The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution

Nenad Ban,[1]* Poul Nissen,[1]* Jeffrey Hansen,[1] Peter B. Moore,[1,2] Thomas A. Steitz[1,2,3]†

The large ribosomal subunit catalyzes peptide bond formation and binds initiation, termination, and elongation factors. We have determined the crystal structure of the large ribosomal subunit from *Haloarcula marismortui* at 2.4 angstrom resolution, and it includes 2833 of the subunit's 3045 nucleotides and 27 of its 31 proteins. The domains of its RNAs all have irregular shapes and fit together in the ribosome like the pieces of a three-dimensional jigsaw puzzle to form a large, monolithic structure. Proteins are abundant everywhere on its surface except in the active site where peptide bond formation occurs and where it contacts the small subunit. Most of the proteins stabilize the structure by interacting with several RNA domains, often using idiosyncratically folded extensions that reach into the subunit's interior.

In the last step of the gene expression pathway, genomic information encoded in messenger RNAs is translated into protein by a ribonucleoprotein called the ribosome (*1*). As in most other organisms, the prokaryotic ribosome (MW $\approx 2.6 \times 10^6$) is about two-thirds RNA and one-third protein and consists of two subunits, the larger of which is approximately twice the molecular weight of the smaller (*2*). The small subunit, which sediments at 30*S* in prokaryotes, mediates the interaction between mRNA codons and tRNA anticodons on which the fidelity of translation depends. The large subunit, which sediments at 50*S* in prokaryotes, includes the activity that catalyzes peptide bond formation—peptidyl transferase—and the binding site for the G-protein (GTP–binding protein) factors that assist in the initiation, elongation, and termination phases of protein synthesis.

Because the structures of several DNA and RNA polymerases have been determined at atomic resolution, the mechanisms of DNA and RNA synthesis are both well understood. Determination of the structure of the ribosome, however, has proven a daunting task. It is several times larger than the largest polymerase, and 100 times larger than lysozyme, the first enzyme to be understood at atomic resolution. Until now an atomic resolution structure for the ribosome has not been available, and as a result the mechanism of protein synthesis has remained a mystery.

Electron microscopy has contributed to our understanding of ribosome structure ever since the ribosome was discovered. In the last few years, three-dimensional (3D) electron microscopic images of the ribosome have been produced at resolutions sufficiently high to visualize many of the proteins and nucleic acids that assist in protein synthesis bound to the ribosome (*3*). Earlier this year, an approximate model of the RNA structure in the large subunit was constructed to fit a 7.5 Å resolution electron microscopic map of the 50*S* subunit from *Escherichia coli* as well as biochemical data (*4*).

Crystallization studies of the ribosome begun two decades ago by Yonath and Wittmann (*5*) and by the group at Pushchino (*6*) opened the possibility of using x-ray crystallography to determine the structure of the ribosome at atomic resolution. The first electron density map of the ribosome that showed features recognizable as duplex RNA was a 9 Å resolution x-ray crystallographic map of the large subunit from *Haloarcula marismortui* published 2 years ago (*7*). A year later, extension of the phasing of that map to 5 Å resolution made it possible to locate several proteins and nucleic acid sequences, the structures of which had been determined independently (*8*). At about the same time, with the use of similar crystallographic strategies, a 7.8 Å resolution map was generated of the entire *Thermus thermophilus* ribosome, showing the positions of tRNA molecules bound to its A, P, and E sites (*9*), and a 5.5 Å resolution map of the 30*S* subunit from *T. thermophilus* was obtained, which allowed the fitting of solved protein structures and the interpretation of some of its RNA features (*10*). Subsequently, an independently determined, 4.5 Å resolution map of the *T. thermophilus* 30*S* subunit was published, which was based, in part, on phases calculated from a model corresponding to 28% of the subunit mass that had been obtained with a 6 Å resolution experimental map (*11*). The interpretation of the subunit packing in the two 30*S* structures is not the same, even though the crystals used by the two groups appear to be identical.

Using a 2.4 Å resolution, experimentally phased, electron density map, we have produced an atomic structure of the *H. marismortui* 50*S* ribosomal. The model includes 2711 of the 2923 nucleotides of 23*S* ribosomal RNA (rRNA), all 122 nucleotides of its 5*S* rRNA, and structures for the 27 proteins that are well ordered in the subunit. Here, we describe the architecture of the subunit, the structure of its RNAs, and discuss the location, structures, and functions of its proteins.

The secondary structures of both 5*S* and 23*S* rRNA are remarkably close to those deduced for them by phylogenetic comparison. The secondary structure of the 23*S* rRNA divides it into six large domains, each of which has a highly asymmetric tertiary structure. The irregularities of their shapes notwithstanding, the domains fit together in an interlocking manner to yield a compact mass of RNA that is almost isometric. The proteins are dispersed throughout the structure and mostly concentrated on its surface, but they are largely absent from the regions of the subunit that are of primary functional significance to protein syntheses: the 30*S* subunit interface and the peptidyl transferase active site. The most surprising feature of many of these proteins is the extended, irregular structure of their loops and termini, which penetrate between RNA helices. The primary role of most proteins in the subunit appears to be stabilization of the 3D structure of its rRNA.

Structure determination. Several experimental approaches were used to extend the resolution of the electron density maps of the *H. marismortui* 50*S* ribosomal subunit from 5 to 2.4 Å. A back-extraction procedure was developed for reproducibly growing crystals that are much thicker than those available earlier and that diffract to at least 2.2 Å resolution. The twinning of crystals, which obstructed progress for many years (*8*), was eliminated by adjusting crystal stabilization conditions (*12*). All of the x-ray data used for high-resolution phasing were collected at the Brookhaven National Synchrotron Light Source except for two native data sets, which were collected at the Advanced Photon Source at Argonne (*13*) (Table 1). Osmium pentamine (132 sites) and iridium hexamine (84 sites) derivatives proved to be the most effective in producing isomorphous replacement and anomalous scattering phase information to 3.2 Å resolution (*14*). Intercrystal density averaging, which had contributed significantly at lower resolution, was not helpful beyond about 5 Å resolution. Electron density maps were dramatically improved, and

[1]Department of Molecular Biophysics & Biochemistry, and [2]Department of Chemistry, Yale University, and [3]Howard Hughes Medical Institute, New Haven, CT 06520–8114, USA.

*These two authors contributed equally to this work.
†To whom correspondence should be addressed.

their resolutions were eventually extended to 2.4 Å with the solvent-flipping procedure in the CNS program (*15*, *16*).

Except for regions obscured by disorder, the experimentally phased 2.4 Å resolution electron density map was of sufficient quality that both protein and nucleic acid sequencing errors could be identified and corrected. Each nucleotide could be fitted individually, and the difference between A and G was usually clear without having to refer to the chemical sequence, as was the distinction between purines and pyrimidines (Fig. 1). Only a few of the many water molecules and metal ions evident in the electron density have been positioned so far.

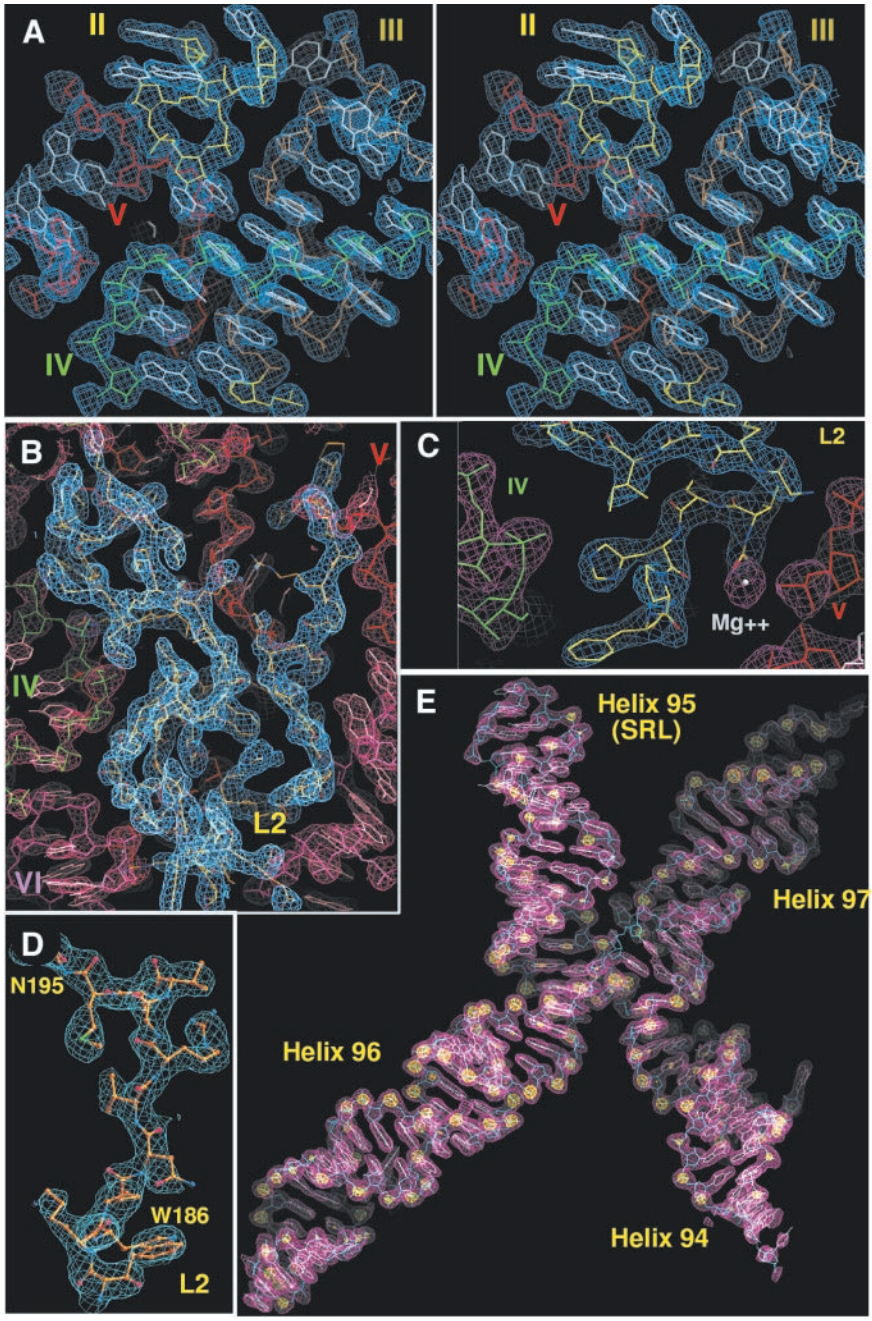

Fig. 1. Portions of the experimental 2.4 Å resolution electron density map. (**A**) A stereo view of a junction between 23*S* rRNA domains II, III, IV, and V having a complex structure that is clearly interpretable. The electron density is contoured at 2σ. The bases are white and the backbones are colored by domain as specified in Fig. 4. (**B**) The extended region of L3 interacting with its surrounding RNA, where the red RNA density is contoured at 2σ and the blue protein density is contoured at 1.5σ. (**C**) Detail in the L2 region showing a bound Mg^{2+} ion. (**D**) Detail from L2 showing amino acid side chains. (**E**) Helices 94 through 97 from domain VI. The red contour level is at 2σ, and the yellow contour at 6σ shows the positions of the higher electron density phosphate groups.

Subtraction of the atomic model from the experimental electron density map leaves no significant density except water and ions, showing that the model accounts for all the macromolecular density. Preliminary refinement of the model was achieved with experimental phase restraint in the program CNS (*16*). The model was further refined in real space against the 2.4 Å electron density map with the program TNT (*17*), which yielded a model with an *R* factor of 0.33. One additional round of mixed target refinement of both atomic positions and *B* factors with CNS led to the structure described here. The current free *R* factor is 0.26 (Table 1).

Sequence fitting and protein identification. Guided by the information available on the secondary structures of 23*S* rRNAs (*18*), the sequence of 23*S* rRNA was fit into the electron density map nucleotide by nucleotide starting from its sarcin/ricin loop sequence [A2691 to A2702 (*E. coli* numbers A2654 to A2665)] whose position had been determined at 5 Å resolution (*8*). The remaining RNA electron density neatly accommodated 5*S* rRNA. The interpretation of electron density corresponding to protein was more complicated because each protein region had to be identified chemically before the appropriate sequence could be fit into it; with the assistance of D. Klein, L. Min, S. Antolić, and M. Schmeing, ~4000 amino acid residues of 27 proteins were fit into electron density.

