RNA sectors and allosteric function within the ribosome

Allison S. Walker[a,1], William P. Russ[b], Rama Ranganathan[c,d,e,1], and Alanna Schepartz[a,f,g,h,1]

[a]Department of Chemistry, Yale University, New Haven, CT 06520; [b]Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390; [c]Center for Physics of Evolving Systems, The University of Chicago, Chicago, IL 60637; [d]Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL 60637; [e]The Pritzker School for Molecular Engineering, The University of Chicago, Chicago, IL 60637; [f]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520; [g]Department of Chemistry, University of California, Berkeley, CA 94720; and [h]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Contributed by Alanna Schepartz, June 18, 2020 (sent for review June 10, 2019; reviewed by Jason W. Chin, Lila M. Gierasch, and Loren Dean Williams)

The ribosome translates the genetic code into proteins in all domains of life. Its size and complexity demand long-range interactions that regulate ribosome function. These interactions are largely unknown. Here, we apply a global coevolution method, statistical coupling analysis (SCA), to identify coevolving residue networks (sectors) within the 23S ribosomal RNA (rRNA) of the large ribosomal subunit. As in proteins, SCA reveals a hierarchical organization of evolutionary constraints with near-independent groups of nucleotides forming physically contiguous networks within the three-dimensional structure. Using a quantitative, continuous-culture-with-deep-sequencing assay, we confirm that the top two SCA-predicted sectors contribute to ribosome function. These sectors map to distinct ribosome activities, and their origins trace to phylogenetic divergences across all domains of life. These findings provide a foundation to map ribosome allostery, explore ribosome biogenesis, and engineer ribosomes for new functions. Despite differences in chemical structure, protein and RNA enzymes appear to share a common internal logic of interaction and assembly.

translation | synthetic biology | ribosome evolution | genetic code expansion

Allostery is a fundamental element of complex macromolecular function, representing the essential process by which events occurring at one site are transmitted to distal sites to regulate activity (1). Allosteric mechanisms have been studied using both experiment (2, 3) and computation (4). While multiple studies recognize that allostery contributes to complex function in RNAs such as riboswitches (5) and the spliceosome (6), most studies of allostery focus on proteins. Thus, the capacity of RNA structures to support allosteric function remains underappreciated (5). Indeed, the fundamental discovery that small ribozymes populate unique and well-defined tertiary structures that support function (7, 8) and the more recent structure of the ribosome, revealing an active site composed entirely of RNA (9), reinforce the parity of protein and RNA structural and functional complexity. Despite this conceptual similarity, the mechanisms that underlie the propagation of signals within structured RNAs remain enigmatic (10). Mapping allosteric pathways in large, functional RNAs such as the ribosome could improve our understanding of the fundamental energetic relationships that support translation, inform current models of ribosome biogenesis, evolution (11), and assembly (12), deepen our understanding and prediction of cooperative interactions between ribosome-binding antibiotics in drug discovery (13), and improve our ability to engineer RNA macromolecular machines for desirable new functions (14).

How can we investigate allosteric mechanisms within structured RNAs such as the ribosome? In proteins, powerful insights into allosteric mechanisms have accrued from statistical analysis of amino acid coevolution over a deep sampling of homologous sequences. In general, coevolution methods fall into two categories—those designed to identify local contacts in tertiary structure [direct coupling analysis, or DCA and its derivatives (15, 16)] and those designed to identify global networks of

evolutionarily correlated positions [statistical coupling analysis, or SCA (17)]. The two methods are complementary, as they sample different parts of the information content within protein sequences and uncover different elements of protein structure. For example, DCA can map nontrivial tertiary contacts in proteins from sequence statistics, in some cases supplying sufficient information to deduce a three-dimensional (3D) structure to good accuracy (18). In contrast, SCA identifies networks of coevolving residues that support protein function and underlie the capacity to evolve. These networks (termed "sectors") usually comprise a small fraction of the total protein sequence, form contiguous networks in tertiary structure, and often display an architecture in which residues in a catalytic or ligand binding site connect to distant, functionally important allosteric sites (17, 19). Sectors have been connected to allosteric mechanisms in several model proteins (20–22), and employed to engineer new allosteric control (23, 24) and design new synthetic proteins that fold and display in vitro biochemical functions that are indistinguishable from their natural counterparts (25). Although conservation-based approaches have been used to identify functionally important regions of the ribosome (26), and several coevolution-based approaches including DCA have been used to predict secondary and tertiary contacts in RNAs as large as the 16S ribosomal RNA (rRNA) (1,543 nucleotides) (16), global coevolution methods such as SCA have not, to our knowledge, been

Significance

The ribosome is a large, complex, and essential biomolecular machine. Despite its importance in all domains of life, the long-range interactions that regulate ribosome function are poorly understood. Here we report that statistical coupling analysis (SCA), a method previously only applied to proteins, can identify networks of coevolving bases within the 23S rRNA of the large ribosomal subunit. Using a continuous culture assay, we show that 23S rRNA bases predicted to be important network members cause growth defects when mutated. These findings inform our understanding of long-range functional couplings in the ribosome and provide a tool for studying the same types of interactions in other large RNAs.

applied to predict long-range allosteric networks in large functional RNA molecules.

In this work, we extend the SCA method for application to large functional RNA molecules and validate its ability to predict previously unrecognized functional relationships within the 23S rRNA of the bacterial 70S ribosome (Fig. 1A). We chose the 23S subunit because it comprises the entire peptidyl transferase center (PTC), where peptide bond formation occurs, and is a promising target for ribosome engineering efforts (27, 28). We show that SCA exposes several distinct groups of coevolving bases (RNA "sectors") within the 23S subunit. One of these sectors contains bases that surround the PTC and extends to functionally significant regions located as far as 117 Å away. To experimentally test the sequence-based predictions, we generated a library of *Escherichia coli* harboring multiple single 23S rRNA mutations and developed a quantitative and high-throughput continuous-culture-with-deep-sequencing assay to probe the effects of these mutations on ribosome function. This assay revealed that mutations of bases in the top two sectors significantly impact fitness, while mutations at positions outside sectors 1 and 2 showed no such correlation. We used the wealth of high-resolution structural data for the ribosome to track these sectors throughout the translational cycle to learn more about their functions. As in proteins, the hierarchy of coevolution that defines RNA sectors is associated with deep phylogenetic divergences in the ribosome family, suggesting a mechanism by which changes in selective pressures during evolution were accommodated. Sector-based inferences of ribosome evolution are consistent with existing models of ribosome evolution.

