

Revealing unique properties of the ribosome using a network based analysis

Hilda David-Eden and Yael Mandel-Gutfreund*

Department of Biology, Technion-Israel Institute of Technology, Haifa 32000, Israel

Received April 22, 2008; Revised June 4, 2008; Accepted June 23, 2008

ABSTRACT

The ribosome is a complex molecular machine that offers many potential sites for functional interference, therefore representing a major target for antibacterial drugs. The growing number of high-resolution structures of ribosomes from different organisms, in free form and in complex with various ligands, provides unique data for structural and comparative analyses of RNA structures. We model the ribosome structure as a network, where nucleotides are represented as nodes and intermolecular interactions as edges. As shown previously for proteins, we found that the major functional sites of the ribosome exhibit significantly high centrality measures. Specifically, we demonstrate that mutations that strongly affect ribosome function and assembly can be distinguished from mild mutations based on their network properties. Furthermore, we observed that closeness centrality of the rRNA nucleotides is highly conserved in the bacteria, suggesting the network representation as a comparative tool for the ribosome analysis. Finally, we suggest a global topology perspective to characterize functional sites and to reveal the unique properties of the ribosome.

INTRODUCTION

The ribosome is a large complex of proteins and ribosomal RNA (rRNA) that is responsible for protein biosynthesis in all organisms. The ribosome is made up of two subunits: a small subunit (30S) and a large subunit (50S). In *Escherichia coli*, the small subunit consists of the 16S rRNA (1542 nt) and 21 proteins, whereas the large subunit contains the 23S rRNA (2904 nt), the 5S rRNA (120 nt) and 33 proteins. The small and the large ribosomal subunits associate into an active 70S complex, which catalyzes protein synthesis. The small subunit contains the decoding center (A-site), known to mediate the correct

interaction between the tRNA anticodon and the mRNA, which is being translated. The large subunit contains the peptidyl transferase center (PTC), which catalyzes the peptide bond formation in the growing polypeptide (1). It is well established that the major functional sites in the ribosome that are involved in peptide bond formation are composed mainly of rRNAs, which is in the heart of the catalytic process (2). Though the ribosome is commonly considered a ribozyme (1), the involvement of a protein in the catalytic site has been demonstrated in *E. coli* (3).

High resolution structures of 30S and 50S ribosomal subunits have been solved by X-ray crystallography. These include the 30S subunit from the eubacteria *Thermus thermophilus* (4,5), the 50S subunit from the archaeon *Haloarcula marismortui* (6) and the eubacteria *Deinococcus radiodurans* (7). Several high-resolution structures of the 70S ribosome are also available, including the *E. coli* ribosome at 3.5 Å (8), and *T. thermophilus* at 2.8 Å (9) and 3.7 Å (10). The 70S ribosome structures provide unique information on the interface between the subunits, as well as the conformation of the active site in the context of the entire ribosome complex. Overall, the growing number of high-resolution structures of ribosomes from different organisms, in the free form as well as in complex with various ligands and antibiotics, provides a unique data set for structural and comparative analysis of the ribosome, specifically of the rRNA (4,6,8,9).

Different computational methods, which are based on force field, have been used to study macromolecular structures, specifically proteins (11). However, the computational cost of these methods for studying long and short range interactions in a large-scale system such as the ribosome is extremely high. Therefore, coarse grained methods have been developed (12,13). Furthermore, graph theory has been found to be a useful tool to investigate different properties of macromolecules such as folding, stability, function and dynamics (14). These studies have predominately concentrated on protein structures. In order to model structure as a graph (network), several representations have been developed, ranging from coarse representation, in which each node represents a secondary

*To whom correspondence should be addressed. Tel: +972 4 8293958; Fax: +972 4 8225153; Email: yaelmg@tx.technion.ac.il

structure element (15), to finer modeling methods, in which each node represents an atom (16).