The *H. marismortui* 50*S* subunit appears to contain 31 proteins, and sequences for 28 of them exist in the Swiss-Prot data bank, including one called HMS6 or L7ae, which originally had been assigned to the small ribosomal subunit (*19*). The three remaining proteins were identified using the sequences of the ribosomal proteins from eukaryotes and other archaeal species as guides. No electron density was found for one of the *H. marismortui* large ribosomal subunit proteins in the sequence database, LX. Either the assignment of LX to the large subunit is in error, or LX is associated with a disordered region of the subunit. It is also possible that LX is absent from the subunits examined altogether.

The 2.4 Å resolution electron density map lacks clear electron density for proteins L1, L10, L11, and L12, the positions of which are known from earlier low-resolution x-ray and/or electron microscopic studies. These proteins are components of the two lateral protuberances of the subunit, which are both poorly ordered in these crystals. L1 is the sole protein component of one of them (*20*) and is visible in 9 Å resolution density maps of the subunit (*7*), but not at higher resolutions. L10, L11, and L12 are components of the other protuberance, which is often referred to as the L7/L12 stalk (*20*). L11 and the RNA to

which it binds were located in the 5 Å resolution electron density map of the *H. marismortui* large subunit (*8*) using the independently determined crystal structures of that complex (*21*, *22*). A protein fragment (~100 residues) associated with the RNA stalk that supports the L11 complex can be seen in the 2.4 Å resolution map. On the basis of its location, the fragment must be part of L10. No electron density corresponding to L12 was seen at any resolution, but the L12 tetramer is known to be attached to the ribosome through L10, and the L10/L12 assembly is known to be flexible under some circumstances (*23*), which may explain its invisibility here.

The structures of eubacterial homologs of proteins L2, L4, L6, L14, and L22 have previously been determined in whole or in part (Table 2). L2, L6, and L14 were initially located in the 5 Å resolution map (*8*). L4 and L22 have now been identified and positioned the same way. Electron density corresponding to most of the remaining proteins was assigned by comparing chain lengths and sequence motifs deduced from the electron density map with known sequence lengths. Occasionally, these comparisons were assisted by the information available on relative protein positions (*24*) and protein interactions with 23*S* rRNA and 5*S* rRNA (*25*). Each of the protein electron density regions so identified is well accounted for by the amino acid sequence assigned to it.

The most interesting of the proteins identified by sequence similarity was L7ae, which first appeared to be L30e. The L30e identification seemed plausible because the structure of yeast L30e superimposes neatly on the electron density of L7ae, and the structure of the RNA to which L7ae binds resembles that of the mRNA element to which yeast L30e binds (*26*). Nevertheless, the sequence of HMS6, which by sequence similarity is a member of the L7ae protein family, better fits the electron density. Four of the other proteins identified by sequence similarity, L24e, L37e, L37ae, and L44e, contain zinc finger motifs. The rat homologs of L37e and L37ae were predicted to be zinc finger proteins on the basis of their sequences (*27*), and this prediction helped identify their homologs in *H. marismortui*. Even though no *H. marismortui* sequences were available for the proteins L10e, L15e, and L37ae, they could be identified using the alignments of other available archaeal sequences.

General appearance of the subunit. In its rotated crown view (Fig. 2), the large ribosomal subunit, which is about 250 Å across, presents its surface that interacts with the small subunit to the viewer with the three projections that radiate from that surface pointed up. Although the protuberance that includes L1 is not visible in the 2.4 Å resolution electron density map, the structure of L1, which has been determined independently (*28*), has been positioned approximately in lower resolution maps (*7*) and is included here to orient the reader. It is evident that, except for its two lateral protuberances, the large ribosomal subunit is monolithic. There is no hint of a division of its structure into topologically separate domains. In addition, partly because it lacks obvious domain substructure but also because it is so large, it is impossible to comprehend looking at it as a whole. To convey a sense of how it is put together, the subunit must be dissected into its chemical components.

RNA secondary structure. All the base pairs in *H. marismortui* 23*S* rRNA stabilized by at least two hydrogen bonds were identified with a computer program that searched the structure for hydrogen bond donors and acceptors separated by less than 3.2 Å. Bases linked by at least two such bonds were considered paired if the angle between their normals was less than 45° and if the angle between bonds and base normals was also less than 45°. On the basis of the results of this analysis, R. Gutell and colleagues prepared a secondary structure diagram (Fig. 3) in the format standard for 23*S*/28*S* rRNAs. The secondary structure predicted for this molecule by phylogenetic comparison was remarkably accurate, but it did not find all of the tertiary pairings and failed to identify interactions involving conserved bases. In addition to base pairs of nearly every type, the RNA contains numerous examples of well-known secondary structure motifs such as base triples, tetraloops, and cross-strand purine stacks, but no dramatically new secondary structure motifs have been identified so far.

The secondary structure of this 23*S* rRNA consists of a central loop that is closed by a terminal stem, from which 11 more or less complicated stem-loops radiate. It is customary to describe the molecule as consisting of six domains and to number its helical stems sequentially starting from the 5′ end (Fig. 4) (*29*). The division of the molecule into domains as shown in Fig. 4 deviates from standard practice with respect to helix 25, which is usually considered part of domain I. Here, it is placed in domain II because it interacts more strongly with domain II than the other elements of domain I.

There are five sequences longer than 10 nucleotides in 23*S* rRNA whose structures cannot be determined from the 2.4 Å resolution map because of disorder. Together they account for 207 out of the 232 nucleotides missing from the final model. The disordered regions are: all of helix 1, the distal end of helix 38, helix 43/44 to which ribosomal protein L11 binds, the loop end of stem-loop 69, and helix 76/77/78, which is the RNA structure to which L1 binds. For completeness, these regions are included in Fig. 3 (in gray) with the secondary structures determined for them phylogenetically.

Overall architecture of rRNA. The six domains of 23*S* rRNA and 5*S* rRNA all have complicated, convoluted shapes that fit together to produce a compact, monolithic RNA mass (Fig. 4, A and B). Thus, despite the organization of its RNAs at the secondary structure level, in three dimensions the large subunit is a single, gigantic domain. In this respect, it is quite different from the small subunit. Even in low-resolution electron micrographs the small subunit consists of three structural domains, each of which contains

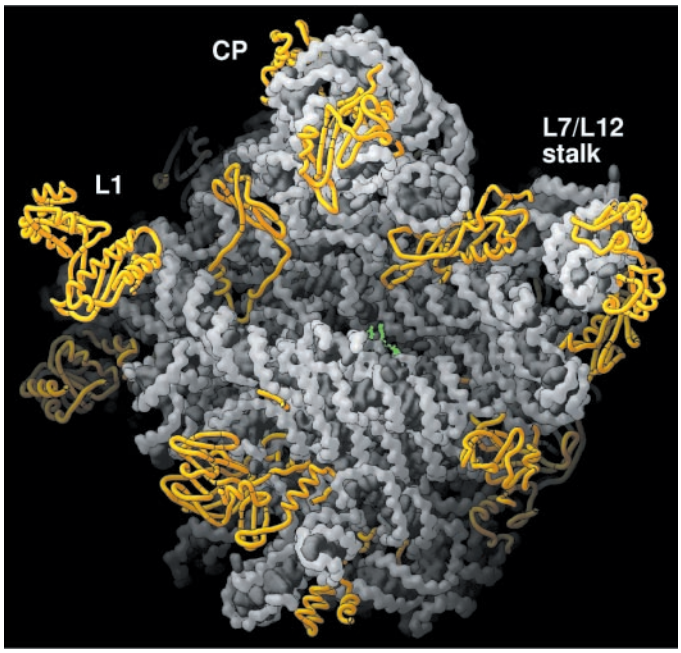

Fig. 2. The *H. marismortui* large ribosomal subunit in the rotated crown view. The L7/L12 stalk is to the right, the L1 stalk is to the left, and the central protuberance (CP) is at the top. In this view, the surface of the subunit that interacts with the small subunit faces the reader. RNA is shown in gray in a pseudo–space-filling rendering. The backbones of the proteins visible are rendered in gold. The Yarus inhibitor bound to the peptidyl transferase site of the subunit is indicated in green (*64*). The particle is approximately 250 Å across.

Secondary Structure: large subunit ribosomal RNA - 5' half

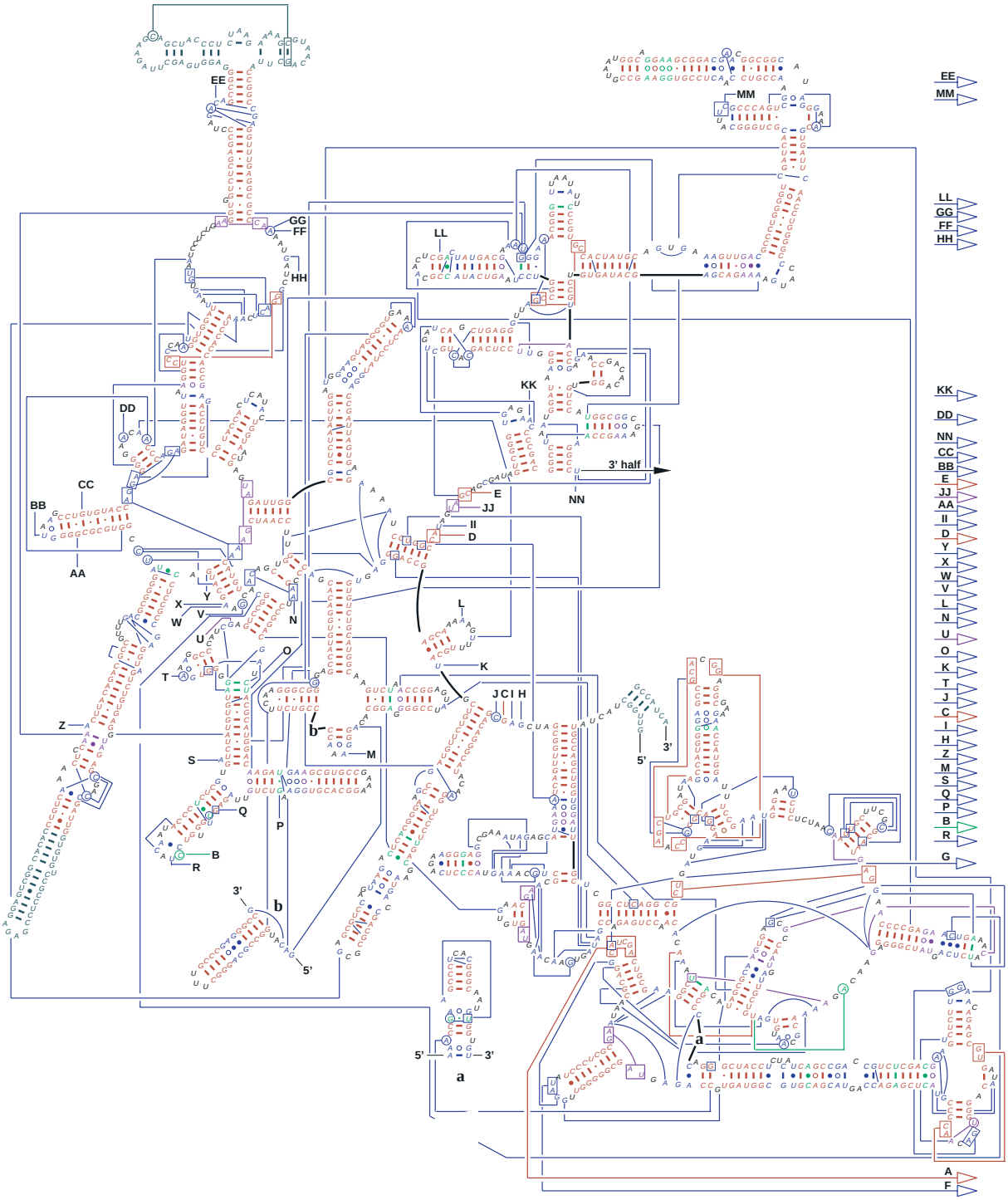

Fig. 3.

one of the three secondary structure domains of its RNA (*30*). This qualitative difference between the two subunits may reflect a requirement for conformational flexibility that is greater for the small subunit.

Domain I, which looks like a mushroom (Fig. 4E), lies in the back of the particle, behind and below the L1 region. The thin part of the domain starts in the vicinity of domain VI, which is the location of its first and last residues. Helices 1 and 25 span the particle in the back and then the domain expands into a larger, more globular structure below and behind the L1 region.