Results

Identifying Coevolution Networks within the 23S RNA. We generated a multiple sequence alignment of the 2,944 23S rRNA sequences found in the Comparative RNA Web database (29), which includes sequences from bacteria (2,540), archaea (119), eukaryotes (106), chloroplasts (269), and mitochondria (16), truncated the alignment to positions present in the *E. coli* 23S sequence, and computed the extent of coevolution between all nucleotide bases using the SCA algorithm (19). We eliminated any sequence that contained gaps at more than 20% of the nucleotide positions or was truncated after position 2,800, leaving a total of 2,464 sequences. We also removed from the alignment any position that was absent in more than 20% of all sequences, leaving
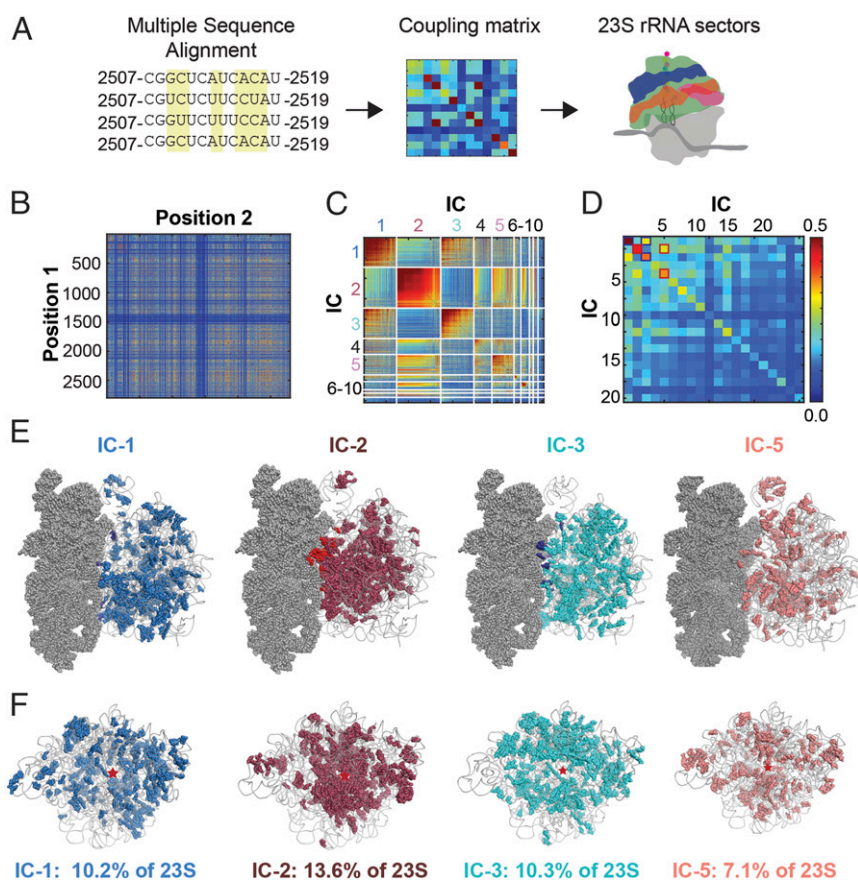
Fig. 1. Extending SCA to study RNA. (A) The input for the SCA algorithm is an MSA of 23S rRNA sequences. With this input, the algorithm generates a coupling (or SCA) matrix by computing the conservation-weighted difference between the frequency at which a specific combination of nucleotides is observed at a pair of positions and the individual frequencies. This process generates an L × L × 4 × 4 matrix (where L is the length of the sequence), which is reduced to an L × L coupling matrix using the Frobenius norm. The L × L matrix can be factored into ICs using spectral decomposition and IC analysis (19) to identify 23S rRNA sectors. (B) The positional coupling matrix generated by applying SCA to the 23S MSA, showing the extent of evolutionary coupling between pairs of positions in the ribosome as a heat map. Greater red indicates a higher degree of coevolution between a pair of positions, whereas more blue indicates lower coevolution. (C) The positional coupling matrix, clustered by the top 10 ICs predicted by SCA. (D) Plot showing the level of coupling between residues in individual ICs. Boxes outlined in blue indicate coupling between IC1 and the other ICs in sector 1; those outlined in red indicate coupling between IC2 and the other ICs in sector 2. (E) Illustration of IC-1, IC-2, IC-3, and IC-5 mapped on the *E. coli* ribosome (Protein Data Bank [PDB] ID code 5JTE). The 16S rRNA is shown as a CPK image in gray; IC1 is shown in blue, IC2 is shown in red, IC3 is shown in cyan, and IC5 is shown in pink. Residues that contact the 16S are shown in dark blue in IC1 and IC3, and in bright red for IC2. (F) View of the ICs looking down on the PTC, the location of which is approximated by a red star. See also *SI Appendix*, Fig. S1 and Tables S1 and S2.

a total of 2,798 positions in the final alignment. Since SCA computes a conservation-weighted covariance matrix that depends on background frequencies of bases as a reference, we modified the code to include these values, which we computed from the 23S alignment. The result is an analysis of pairwise coevolution between all 23S rRNA positions that emphasizes global patterns of interactions between bases (Fig. 1A and *SI Appendix*, Fig. S1A).

To examine the global coevolution of nucleotide bases in 23S rRNA, we used standard methods [spectral decomposition followed by independent component analysis (ICA) (19)] to identify groups of bases that collectively evolve. These groups are found in the top so-called independent components (ICs) of the SCA matrix (Fig. 1 B and C). As in proteins, this analysis reveals a hierarchical organization with two primary near-independent ICs (1, 2) (Fig. 1D) and a number of other components, most of which show partial correlation to IC-1 or IC-2. This organization is deeply connected to the evolutionary history and possible functional specializations of the 23S rRNA and is discussed in more detail below. The top two ICs are sparse, containing only 10.2% and 13.6% of the total 23S rRNA bases, respectively (*SI Appendix*, Fig. S1B), and, together with their associated ICs (IC-3 and IC-5, respectively), form mostly contiguous networks in the tertiary structure (*SI Appendix*, Fig. S1C and Fig. 1 E and F). Conservatively combining ICs that show the highest correlation to each other (30) (Fig. 1D), we identify two sectors in the 23S rRNA: sector 1, consisting of IC-1 and IC-3, and sector 2, consisting of IC-2 and IC-5 (Fig. 2); nonsector ICs are shown in *SI Appendix*, Fig. S2). These two sectors are mapped onto the tertiary structure of the *E. coli* ribosome in Fig. 2, and a ranked list of all positions in each sector is found in *SI Appendix*, Tables S1 and S2. We also mapped the two sectors onto the established 23S rRNA secondary structure (Fig. 3) and analyzed them with respect to A-minor interactions, the most abundant tertiary interaction in the large ribosomal subunit (31) (*SI Appendix*, Table S3). This analysis revealed that bases involved in A-minor motifs tend to segregate to one and only one sector: 15% of all A-minor motifs localize to the same sector, 38% localize the paired bases in the same sector, and 26% localize the nonpaired base in the same sector as at least one of the base-paired bases.

