2583			L14	
26L	C2114	U2115	G2632	A2633	S–E	V	CU/GA	G2471	A2470		G2630	L2	
27L	C2555	C2556	A2599	A2600	S–J	V	CC/AA	G2580, A2577	G2579,	U2554		L14	
28L	G2574	C2575	A2775	A2776	S–I	V–VI	GC/AA	C2559	G2558		G2798	(L6)	
cis RZ													
c1L	C1456	U1457	A1657	A1658	I–I	III	CU/AA	G1489	A1485	C1534	C1533		
Single RZ													
s1L	U233	A234	A436	A437	J–J	I	UA/AA	G206		A204	U205	(L15e)	Type A
s2L	A773	C774	G471	A472	S–E	I–II	AC/GA	U888	G887		G469	L37e (L15e, L39e)	Type A
s3L	G1110	U1111	G579	A580	S–E	II	GU/GA	C1253	A1252		G577		Type A
s4L	C729	G730	G697	A698	S–J	II	CG/GA	G742	C741			L15	Type B
s5L	C2405	U2406	A736	A737	S–E	II–V	CU/AA	A2382	G2405			L21e (L44e)	Type A
s6L	U2297	C2298	A1006	A1007	S–I	II–V	UC/AA	G2310	A2311			L21e, L10e	Type A
s7L	G2491	U2492	A1057	A1058	I–E	II–V	GU/AA	C2530	G2529		G1055	L10e	Type A
s8L	C1786	C1787	G2882	A2883	S–I	IV–VI	CC/GA	G1806	G1805	A2875	A2874	L19e	Type B
s9L	C1892	C1893	A1968	A1969	S–J	IV	CC/AA	G1945	G1944			(L2)	Type A
s10L	C2126	U2127	A1930	A1931	S–I	IV–V	CU/AA	G2267	A2266		G1908	L2 (L15e)	Type A
s11L	G2558	C2559	G2798	A2799	S–I	V–VI	GC/GA	C2575	G2574	A2776	U2774	L6	Type B
s12L	C2822	G2823	G2826	A2827	J–J	VI	CG/GA	G2667	A2914	A2914	A2913	L3	Type B
Reverse single RZ													
rs1L	U1062	G1063	U2306	A2307	S–E	II–V	UG/UA	G1052	C1051	A2301	A2300	L21e	Type B
Naked RZ													
n1L	C930	C931	G911	A912	J–J	II	CC/GA	A1040	G1039	A1294	U1293	(L15, L32e)	
n2L	G2862	G2863	G2755	U2756	S–I	VI	GG/GU	C2894	C2893	A2727	A2896	L3, L24e (L31e)	

(*continued*)

Table 1 Continued

Entry[a]	Residue[b] s5′	s3′	15′	13′	SL[c]	Domain[d]	Type[e]	Base pair[f] S5′	S3′	15′	13′	Protein interaction[g]	Notice[h]
Inter Interaction between 23 S and 5 S													
Canonical Ribose Zipper													
Inter_1L	A955	G956	A80	C81	I–I	II-5S	AG/AC	A1012	C1011	G102	G101		
Naked ribose zipper													
Inter_n1L	A1012	A1013	G102	A103	I-I	II-5S	AA/GA	A955	U954	A80	U79		

[a] Assigned name for each ribose zipper.
[b] Residue types and numbers found in ribose zippers at the corresponding locations of s5′, s3′ 15′, and 13′ (Figure 2(b)). The residue number of each RNA residue is the same as those in the corresponding PDB file. The sequence of *T. thermophilus* ribosomal RNA in the original PDB file are numbered according to those of *E. coli*, however, we treated residues 1168 and 1169 of **11S** in the original PDB file (1FJF) as 1169 and 1170, respectively, in the *E. coli* sequence to adjust a gap between these sequence numbers. For the interchain interaction between 23 S and 5 S rRNA, the first two residues are in 23 S rRNA and the second two are in 5 S rRNA.
[c] Interacting secondary structure elements mediated by the ribose zipper. In this column, S represent stem, I represents internal loop, E represents external loop, and J represents junction loop.
[d] Domains where the ribose zipper is located are presented. If the ribose zipper mediates an inter-domain interaction, the two domains are connected by a horizontal bar in the column.
[e] The first two letters indicate the residue type in the 5′ to 3′ direction of the stem or stem-like side (residues at 5′ to 3′ in Figure 2(b), respectively) and the residue types following the slash refer to the 5′ to 3′ sequence on the loop-side (residues at 15′ and 13′ in Figure 2(b), respectively). For the interchain interaction between 23 S and 5 S rRNA, residues to the left of the slash are found in 23 S rRNA and to the right in 5 S rRNA.
[f] Base pair residues for each position s3′, s5′, 15′, and 13′ are shown. Underlined residues form non-canonical base pairs.
[g] Proteins, which interact with ribose zipper residues and with base paired residues of the ribose zipper (with parentheses), are listed. An underlined protein bridges a stem or stem-like segment and a loop segment of the ribose zipper.
[h] Additional information for each ribose zipper.



and large ribosomal RNAs is available as supplementary information (Supplementary Table 1).

Out of the 20 RZs in 16 S rRNA and 52 (44 in *H. marismortui* and eight additional in *D. radiodurans*) RZs in 23 S rRNA, 15 (75.0%) and 34 (65.4%), respectively, are found in intradomain interactions. Therefore, ribose zippers primarily mediate tertiary interactions between segments within the same domain. However, the interdomain interaction ratios are high in domain IV of 16 S rRNA as well as domains V and VI of 23 S rRNA. In domain V of 23 S rRNA, there are a large number of interdomain interactions with domain II (6) relative to the total number of RZs (10) in the domain. A summary of the RZs within and between each domain of the small and large ribosomal subunits is given as supplementary information (Supplementary Table 2).

We describe RZs using the terminology shown in Figure 2(b) where s5′ and s3′ correspond to the 5′ and 3′ end residues of the stem or stem-like side (base-paired) and 15′ and 13′ correspond to the 5′ and 3′ end residues of the loop-side. The upper layer and lower layer are defined so that residues in the s3′ and 15′ positions belong to the upper layer while those in the s5′ and 13′ positions belong to the lower layer (Figure 2(b)).

Figure 4(a) shows the definition of the pseudotorsion angles (θ and η)[12] in the canonical RZ (**8L**). Figure 4(b) shows the θ–η plot for s5′, s3′, 15′, and 13′ as rectangles, circles, crosses and triangles, respectively, in canonical RZs mediating stem–loop (black) and loop–loop (red) interactions. The gray-colored vertical and horizontal areas correspond to the distribution of either θ or η for these nucleotides and the intersection of these areas corresponds to the pseudotorsion angles of the typical helical structure.[12] For the stem–loop and loop–loop cases, θ of the 5′ residues and η of the 3′residues are confined to helical values. On the other hand, pseudotorsion angles for flexible linkers between the RZ and adjacent regions (θ of the 3′ residues and η of the 5′ residues) show a distribution of values. Except for the pseudo cis RZ, pseudotorsion angles for non-canonical RZs are distributed in the same way as in the canonical RZs (plots in Supplementary Figure 2).

Figure 2. (a) Schematic representation of the 12 proposed classes of ribose zippers (RZ). Light blue colored broken lines represent hydrogen bonds. (b) Schematic representation and nomenclature of the s5′, s3′ (stem-side) and 13′, 15′ (loop-side) positions and for the upper and lower layers of ribose zippers. (c) Schematic diagram of parallel and antiparallel double RZs. (d) Schematic representations of Type A and Type B single base ribose zippers are shown. Type A single base RZs have a base–backbone hydrogen bond in the upper layer and Type B single base RZs have a base–backbone hydrogen bond in the lower layer.