Domain II is the largest of the six 23*S* rRNA domains, accounting for most of the back of the particle. It has three protrusions that reach toward the subunit interface side of the particle (Fig. 4F). One of them (helix 42 to 44) is the RNA portion of the L7/L12 stalk, which is known to interact with elongation factors, is not well ordered in these crystals. The second domain II protrusion is helix 38, which is the longest, unbranched stem in the particle. It starts in the back of the particle, bends by about 90° and pro-

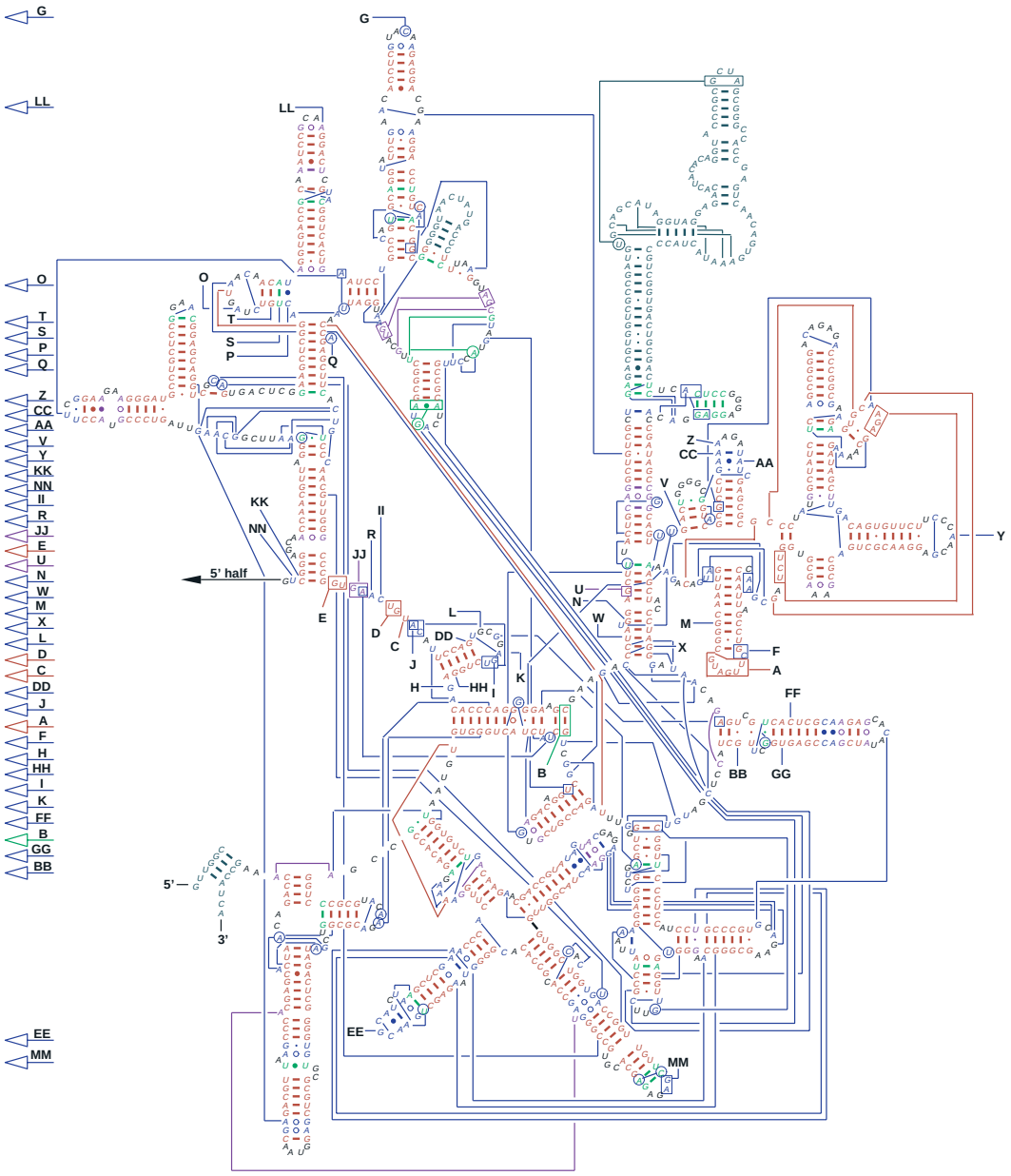

Fig. 3. The secondary structure of the 23*S* rRNA from *H. marismortui* is shown in a format made standard by R. Gutell and colleagues (*65*). It was prepared by Dr. Gutell to show all the base pairings seen in the crystal structure of the large subunit that are stabilized by at least two hydrogen bonds. Pairings shown in red were predicted and were observed. Those shown in green were predicted, but were not observed. Interactions shown in blue were observed, but were not predicted. Bases shown in black were not involved in pairing interactions. Sequences that cannot be visualized in the 2.4 Å resolution electron density map are depicted in gray with the secondary structures predicted for them.

trudes toward the small subunit between domains V and 5*S* rRNA. The third region (helix 32 to 35.1) points directly toward the small subunit and its terminus, the loop of stem-loop 34, interacts directly with the small ribosomal subunit (*31*). This loop emerges at the subunit interface between domains III and IV.

Domain III is a compact globular domain that occupies the bottom left region of the subunit in the crown view (Fig. 4G). It looks like a four-pointed star with the origin of the domain (stem-loop 48) and stem-loops 52, 57, and 58 forming the points. The most extensive contacts of domain III are with domain II, but it also interacts with domains I, IV, and VI. Unlike all the other domains, domain III hardly interacts with domain V at all; the sole contact is a van der Waals interaction involving a single base from each domain.

Domain IV accounts for most of the interface surface of the 50*S* subunit that contacts the 30*S* subunit (Fig. 4H). It forms a large diagonal patch of flat surface on that side of the subunit and connects to domains III and V in the back of the particle. Helices 67 through 71 constitute the most prominent feature of domain IV and form the front rim of the active site cleft, which is clearly visible at low resolution (Fig. 2). This is one of the few regions of the 23*S* rRNA that is not extensively stabilized by ribosomal proteins. Helix 69 in the middle of this ridge interacts with the long penultimate stem of 16*S* rRNA in the small ribosomal subunit (*9*).

Domain V, which is sandwiched between domains IV and II in the middle of the subunit, is known to be intimately involved in the peptidyl transferase activity of the ribosome (*32*). Structurally, this domain can be divided into three regions (Fig. 4, I and J). The first starts with helix 75 and ultimately forms the binding site for protein L1. The second, which consists of helices 80 through 88, forms the bulk of the central protuberance region and is supported in the back by the 5*S* rRNA and domain II. The third region, which includes helices 89 through 93, extends toward domain VI and

helps stabilize the elongation factor-binding region of the ribosome.

The smallest domain in 23*S* rRNA, domain VI, which forms a large part of the surface of the subunit immediately below the L7/L12 stalk, resembles a letter X with a horizontal bar at the bottom (Fig. 4K). The most interesting region of this domain is the sarcin-ricin loop (SRL) (stem-loop 95), the structure of which has been extensively studied in isolation (*33*, *34*). The SRL is essential for factor binding, and ribosomes can be inactivated by the cleavage of single covalent bonds in this loop (*35*). As suggested by nucleotide protection data, the major groove of this loop is exposed to solvent (*36*), and its conformation is stabilized by proteins and through interaction with domain V.

Table 1. Statistics for data collection, phase determination, and model construction. HA, heavy-atom concentration; ST, soaking time; Res, resolution; λ, wavelength; Obs, observations; Redun, redundancy; Compl, completeness; (*) last-resolution shell. R_{iso}: $\Sigma|F_{PH} - F_P|/F_{PH}$, where F_{PH} and F_P are the derivative and the native structure factor amplitudes, respectively. R_{sym}: $\Sigma\Sigma_i|I_{(h)i} - I_{(h)i}|/\Sigma\Sigma$: $I_{(h)i}$, where $I_{(h)}$ is the mean intensity after reflections. Phasing power: rms isomorphous difference divided by the rms residual lack of closure. R_{cullis}: $\Sigma(\|F_{PH} - F_P| - |F_{H(calc)}\|)/\Sigma|F_{PH} - F_P|$, where F_{PH} is the structure factor of the derivative and F_P is that of the native data. The summation is valid only for centric reflection. FOM (figure of merit): mean value of the cosine of the error in phase angles. Abbreviations: MIRAS, multiple isomorphous replacement, anomalous scattering; SAD, single wavelength anomalous diffraction.

Data statistics

	MIRAS1			MIRAS2			
	Native1	$Os(NH_3)_5^{2+}$	$UO_2F_5^{3-}$	Native2	$Ir(NH_3)_6^{3+}$	$Os(NH_3)_6^{3+}$	$Ta_6Br_{12}^{2+}$
HA (mM)	–	30.0	0.5	–	20.0	4.5	3.0
ST (hours)	–	1.5	4	–	24 hours	24 hours	24 hours
Sites no.	–	132	20	–	84	38	9
Res (Å)	90–2.4	40–3.5	40–3.8	30–2.9	30–3.2	30–3.5	30–3.8
(*)	(2.5–2.4)	(3.6–3.5)	(3.9–3.8)	(3.32–3.22)	(3.27–3.20)	(3.6–3.5)	(3.97–3.80)
λ (Å)	1.00	1.14	1.30	1.00	1.075	1.14	1.255
Obs	6,089,802	1,308,703	596,166	2,832,360	1,823,861	1,646,468	1,288,524
Unique	665,928	429,761	313,863	390,770	541,488	488,275	346,745
Redun (*)	9.1 (6.5)	3.0 (2.5)	1.9 (1.6)	7.2	3.4	4.3 (4.2)	3.7
Compl (*)	95.6 (71.0)	99.4 (96.8)	92.0 (54.1)	97.1	93.8	98.1 (99.0)	99.5
$I/\sigma I$ (*)	25.5 (1.9)	13.5 (3.3)	8.9 (1.6)	18.0 (6.4)	12.0 (2.6)	10.6 (2.7)	10.8 (3.2)
R_{merge} (*)	8.6 (69.1)	7.2 (32.0)	9.1 (37.9)	11.2 (36.9)	8.5 (29.5)	12.1 (46.0)	12.1 (40.5)
χ^2 (ano) (*)	–	2.8 (1.0)	1.5 (1.0)	–	2.63 (1.48)	1.8 (1.0)	2.42 (1.18)
R_{merge} (ano)	–	6.2	8.0	–	6.7	6.9	
R_{iso} (*)	–	14.1 (22.7)	26.4 (47.0)	–	12.9 (28.1)	19.5 (39.4)	

Phasing statistics

	Resolution shells (Å): ~73,200 reflections per bin					
	30.0	5.1	4.0	3.5	3.2	Total
MIRAS1 (FOM)	0.52	0.31	0.14	–		0.32
$Os(NH_3)_5^{2+}$						
Phasing power	0.87	0.72	0.66	–		0.75
Phasing power (SAD)	1.40	0.58	0.26	–		0.75
R_{cullis} (centric)	0.62	0.65	0.67	–		0.65
$UO_2F_5^{3-}$						
Phasing power	0.47	0.33	0.28	–		0.36
Phasing power (SAD)	0.46	0.25	–	–		0.36
R_{cullis} (centric)	0.72	0.77	0.75	–		0.75
MIRAS2 (FOM)	0.48	0.40	0.28	0.12		0.33
$Ir(NH_3)_6^{3+}$						
Phasing power	1.02	0.92	0.78	0.66		0.89
Phasing power (SAD)	2.02	1.60	1.22	0.83		1.47
R_{cullis} (centric)	0.58	0.63	0.70	0.74		0.63
$Os(NH_3)_6^{3+}$						
Phasing power	0.62	0.57	0.58	0.58		0.59
Phasing power (SAD)	0.47	0.39	–	–		0.42
R_{cullis} (centric)	0.78	0.78	0.78	0.76		0.78
$Ta_6Br_{12}^{2+}$ (used for SAD phasing only)						
Phasing power (SAD)	2.77	0.35	0.13	–		1.19
$FOM_{(MIRAS1+MIRAS2+SAD)}$	0.76	0.51	0.31	0.14		0.37

Model statistics

Resolution range (Å)	90.0–2.4	rms deviations:		Average *B* factors (Å^2)	
Reflections	577,304	Bonds (Å)	0.0064	All atoms	37.4
R_{cryst} (%)	25.2	Angles (°)	1.19	23*S* rRNA	32.3
R_{free} (%)	26.1	Dihedrals (°)	28.8	5*S* rRNA	43.2
		Impropers (°)	1.68	Minimum/Max *B* factors (Å^2)	7.0/107.9