If the 23S RNA sectors identified by SCA identify important allosteric networks, then they should contain bases within multiple essential functional regions of the ribosome, such as the PTC, the decoding center, the exit tunnel, and the intersubunit interface. Perhaps the most fundamental ribosome function is peptide bond formation within the PTC, and, as a result, one would expect at least one sector to contain bases within this essential 23S rRNA region. Although only 2 of the 599 bases within sector 1 (2453 and 2604) are located within the core PTC (within 8 Å of the 3′ end of either A- or P-site tRNA, 32 bases total) (Fig. 2 and *SI Appendix*, Fig. S3A), 25 of the 602 bases in sector 2 are located in this region (2063 to 2065, 2251 to 2253, 2439, 2450 to 2452, 2493, 2505 to 2508, 2553 to 2554, 2574, 2583 to 2585, and 2600 to 2603) (Fig. 2 and *SI Appendix*, Fig. S3A). Five of the highly and/or universally conserved bases within the PTC (C2063, U2506, U2585, A2602, and A2451) are located within sector 2, an assignment that naturally emerges from statistical adjustments required to correct for the limited sampling of sequences (Fig. 2 legend). This assignment should be considered a hypothesis that such bases act collectively with other conserved coevolving positions in mediating fitness. We discuss the meaning of this assignment with regard to ribosome function below.

Accurate and rapid translation also requires dynamic interactions between the 23S and 16S rRNAs that link the decoding center in the small subunit with the PTC and GTPase center in the large subunit. The 23S/16S interface encompasses 8,474 Å² (2.2%) of the total 23S rRNA surface and includes direct contacts to more than 90 23S rRNA bases. Multiple bases within

sectors 1 and 2 directly contact 16S rRNA bases at the 23S/16S interface. Sector 1 contains 599 bases, of which 19 are located at the 23S/16S interface (< 8 Å away from a residue in the 16S) (Fig. 2 and *SI Appendix*, Fig. S3B). These 19 bases include 5 within intersubunit bridges B2c and B3 (positions 1832, 1948, 1949, 1950, 1958). Sector 2 contains 602 bases, of which 34 are located at the 23S/16S interface; these 34 bases include 12 within intersubunit bridges B2a/d, B2b, B2c, B3, B4, B5, and B7a (716, 1700, 1833, 1836, 1894, 1912, 1914, 1919, 1920, 1931, 1947, 1959). Sectors 1 and 2 also contain bases within other essential ribosome regions, such as within the exit tunnel (*SI Appendix*, Fig. S3C), the L1 stalk (*SI Appendix*, Fig. S3D), and the sarcin−ricin loop (SRL) that interacts with GTPases (*SI Appendix*, Fig. S3E) (see *SI Appendix*, Table S2 for a complete listing). Based on these observations, we hypothesize that sectors 1 and 2 predicted by SCA identify networks of 23S rRNA nucleotides that support fundamental and evolutionarily conserved ribosome functions.

Experimentally Verifying Coevolution Networks within the 23S RNA. To test the hypothesis that sectors 1 and 2 support essential ribosome functions, we evaluated, in high-throughput, whether SCA could accurately predict the functional relevance of bases within the ribosomal large subunit. We designed a high-throughput growth competition assay that systematically analyzed the relative effects of multiple mutations in a single, well-controlled experiment. This assay, which combines continuous culture with deep sequencing, measures the relative growth rates of cells whose ribosomes harbor unique 23S mutations by allowing the population to compete under tightly defined growth conditions; relative fitness values resulting from the 23S mutations are determined experimentally by deep sequencing the population as a function of time. As the cells contain
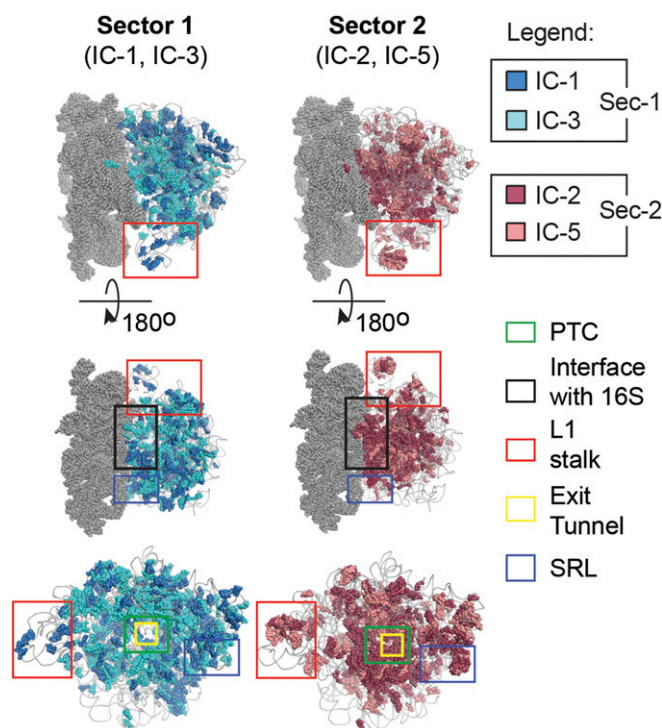
Fig. 2. Sectors contain residues within multiple essential ribosome regions. Three views of the *E. coli* ribosome illustrating the locations of residues within sectors 1 and 2. Functional ribosome regions are identified by boxes. Statistical corrections required for limited sampling of sequences will naturally cause universally conserved bases to be assigned to a sector based on their degree of conservation. This effectively represents a hypothesis that such bases act collectively with other similarly conserved positions to influence fitness. See also *SI Appendix*, Fig. S2.
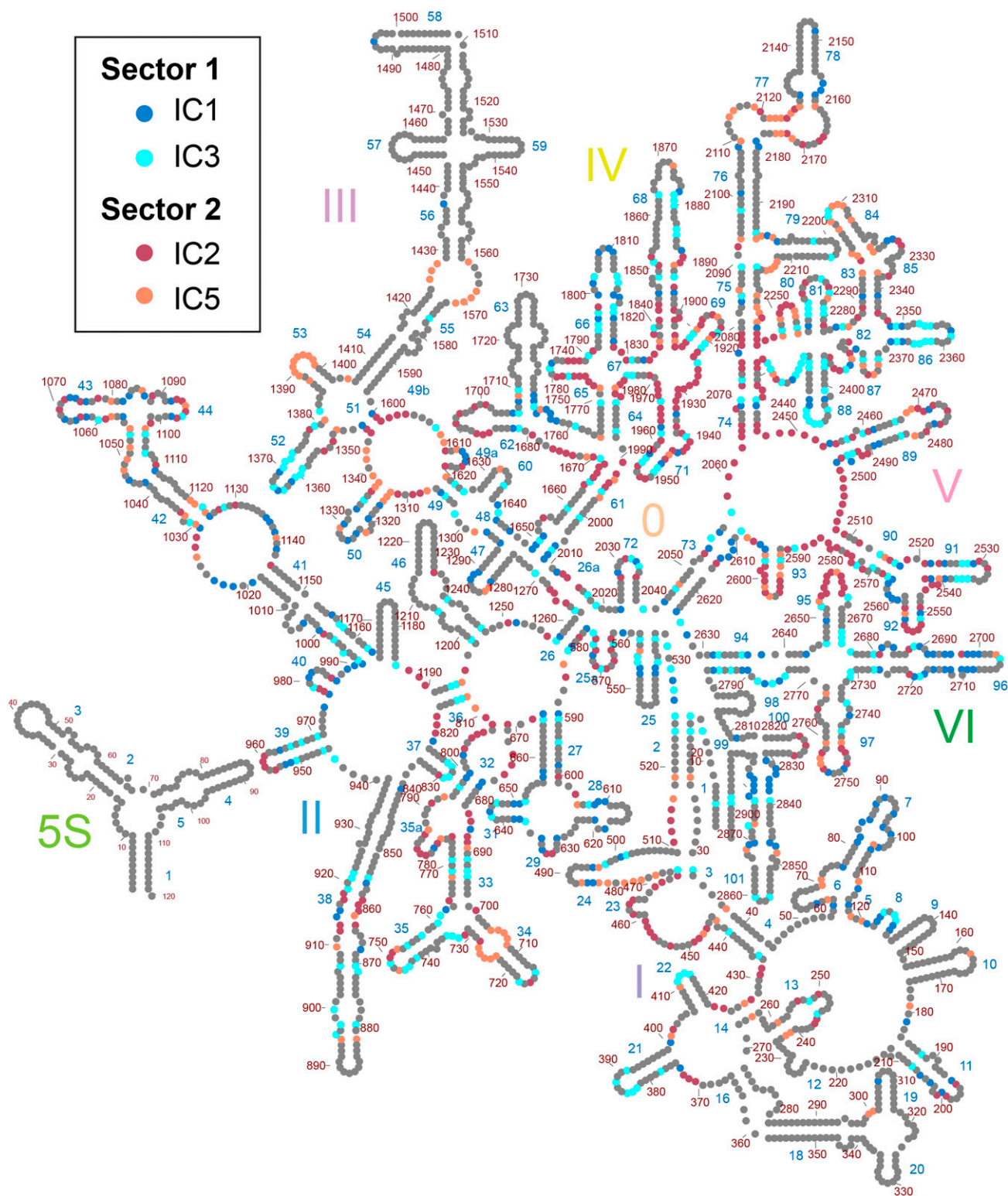
Fig. 3. Sectors mapped onto the secondary structure of the 23S rRNA. Bases in sectors 1 and 2 are identified by blue and red circles, respectively. Image created using RiboVision (32).