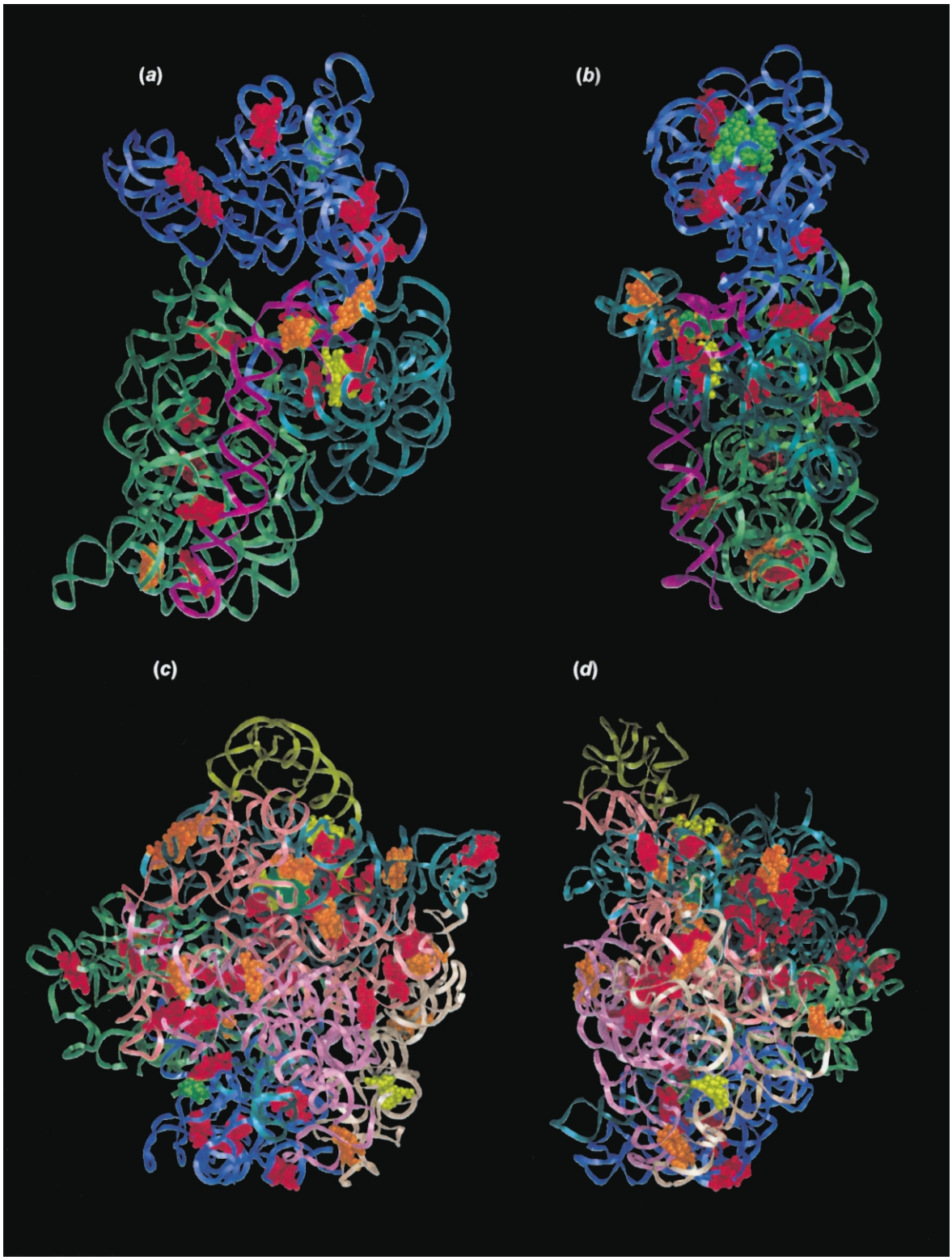


Figure 3. The atoms included in ribose zipper residues are drawn as colored spheres. The canonical RZs are red, the cis RZs are yellow green, the pseudo cis RZ is firebrick red, the single base RZs are orange, the cis single base RZ is green, and the naked RZs are yellow. (a) and (b) Ribbon drawings of *T. thermophilus* small ribosomal subunit RNA (16 S rRNA) (a) as viewed into the face interacting with 23 S rRNA and (b) rotated by 90° about the vertical axis. Each colored region represents an rRNA domain (domain I is lime, domain II is teal, domain III is slate, and domain IV is pink). (c) and (d) Ribbon drawings of *H. marismortui* large subunit ribosomal RNA (23 S and 5 S rRNA) (c) as viewed into the face that interacts with 16 S rRNA and (d) rotated by 90° about the vertical axis. The color of the ribbon represents the 23 S domains and 5 S rRNA (domain I: lime, domain II: teal, domain III: slate, domain IV: pink, domain V: salmon, domain VI: wheat, and 5 S rRNA: olive).

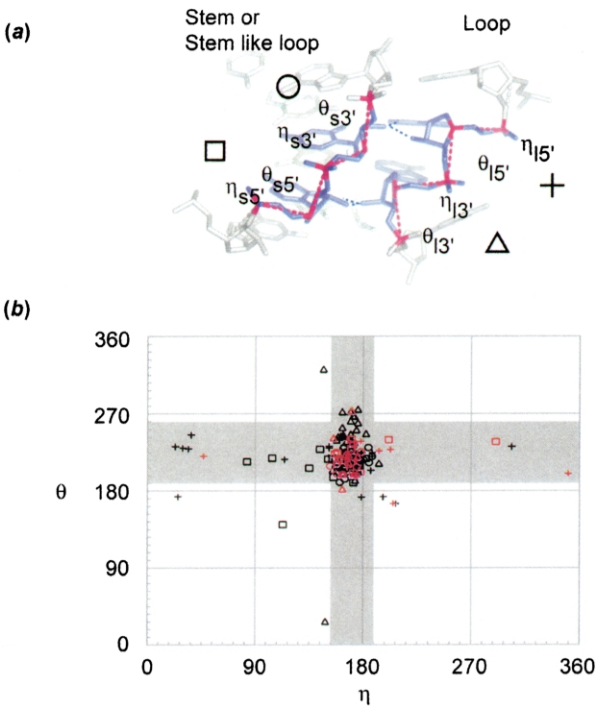


Figure 4. (a) Pseudotorsion angles in the canonical RZ **8L**, where the pseudo-bonds are represented by magenta broken lines and corresponding C4′ and P atoms are colored magenta and other atoms of **8L** are colored slate. Rectangles, circles, crosses, and triangles corresponding to residues in the s5′, s3′, 15′, and 13′ positions are shown in θ–η plots. (b) θ–η plots of canonical RZs mediating stem–loop (black) and loop–loop (red) interactions in 16 S and 23 S rRNA. Gray colored vertical and horizontal areas correspond to the helical values of either θ or η for these nucleotides. Rectangles, circles, crosses, and triangles correspond to the s5′, s3′, 15′, and 13′ residues, respectively.

The canonical ribose zipper

There are a total of 40 canonical RZs in the 16 S and 23 S ribosomal RNAs, 30 (75.5%) of which are involved in stem–loop interactions. There are ten loop–loop interactions observed, however, in each of these, residues from one of the loop segments participate in non-canonical base-pairing. This loop segment is therefore referred to as the "stem-like" side.

Sequence specificity of canonical ribose zippers in ribosomal RNA

Table 2 shows the residue identities for the canonical ribose zippers found in 16 S and 23 S rRNA, respectively. The left side corresponds to the residues on the stem or stem-like side and the right side corresponds to the residues in a loop. An asterisk (∗) indicates that any residue (A, U, G, or C) can occupy that position. The order of the residues on both sides of the RZ is 5′ to 3′ separated by a slash, where the interacting residues are between the 5′ end and 3′ end residues (i.e. s5′ s3′/15′13′, Figure 2(b)). For the canonical RZ, we observe that 20 of the 40 examples obey the sequence pattern CC/AA.

Structural basis of sequence specificity

Figure 5(a) shows a typical example of a CC/AA canonical RZ mediating a stem–loop interaction in 23 S rRNA (**6L**). Figure 5(b) is a view of **6L** from above the upper layer of the RZ and Figure 5(c) and (d) shows the separate lower and upper layers of **6L**, respectively. We find a base triple between the CG base-pair in the stem segment with an adenosine in the loop segment of the lower layer in almost all CC/AA type RZs, in which there are hydrogen bonds between N3 of A and N2 of G and between N1 of A and ribose O2′ of G. This type of base triple has been previously described as a type I A-minor motif.[5]

In the upper layer of nine of the 13 CC/AA canonical RZs in *H. marismortui* 23 S rRNA (69.2%), we find that there is a water molecule in the minor groove of the CG base-pair (i.e. **6L** in Figure 5(d)) and this water makes hydrogen bonds bridging the N1 of A and the N2 and N3 of G, thus forming a water-mediated base triple for the nine RZs. This type of triple (without bridging water molecules indicated) was previously described as a type II A-minor motif.[5]

The ∗∗/∗A sequence pattern

There are a total of 34 ∗∗/∗A pattern canonical RZs (85.0% of the total) in 16 S and 23 S rRNAs.) All base types are observed in the s5′ position of canonical RZs such as C∗/∗A, U∗/∗A, G∗/∗A, and A∗/∗A. In C∗/∗A and G∗/∗A pattern canonical RZs, we observe a base triple in the lower layer except for **19L**. However, in A∗/∗A and U∗/∗A patterns, the adenosine packs tightly in the minor groove without formation of hydrogen bonds. These observations may reflect the order of energetic preference for the interaction between an adenosine in the minor groove of each Watson–Crick base-pair as a type I A-minor motif: CG > GC > UA, AU.[9]