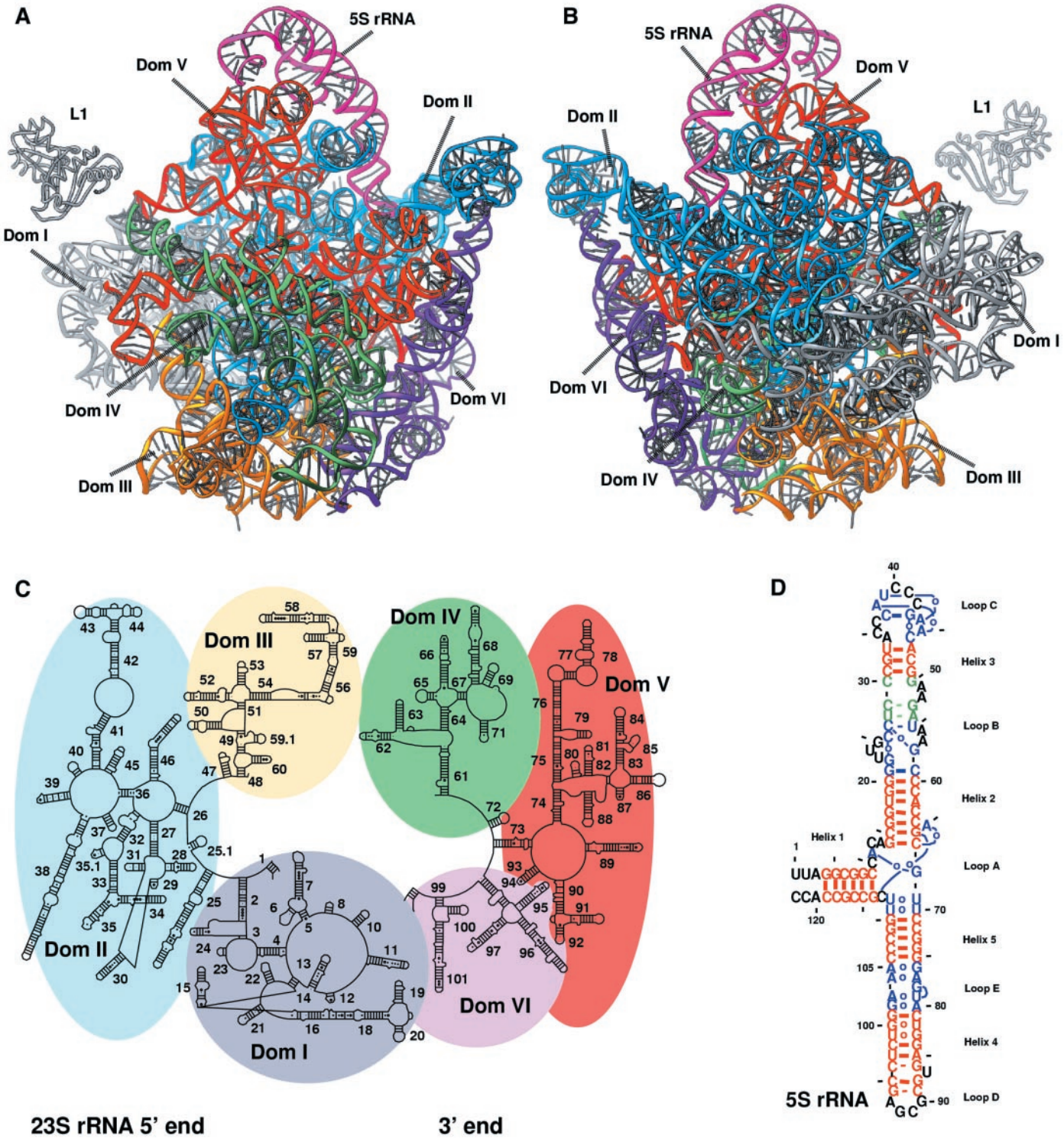

Fig. 4. The tertiary and secondary structures of the RNA in the *H. marismortui* large ribosomal subunit and its domains. (**A** and **B**) The RNA structure of the entire subunit. Domains are color-coded as shown in the schematic (**C**). (A) The subunit particle in its crown view. (B) The crown rotated by 180° about a vertical axis in the plane of the image. (C) Schematic secondary structure diagram of 23*S* rRNA with the domain coloring used throughout the figures and the helices numbered according to Leffers *et al.* (*29*). (**D**) The secondary structure of 5*S* rRNA from *H. marismortui*. Bases joined by thick lines represent Watson-Crick pairing, and those joined by a lower case "o" indicate non–Watson-Crick pairing. Bases joined by thin lines interact via a single hydrogen bond, whereas those in black are unpaired. Base pairings shown in red are phylogenetically predicted pairings that are now confirmed (*66*). Pairs shown in blue were observed but were not predicted, and pairs shown in green were predicted but were not observed. (**E** through **L**) Stereo views of the RNA domains of the 23*S* rRNA and of 5*S* rRNA. Each domain is color-coded from its 5′ end to its 3′ end to help the viewer follow its trajectory in three dimensions. The backbones are shown as ribbons and the bases as sticks. The surfaces where the most important interdomain interactions occur are shown in mono to the right. (E), Domain I; (F), domain II; (G), domain III; (H), domain IV; (I), domain V, crown view; (J), domain V, back view; (K), domain VI; and (L), 5*S* rRNA.

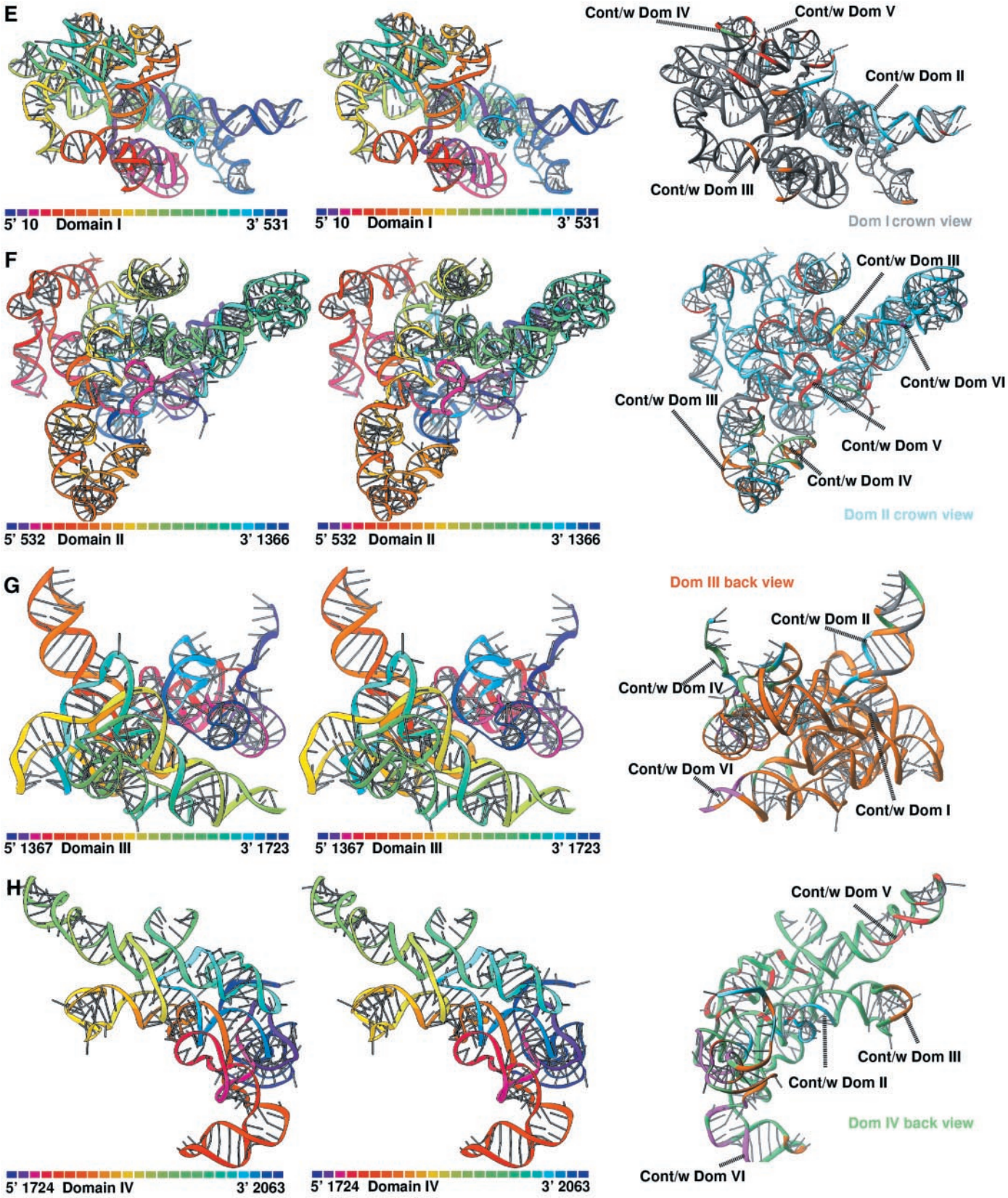

Fig. 4. (continued)

5*S* ribosomal RNA, which is effectively the seventh RNA domain in the subunit, consists of three stems radiating out from a common junction called loop A (Fig. 4D). In contrast to what is seen in the crystal structure of fragment 1 from *E. coli* 5*S* rRNA (*37*), the helix 2/3 arm of the molecule stacks on its helix 4/5 arm, not helix 1 (Fig. 4L). This arrangement results from a contorted

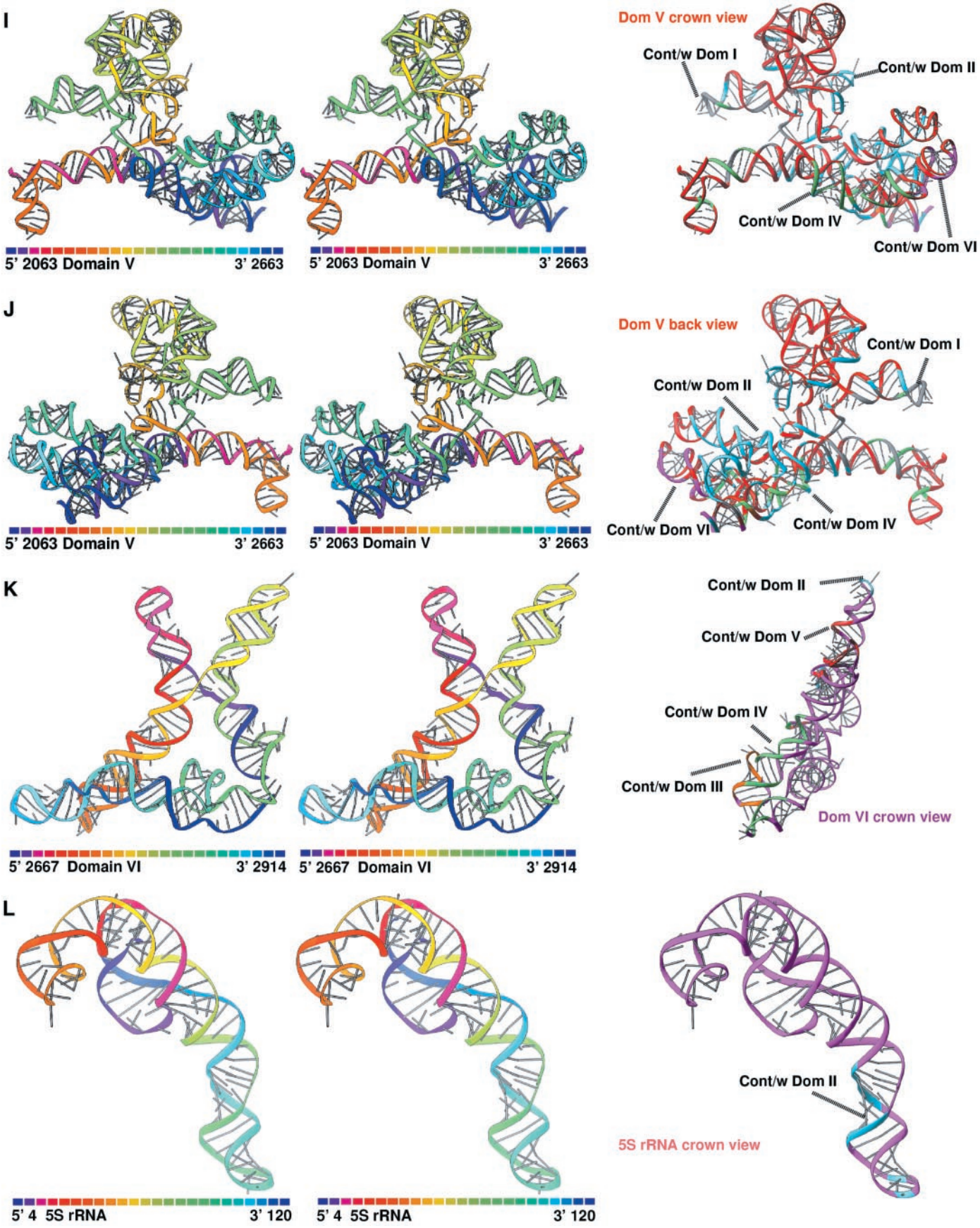

Fig. 4. (continued)

conformation of loop A residues that involves two stacked base triples. Indeed, from the secondary structure point of view, the loop A–helix 2/3 arm of 5*S* rRNA is remarkable, with a high concentration of unusual pairings leading to a convoluted RNA secondary structure.