only mutant ribosomes, the relative growth rates provide a direct readout of ribosome function—the ability to synthesize proteins to support life and cell division (Fig. 4A).

We generated a library of single-mutant *E. coli* ribosomes containing substitutions at 24 23S rRNA bases located in or near

the exit tunnel (2056 to 2059, 2064, and 2079), L1 stalk (2110, 2112 to 2115, 2154, and 2160), P site (2249), A site (2465, 2483 to 2486, 2497, and 2499), and the PTC itself (2502, 2503, 2505, and 2507) as input for the continuous culture assay. The bases chosen include those possessing a range of predicted functional importance;
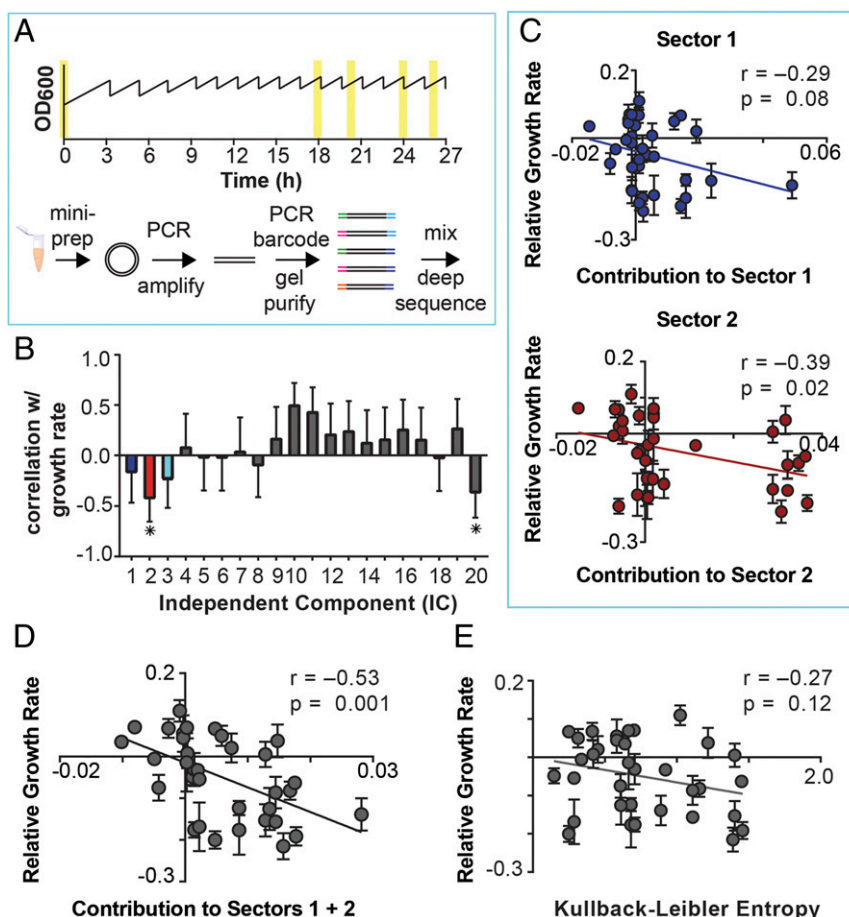
Fig. 4. A quantitative continuous-culture-with-deep-sequencing assay to validate SCA predictions. (*A*) A mixture of SQZ *E. coli* (33) harboring PLK35 plasmids with single or double mutations within the 23S rRNA gene (rrnB) (a total of 59 mutations) was maintained at $OD_{600} = 0.01$, and aliquots (5 mL) were withdrawn at 0, 18, 20, 24, and 26 h. The plasmid DNA contained in each aliquot was isolated, and the region of interest was deep-sequenced. (*B*) Mutations in residues within only certain ICs correlate negatively with growth. Bar graph showing the correlation values calculated from plots of growth rate versus the average contribution of the mutated position to each IC. Correlation values were determined from plots of relative growth rate versus contribution to IC. (*C–E*) Plots of the average relative growth rate of SQZ *E. coli* whose ribosomes contain 1 of 34 different 23S mutations (*C* and *D*) versus the weighted average contribution of the mutated residue to (*C*) sector 1 or sector 2 and (*D*) sectors 1 and 2, and (*E*) versus conservation of the mutated residue as measured by the Kullback–Leibler entropy. In all plots, r is the Pearson's correlation coefficient, a measure of the strength of the correlation, and p is the statistical significance associated with the Pearson's correlation—the probability of observing the r value if there was, in fact, no correlation between the two variables. The correlations we determined to be significant are moderate (r between −0.39 to −0.53) but have strong statistical significance ($P < 0.05$, with p as low as 0.001 for the correlation between the weighted contribution to the two sectors and growth rate). The modest value of the Pearson's correlation likely reflects the random noise in both variables (statistical correlations and experimental data).