There are 31 (of 40) ∗∗/AA patterns in small and large subunit ribosomal RNAs, comprising 77.5% of all canonical RZs found. In all of these, the upper adenosine forms a type II A-minor motif where the adenosine interacts most frequently with a CG Watson–Crick base-pair (22 cases), forming a water-mediated base triple. No type I A-minor motifs are observed in the upper layer. However, in other Watson–Crick base-pairs (∗U/AA, ∗A/AA, ∗G/AA), base triples and water-mediated base triples are observed (Supplementary Figure 1). Thus, the flexibility in formation of water-mediated base triples explains why there is less sequence specificity observed and less energetic differences between sequences forming the type II A-minor motif.[9] There are three other ∗∗/∗A pattern RZs (Supplementary Figure 1). Interestingly, in the ∗U/GA type upper layer

Table 2. Residue types observed in canonical ribose zippers of 16 S and 23 S rRNA according to their secondary structure interactions

		Residue Type[a]													
Interaction[b]	Total[c]	**/*A, **/AA CC/AA	**/*A, **/AA CU/AA	**/*A, **/AA CG/AA	**/*A, **/AA CA/AA	**/*A, **/AA UG/AA	**/*A, **/AA (GC/AA)	**/*A CU/GA	**/*A CU/CA	**/*A AU/GA	**/A* UU/AG	**/A* AU/AG	**/A* AC/AC	**/A* AG/AU	AA/** AA/CG
S–S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S–I	13	8	1	1	1	0	2	0	0	0	0	0	0	0	0
S–E	12	7	2	1	0	0	0	1	1	0	0	0	0	0	0
S–J	5	4	1	0	0	0	0	0	0	0	0	0	0	0	0
I–I	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
E–E	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
J–J	6	1	0	0	0	1	0	0	0	0	0	2	1	1	0
I–E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I–J	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
E–J	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Sum[d]	40	20	4	3	1	1	2	1	1	1	1	2	1	1	1

[a] The first two letters indicate the residue type in the 5′ to 3′ direction of the stem or stem-like side (residues at 5′ to 3′ in Fig. 2(b), respectively) and the residue types following the slash refer to the 5′ to 3′ sequence on the loop side (residues at 15′ and 13′ in Fig. 2(b), respectively). An asterisk (∗) indicates that any residue (A, U, G, or C) can occupy that position.
[b] Interacting secondary structural elements: S represents stem, I represents internal loop, E represents external loop, and J represents junction loop.
[c] Total number of ribose zippers mediating the corresponding secondary structural elements.
[d] Sum of the corresponding columns.

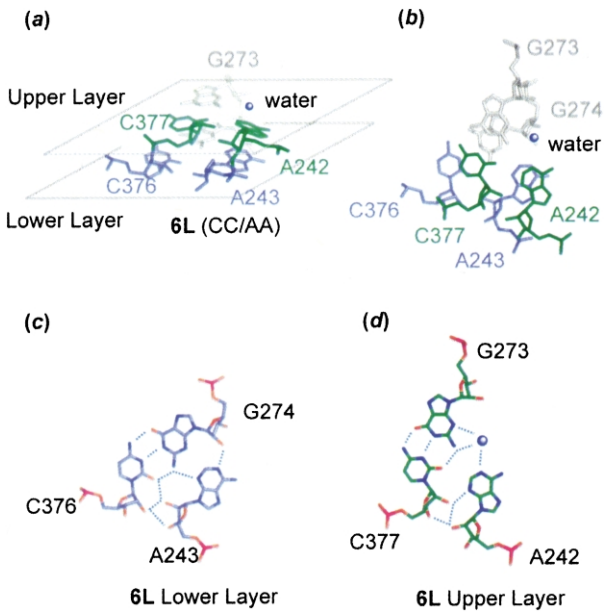


Figure 5. Stick diagrams for CC/AA canonical RZs. (a) and (b) The canonical RZ **6L** where C377 and A242 belong to the upper layer (blue green) and C376 and A243 belong to the bottom layer (slate) (a) viewed perpendicular to the backbone, showing the ribose–ribose interactions and (b) from above the upper layer. The residues on the left side are in a stem region and the residues on the right side are in a loop region. (c) Hydrogen bond networks in the lower layer of **6L**. Light blue broken lines represent the hydrogen bonds. A type I A-minor motif, in which there are hydrogen bonds between N3 of A243 and N2 of G274 and between N1 of A243 and ribose 2′-hydroxy group of G274, is observed. Such a type I A-minor motif is the common feature at the lower layer for the C*/*A type RZ. (d) Hydrogen bond networks in the upper layer of **6L**. There are water-mediated hydrogen bonds between N1 of A242 and N2 of G273. A242 interacts with the minor groove of the GC Watson–Crick pair as in a type II A-minor motif.

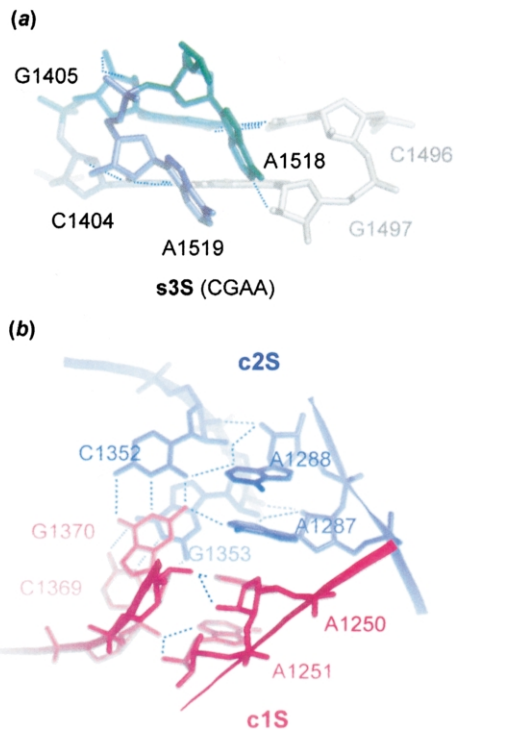


Figure 6. (a) Stick diagrams showing the cis RZ **c3S** from the minor groove side of the Watson–Crick base-pairs. The residues of the upper layer are blue green and those of the lower layer are slate. A1518–A1519 base stacking is inclined by over 45° to the C1404–G1405 base stacking direction. The base–backbone interaction in the upper layer is formed between the 3′ stem-side residue (G1405) and the 5′ loop-side residue A1518 (s3′–15′) instead of the 5′ loop and 3′ stem-side (s5′–13′) residues observed in canonical RZs. (b) The antiparallel double RZ formed by **c1S** (magenta) and **c2S** (blue) as viewed from the **s1S** side. Light blue broken lines represent hydrogen bonds.

pattern, there is also a direct hydrogen bond between N6G and O2U and this pattern is energetically favorable.[9]

There are only six canonical RZs that do not follow the **/*A sequence pattern. Except for **12L**, their sequence follows the **/A* pattern where the 15′ adenosine interacts with the minor groove of another side in a Type II A-minor motif. **12L** is the only example of all canonical RZs that does not have any adenosine on the loop-side (AA/CG). These six canonical RZs form a Watson–Crick or non-Watson–Crick base-pair with another residue (except for **12L**) and thus have a unique structural arrangement.

The cis-ribose zipper

We find three (all CG/AA) cis-RZs in 16 S rRNA mediating stem–loop interactions and one cis-RZ (CC/AA) in 23 S rRNA mediating a loop–loop interaction. Figure 6(a) shows a cartoon of the cis-ribose zipper **c3S**. The loop-side base stacking (A1518-A1519) is inclined at about 45° to the stem-side base stacking direction even though there are adenosines on the loop-side.

For the helical base-pairs C1352-G1370 and G1353-C1369, each side of the helix forms a cis-RZ with opposite sides of the same internal loop (**c1S**, **c2S**, Figure 6(b)) resulting in an antiparallel double ribose zipper that mediates the same stem–loop interaction. The loop-side bases of **c1S** and **c2S** are stacked on each other with the direction of this stacking at almost right angles to that of the bases on the stem-sides making it possible to form this packed structure.