A common methodology to represent a macromolecular structure as a network considers amino acids/bases as nodes and inter-residue interactions as edges (17–20). In protein structures, it has been shown that the network of amino acid interactions is highly clustered and has properties of a small-world network. Such a network is characterized by the presence of a small number of central nodes (21). Interestingly, the central nodes defined by the small-world network description were found to be associated with key residues in protein folding and dynamics (18,22–25), functional sites such as enzymes catalytic sites, ligand-binding sites (17,26–29) and hot spots in protein–protein interactions (20,30).

Previously, a small-world network approach has been applied to study the conformational space of tRNA secondary structure (31). In addition, a network approach has been utilized for RNA structure characterization, in which different types of interactions (i.e. Watson–Crick, Hoogsteen and Sugar-edge) were applied to present the complexity of the structure (32). Here, we represented the rRNA 3D structure as a network with nucleotides as nodes and inter-nucleotides interactions as edges. We revealed that the rRNA structure-derived network fits the model of geometric random graphs with characteristics of a small-world network. Though the network parameters were directly derived from the structure of the ribosome complex, they were not found to simply correlate with classical structural parameters (e.g. solvent surface accessibility). The lack of strong correlation between structural properties and network parameters suggests that the latter can provide extra insights on the ribosome structure–function relationship. Specifically, we found that nucleotides with significantly high centrality values in the network correspond to the major functional regions in the small and large subunits of the ribosome. Additionally, we observed that rRNA mutations that cause a strong deleterious effect to the ribosome function, exhibit high centrality. In summary, applying a network-based analysis, we show that critical sites in the bacterial ribosome exhibit high values of local and global structural parameters. Moreover, these functional sites can be identified based solely on the network parameters without considering evolutionary information.

MATERIALS AND METHODS

Network analysis

The ribosome structure was presented as an undirected graph (network), in which nodes represent nucleotides or amino acids, and edges represent contacts. In order to generate the network, all atomic contacts were calculated using the CSU program (33). Nucleotides or amino acids were considered to be in contact if at least one of the corresponding atoms was in surface complementarity, as defined in ref. (33).

Two parameters were calculated to characterize the network: the average shortest path length, and the average

clustering coefficient. The average shortest path length of a network with N nodes is defined by Equation (1):

$$L = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N L_{ij} \quad 1$$

where, L_{ij} is the shortest path length between nodes i and j .

The average clustering coefficient of a network with N nodes is defined by Equation (2):

$$C = \left(\frac{1}{N}\right) \sum_{i=1}^N C_i \quad 2$$

where, C_i is the clustering coefficient of node i , defined as the fraction of contacts that exist among its nearest neighbors relative to the maximum contacts among all neighbors.

In order to test the network properties, we calculated the average shortest path length and the average clustering coefficient for random and regular networks with the same number of nodes, and the same average number of edges (degree) (24). In a random network, $L_{\text{rand}} \sim \ln N / \ln K$ and $C_{\text{rand}} \sim K/N$, while for a regular network, $L_{\text{reg}} = N(N+K-2)/[2K(N-1)]$ and $C_{\text{reg}} = 3(K-2)/[4(K-1)]$, N denotes number of nodes and K number of edges.

For each node in the network, we calculated three centrality measures, i.e. degree, closeness and betweenness. The degree of a node i is defined as the number of edges connected to i . The closeness of a node i is defined as the inverse average length of the shortest paths to all other nodes in the graph (34). The closeness of node i is given in Equation (3):

$$\frac{|N-1|}{\sum_{j \neq i} d_{ij}} \quad 3$$

where, N is the total number of nodes in the network, and d_{ij} is the shortest path length to node i .

The betweenness of a node i is defined by the number of the shortest paths that cross node i . The betweenness of node i is given in Equation (4):

$$\sum_{j \neq k, i \neq j, k \neq i} \frac{g_{jik}}{g_{jk}} \quad 4$$

where, g_{jik} is the number of the shortest paths from j to k that pass through i .