17 are located within sectors 1 and 2, while 7 (2057, 2112, 2113, 2154, 2160, 2484, and 2486) are not. We transformed this library into the *E. coli* strain SQZ10, which lacks any genomic copies of the rrnB operon but instead harbors a single rrnB gene on a plasmid containing a sucrose sensitivity gene (Suc^S). This sensitivity allows the plasmid to be exchanged with one containing the rrnB mutant of interest (33). To measure the relative growth rates of all clones in high-throughput, we grew the SQZ10 library, along with an SQZ10 strain transformed with a plasmid encoding a wild-type (WT) rrnB gene, in a turbidostat (34) that tightly controlled the growth of the population by diluting the culture whenever it reached a specified threshold optical density (OD). This feature greatly extends the time over which population changes can be evaluated, and thus allows the assay to detect more-significant changes in mutational distributions.

The populations were evaluated by sampling the growths at 0, 18, 20, 24, and 26 h (Fig. 4*A*); the rrnB-containing plasmids were isolated, subjected to PCR to amplify the region containing mutations (2036 to 2550), and subjected to PCR again to barcode the amplicons, and the products were sequenced at high depth using MiSeq or HiSeq (*SI Appendix*, Fig. S4 *A* and *B*) (35). The frequency of each mutant relative to its frequency at the start of the experiment is a measure of how well that mutant competed for growth while incubated in the turbidostat. Four turbidostat runs were performed; in each case, the growth rate of each mutant relative to WT was obtained by calculating the slope of the log of the relative enrichment as a function of time. To utilize the highest-quality data from each run, we chose quality thresholds for the number of reads and the quality of the fit, as measured by R^2 of the growth curve for each mutant, that were necessary to ensure high replicability. In general, as either threshold was increased, the correlation between different replicates of the experiment improved (R_{12}^2) (*SI Appendix*, Fig. S4C). Based on the observed replicability, we chose to quantify fitness only for sequences represented by more than 300 reads and whose growth curve exhibited a fit with an $R^2 > 0.6$. Thirty-four mutants exceeded these thresholds and were included in our validation set. This dataset exhibited a high overall correlation between different replicates, with a Pearson's correlation coefficient of 0.95 and an R^2 of 0.90 (*SI Appendix*, Fig. S4D).

To assess how accurately SCA identifies functionally important 23S rRNA bases, we compared the relative growth rate of each mutant in the library to the mutated residue's contribution to sectors 1 and 2 using a weighted average of the contribution to the IC in each sector. Bases in sector 2 show a significant negative correlation with growth, with a Pearson's coefficient of –0.39 ($P = 0.02$) (Fig. 4C). Sector 1 is also negatively correlated with growth, with Pearson's coefficient = –0.25, but this correlation is not statistically significant ($P = 0.08$) (Fig. 4C). Notably, there is a stronger correlation between growth rate and the weighted average contribution to both sectors (Pearson's coefficient = –0.53, $P = 0.001$) (Fig. 4D). The growth rate is more correlated with contribution to SCA sectors than with conservation (as measured by the Kullback–Leibler entropy) (36) (Pearson's coefficient = –0.27, $P = 0.12$) (Fig. 4E). Finally, a plot of the relative growth rate versus the weighted average of the contribution to ICs that are not included in sectors 1 and 2 is positive (Pearson's coefficient = +0.33, $P = 0.06$) (*SI Appendix*, Fig. S4E), which means that likely only the ICs in sectors 1 and 2 play a role in essential ribosome functions (ICA analysis maximizes the independence of ICs, meaning that positions in sectors 1 and 2 will generally not contribute strongly to ICs not in the sector, leading to the positive correlation). In addition, SCA was able to predict several bases far from the PTC at which mutations significantly impacted growth rate. For example, mutation of bases 2079 and 2498, to A and T, respectively, which are both located within sector 2 but more than 20 Å from the PTC, causes a large growth defect with average relative growth rates of –0.22 and –0.19. Indeed, 2079A has the slowest growth rate of any mutant tested. We conclude that SCA provides relevant information about the importance of 23S rRNA bases to ribosome function beyond what could predicted based on the location or conservation of individual bases.

Discussion

Sectors 1 and 2 Support Subunit Association, Communication, and Translocation. One overarching function of the 23S rRNA is to interact with the 16S rRNA found in the small subunit. The structure of the *E. coli* ribosome shows 90 of the 2,904 23S bases (3%) within 8 Å of at least one 16S residue. Of these 90 bases, 53 (59%) are found within sector 1 (19 bases) or 2 (34 bases) (Fig. 5A). As only 41% of the total 23S bases are found within sectors 1 and 2, we conclude that 23S bases that contact the 16S rRNA are significantly overrepresented in the sectors. One region of the intersubunit interface, the A-site finger located in H38, contains five bases in sector 1 (871, 879, 896, 898, and 900) and two in sector 2 (881 and 895) (Fig. 5B). Previous work has shown that, while deletion of the A-site finger has a small effect on subunit association, growth rate, and the rate of peptide bond formation, it dramatically increases both +1 frameshift read through and translocation within the decoding center of the small subunit (37, 38). In addition, chemical probing experiments have revealed that deletion of the A-site finger induces structural changes in the 5S and the P loop located within the PTC (38). All of the P-loop 23S nucleotides that exhibit changes in the reactivity to chemical probes upon deletion of the A-site finger (2249, 2250, 2255, and 2529) (38) are located in sector 2 and are adjacent to bases in sector 1 (Fig. 5B). In addition, mutation of a sector 1 residue in the PTC (2453) leads to markedly diminished subunit association (39). These observations suggest that bases within sectors 1 and 2 provide an allosteric link between the intersubunit interface, the decoding center in the small subunit, and the PTC, and imply direct communication between sectors 1 and 2.

Deletions within helices H34 and H68, a second region of the intersubunit interface, have an even larger impact on ribosome function, causing a decrease in subunit association and growth (37). Deletions in H34 that negatively impact subunit association (Δ709-710 and Δ721-722) bracket positions in sector 1 (714 and