The single base ribose zipper

There are a total of 16 single base ribose zippers in 16 S and 23 S rRNA and 11 of these (68.8%) are stem–loop interactions. In each of the four loop–loop interactions one of the loop segments is base-paired and forms a stem-like structure. However in **s1L**, these pairings are non-Watson–Crick and

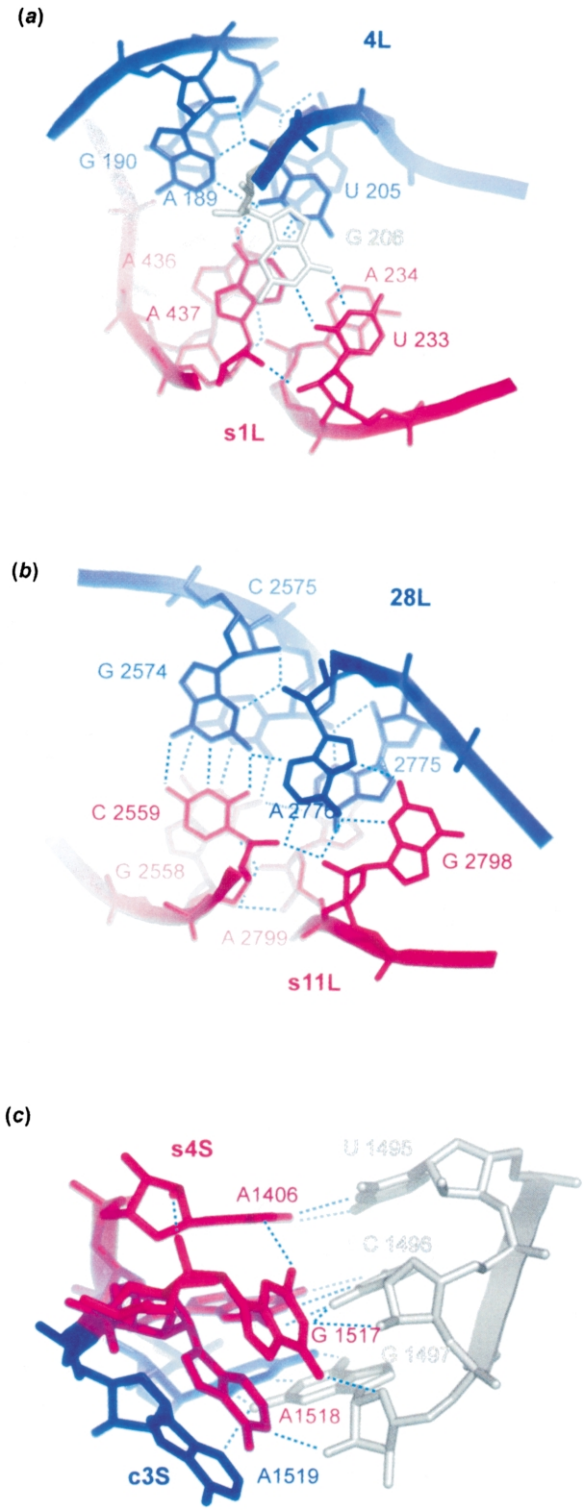


Figure 7. (a) Parallel double RZ formed by a single base RZ **s1L** (magenta) and a canonical RZ **4L** (blue) viewed from above the upper layer of **s1L**. (b) Antiparallel double RZ formed by a single base RZ **s11L** (magenta) and a canonical RZ **28L** (blue) viewed from above the upper layer of **s11L**. (c) Continuous RZ formed by a single base RZ **s4S** (magenta and violet) and a cis RZ **c3S** (violet and blue), where each RZ shares violet colored residues (G1405 and A1518) viewed from the minor groove side.

oriented in a parallel manner. Thus, we treated the paired AA residues as the loop-side.

Sequence specificity of single base ribose zippers in ribosomal RNA

In contrast to the canonical RZs, there is no observed preference in single base RZs for CC/AA in either 16 S or 23 S rRNA. Residue identities for the single base ribose zippers found in 16 S and 23 S rRNA are given as supporting information (Supplementary Table 3).

There are two types of single base RZs: one with a base–backbone hydrogen bond in the upper layer (Type A) and the other with a base–backbone hydrogen bond in the lower layer (Type B) (Figure 2(d)). Ten Type A and five Type B (see Table 1) examples of single base RZs are observed. Except for **s2S**, all residue patterns are of the $^{*}{}^{*}/{}^{*}$A type, where the adenosine in the lower layer interacts with the minor groove through a type I A-minor motif.

In the Type A single base RZ, the 5′ stem base (s5′) of the lower layer is rotated so that the hydrogen bond between O2′ and the base cannot be formed. Even though this base–backbone hydrogen bond is missing in the lower layer, base triples or water-mediated base triples still form an A-minor motif in most cases. A unique case is **s1L** (Figure 7(a)) where non-Watson–Crick base-paired residues of the "loop" segment interact with a fourth segment of the same junction loop by a canonical RZ **4L** to form a parallel type double ribose zipper motif (Figure 2(c)). The residues base-paired to the stem-side of **s3S** form a canonical RZ (**5 S**) with residues from the same junction loop as the loop-side residues of **s3S**, resulting in an antiparallel double ribose zipper.

In the Type B single base RZs, the loop-side always follows the $^{*}{}^{*}$/GA sequence pattern (see Table 1). An interesting antiparallel double RZ is formed when A2776, which is a part of RZ **28L**, is inserted into **s11L**, and these two RZs mediate a common stem–loop interaction (Figure 7(b)). We observed only one RZ example of three consecutive O2′–O2′ hydrogen bonds. We classify this interaction as an "overlapping" double RZ, where the shared nucleotide G1405 serves as the s5′ and s3′ residue for **s4S** and **3S**, respectively, and the shared nucleotide A1518 serves as the 13′ and 15′ residue for **s4S** and **3S**, respectively (Figure 7(c)). In this case, the stacking direction of the loop-side is also inclined at approximately 45° to the stem-side base stacking direction. This overlapping RZ is located at the A-site of 16 S rRNA. Single mutations of the residues of the RZs from the Ribosomal RNA Mutation Database† are summarized in supplementary information. Aminoglycoside antibiotics such as paromomycin, neomycin and

† http://ribosome.fandm.edu

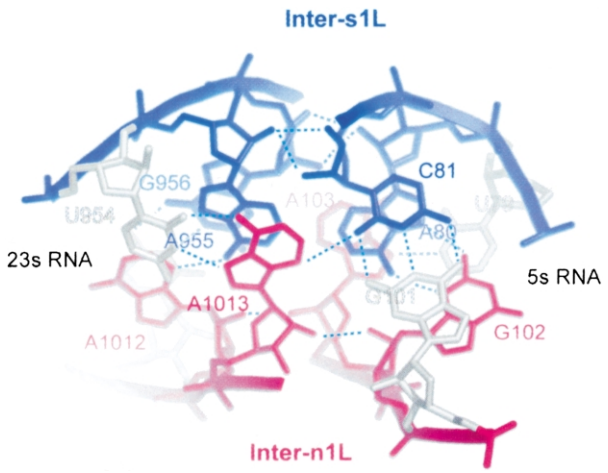


Figure 8. The antiparallel double RZ, which connects an internal loop of 23 S and loop E of 5 S rRNA, by an interchain naked RZ **Inter_n1L** (magenta) and an interchain single base RZ **Inter_1L** (blue). The left side segments belong to 23 S and the right side segments belong to 5 S rRNA. The right segment of each RZ belongs to 5 S rRNA (loop E) and the left segment of each RZ belongs to 23 S rRNA.

gentamicin target this highly conserved decoding region.[13,14]

The reverse single ribose zipper

We have identified only one member of this class, **rs1L**, which mediates a stem–loop interaction. Hydrogen bonds between the base and backbone occur only in the lower layer (Type B), while in the upper layer, there are multiple base interactions.

The naked ribose zipper

There are two naked ribose zippers that meditate stem–loop interactions. Residues of the loop-side segments also form Watson–Crick or non-Watson–Crick base-pairs, however, the direction of loop-side base-pair stacking is almost perpendicular to that of stem-side base stacking.

The RZ **n1L** is involved in a complex with the canonical ribose zipper **14L** forming an antiparallel double ribose zipper where one non-canonical base-pair (C931-A1294) is shared by the RZs. The absence of the structural constraints of base–base and base–backbone interactions in naked ribose zippers allows formation of these complex double RZ structures.

The pseudo cis ribose zipper

We find only one instance (**pc1S**) of this class in Domain I of the 16 S rRNA (Table 1 and Supplementary Figure 4). This zipper mediates the interaction between residues in the same junction loop. The pseudotorsion angles of the stem-side residues are distributed in the helical structure region. However, those of loop-side residues are not in the helical region (Supplementary Figure 2).

Interchain interaction

In the large subunit ribosomal RNAs, we find a canonical RZ, **Inter_1L** (AG/AC), and a naked ribose zipper, **Inter_n1L** (AA/CA), between an internal loop of the 23 S chain and the internal loop E of 5 S rRNA, forming an antiparallel double ribose zipper. The non-Watson–Crick paired bases A80 and G102 of 5 S rRNA loop E are part of **Inter_1L** and **Inter_n1L**, respectively, and the non-Watson–Crick paired bases A955 and A1012 of 23 S rRNA (stem 38) are also part of **Inter_1L** and **Inter_n1L**, respectively. Thus, these two RZ complexes mediate the interaction between the two stem-like structures: 5 S rRNA loop E and 23 S rRNA stem 38 (Figure 8(a)).