Sequence conservation and interactions in 23*S* rRNA. Although 23*S*/28*S* rRNAs contain many conserved sequences, they also vary substantially in chain length. Shorter 23*S*/28*S* rRNAs are distinguished from their longer homologs by the truncation of, or even the elimination of, entire stem-loops, and by comparing sequences, one can identify a minimal structure that is shared by all (*38*). The expansion sequences in the 23*S* rRNA of *H. marismortui*, i.e., the sequences it contains that are larger than the minimum, are shown in Fig. 5 in green. They are largely absent from the subunit interface surface of the particle, but they are abundant on its back surface, far from its active sites. This is consistent with low-resolution electron microscopic observations, suggesting that the region of the large subunit whose structure is most conserved is the surface that interacts with the small subunit (*39*).

There are two classes of conserved sequences in 23*S* rRNA. One contains residues concentrated in the active site regions of the large subunit. The second class consists of much shorter sequences scattered throughout the particle (Fig. 5, red sequences). The SRL sequence in domain VI and the cluster of conserved residues belonging to domain V located at the bottom of the peptidyl transferase cleft are members of the first class. They are conserved because they are essential for substrate binding, factor binding, and catalytic activity. Most of the residues in the second class of conserved residues are involved in the inter- and intradomain interactions that stabilize the tertiary structure of 23*S* rRNA. Adenosines are disproportionately represented in this class. The predominance of adenosines among the conserved residues in rRNAs has been pointed out previously (*40*). Throughout the particle, adenosines are observed to participate in tertiary interactions by exploiting the smooth N1-C2-N3 face of the adenine base, which allows for very close packing and additional backbone-backbone interactions. In particular, a reoccurring pattern of two or more stacked adenosines that dock into the minor grooves of receptor helices seems to reveal a very basic principle in tertiary RNA structure formation and could be regarded as an equivalent of a hydrophobic core formation in glob-

Table 2. Large-subunit proteins from *Haloarcula marismortui*. The top block of proteins include all those known to have eubacterial homologs of the same name. The second block lists proteins found in the *H. marismortui* large ribosomal subunit that have only eukaryotic homologs (*19*). Their names are all followed by the letter "e" to distinguish them from eubacterial proteins that would otherwise have the same name. The third block are large-subunit proteins for which no *H. marismortui* sequence yet exists. They are identified by sequence homology with standard L names. [1]The structures of all or part of homologs of the following proteins were previously determined: L1 (*28*), L2 (*43*), L4 (*44*), L6 (*58*), L11 (*21, 22, 59*), L12 (*60*), L14 (*61*), L22 (*62*), and L30 (*63*). All other structures, except L10, have been newly determined in this study. [2]Rat homolog. Rat equivalents to *H. marismortui* protein are from (*26*). [3]Sequence chain length. [4]Conformation: glb, globular; ext, extension. [5]The protein interactions with the six domains of 23*S* rRNA, 5*S* rRNA, and other proteins are specified. (+) Implies that the interaction is substantial; (±) implies a weak, tangential interaction. Protein names in parentheses implies that the interactions are weak; otherwise, the interaction is substantial.

Name[1]	Hmlg[2]	Lgth[3]	Conf[4]	Interactions[5]							Proteins
				I	II	III	IV	V	VI	5*S*	
L1*	?	211	glb.					+			None
L2†	RL8	239	glb+ext		+	+	+	+			(L37ae)
L3	RL3	337	glb+ext		±		+	+	+		L14, L24e, (L13)
L4†	RL4	246	glb+ext	+	+			±			(L18e), (L24), (L37e)
L5	RL11	176	glb					+		+	L18
L6	RL9	177	glb		±			±	+		(L13)
L10*	RP0	348	glb?		+						L12
L11*	RL12	161	glb		+						None
L12*	RP1/2	115	glb								L10
L13	RL13a	145	glb		+			±	±		(L3), (L6)
L14	RL23	132	glb				+	+	+		L3, L24e
L15	RL27a	164	glb+ext	+	+			+			(L18e), (L32e)
L18	RL5	186	glb+ext		±			+		+	L5, L21e
L19	RL19	148	glb+ext		+	+	+		±		None
L22	RL17	154	glb+ext	+	±	+	+	+	+		None
L23	RL23a	84	glb	±		+					L29, (L39e)
L24	RL26	119	glb+ext	+							(L4)
L29	RL35	70	glb	+							L23
L30	RL7	154	glb		+					+	None
L18e	RL18	115	glb		+						(L4), (L15)
L21e	RL21	95	glb		+			+		±	L18
L24e	RL24	66	glb				±		+		L3, L14
L31e	RL31	91	glb			+	+		+		None
L32e	RL32	240	glb	±	+						(L15)
L37e	RL37	56	glb+ext	+	+	+	±				(L4)
L39e	RL39	49	ext	+		+					(L23)
L44e	RL36a	92	glb+ext	+			±	+			(L15e)
L7ae	RL7a	110	glb	±							L15e
L10e	RL10	163	glb		+			+		±	None
L15e	RL15	184	glb+ext	+	±	±	±	+			(L44e), L7ae
L37ae	RL37a	72	glb+ext		+	+	+				L2

*All entries so designated describe proteins that are not fully represented in the electron density maps described here. The summary information provided is derived from literature sources and is included here for completeness only. †The structure available for this protein in isolation does not include the extension(s) reported here.

ular protein domains. Common RNA structural motifs, such as the ribose zipper and the tetraloop-tetraloop receptor interaction, depend on this principle of adenosine packing. A manuscript in preparation describes these A-dependent interactions at greater length.

In addition to its reliance on A-dependent motifs, the tertiary structure of the domains of 23*S* rRNA and their relative positions are stabilized by familiar tertiary structure elements like pseudoknots and tetraloop-tetraloop receptor motifs (*41*, *42*). Thus, in many places, base pairs and triples stabilize the interactions of sequences belonging to different components of the secondary structure of 23*S* rRNA.

5*S* rRNA and 23*S* rRNA do not interact extensively with each other. The few RNA/RNA interactions that do occur involve the backbones of the helix 4/5 arm of 5*S* rRNA and of helix 38 of 23*S* rRNA. Most of the free energy and specificity of 5*S* rRNA binding to the large ribosomal subunit appears to depend on its extensive interactions with proteins that act as modeling clay, sticking it to the rest of ribosome.

Proteins. We have determined the structures of 27 proteins found in the large ribosomal subunit of *H. marismortui* (Table 2). Twenty-one of these protein structures have not been previously established for any homologs, and the structures of the six that do have homologs of known structure have been rebuilt into the electron density map with their *H. marismortui* sequences. In addition, there are structures available for homologs of *H. marismortui* L1, L11, and L12, which cannot be visualized in the 2.4 Å resolution electron density map. Only the structure of L10 is still unknown among the 31 proteins of this subunit.

Almost all of these structures are complete. Yet, an entire domain of L5 is missing from the electron density, presumably because of disorder. Further, L32e is also noteworthy. Its NH_2-terminal 97 residues are not seen in the electron density map, and the electron density map suggests that its COOH-terminal residue may be covalently bonded to the most NH_2-terminal of its visible residues.

Of the 30 large subunit ribosomal proteins whose structures are known, 17 are globular proteins, similar in character to thousands whose structures are in the Protein Data Bank (Table 2). The remaining 13 proteins either have globular bodies with extensions protruding from them ("glb+ext") or are entirely extended ("ext"). Their extensions often lack obvious tertiary structure and in many regions are devoid of significant secondary structure as well (Fig. 6). These extensions may explain why many ribosomal proteins have resisted crystallization in isolation. The exceptions that prove the rule are L2 and L4, both of which are proteins belonging to the "glb+ext" class. Protein L2 was crystallized and its structure solved only after its extensions had been removed (*43*), and the large loop of L4 that is extended in the ribosome is disordered in the crystal structure of intact L4 (*44*).

Except for proteins L1, L7, L10, and L11, which form the tips of the two lateral protuberances, the proteins of the 50*S* subunit do not extend significantly beyond the envelope defined by the RNA (Fig. 7). Their globular domains are found largely on the particle's exterior, often nestled in the gaps and crevices formed by the folding of the RNA. Thus, unlike the proteins in spherical viruses, the proteins of the large ribosomal subunit do not form a shell around the nucleic acid with which they associate, and unlike the proteins in nucleosomes, they do not become surrounded by nucleic acid, either. Instead, the proteins act like mortar filling the gaps and cracks between "RNA bricks."

The distribution of proteins on the subunit surface is nearly uniform, except for the active site cleft and the flat surface that interacts with the 30*S* subunit. In the crown view, the proteins lie around at the periphery of the subunit (Fig. 7A), but when viewed from the side opposite the 30*S* subunit binding site (the "back side"), they appear to form an almost uniform lattice over its entire surface (Fig. 7B). Similarly, the bottom surface of the subunit, which includes the exit of polypeptide tunnel, is studded with proteins (Fig. 7C). Indeed, the six proteins that surround the tunnel exit may play a role in protein secretion because they are part of the surface that faces the membrane and the translocon when membrane and secreted proteins are being synthesized (*45*).

Although Fig. 7 shows protein chains disappearing into the ribosome interior, the degree to which proteins penetrate the body of the particle can be fully appreciated only when the RNA is stripped away. The interior of the particle is not protein-free, but it is protein-poor compared with the surface of the particle. Extended tentacles of polypeptide, many of which emanate from globular domains on the surface, penetrate into the interior, filling the gaps between neighboring elements of RNA secondary structure (Fig. 8E). The bizarre structures of these extensions are explained by their interactions with RNA. A detailed analysis of these proteins and their interactions with RNA will be presented elsewhere.

Although extended, nonglobular structures are rare in the protein database, they are not unknown. Extended protein termini often form interprotein contacts, e.g., in viral capsids, presumably adopting fixed structures only upon capsid formation (*46*). The basic

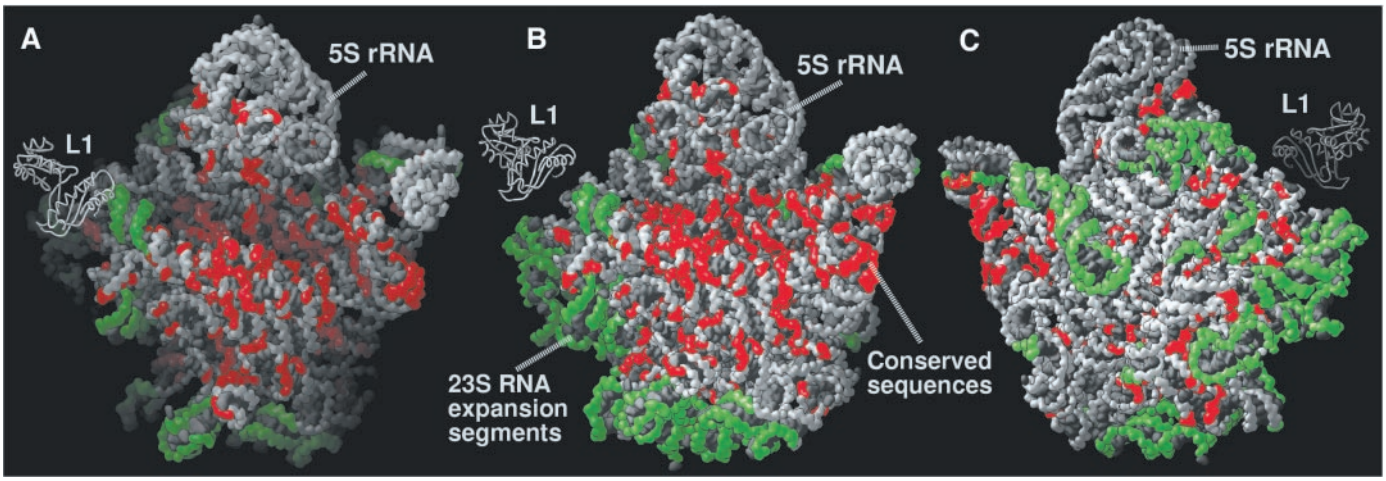

Fig. 5. Conserved residues and expansion sequences in the 23*S* rRNA of *H. marismortui*. The general, nonconserved RNA in these images is gray. Sequences that are found to be >95% conserved across the three phylogenetic kingdoms are shown in red. Sequences where expansion in the basic 23*S* structure is permitted are shown in green (*65*). (**A**) The particle rotated with respect to the crown view so that its active site cleft can be seen. (**B**) The crown view. (**C**) The back view of the particle, i.e., the crown view rotated 180° about its vertical axis.