718) and sector 2 (716) (37) (Fig. 5C); deletions in H68 that significantly increase doubling time (Δ1845-1895) (37) correspond to many bases in sector 1 (1845, 1849, 1852, 1864, 1877 to 1880, 1884, 1885, 1888, 1893, and 1895) and some bases from sector 2 (1851, 1853, 1872, 1889, 1890, and 1894) (Fig. 5D). A third region of the intersubunit interface comprises H69 (bridge B2a); the two bases in this helix that are essential for ribosome function (A1912 and U1917) (41) are both located within sector 2 (Fig. 5E) and contribute strongly to that sector—these two positions are associated with the 164th and 30th highest (out of 602) eigenvalues in sector 2, respectively. The demonstrated importance of multiple bases in sectors 1 and 2 that directly contact the 16S rRNA implies that these two sectors, in particular, sector bases in H38, H34, H68, and H69, provide an
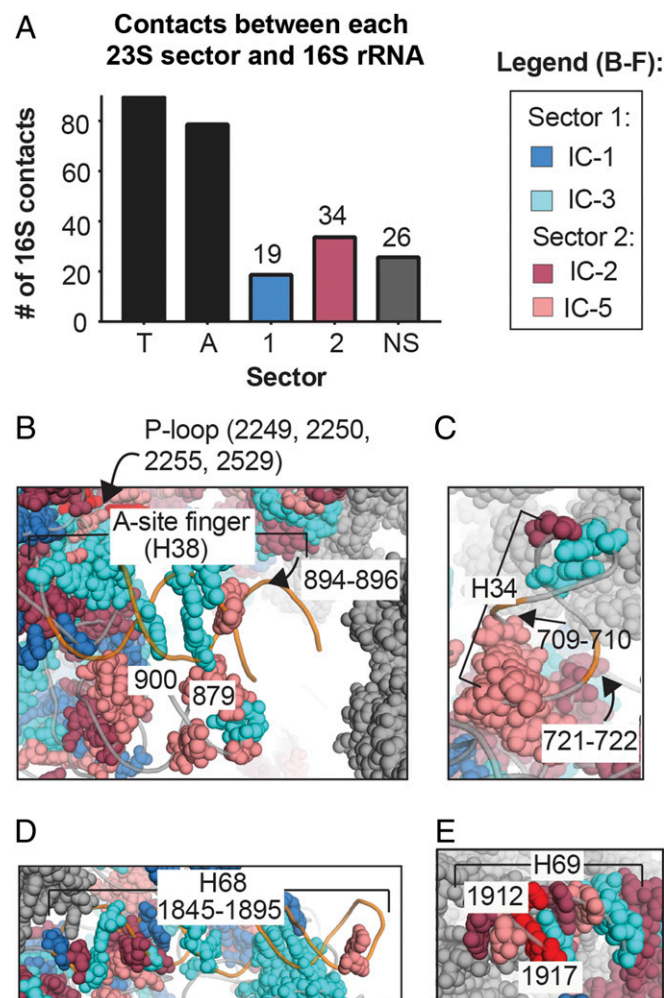
Fig. 5. Residues in sectors 1 and 2 contact and link multiple essential ribosome regions. (*A*) Bar graph showing the number of residues in each sector that are located within 8 Å of any residue in the 16S rRNA. T, total number of 23S residues that contact the 16S; A, total number of residues in all ICs that contact the 16S, NS, total number of residues in nonsector ICs that contact the 16S. Plots and images are color-coded: sector 1 (blues) and sector 2 (pinks). (*B*) Sectors surrounding the A-site finger (H38). The regions deleted in two studies (37, 38) are shown in orange; residues in sector 2 that showed changes in susceptibility to chemical probing upon deletion of these regions (38) are shown in bright red. (*C*) Sectors surrounding H34. Deletions deleterious to subunit association are shown in orange (37). (*D*) Sector residues surrounding H68. Deletions that significantly increase doubling time are shown in orange (37). (*E*) Sector residues surrounding H69. Residues in sector 2 that were intolerant to mutation (40) are shown in bright red.
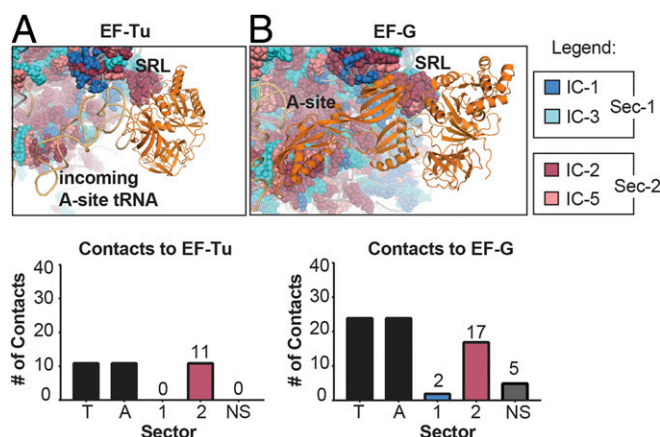
Fig. 6. Residues in sector 2 contact many core translation factors and link them to essential ribosome regions. (*A*) Image showing contacts between sectors and EF-Tu (PDB ID code 5AFI) and a bar graph showing the number of contacts to each sector. (*B*) Image showing contacts between sectors and EF-G (PDB ID code 4V7B) and a bar graph showing the number of contacts to each sector. Factors are shown in orange in both images. T, total number of 23S residues that contact the factor; A, the total number of residues in all ICs that contact the factor; NS, total number of residues in nonsector ICs that contact the factor.

allosteric link to mediate information flow between the large and small ribosomal subunits.

Sector 2 Supports Catalysis within the PTC. Peptide bonds form when the α-amino group of an aminoacyl transfer RNA (tRNA) in the A site attacks the carbonyl carbon of a peptidyl tRNA in the P site. The rate of this reaction is enhanced 10-million-fold within the PTC (42). Peptide bond formation transfers the growing peptide from the P-site tRNA to the A-site tRNA, which, after translocation, is transformed into a new peptidyl tRNA. Peptide bond formation within the PTC is facilitated by direct interactions with 23S rRNA bases within sector 2 at the center of the PTC: A2451, U2506, U2585, C2452, and A2602 (43). As noted above, the highly conserved nature of the bases 2506, 2585, and 2602 makes their assignment to sector 2 a statistical hypothesis that they work collectively with other coevolving positions within this sector. Indeed, these bases occur within the environment of many spatially proximal, moderately conserved portions of the PTC, and other experimental findings support their assignment in sector 2. In particular, all three waters in the proposed "proton wire" (44) required for proton transfer in or near the transition state are positioned by highly ranked bases in sector 2. Water 1 lies in a cavity formed by bases in sector 2, A2602 and A2451, and remains hydrogen-bonded to these bases in both the preattack and postattack state; water 2 interacts with A2602 (through N1) (sector 2) as well as U2584 (sector 2); and water 3 interacts with C2063 (sector 2).

Sector 2 Interacts with Elongation Factors. In order to learn more about the roles sectors play in the translation cycle, we examined their presence in 23S rRNA regions that associate with EF-Tu and EF-G. EF-Tu, the factor involved in tRNA accommodation into the A site (45), contacts only bases in sector 2 (Fig. 6*A*). EF-G, the factor that catalyzes translocation (46), contacts 2 bases in sector 1 and 17 bases in sector 2 (Fig. 6*B*). Interactions between sector 2 and these two elongation factors are mediated by bases in the essential SRL (45–47). These results suggest that there is a network connecting the SRL to the PTC and exit tunnel and other distant sites in the ribosome through sector 2 that allows

for communication between elongation factors and functional sites in the ribosome.