Phylogenetic conservation in 16 S rRNA

Figure 9 shows sequence logos of 16 S ribosomal RNA using all aligned prokaryotic sequences (16,277) from positions 323 to 346 and 1422 to 1445 for the canonical RZ **3S** and from 389 to 412 and 610 to 633 for the canonical RZ **4S**. The sequence logo bit scores for canonical RZs in 16 S rRNA using all prokaryotic sequences are much higher than the average value of 1.33 over all the residues except for RZs **2S**, **5S**, **10S**, and **12S** (summarized in Supplementary Information, Table 4). These results suggest that the sequences of RZs are strongly conserved under evolutionary pressure to preserve the RZ tertiary interactions. The average bit scores for sequences on the loop-side are slightly higher than those of the stem or stem-like side. This phenomenon suggests that while the stem may co-vary in sequence, the loop residues are subject to additional evolutionary influences beyond formation of a canonical RZ triple. This may be due to the flexibility and different types of RZs that can be formed using structural water molecules.

Figure 10(a)–(c) shows sequence logos corresponding to **12S** from 1316 to 1335 and 1259 to 1278 using (a) all prokaryotic sequences, (b) only the archaeal sequences, and (c) only bacterial sequences. Among archaeal sequences, CC/AA is the major pattern for the **12S** sequence position. The bit scores of the residues are 1.83, 1.86, 1.42, and 1.98, respectively, which are higher than the average value of the total sequence in archaea (1.29). This shows that the CC/AA sequence of **12S** is highly conserved within archaea. On the other hand, CU/GA is the major pattern among the bacterial sequences (Figure 11(c)) and the bit scores are (stem 3′, 5′, loop 3′, 5′) 0.92, 1.40, 1.40, and 1.97, respectively. Since there is a suitable hydrogen bond network for both C–A and U–G interactions in the upper layer (Figure 5(d) and (f), respectively) of CC/AA and CU/GA sequence

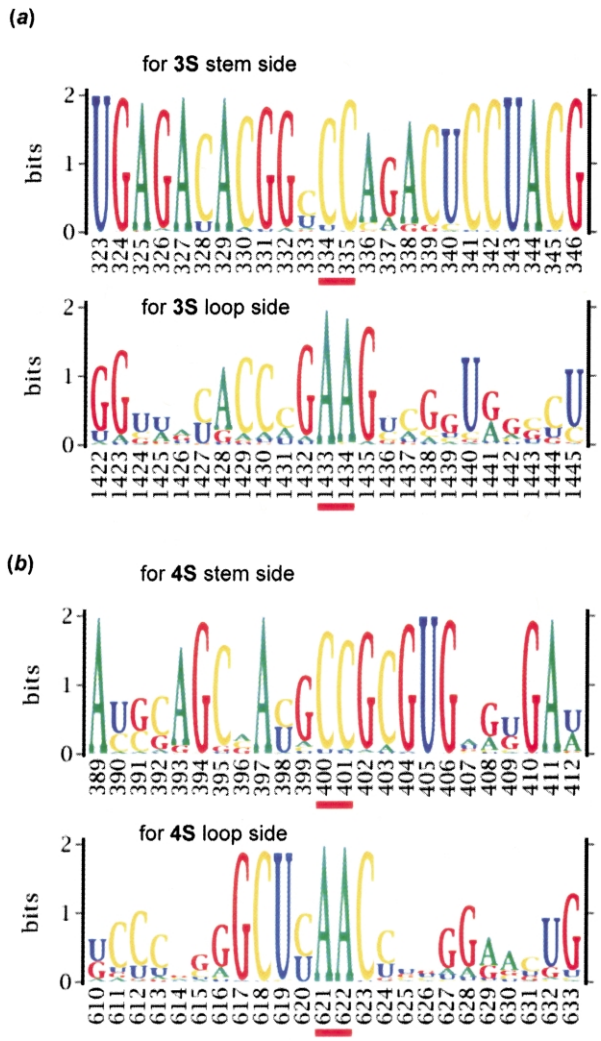


Figure 9. Sequence logos (a) for the **3S** RZ from residues 323 to 346 and 1422 to 1455 and (b) for the **4S** canonical RZ from 389 to 412 and 610 to 633 of *T. thermophilus* small ribosomal RNA. The sequence numbers referred to are those of *E. coli*. A higher bit score (information content) indicates a greater degree of sequence conservation at that position. A bit score of 2.0 is the maximum, meaning the residue is completely conserved. The bit scores of residues in these RZs have high values even if those of neighboring sequences are low. These sequence logos were constructed by using 16,277 prokaryotic aligned sequences (1173 archaeal and 15,104 bacterial 16 S rRNA sequences) from the Ribosomal Database Project Release 8.1. The residues of the RZ are underlined in bold red.

patterns, respectively, the difference of the major sequence patterns in the upper layer of the **12S** RZ between archaea (C–A) and bacteria (U–G) (Figure 10(b) and (c)) may be the result of co-evolution of the RZ residues even though these residues are distant from each other in the primary and secondary structure. Consequently, this means that the RZ mediating a tertiary interaction is conserved between kingdoms through covariation. However, the bit score of the 3′ stem residue of **12S** within bacterial sequences is still small. Table 3 shows the total bit scores and the individual contributions (information content) of each base, the number of examples, and the percentages of the 3′ stem residue of **12S** (1326) when the sequence logo is built under conditions where the corresponding loop-side (loop 5′, 1268) is G, A or any residue. Within sequences where residue 1268 is G, the information content and the percentage of U (0.973, 82.06%) is higher and the information content and percentage of C (0.169, 14.21%) is lower than those of the total bacterial sequences (0.606, 69.97% and 0.214, 24.67%, respectively). On the other hand, within sequences where residue 1268 is A, the contribution content and percentage of U (0.219, 32.46%) is lower and the contribution content and percentage of C (0.404, 59.95%) is higher than that of the total bacterial sequence set.

The average values of each position for cis RZs, pseudo cis RZs, and naked RZs are also much higher than 1.33 (average value of all 16 S residues) (Supplementary Table 4). In the single ribose zipper, the loop-side residues are highly conserved, but the stem-side residues are more variable. This was also observed for the canonical RZs as discussed above.

Protein interactions with ribosomal ribose zippers

Out of a total of 66 ribose zippers in the small ribosomal subunit of *T. thermophilus*, and the large ribosomal subunit of *H. marismortui*, 43 ribose zippers (65.2%) interact with ribosomal proteins. This is especially true for canonical RZs, where 30 (75.0%) of the 40 form hydrogen bonds between the RNA backbone atoms and residues of a neighboring protein or several proteins (see Table 1). Arginine and lysine are the most common protein residues for hydrogen bonding to the ribose zipper backbone, thus providing charge neutralization. As judged from the structure of the large ribosomal subunit, water-mediated RNA–protein hydrogen bonds are much more frequently observed than direct hydrogen bonds. There are also a few cases in which nucleotide base atoms are used for hydrogen bonding with protein.

In several *T. thermophilus* small subunit ribosomal proteins, there are extended regions involved in ribosomal RNA interaction.[15] Canonical ribose zippers **1S**, **4S** and **6S** interact with extended regions of *T. thermophilus* small subunit ribosomal proteins S5, S4 and S12, respectively (Table 1), and are clustered near the tRNA and mRNA binding sites.[15] In addition to a role in the functional sites of the small subunit, protein–RZ interactions may be important in the ribosomal assembly process. Three of the six proteins identified as primary binders[16] (able to bind naked 16 S rRNA), S 4, S 7 and S 20, also contact ribose zippers (Table 1) and may initiate folding and compaction of the RNA by stabilization of these long-range tertiary interactions. S 4 and S 7 specifically have