"tails" of histones may behave the same way when nucleosomes form (*47*). The NH_2-terminal sequences of capsid proteins are often positively charged, and in virus crystal structures, the electron density for these sequences often disappears into the interior of the virus where they presumably interact with asymmetrically arranged nucleic acid. The interactions observed in the ribosome could be useful models for these viral interactions.

The interactions between extended polypeptides and RNA in the large subunit, which stabilize its massive nucleic acid structure, result in an intertwining of RNA and protein in the center of the subunit (Fig. 8, A and B). It is hard to imagine such an object assembling from its components efficiently in anything other than a highly ordered manner. Chaperones may well be required to prevent the aggregation of the extended regions of these proteins, which are likely to be disordered outside the context provided by rRNA, and to manage the folding of rRNA.

Mutations in some ribosomal proteins render bacteria resistant to certain antibiotics. One such example is a deletion of three amino acids in the β hairpin loop of protein L22 that renders bacteria resistant to erythromycin (*48*). Because this β hairpin is forming part of the surface of the tunnel wall, the mutation changes the surface properties of the polypeptide exit tunnel and may prevent the antibiotic from binding; alternatively, the mutation could be acting indirectly through RNA.

Protein and RNA interactions. Because protein permeates the large subunit extensively, there are only a few segments of the 23*S* rRNA that do not interact with protein at all. Of the 2923 nucleotides in 23*S* rRNA, 1157 make at least van der Waals contact with protein (Fig. 8D), and there are only 10 sequences longer than 20 nucleotides in which no nucleotide contacts protein. The longest such sequence contains 47 nucleotides, and is the part of domain IV that forms the ridge of the active site cleft.

The extent of the interactions between RNA and protein that occur when the large subunit assembles can be estimated quantitatively. Using the Richards algorithm (*49*) and a 1.7 Å radius probe to compute accessible surface areas, it can be shown that 180,000 $Å^2$ of surface become buried when the subunit forms from its isolated, but fully structured components. This is about half their total surface area. The average is about 6000 $Å^2$ per protein. Although this is an enormous amount compared with the surface buried when most protein oligomers form, it should be recognized that ribosome assembly must be accompanied by a large loss in conformational entropy that does not occur when most proteins oligomerize. The extended protein termini and loops of the ribosomal proteins are almost certainly flexible in isolation, and in the absence of protein, the RNA is probably quite flexible as well. Thus, the burial of a large amount of surface area may be required to provide the free energy required to immobilize the structures of these molecules.

All of the proteins in the particle except L12 interact directly with RNA, and all but 7 of the remaining 30 proteins interact with two rRNA domains or more (Table 2). The "champion" in this regard is L22, which is the only protein that interacts with RNA sequences belonging to all six domains of the 23*S* rRNA (Fig. 8C). The protein-mediated interactions between 5*S* rRNA and 23*S* rRNA are particularly extensive. Protein L18 attaches helix 1 and helix 2/3 of 5*S* rRNA to helix 87 of 23*S* rRNA. Protein L21e mediates an interaction between the same part of 5*S* rRNA and domains II and V. Protein L30 binds helix 4/5 region of 5*S* RNA to domain II. Loop C is linked to domain V by protein L5, and loop D is attached to domains II and V by protein L10e. Whatever else they may do, it is evident that an important function of these proteins is stabilization of the relative orientations of adjacent RNA domains. Several also help secure the tertiary structures of the domains with which they interact.

Because most ribosomal proteins interact with many RNA sequences and the number of proteins greatly exceeds the number of RNA domains, it can hardly come as a surprise that every rRNA domain interacts with multiple proteins (Table 2). Domain V, for example, interacts with 15 proteins, some intimately and a few in passing.

It is clear that the oligonucleotide binding experiments long relied on for information about the RNA binding properties of ribosomal proteins have underestimated their potential for interacting with RNA. The high-affinity RNA binding site identified on a protein by such an experiment may indeed be

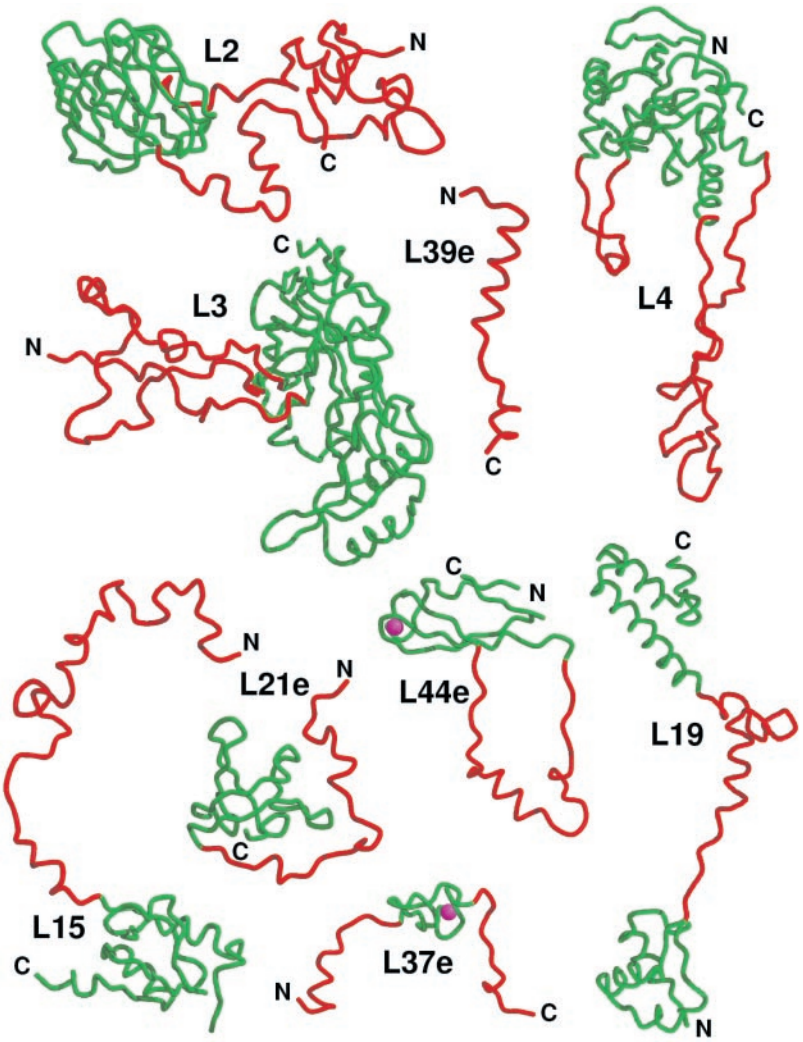

Fig. 6. The backbone structures of some large subunit ribosomal proteins that have nonglobular extensions. The globular domains of these proteins are shown in green, and their nonglobular extensions are depicted in red. The positions of the zinc ions in L44e and L37e are indicated by large dots in red.

important for ribosome assembly, but its many, weaker interactions with other sequences are likely to be missed, and they too may be vital for ribosome structure. Most ribosomal proteins crosslink RNA, and crosslinking is impossible without multiple interactions. Similar considerations may apply to proteins that are components of other ribonucleoproteins, such as the spliceosome.

Of the seven proteins that interact with only one domain, three (L1, L10, and L11) participate directly in the protein synthesis process. Rather than being included in the ribosome to ensure that the RNA adopts the proper conformation, it seems more appropriate to view the RNA as being structured to ensure the correct placement of these proteins. Another three (L24, L29, and L18e) interact with several secondary structure elements within the domains to which they bind, and presumably they function to stabilize the tertiary structures of their domains. The last of the single RNA domain proteins, L7ae, is puzzling. It cannot function as an RNA stabilizing protein because it interacts with only a single sequence in domain I, but it is far from the peptidyl transferase and factor binding sites. It is quite close to L1, however, which appears to be important for E-site function (*50*), and maybe it is involved in that activity. It could also be involved in the 70*S* assembly, because L7ae was originally assigned as a small subunit protein (HMS6).

While many ribosomal proteins interact primarily with RNA, a few interact significantly with other proteins. The most striking

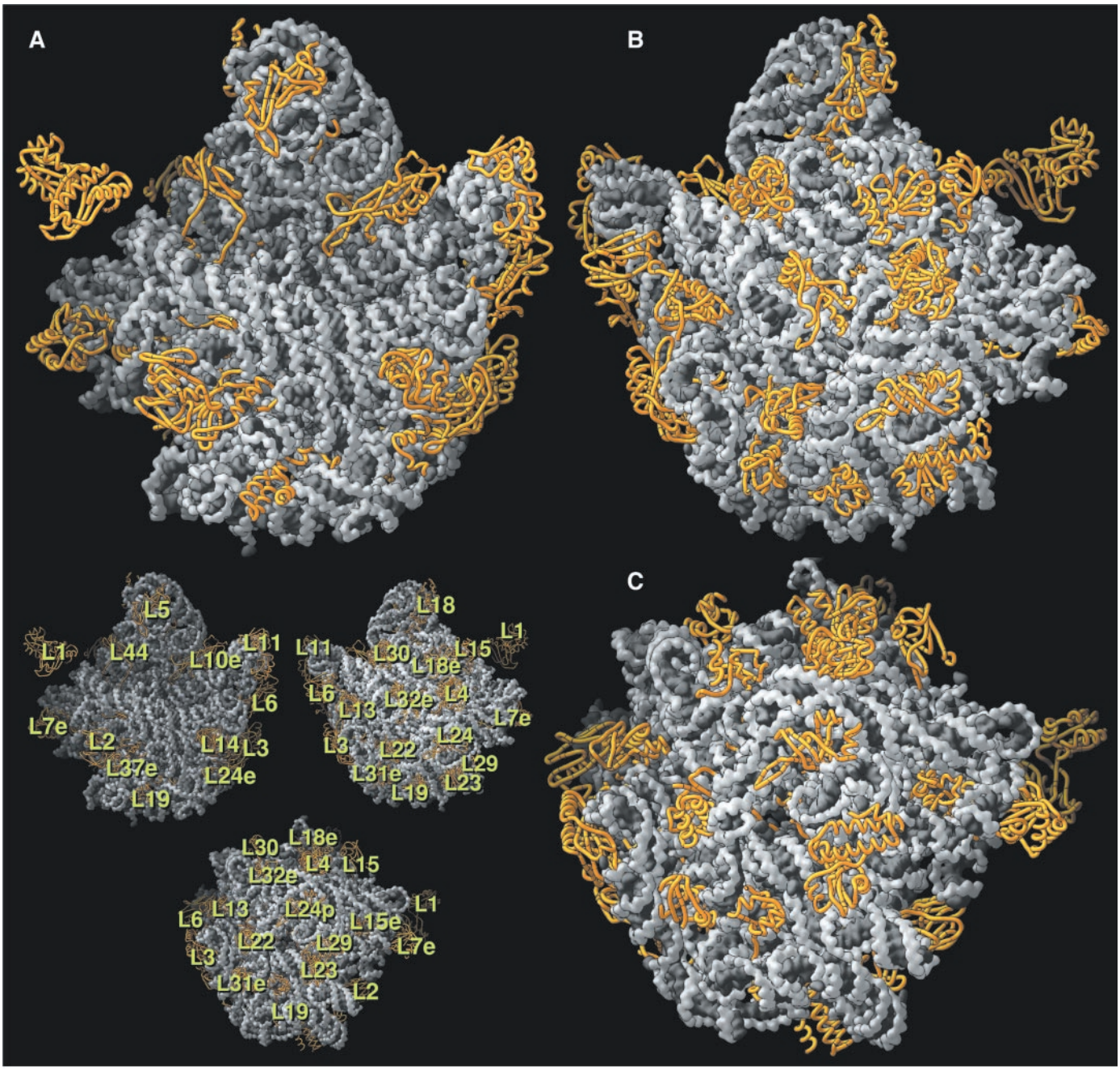

Fig. 7. Proteins that appear on the surface of the large ribosomal subunit. The RNA of the subunit is shown in gray and protein backbones are shown in gold. (**A**) The crown view of the subunit. (**B**) The back side of the subunit in the 180° rotated crown view orientation. (**C**) A view from the bottom of the subunit down the polypeptide tunnel exit which lies in the center. The proteins visible in each image are identified in the small images at the lower left of the figure. Figures were generated using RIBBONS (*67*).