Statistical Coupling Provides Insight into the Evolution of rRNA. SCA allows for projection from sectors to the architecture of subfamilies within a multiple sequence alignment, revealing the extent to which bases in each sector vary between different phylogenetic groups (30). This analysis can provide further insight into the functional relevance of sectors (21). Examination of the projection of sectors from different phylogenetic groups on to the multiple sequence alignment (MSA) reveals that sequences separate along phylogenetic lines along coordinates for the two ICs that comprise sector 1 (IC-1 and IC-3) (Fig. 7*A*) but not sector 2 (IC-2 and IC-5) (Fig. 7*B*). All phylogenetic groups are clustered separately along the projected IC corresponding to IC-1 (*x* axis in Fig. 7*A*), while the projected IC-3 separates bacteria, mitochondria, and chloroplasts from eukaryotes and archaea (*y* axis in Fig. 7*A*). By contrast, unlike ICs that comprise sector 1, ICs in sector 2 do not clearly cluster along phylogenetic lines: All phylogenetic groups exhibit a center of mass at approximately the same coordinate along the projected ICs corresponding to sector 2 (Fig. 7*B*), although a small number of mitochondrial sequences do cluster separately from the remaining sequences (Fig. 7*B*). The lack of clustering along phylogenetic lines suggests that sector 2 is more ancient than sector 1, predating the last universal common ancestor (LUCA), and that any further evolution of these sectors was based on functional constraints. Conversely, the strong relationship between sector 1 bases and phylogeny suggests that, although initial evolution of sector 1 might predate the LUCA, its evolution continued at least through the common ancestor of each kingdom.
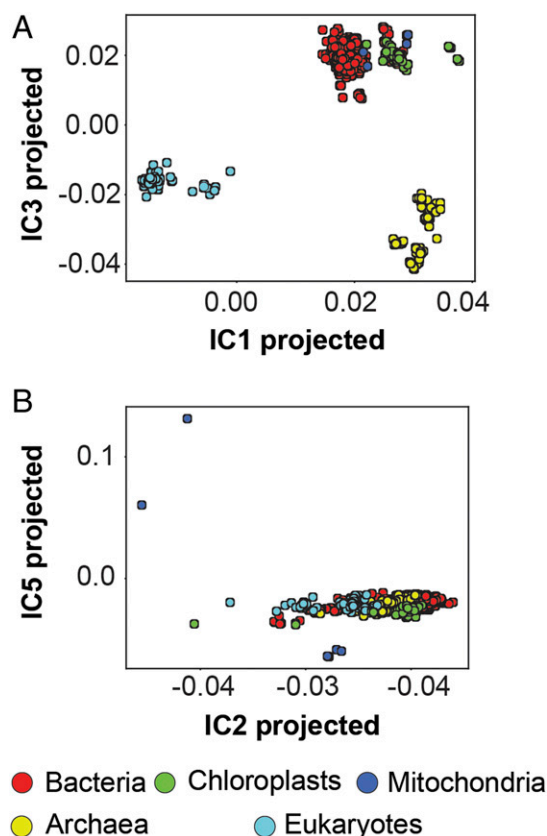
Fig. 7. Statistical coupling provides insight into the evolution of the large ribosome subunit. Projections of ICs onto sequence space (30). (*A*) Projection of IC1 versus projection of IC3. (*B*) Projection of IC2 versus projection of IC5.

BIOCHEMISTRY

CHEMISTRY

To learn more about the relative ages of the sectors, we tracked the sectors through a current model of ribosome evolution (48). In this model, the ribosome evolves by accretion, the addition of expansion segments to established helices in a manner that does not perturb the existing ribosome structure (referred to here as the accretion model) (48, 49). The accretion model was developed by identifying insertion fingerprints, junctions where the insertion of a more recent branch does not perturb the more ancient trunk. Insertion fingerprints were first identified by comparing the structures of the bacterial and eukaryotic ribosome (48). We examined the relationship of more recent evolutionary events to the sectors by comparing the structures of the *E. coli* and *Saccharomyces cerevisiae* ribosomes. We examined each of the expansion segments present in the large subunit of the yeast ribosome and used in the development of the accretion model—corresponding to insertions at H15, H25, H30, H38, H52, H54, H63, H78, H79, H98, and H101 (48). The majority of the junctions for these insertions (6 of 11) are not proximal to any residue in sectors 1 or 2. This observation is consistent with the prediction that sectors support fundamental elements of ribosome function and therefore are less likely to be the site of a potentially perturbing insertion. Sector 1 is the only sector that contains bases at the site of the remaining five insertions, which is consistent with the prediction that sector 1 is more modern than sector 2 and does not play a direct role in catalysis.

Conclusions

RNA and protein differ in many ways. RNA, unlike protein, folds through interactions of small independent and hyperstable elements, contains a charged and self-repulsive backbone, and, lacking side chains, cannot form protein-like hydrophobic cores. RNA also possesses greater backbone degrees of freedom than protein, and its local interactions are constrained by reasonably strict base-pairing interactions. Among biological RNAs, the ribosome is unique, having evolved over 3 billion to 4 billion years to execute a single chemical reaction, the templated polymerization of α-amino acids. The size and complexity of its structure and the diversity of covalent and noncovalent interactions required to execute 10 to 20 translation cycles per second with high fidelity (50) demand the coordinated interactions of bases that span more than 390,000 Å of molecular surface. But how does the ribosome actually work? How is molecular information transferred between functional centers? What are the allosteric pathways and conduits of communication, and how do specific residues in the structure contribute to this communication? In this work, we apply a global coevolution method (SCA) to show that, despite differences in fundamental chemistry, RNA and protein appear to use a common logic of interaction and assembly. SCA revealed that ribosome function is defined by discrete sets of coevolving residue networks (sectors) within the large ribosomal subunit whose significance was verified experimentally.

The 23S rRNA sectors revealed by SCA comprise near-independent and physically contiguous networks within the 3D structure that link the PTC with multiple functional regions and can be traced to phylogenetic divergences across all domains of life. These findings inform our understanding of long-range functional couplings in the ribosome and provide a tool for studying the same types of interactions in other large RNAs.

Material and Methods

Materials. The *E. coli* strain used in this study, Squires strain (SQZ10), was obtained from the Joseph laboratory (University of California San Diego). All PCR reactions were performed using either the Phusion polymerase kit (New England Biolabs) or the KOD polymerase kit (Millipore Sigma). All oligonucleotides were purchased from Integrated DNA Technologies.

Data and Software Availability. Scripts used to generate multiple sequence alignments and perform analysis of experimental validation are available at GitHub, https://github.com/schepartzlab/Ribosomal-SCA-analysis. Code for SCA 6 is available at GitHub, https://github.com/ranganathanlab/pySCA. Raw sequencing data are available at National Center for Biotechnology Information Bioproject under accession number 511591.

Computational Methods. We used a custom Python script to generate our MSA and a modified version of the SCA version 6 script previously described (19) to calculate the SCA coupling matrix and ICs. We used a custom script to determine SCA sectors from ICs. A full description of all computational methods is available in *SI Appendix*.

Experimental Methods. We generated a library of 23S rRNA mutants in a PLK35 plasmid (51) background using Gibson Assembly (52). A detailed description of PCR and Gibson Assembly protocols is available in *SI Appendix*. Mutants were grown in a turbidostat clamped at an OD_{600} of 0.1. A full description of the protocol for the continuous culture experiment is available in *SI Appendix*.

Processing of Next-Generation Sequencing Data. Two custom Python scripts were developed to process the next-generation sequencing data: Miseq_data_analysis.py or Hiseq_data_analysis.py. These scripts are available on GitHub, and a description of all processing steps is available in *SI Appendix*. A description of the methods used to calculate the correlation between experimental results and SCA coupling predictions is also available in *SI Appendix*.