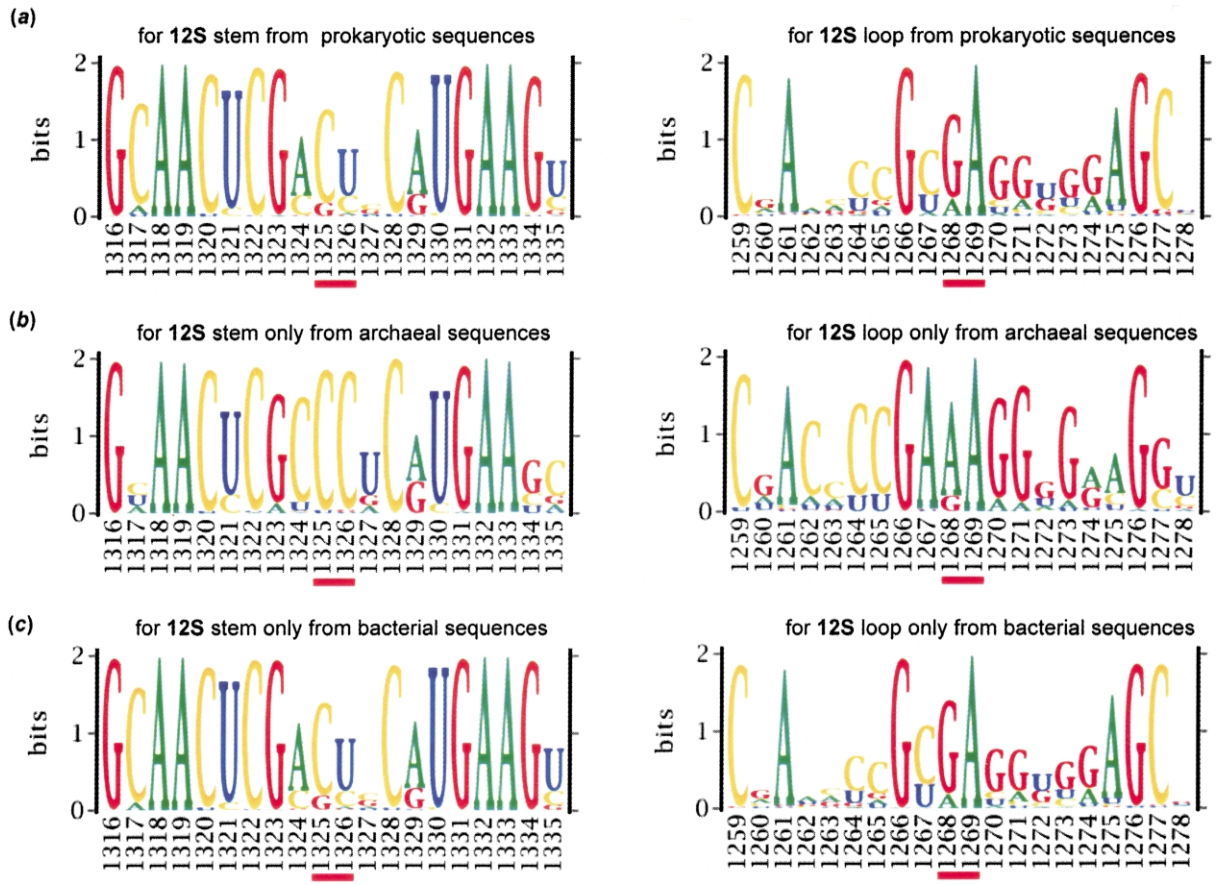


Figure 10. The sequence logos for 12S from 1316 to 1335 and from 1259 to 1278 calculated using (a) all prokaryotic sequences (16,277) in the Ribosomal Database Project Release 8.1, (b) only archaeal sequences (1173) and (c) only bacterial sequences (15,104). Residues corresponding to the RZ are underlined in bold red.

been proposed to nucleate assembly of the body and head of the 30 S subunit.[17]

In canonical RZs of the large subunit ribosomal RNA, protein bridging between the stem or stem-like side and loop-side is frequently observed (14 cases, 50.0%). However, this kind of protein bridging is uncommon in non-canonical RZs. Figure 11(a) shows **3L** and neighboring proteins where L15e (green) interacts with the residues of L3 (magenta) and L4 (brown) and L37e (yellow) interact with the base-paired residues of the **3L** stem-side. Interestingly, in order to interact with RZ, the extended residues of the L15e and L4 ribosomal proteins reach deep inside the folded RNA from the RNA surface where the main body of these proteins are located. L3 (and L24) have been identified as "initiator" proteins for 50 S ribosomal subunit assembly.[18]

Figure 11(b) shows the hydrogen bonds between **3L** residues and an L15e residue, which bridges the RZ stem-side and loop-side. This kind of interaction between protein residues buried deeply inside ribosomal RNA through an extended structure is observed for **8L**, **13L**, **14L**, and **26L** in

Table 3. Phylogenetic analysis of the **8S** ribose zipper residue 1326

Condition[a]	Total bit[b]	Sample number[c]	Bit[d]	A Number (%)[e]	Bit[d]	U Number (%)[e]	Bit[d]	G Number (%)[e]	Bit[d]	C Number (%)[e]
Any	0.865	12441	0.031	439 (3.53)	0.606	8705 (69.97)	0.016	228 (18.33)	0.214	3069 (24.67)
G	1.186	9904	0.044	367 (3.71)	0.973	8127 (82.06)	0.000	3 (0.03)	0.169	1407 (14.21)
A	0.674	1568	0.021	50 (3.19)	0.219	509 (32.46)	0.030	69 (4.40)	0.404	940 (59.95)

[a] Residue type at position 1268 for the sequences used to build the sequence logo.
[b] Bit score of residue 1326.
[c] Number of sample sequences used for the sequence logo under the corresponding condition.
[d] Individual contribution (information content) of each base to the total bit score.
[e] Number of each base type and its percentage over all sequences.

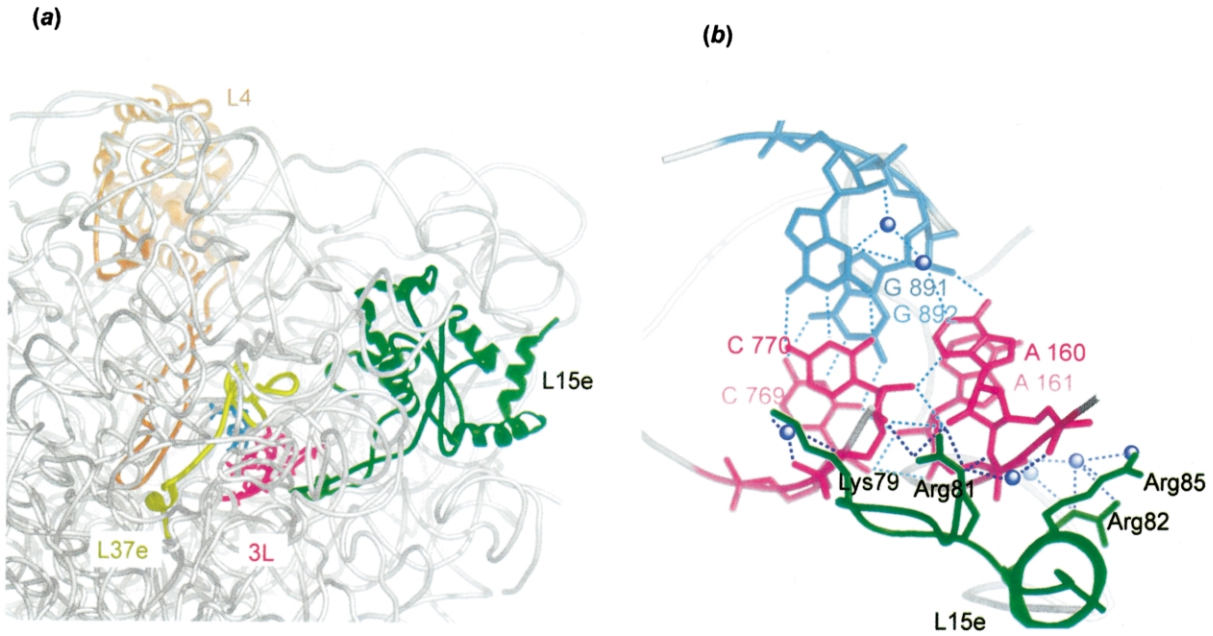


Figure 11. (a) Tube and ribbon drawings of *H. marismortui* 23 S rRNA around **3L**, where residues included in **3L** (magenta) and its base-pair residues on the stem-side (slate) are drawn as sticks, and its neighboring ribosomal proteins L15e, L37e, and L4, which interact with **3L** or its base-pair residues on the stem-side. (b) Stick and tube diagrams of **3L** and its neighboring residues of L15e, with only the residues used in hydrogen bonding drawn as sticks. Hydrogen bonds are shown as broken blue lines.

H. marismortui large ribosomal RNA. In all of these cases, the interacting protein bridges between the stem or stem-like side and the loop-side of these RZs. The interaction between these RZs and ribosomal proteins may be an important initial step in the ribosomal subunit folding process. For example, ribosomal protein L37e in the *H. marismortui* large ribosomal subunit is a small protein and completely buried in the 23 S rRNA structure. This protein interacts with **2L** and **s2L**, bridging the stem-side and loop-side, and the base-paired residues of **3L** and **11L** inside the large ribosomal subunit.