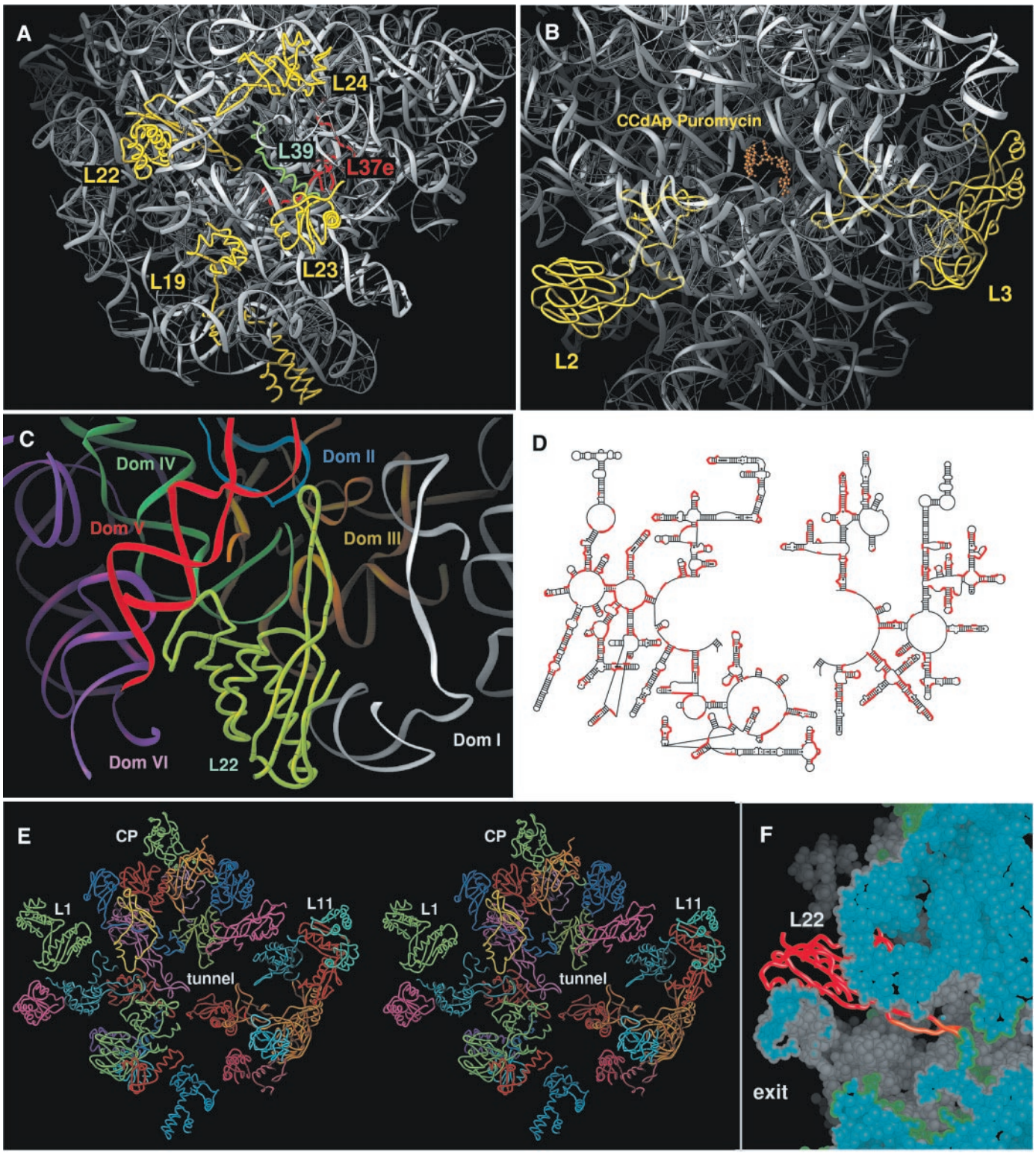

Fig. 8. The protein extensions into the RNA and multiple domain interactions. (**A**) Some of the proteins in the neighborhood of the polypeptide tunnel exit, showing the unusual extended structure of L39e (green) that enters the tunnel and L37e (red) that interpenetrates the RNA. L29, which is on top of L37e, has been removed. Protein L22 extends a long β hairpin extension inside the 23*S* rRNA. L24 has a similar extension but the entire protein is on the surface of the particle. L39 is the only protein in the subunit that lacks tertiary structure, whereas L37e has both NH_2- and COOH-terminal extensions. L19 is unique in having two globular domains on the surface of the subunit connected by an extended sequence that weaves through the RNA, shown as gray ribbons. (**B**) The nonglobular extensions of L2 and L3 reaching through the mass of 23*S* rRNA toward the peptidyl transferase site, which is marked by a CCdAp-puromycin molecule, the Yarus inhibitor (*64*). (**C**) L22 interacting with portions of all six of the domains of 23*S* rRNA. (**D**) Schematic of the 23*S* rRNA secondary structure showing the locations sequences (red) that make contact with protein. (**E**) Stereo view of the proteins of the large ribosomal subunit without the RNA. Proteins are colored as an aid to visualization only. (**F**) A cross section of the subunit in the area of the tunnel exit. Protein L22 is shown as ribbons in red, and the β hairpin loop where mutations confer erythromycin resistance is in orange. Atoms on the surface are gray, protein atoms are green, and atoms at the slice interface are blue.

structure generated by protein-protein interactions is the protein cluster composed of L3, L6, L13, L14, and L24e that is found close to the factor binding site. The surface of these proteins provides important interactions with factors. It may prove to be more generally the case that ribosomal proteins interacting primarily with RNA are principally stabilizing RNA structure, whereas some of those showing extensive protein-protein interactions may have additional binding functions.

The structure presented above illuminates both the strengths and weaknesses of approaches to complex assemblies that depend on determining the structures of components in isolation. The structures of the globular domains of homologs of the proteins in the large ribosomal subunit from *H. marismortui* are largely the same as those of the corresponding domains in the intact subunit, though adjustments in domain positions are sometimes required. Consequently, these structures were very useful for locating proteins and interpreting lower resolution electron density maps. However, for obvious reasons, the structures of the extended tails and loops of ribosomal proteins cannot be determined in the absence of the RNAs that give them structure, and the feasibility of strategies that depend on producing low–molecular weight RNA-protein complexes that have all the RNA contacts required to fix the structures of such proteins seems remote. The structures of RNA fragments also depend on their context. Whereas the sarcin/ricin loop has much the same structure in isolation (*33, 34*) as it does in the ribosome, the structure of 5*S* rRNA in isolation (*37*) differs in some respects from what is seen in the ribosome, and the structure of the isolated P loop (*51*) shows no resemblance to the structure of the P loop in the ribosome. Clearly, a "structural genomics" approach to the ribosome, which would have entailed determining the structures of all of the proteins and all possible rRNA fragments, neither would have provided the relevant structures of all of the pieces nor would it have shown their relative positions. Indeed, the structure of the large ribosomal subunit highlights the importance of structural studies of entire assemblies that show biological activity.

The analysis of the 50*S* ribosomal subunit structure presented here describes the overall architectural principles of RNA folding and its interaction with proteins, but many exciting details remain to be explored. The principles of protein-RNA interaction that should emerge from the 27 protein complexes with RNA have yet to be developed. On average, each of the 27 proteins has 3000 Å^2 of surface area in contact with RNA, which is comparable to the 2700 Å^2 of glutaminyl-tRNA synthetase that contact tRNA^{Gln} (*52*), so the number of interactions between RNA and protein to be analyzed in the large subunit structure is 30 times the number in this synthetase complex. Further, because the RNA structure of the large subunit will increase the RNA structural database by a factor of 4 to 5, most of the important RNA secondary and tertiary structural motifs to be found in nature may be represented. It will be interesting to see whether a complete analysis of this RNA structural database will enable the prediction of structures for other RNA sequences. Unknown at this time is the ease with which it will be possible to model by sequence homology the 50*S* ribosomal subunit rRNA from other species and kingdoms. However, the extensive sequence conservation in the 23*S* rRNA that forms the core active site and peptide tunnel regions suggests that reasonably accurate homology modeling based on this *H. marismortui* subunit structure may be feasible. Finally, enormous numbers of monovalent and divalent metal ions as well as water molecules are visible in this map. Analysis of their interactions with RNA should elucidate their roles in the formation and stabilization of RNA structure.

References and Notes

1. R. A. Garrett *et al.*, Eds., *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions* (American Society for Microbiology, Washington, DC, 2000).
2. B. Wittmann-Liebold, in *Structure, Function, and Genetics of Ribosomes*, B. Hardesty and G. Kramer, Eds. (Springer-Verlag, New York, 1986), pp. 326–361.
3. R. K. Agrawal, P. Penczek, R. A. Grassucci, J. Frank, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6134 (1998); H. Stark, M. V. Rodnina, H.-J. Wieden, M. van Hell, W. Wintemeyer, *Cell* **100**, 301 (2000).
4. F. Mueller *et al.*, *J. Mol. Biol.* **298**, 35 (2000).
5. A. Yonath *et al.*, *Biochem. Int.* **2**, 428 (1980).
6. S. D. Trakanov *et al.*, *FEBS Lett.* **220**, 319 (1987).
7. N. Ban *et al.*, *Cell* **93**, 1105 (1998).
8. N. Ban *et al.*, *Nature* **400**, 841 (1999).
9. J. H. Cate, M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, H. F. Noller, *Science* **285**, 2095 (1999).
10. W. M. Clemons Jr. *et al.*, *Nature* **400**, 833 (1999).
11. A. Tocilj *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14252 (1999).
12. *H. marismortui* (American Type Culture Collection 43049) was grown, and ribosomes were prepared from it as described previously (*7*, *53*). The buffers and precipitants used for crystal growth were those described earlier (*7*, *54*), but with the following modifications. A crystallization solution was obtained by back-extraction of precipitated subunit at saturation into the crystallization buffer [1.2 M KCl, 0.5 M NH_4Cl, 100 mM Kacetate, 30 mM $MgCl_2$, 7% polyethylene glycol (PEG) 6000, 15 mM tris, 15 mM MES, and 1 mM $CdCl_2$ (pH 7.1)]. The crystals that resulted had maximum dimensions of 0.5 mm by 0.5 mm by 0.2 mm and were harvested after ~2 weeks. Crystals were stabilized by gradual transfer into a solution containing 12% PEG 6000, 22% ethylene glycol, 1.7 M NaCl, 0.5 M NH_4Cl, 100 mM potassium acetate, 30 mM $MgCl_2$, and 1 mM $CdCl_2$ (pH 6.2) at 4°C. The crystals were flash-frozen in liquid propane.
13. All data, except the two native data sets, were collected at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) from crystals frozen at 100 K using beamlines X12b and X25 and recorded using a 345-mm MAR imaging plate. For each heavy-atom derivative, anomalous diffraction data were collected at the wavelength corresponding to the peak anomalous scattering, except for uranium, where a low energy was used. The beam size was 100 μm by 100 μm for most data collections at beamline X25 and 200 μm by 200 μm at beamline X12b. The crystals were aligned along the long axis of the unit cell (~575 Å) so that 1.0° oscillations could be used to collect reflections out to a maximum of 2.7 Å resolution at the edge of the MAR detector. At beamline X12b, the crystal-to-detector distances were varied depending on wavelength, crystal quality, and beam divergence so that maximum resolution data could be collected while avoiding overlapping of spots. At beamline X25, the detector was positioned on a rigid platform at 480 mm, which allowed data collection to 3.2 Å for iridium and osmium derivatives with the wavelength set at the anomalous edge. Native data to 2.4 Å resolution were collected at the structural biology beamline 19ID of the Advanced Photon Source (Argonne National Laboratory, Argonne, IL) using a 3 by 3 charge-coupled device (CCD) detector, an 80 μm by 80 μm beam size, and 0.4° oscillations. Data sets were processed with DENZO and SCALEPACK (*55*).
14. Heavy-atom–based phasing was extended to 3.2 Å resolution by combining MIR (multiple isomorphous replacement) phases calculated for two different isomorphous groups of data (MIR1 and MIR2) (Table 1) with single-wavelength anomalous dispersion (SAD) phases. Weights on the individual phase sets were adjusted on the Hendrickson-Lattman coefficients to prevent phase bias due to nonisomorphism. The best two derivatives were osmium pentamine and iridium hexamine, each of which contained a large number of binding sites (Table 1). Other derivatives with smaller numbers of sites further improved map quality. All phasing was done by maximum likelihood methods implemented in CNS with the exception of the Ta_6Br_{12} derivative, which was refined in SHARP (*56*), represented as spherically averaged electron density (Table 1). Phases were improved and extended from 3.3 to 2.4 Å by solvent flipping (*15*), with a scheme that progressively refined the solvent mask from 12 to 2.6 Å detail. RNA structure was built by positioning of the individual residues, and their rotomers were picked from a library derived from high-resolution structures (M. Kjeldgaard, unpublished data). Models were built with O (*57*).
15. J. P. Abrahams and A. G. W. Leslie, *Acta Crystallogr.* **D52**, 30 (1996).
16. A. T. Brunger *et al.*, *Acta Crystallogr.* **D54**, 905 (1998).
17. D. E. Tronrud, *Methods Enzymol.* **277** 306 (1997).
18. R. R. Gutell, in *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Biosynthesis*, A. Dahlberg and R. Zimmerman, Eds. (CRC Press, Boca Raton, FL, 1996), pp. 111–128.
19. B. Wittmann-Liebold *et al.*, in *The Ribosome: Structure, Function, and Evolution*, W. E. Hill *et al.*, Eds. (American Society for Microbiology, Washington, DC, 1990), pp. 598–616.
20. M. Oakes, E. Henderson, A. Scheinman, M. Clark, J. A. Lake, in *Structure, Function and Genetics of Ribosomes*, B. Hardesty and G. Kramer, Eds. (Springer-Verlag, New York, 1986), pp. 47–67; G. Stoeffler and M. Stoeffler-Meilicke, in *Structure, Function, and Genetics of Ribosomes*, B. Hardesty and G. Kramer, Eds. (Springer-Verlag, New York, 1986), pp. 28–46.
21. G. L. Conn, D. E. Draper, E. E. Lattman, A. G. Gittis, *Science* **284**, 1171 (1999).
22. B. T. Wimberly, R. Guymon, J. P. McCutcheon, S. W. White, V. R. Ramakrishnan, *Cell* **97**, 491 (1999).
23. W. Moller and J. A. Maassen, in *Structure, Function, and Genetics of Ribosomes*, B. Hardesty and G. Kramer, Eds. (Springer-Verlag, New York, 1986), pp. 309–325.
24. J. Walleczek, D. Schuler, M. Stoeffler-Meilicke, R. Brimacombe, G. Stoeffler, *EMBO J.* **7**, 3571 (1988).
25. P. Ostergaard *et al.*, *J. Mol. Biol.* **284**, 227 (1998).
26. H. Mao, S. A. White, J. R. Williamson, *Nature Struct. Biol.* **6**, 1139 (1999).
27. I. G. Wool, Y.-L. Chan, A. Gluck, *Biochem. Cell Biol.* **73**, 933 (1995).
28. N. Nevskaya *et al.*, *Struct. Fold. Des.* **8**, 363 (2000).
29. H. Leffers, J. Kjems, L. Ostergaard, N. Larsen, R. A. Garrett, *J. Mol. Biol.* **195**, 43 (1987).
30. H. F. Noller *et al.*, in *The Ribosome: Structure, Function, and Evolution*, W. E. Hill *et al.*, Eds. (American Society for Microbiology, Washington, DC, 1990), pp. 73–92.