Quantification and Statistical Analysis. We used Pearson's correlation coefficient to quantify the degree of correlation between contribution to SCA sectors and growth rates determined in continuous culture experiments. Pearson's r and associated *P* values are listed in the main text and displayed in Fig. 4 and *SI Appendix*, Fig. S4.

1. J. Monod, J. Wyman, J. P. Changeux, On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
2. G. P. Lisi, J. P. Loria, Solution NMR spectroscopy for the study of enzyme allostery. *Chem. Rev.* **116**, 6323–6369 (2016).
3. E. Lerner *et al.*, Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **359**, eaan1133 (2018).
4. N. Plattner, F. Noé, Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **6**, 7653 (2015).
5. W. C. Winkler, C. E. Dann 3rd, RNA allostery glimpsed. *Nat. Struct. Mol. Biol.* **13**, 569–571 (2006).
6. D. A. Brow, Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**, 333–360 (2002).
7. H. W. Pley, K. M. Flaherty, D. B. McKay, Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994).
8. J. H. Cate *et al.*, Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273**, 1678–1685 (1996).
9. N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
10. G. J. Narlikar, D. Herschlag, Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.* **66**, 19–59 (1997).
11. A. M. Mulder *et al.*, Visualizing ribosome biogenesis: Parallel assembly pathways for the 30S subunit. *Science* **330**, 673–677 (2010).
12. J. H. Davis *et al.*, Modular assembly of the bacterial large ribosomal subunit. *Cell* **167**, 1610–1622.e15 (2016).
13. A. Yonath, Antibiotics targeting ribosomes: Resistance, selectivity, synergism and cellular regulation. *Annu. Rev. Biochem.* **74**, 649–679 (2005).
14. A. Schepartz, Foldamers wave to the ribosome. *Nat. Chem.* **10**, 377–379 (2018).
15. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa, Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 67–72 (2009).
16. C. Weinreb *et al.*, 3D RNA and functional interactions from evolutionary couplings. *Cell* **165**, 963–975 (2016).
17. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
18. H. Kamisetty, S. Ovchinnikov, D. Baker, Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15674–15679 (2013).

19. K. A. Reynolds, W. P. Russ, M. Socolich, R. Ranganathan, Evolution-based design of proteins. *Methods Enzymol.* **523**, 213–235 (2013).

20. M. E. Hatley, S. W. Lockless, S. K. Gibson, A. G. Gilman, R. Ranganathan, Allosteric determinants in guanine nucleotide-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14445–14450 (2003).

21. R. G. Smock et al., An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Mol. Syst. Biol.* **6**, 414 (2010).

22. K. A. Reynolds, R. N. McLaughlin, R. Ranganathan, Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–1575 (2011).

23. M. Novinec et al., A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods. *Nat. Commun.* **5**, 3287 (2014).

24. J. Lee et al., Surface sites for engineering allosteric control in proteins. *Science* **322**, 438–442 (2008).

25. W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffe, R. Ranganathan, Natural-like function in artificial WW domains. *Nature* **437**, 579–583 (2005).

26. S. M. Doris et al., Universal and domain-specific sequences in 23S-28S ribosomal RNA identified by computational phylogenetics. *RNA* **21**, 1719–1730 (2015).

27. L. M. Dedkova et al., β-Puromycin selection of modified ribosomes for in vitro incorporation of β-amino acids. *Biochemistry* **51**, 401–415 (2012).

28. C. Melo Czekster, W. E. Robertson, A. S. Walker, D. Söll, A. Schepartz, In vivo biosynthesis of a β-Amino acid-containing protein. *J. Am. Chem. Soc.* **138**, 5194–5197 (2016).

29. J. J. Cannone et al., The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**, 2 (2002).

30. O. Rivoire, K. A. Reynolds, R. Ranganathan, Evolution-based functional decomposition of proteins. *PLOS Comput. Biol.* **12**, e1004817 (2016).

31. P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4899–4903 (2001).

32. C. R. Bernier et al., RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss.* **169**, 195–207 (2014).

33. T. Asai, D. Zaporojets, C. Squires, C. L. Squires, An *Escherichia coli* strain with all chromosomal rRNA operons inactivated: Complete exchange of rRNA genes between bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1971–1976 (1999).

34. E. Toprak et al., Building a morbidostat: An automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nat. Protoc.* **8**, 555–567 (2013).

35. A. Walker, Impact of 23S rRNA single mutations on growth rates. NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/?term=511591. Deposited 23 December 2018.

36. S. Kullback, R. A. Leibler, On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).

37. T. Komoda et al., The A-site finger in 23 S rRNA acts as a functional attenuator for translocation. *J. Biol. Chem.* **281**, 32303–32309 (2006).

38. P. V. Sergiev et al., The conserved A-site finger of the 23S rRNA: Just one of the intersubunit bridges or a part of the allosteric communication pathway? *J. Mol. Biol.* **353**, 116–123 (2005).

39. M. A. Bayfield, J. Thompson, A. E. Dahlberg, The A2453-C2499 wobble base pair in *Escherichia coli* 23S ribosomal RNA is responsible for pH sensitivity of the peptidyl-transferase active site conformation. *Nucleic Acids Res.* **32**, 5512–5518 (2004).

40. M. O'Connor, Helix 69 in 23S rRNA modulates decoding by wild type and suppressor tRNAs. *Mol. Genet. Genomics* **282**, 371–380 (2009).

41. N. Hirabayashi, N. S. Sato, T. Suzuki, Conserved loop sequence of helix 69 in *Escherichia coli* 23 S rRNA is involved in A-site tRNA binding and translational fidelity. *J. Biol. Chem.* **281**, 17203–17211 (2006).

42. A. Sievers, M. Beringer, M. V. Rodnina, R. Wolfenden, The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7897–7901 (2004).

43. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).

44. Y. S. Polikanov, T. A. Steitz, C. A. Innis, A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat. Struct. Mol. Biol.* **21**, 787–793 (2014).

45. N. Fischer et al., Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570 (2015).

46. D. J. Ramrath et al., Visualization of two transfer RNAs trapped in transit during elongation factor G-mediated translocation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20964–20969 (2013).

47. X. Shi, P. K. Khade, K. Y. Sanbonmatsu, S. Joseph, Functional role of the sarcin-ricin loop of the 23S rRNA in the elongation cycle of protein synthesis. *J. Mol. Biol.* **419**, 125–138 (2012).

48. A. S. Petrov et al., Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10251–10256 (2014).

49. A. S. Petrov et al., History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15396–15401 (2015).

50. H. S. Zaher, R. Green, Quality control by the ribosome following peptide bond formation. *Nature* **457**, 161–166 (2009).

51. M. O'Connor, W. M. Lee, A. Mankad, C. L. Squires, A. E. Dahlberg, Mutagenesis of the peptidyltransferase center of 23S rRNA: The invariant U2449 is dispensable. *Nucleic Acids Res.* **29**, 710–715 (2001).

52. D. G. Gibson et al., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).