Ribose zippers in the large subunit ribosomal RNA of *D. radiodurans*

We find a total of 31 RZs in the *D. radiodurans* large ribosomal subunit RNA compared with 46 in the *H. marismortui* large ribosomal RNA. The gross topology and location of residues of almost all canonical RZs in *H. marismortui* are found in the *D. radiodurans*. Ten additional canonical RZs found in *H. marismortui* have very similar analogs in *D. radiodurans*, however, they are not classified as RZs here, since one, or both, backbone–backbone interactions are outside the H-bonding limit (**1L**, **2L**, **4L**, **7L**, **11L**, **12L**, **19L**, **15L**, **24L** and **inter_1L**) of 3.8 Å. A summary of ribose zippers in *D. radiodurans* is available as supplementary information (Supplementary Table 5). Exceptions are **9L** and **10L** where there is no corresponding structural element in the sequence of *D. radiodurans* and **6L** and **23L** for which there is no structural information about this sequence in the PDB file (1KPJ). Ten canonical RZs (**3L**, **8L**, **13L**, **16L**, **17L**, **18L**, **22L**, **25L**, **26L**, and **28L**) in *H. marismortui* are also found as canonical RZs in *D. radiodurans*. All of these canonical RZs mediate stem–loop interactions. Loop-side residues of **22L** are in an external loop at the end of stem 59 in *H. marismortui*, however, stem 59 of *D. radiodurans* is much longer than that of *H. marismortui* and the loop-side residues corresponding to **22L** are in the stem region near the joint of stem 59 to a parent junction loop. The canonical RZ corresponding to **5L** is shifted by one base compared to the *H. marismortui* secondary structure.[11] Two canonical RZs (**14L**, **27L**) in *H. marismortui* are changed to single-base RZs in *D. radiodurans*. The loop-side residues of **20L** and **21L** in *D. radiodurans* assume different conformations from these RZs in *H. marismortui*.

The residues of almost all non-canonical RZs in the *H. marismortui* large ribosomal RNA are conserved in *D. radiodurans*. There are only three exceptions: **s3L** has no corresponding structural element; corresponding residues of **s1L** are located nearby, but are disordered; and the residues corresponding to the loop-side of **s12L** are not continuous (with one turned out residue between them in *D. radiodurans*). Three single base RZs (**s8L**, **s9L**, and **s10L**) are replaced by canonical RZs in *D. radiodurans* and these RZs mediate stem–loop interactions. Two single base RZs (**s7L** and **s11L**) are conserved, two other non-canonical RZs (**c1L** and **n1L**) are changed to single base RZs, and a naked RZs **n2L** is conserved in *D. radiodurans*.

Table 4. List of sequences of new ribose zippers in *D. radiodurans* large ribosomal RNA

Entry[a]	Residue[b] s5′	s3′	15′	13′	SL[c]	Domain[d]	Residue in *Haloarcula marismortui*[e] s5′	s3′	15′	13′
Canonical RZ										
29L	C 113	C 114	A 124	A 124	S–E	I	C 110	C 111	A 120	U 121
30L	C2421	C2422	A 688	A 689	S–I	II–V	C2477	U2478	A766	A767
31L	C2494	G2495	A1139	A1140	S–I	II–V	U2550	C2551	A1232	A1233
Single RZ										
s13L	C1327	C1328	A1405	A1406	S–E	III	C1420	C1421	A1501	A1502
Pseudo cis RZ										
pc1L	G1476	C1477	A2681	C2682	E–E	III–VI	G1563	C1564	G2738	A2739
Pseudo single RZ										
ps1L	G1993	U1994	G1338	U1339	E–J	III–IV	G2051	U2052	C1431	U1432
ps2L	U1695	C1696	C1973	U1974	J–J	IV	A2030	C2031	G1756	U1757
Naked RZ										
n3L	G 171	A 172	G 227	A 228	S–E	I	A165	A165	G221	A222
n4L	G1652	C1653	A1751	U1752	E–J	III–IV	A1712	G1713	C1816	U1817

[a] Assigned names for each new ribose zipper in *D. radiodurans*.
[b] Residue types and numbers found in RZs at corresponding locations s5′, s3′, 15′, and 13′ (Fig. 2(b)).
[c] Interacting secondary structure elements mediated by the ribose zipper. In this column, S represent stem, I represents internal loop, E represents external loop, and J represents junction loop.
[d] Domains where the ribose zipper is located are presented. If the RZ mediates an inter-domain interaction; the two domains are connected by a horizontal bar in the column.
[e] Residue types and numbers in the *H. marismortui* sequence corresponding to the residues at s5′, s3′, 15′, and 13′ of each new RZ entry.

The segments corresponding to two inter-chain RZs in *H. marismortui* are also observed in *D. radiodurans* with a similar topology. Hydrogen-bonding interactions of **inter_n1L** in *D. radiodurans* are conserved in *H. marismortui*. However, in *D. radiodurans*, both O2′–O2′ distances between the residues corresponding to **inter_1L** are beyond the hydrogen-bonding cutoff distance. Nevertheless, the overall conformation of the antiparallel double ribose zipper between **Inter_1L** and **Inter_n1L** mediating an internal loop of the 23 S chain and the internal loop E of 5 S rRNA is well conserved in *D. radiodurans*.

In *D. radiodurans* large ribosomal RNA, we find nine new RZs (Table 4), which were not observed in *H. marismortui*. Stem or stem-like and loop segments corresponding to all new RZs in *D. radiodurans* were also located spatially near each other in *H. marismortui*. Backbone conformations of these new RZs are also almost the same between these two species except for **29L**, in which the adenosine in the 15′ position is turned out in *H. marismortui*. The sequence patterns of new canonical RZs (**29L**–**31L**) and one single base RZ (**s13L**) match the C*/AA pattern and these adenosines form Type I and Type II A-minor motifs in the lower and upper layer, respectively. Almost all these adenosines and their A-minor motifs are conserved in *H. marismortui* except for those in **29L**.

Ribose zippers in other RNAs

We have searched all the RNA entries in the PDB database for RZs. We did not find any previously unreported RZs. Our program finds a previously reported canonical RZ in the hepatitis delta virus ribozyme[19] and two in the group I intron,[4] as well as an intermolecular canonical RZ in the hammerhead ribozyme.[7] In both the hepatitis delta virus ribozyme and the hammerhead ribozyme, the RZs mediate loop–loop interactions with the residue pattern being CC/AA. In the group I intron, both RZs are between loop regions with sequence patterns GC/AA and CU/AA (Figure 1). Our program also finds a previously reported canonical ribose zipper in a highly conserved 58-nucleotide domain of *Thermotoga maritima*[20] or *Escherichia coli*[21] 23 S rRNA[22,23] in complex with ribosomal protein L11. The sequence pattern in *T. maritima* is CC/AA and that in *E. coli* is CU/AA. This ribose zipper corresponds to the canonical RZ **19L** (CC/AA) of *H. marismortui* 23 S rRNA. In *D. radiodurans* 23 S rRNA this site is missing one backbone–backbone hydrogen bond. This highly conserved domain is called the GTPase-center or GTPase-associated region,[20] which plays a crucial GTPase-related role involving two elongation factors (EF-Tu and EF-G).[24]

Discussion

We have searched the coordinate files of the large and small subunit ribosomal RNAs for ribose zipper tertiary interactions and identified 97 examples: 20 RZs in the small ribosomal subunit of *T. thermophilus*, 46 RZs in the large ribosomal subunit of *H. marismortui* and 31 RZs in the large ribosomal subunit of *D. radiodurans*. The average frequency of occurrence of RZs in *H. marismortui*

23 S rRNA, the highest resolution structure, is about one RZ for each 63 residues. In *T. thermophilus* 16 S rRNA and 23 S rRNA from *D. radiodurans*, the average frequencies are one RZ for each 77 and 93 residues, respectively, likely reflecting the lower resolution of these structures. These frequencies generally agree well with the single RZ found in the 72 residue HDV ribozyme structure and the two RZs found in the 158 residue P4–P6 domain of the group I intron. On the other hand, we observe RZ formation between sequences separated by as few as three residues (**L24**).

These 97 examples were categorized into seven classes based on the type and number of ribose–base interactions. All of these are antiparallel backbone interactions and only one is longer than two consecutive residues. We also observe double RZ in both parallel and antiparallel orientations and a single "overlapping" RZ consisting of three consecutive ribose–ribose interactions at the small ribosomal subunit decoding site. Only ~1/3 of the observed RZs bridge between rRNA domains, while the rest are intradomain. An antiparallel double RZ mediates the interchain interaction between 5 S and 23 S rRNA.