31. G. M. Culver, J. H. Cate, G. Zh. Yusupova, M. M. Yusupov, H. F. Noller, *Science* **285**, 2133 (1999).
32. R. A. Garrett and C. Rodriguez-Fonseca, in *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Biosynthesis*, R. A. Zimmerman and A. E. Dahlberg, Eds. (CRC Press, Boca Raton, FL, 1996), pp. 327–355.
33. A. A. Szewczak and P. B. Moore, *J. Mol. Biol.* **247**, 81 (1995).
34. C. C. Correll *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13436 (1998).
35. I. G. Wool, A. Gluck, Y. Endo, *Trends Biochem. Sci.* **17**, 266 (1992).
36. D. Moazed, J. M. Roberston, H. F. Noller, *Nature* **334**, 362 (1988).
37. C. C. Correll, B. Freeborn, P. B. Moore, T. A. Steitz, *Cell* **91**, 705 (1997).
38. S. Gerbi, in *Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Biosynthesis*, R. A. Zimmermann and A. E. Dahlberg, Eds. (CRC Press, Boca Raton, FL, 1996), pp. 71–88.
39. P. Dube *et al.*, *Structure* **6**, 389 (1998).
40. V. C. Ware *et al.*, *Nucleic Acids Res.* **22**, 7795 (1983).
41. P. B. Moore, *Annu. Rev. Biochem.* **68**, 287 (1999)
42. E. Westhof and V. Fritsch, *Structure* **8**, R55 (2000).
43. A. Nakagawa *et al.*, *EMBO J.* **18**, 1459 (1999).
44. M. Wahl, R. Huber, M. C. Wahl, *EMBO J.* **19**, 807 (2000).
45. R. Beckmann *et al.*, *Science* **278**, 2123 (1997).
46. M. G. Rossmann and J. E. Johnson, *Annu. Rev. Biochem.* **58**, 533 (1989); A. Liljas, *Int. Rev. Cytol.* **124**, 103 (1991).
47. K. Lugor *et al.*, *Nature* **389**, 251 (1997).
48. H. S. Chittum and W. S. Champney, *J. Bacteriol.* **176**, 6192 (1994).
49. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
50. R. K. Agrawal *et al.*, *J. Biol. Chem.* **274**, 8723 (1999).
51. E. V. Puglisi, R. Green, H. F. Noller, J. D. Puglisi, *Nature Struct. Biol.* **4**, 775 (1997).
52. M. A. Rould, J. J. Perona, D. Söll, T. A. Steitz, *Science* **246**, 1135 (1989).
53. A. Shevack, H. S. Gewitz, B. Hennemann, A. Yonath, H. G. Wittmann, *FEBS Lett.* **184**, 68 (1985).
54. K. vanBohlen *et al.*, *J. Mol. Biol.* **222**, 11 (1991).
55. Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, D. Bailey, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1993), pp. 52–62.
56. E. de La Fortelle and G. Bricogne, *Methods Enzymol.* **276**, 472 (1997).
57. T. A. Jones, S. Cowan, J.-Y. Zou, M. Kjeldgaard, *Acta Crystallogr.* **A46**, 110 (1991).
58. B. L. Golden, V. Ramakrishnan, S. W. White, *EMBO J.* **12**, 4901 (1993).
59. M. A. Markus, A. P. Hinch, S. Huang, D. E. Draper, D. E. Torchia, *Nature Struct. Biol.* **4**, 70 (1997).
60. M. Leijonmarck, S. Eriksson, A. Liljas, *Nature* **286**, 824 (1980).
61. C. Davies, S. W. White, V. Ramakrishnan, *Structure* **4**, 55 (1996).
62. J. Unge *et al.*, *Structure* **6**, 1577 (1998).
63. K. S. Wilson, K. Appelt, J. Badger, I. Tanaka, S. W. White, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7251 (1986).
64. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **289**, 920 (2000).
65. R. R. Gutell *et al.*, in preparation. (Data can be found at www.rna.icmb.utexas.edu.)
66. M. Symanski, T. Specht, M. C. Barciszewska, J. Barciszewski, V. A. Erdmann, *Nucleic Acids Res.* **26**, 156 (1998).
67. M. Carson, *Methods Enzymol.* **277**, 493 (1997).

68. We thank B. Freeborn for her skilled technical assistance in preparing 50*S* ribosomal subunit material and crystals. We are indebted to D. Klein, M. Lu, S. Antoć, and M. Schmeing for their help with the fitting of protein sequences into electron density. We thank M. Kjeldgaard for providing us with a prerelease version of O adapted for RNA model building, J. Cate for contributing the iridium hexamine, R. Gutell and J. Cannone for preparing the secondary structure diagram for 23*S* rRNA, M. Wahl for sending us coordinates for L4 before their release, J. Williamson for sending us coordinates of the L30e-RNA complex before their release, R. Sweet, L. Berman, and M. Capel for their assistance with data collection at the National Synchrotron Light Source, and A. Joachimiak and the staff of 19ID at the Advanced Photon Source. Supported by grants from NIH to T.A.S. (GM22778) and P.B.M. (GM54216) and by a grant from the Agouron Institute to T.A.S. and P.B.M. N.B. is supported by a Burroughs Welcome Fund Career Award. Complete coordinates for 23*S* and 5*S* rRNAs and α carbon coordinates for the 27 proteins discussed have been deposited in the Protein Data Bank. The accession number is 1FFK for the amplitudes, experimental phases, and coordinates.

29 June 2000; accepted 24 July 2000

The Structural Basis of Ribosome Activity in Peptide Bond Synthesis

Poul Nissen,[1]* Jeffrey Hansen,[1]* Nenad Ban,[1]* Peter B. Moore,[1,2] Thomas A. Steitz[1,2,3]

Using the atomic structures of the large ribosomal subunit from *Haloarcula marismortui* and its complexes with two substrate analogs, we establish that the ribosome is a ribozyme and address the catalytic properties of its all-RNA active site. Both substrate analogs are contacted exclusively by conserved ribosomal RNA (rRNA) residues from domain V of 23*S* rRNA; there are no protein side-chain atoms closer than about 18 angstroms to the peptide bond being synthesized. The mechanism of peptide bond synthesis appears to resemble the reverse of the acylation step in serine proteases, with the base of A2486 (A2451 in *Escherichia coli*) playing the same general base role as histidine-57 in chymotrypsin. The unusual pK_a (where K_a is the acid dissociation constant) required for A2486 to perform this function may derive in part from its hydrogen bonding to G2482 (G2447 in *E. coli*), which also interacts with a buried phosphate that could stabilize unusual tautomers of these two bases. The polypeptide exit tunnel is largely formed by RNA but has significant contributions from proteins L4, L22, and L39e, and its exit is encircled by proteins L19, L22, L23, L24, L29, and L31e.

It has been known for 35 years that the peptidyl transferase activity responsible for the peptide bond formation that occurs during messenger RNA (mRNA)–directed protein synthesis is intrinsic to the large ribosomal subunit (*1–4*), and it has been understood for even longer that the ribosome contains proteins as well as RNA. In bacteria, for example, the large ribosomal subunit contains ~35 different proteins and two RNAs (*5, 6*). These findings pose three related questions: (i) which of the macromolecular components of the large ribosomal subunit contribute to its peptidyl transferase site, (ii) where is that site located, and (iii) how does it work?

By 1980, the list of components that might be part of the ribosome's peptidyl transferase center had been reduced to about a half dozen proteins and 23*S* rRNA [for reviews, see (*7, 8*)]. Following the discovery of catalytic RNAs (*9, 10*), the hypothesis that 23*S* rRNA might be its sole constituent, which had been proposed years earlier (*11*), began to gain favor. In 1984, Noller and colleagues published affinity-labeling results that showed that U2619 and U2620 (U2584 and U2585, respectively, in *E. coli*; hereafter, bases in parenthesis indicate the corresponding position in *E. coli* rRNA) are adjacent to the CCA-end of P site–bound transfer RNA (tRNA) (*12, 13*). These nucleotides are part of a highly conserved internal loop in the center of domain V of 23*S* rRNA. The hypothesis that this loop is intimately involved in the peptidyl transferase activity was supported by the observation that mutations in that loop render cells resistant to many inhibitors of peptidyl transferase, and evidence implicating it in this activity has continued to mount (*14, 15*).

Definitive proof that the central loop in domain V is the sole component of the ribosome involved in the peptidyl tranferase activity has remained elusive, however. In the 1990s, Noller and colleagues prepared particles that retain peptidyl transferase activity by increasingly vigorous deproteinizations of large ribosomal subunits, but active particles that were completely protein-free could not be produced (*16, 17*). Nevertheless, combined with earlier reconstitution results (*18*), this work reduced the number of proteins that might be involved to just two: L2 and L3 (*19*). More recently, Watanabe and co-workers reported success in eliciting peptidyl transferase activity from in vitro–synthesized, protein-free 23*S* rRNA (*20, 21*), but their observations have not withstood further scrutiny (*22*). Thus, the question still re-

[1]Department of Molecular Biophysics and Biochemistry and [2]Department of Chemistry, Yale University, and [3]Howard Hughes Medical Institute, New Haven, CT 06520–8114, USA.

*These authors contributed equally to this work.