The canonical RZ is not only characterized by consecutive ribose–ribose interactions but also by interactions between the minor grooves of base-pairs formed on one side of the RZ (stem or stem-like side) and the bases (often adenosine) on the other side (loop-side). These hydrogen-bonding interactions play an important role in stabilizing the RZs conformation, and are the basis for the observed secondary structure and sequence specificity of ribose zippers. Ribose zippers are predominantly observed as stem–loop interactions. Even when an RZ is apparently found in a loop–loop interaction, most of these have a stem-like structure on one side with several hydrogen bonds between the atoms in the minor groove of the stem-like side and the bases on the other side. The CC/AA pattern RZ is the most common in both small and large subunit ribosomal RNAs. In this pattern, we observed suitable geometry for a minor groove-base hydrogen bond network forming a water-mediated base triple (type II A-minor motif) in the upper layer and a type I A-minor motif in the lower layer. The sequences of the canonical RZ show phylogenetic conservation, suggesting that RZ-mediated tertiary interactions are preserved during evolution.

The sequences of RZs in *T. thermophilus* 16 S rRNA were highly conserved among prokaryotic 16 S sequences compared to the average conservation of the entire sequence. This suggests that tertiary interactions mediated by these RZs are also conserved in other prokaryotic 16 S rRNA structures. Loop-side residues are more conserved than stem or stem-side residues, implying that the interaction between the minor groove of the stem or stem-like side residues and the base on the loop-side is more important for stabilization of the ribose zipper. Silverman *et al.*[8] and Doherty *et al.*[9] have shown that adenosine is the energetically most suitable residue for this minor groove interaction. In the ribose zippers of small and large subunit ribosomal RNAs, adenosine is highly preferred in loop-side residues to form an A-minor motif. There are 33 ** /AA pattern canonical ribose zippers covering 76.7% of the total of 43 canonical RZs found in *T. thermophilus* 16 S rRNA, and *H. marismortui* and *D. radiodurans* large ribosomal subunit RNAs. Other residues are also found in loop-side positions and are conserved in prokaryotic 16 S rRNA sequences, especially if the 13′ position is not adenosine. When adenosine is replaced by another base at the 13′ position, more diverse RZ structures are observed.

Our results suggest that the sequence and secondary structure conservation found in ribose zippers can be used for the prediction of tertiary structure of RNA. Given a well-characterized secondary structure and a set of related sequences, candidate RZs can be postulated by searching for CC regions in double helical stems and AA regions in loops that are conserved or show covariation with other possible RZ base-pair triples.

Two-thirds of the ribosomal RNA ribose zippers interact with ribosomal proteins by hydrogen bonding and charge neutralization. These proteins often bridge the backbones of the RNA chain segments stabilizing these important tertiary interactions. Protein–RZ interactions are found among primary binders in ribosomal assembly, in regions critical for ribosome function, and at sites of antibiotic binding and resistance mutations.

Materials and Methods

The coordinates of 16 S rRNA found in the small ribosomal subunit of *T. thermophilus* were from the PDB file 1J5E determined at 3.0 Å resolution[25] and the coordinates of the large subunit ribosomal RNAs (5 S and 23 S) of *H. marismortui* at 2.4 Å resolution and *D. radiodurans* determined at 3.1 Å, were from the PDB files 1JJ2[26] and 1KPJ, respectively. The algorithm used to search for RZs has two steps. The first step is to find all ribose–ribose and ribose–base hydrogen bonds and the second step is to search this hydrogen-bonding list for the specific pattern associated with an RZ. If the distance between hydrogen bond donor and acceptor atoms was less than 3.6 Å for 1JJ2, or 3.8 Å for the lower resolution 1J5E and 1KPJ, these atoms were counted as a hydrogen-bonded pair. As the template for RZs, we searched for two or more consecutive 2′-hydroxyl to 2′-hydroxyl hydrogen bonds between residues separated in the primary sequence. No restrictions for chain direction or involvement of the bases in the hydrogen bonding between chain segments were used in the search. After all potential RZs were identified, they were classified according to type as discussed below. The program for ribose zipper searches is easily applied to any PDB file and available from the authors and on the Structural Classification of RNA (SCOR) website†.

† http://scor.lbl.gov

The sequence conservation for the canonical RZ in 16 S rRNA was analyzed by sequence logos[27] using software developed in-house. We use aligned 16 S prokaryotic sequences from Ribosomal Database Project Release 8.1,[10] which contains 1173 archaeal and 15,104 bacterial 16 S rRNA sequences.

The conformation of RZs was analyzed by use of pseudotorsion angles as defined by Pyle and co-workers.[12] The pseudotorsion angles simplify the six backbone torsions of each nucleotide to two angles, θ and η defined by the pseudo-bonds (P_n–$C4_n$) and ($C4_n$–P_{n+1}), respectively.

Ribosomal protein interactions with RZs were analyzed by computing the distances between all RNA atoms involved in RZs and all protein atoms as well as water molecules. We accepted potential hydrogen bonds within 3.8 Å and potential hydrophobic interactions between carbon atoms within 4.2 Å. All molecular graphics were drawn with the PyMOL molecular graphics software package.[28]

Acknowledgments

A Laboratory Directed Research and Development Grant from Lawrence Berkeley National Laboratory to S. Holbrook and S. Brenner to develop the SCOR database supported this research in part. M. Tamura was supported by a grant from Uehara Memorial Foundation of Japan. We are grateful to Professor Steven Brenner for his valuable suggestions and Richard Meraz for his technical assistance.

References

1. Batey, R. T., Rambo, R. P. & Doudna, J. A. (1999). Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed. Engl.* **38**, 2326–2343.
2. Kim, S.-H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C. & Rich, A. (1974). Three-dimensional structure of yeast phenylalanine transfer RNA. *Science*, **185**, 435–440.
3. Chang, K.-Y. & Tinoco, I., Jr (1994). Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. *Proc. Natl Acad. Sci. USA*, **91**, 8705–8709.
4. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
5. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4889–4903.
6. Pleij, C. W. A. (1994). RNA pseudoknots. *Curr. Opin. Struct. Biol.* **4**, 337–344.
7. Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, **372**, 111–113.
8. Silverman, S. K. & Cech, T. R. (1999). Energetics and cooperativity of tertiary hydrogen bonds in RNA structure. *Biochemistry*, **38**, 8691–8702.
9. Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. (2001). A universal mode of helix packing in RNA. *Nature Struct. Biol.* **8**, 339–343.
10. Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T. J., Saxman, P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M. & Tiedje, J. M. (2001). The RDP-II (Ribosomal Database Project). *Nucl. Acids Res.* **29**, 173–174.
11. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, D'Souza, L. M., Du, Y. *et al.* (2002). The comparative RNA web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics*, 3.
12. Duarte, C. M. & Pyle, A. M. (1998). Stepping through RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.* **284**, 1465–1478.
13. Lynch, S. R. & Puglisi, J. D. (2001). Structural origins of aminoglycoside specificity for prokaryotic ribosomes. *J. Mol. Biol.* **306**, 1037–1058.
14. Fourmy, D., Yoshizawa, S. & Puglisi, J. D. (1998). Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA. *J. Mol. Biol.* **277**, 333–345.
15. Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Wimberly, B. T. & Ramakrishnan, V. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* **316**, 725–768.
16. Held, W. A., Ballou, B., Mizushima, S. & Nomura, M. (1974). Assembly mapping of 30 S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.* **249**, 3103–3111.
17. Nowotny, V. & Nierhaus, K. H. (1988). Assembly of the 30 S subunit from *Escherichia coli* ribosomes occurs *via* two assembly domains which are initiated by S4 and S7. *Biochemistry*, **27**, 7051–7055.
18. Nowotny, V. & Nierhaus, K. H. (1982). Initiator proteins for the assembly of the 50 S subunit from *Escherichia coli* ribosomes. *Proc. Natl Acad. Sci. USA*, **79**, 7238–7242.
19. Ferre-D'Amare, A. R., Zhou, K. & Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567–574.
20. Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell*, **97**, 491–502.
21. Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein–RNA complex. *Science*, **284**, 1171–1174.
22. Thompson, J., Cundliffe, E. & Stark, M. (1978). Binding of thiostrepton to a complex of 23-S rRNA with ribosomal protein L11. *Eur. J. Biochem.* **98**, 261–265.
23. Schmidt, F. J., Thompson, J., Lee, K., Dijk, J. & Cundliffe, E. (1981). The binding site for ribosomal protein L11 within 23 S ribosomal RNA of *Escherichia coli*. *J. Biol. Chem.* **256**, 12301–12305.
24. Wilson, K. S. & Noller, H. F. (1998). Molecular movement inside the translational engine. *Cell*, **92**, 337–349.
25. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30 S ribosomal subunit. *Nature*, **407**, 327–339.
26. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the

large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
27. Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucl. Acids Res.* **18**, 6097–6100.
28. DeLano, W. L. (2002). *The PyMol Molecular Graphics System*, DeLano Scientific, San Carlos, CA.

Edited by J. Doudna

(Received 5 March 2002; received in revised form 3 May 2002; accepted 14 May 2002)