The shortest path length calculations are based on a modified form of Dijkstra's algorithm (35). Network parameters were calculated with the igraph package version 0.1.2 using GNU R statistical software (<http://cneur.ocvs.rmki.kfki.hu/igraph>), and the network python package (<https://networkx.lanl.gov>).

Statistical analysis

Enrichment of mutations within high centrality nucleotides was evaluated based on the Hyper Geometric Distribution using the Fisher's exact test. In addition, we tested the enrichment in 100 random sets of nucleotides

of the same size of each mutations dataset and expended the tested set to include the contacting nucleotides, as applied for the original datasets. The proportion of nucleotides that were randomly sampled from the large and small ribosomal subunit was as in the mutation datasets.

Ribosome structure analysis

In order to define the ribosomal subunits interface, the solvent accessible surface area of each nucleotide in the 16S and the 23S rRNA was calculated using the POPS server (36,37). A nucleotide was considered to be in the interface if it had at least one atom that lost solvent accessibility upon complex formation (PDB codes 2AVY and 2AW4).

The nucleotides of the tunnel were defined using the 3V Channel Extractor (<http://geometry.molmovdb.org/3v>) with a large probe radius of 9.0 Å and a small probe radius of 3.4 Å. Using the 3V Channel Extractor, an output file with the tunnel surface represented as water atoms was generated. We considered a nucleotide to be located in the tunnel if at least one of its atoms was located at a distance <5 Å to the computed water oxygen atom.

Accessible surface area of each nucleotide was calculated using the POPS server (36,37). POPS is based on a fast algorithm for calculating solvent accessible surface areas of large macromolecular assemblies, such as the ribosome. Nucleotides were defined as interface nucleotides if their solvent accessible surface area decreased upon subunit association. The accessible surface area of each nucleotide in the 70S *E. coli* rRNA (PDB codes 2AVY and 2AW4), and the 16S and 23S rRNA were calculated separately.

Theoretical B-factors were downloaded from the iGNM V1.2 database (<http://ignm.cccb.pitt.edu>). The GNM program calculated the theoretical B-factors (1JJ2) for a 1 nt model using a cutoff of 7.3 Å for aa, and 19 Å for nucleotide.

The following nucleotides in the 16S rRNA were defined as the A site: 1405, 1406, 1407, 1408, 1409, 1410, 1490, 1491, 1492, 1493, 1494, 1495 and 1496. The following nucleotides in the 23S rRNA were defined as the PTC nucleotides: 2061, 2062, 2063, 2064, 2251, 2252, 2439, 2450, 2451, 2452, 2453, 2492, 2493, 2494, 2503, 2506, 2507, 2553, 2554, 2555, 2573, 2583, 2584, 2585, 2602 and 2603.

Evolutionary rate

Aligned sequences and a guide tree for both the 16S/18S and the 23S/28S were downloaded from the ARB-SILVA database (release 92) (<http://www.arb-silva.de>) (38). The 16S/18S alignment comprises high-quality sequences with a minimum length of 1200 bases for Bacteria and Eukarya and 900 bases for Archaea. The 23S/28S alignment comprises high quality rRNA sequences with a minimum length of 1900 bases. Positional variability in bacteria was calculated with the ARB package using parsimony function (38), including 168131 and 4115 sequences for the 16S and 23S rRNA, respectively. The evolutionary

rate was calculated for each column as the mutations sum divided by the mutations frequencies sum.

Mutation data

Datasets 1, 2. The two datasets are based on mutation data, collected from various studies. The ribosomal RNA mutation database in *E. coli* (16SMDB and 23SMDB) was downloaded (<http://server1.fandm.edu/departments/Biology/Databases/RNA.html>) (39). We further extended the database by including data from recent studies (40–43). *Dataset 1* includes mutations with strong effect on the ribosome function were selected by using a keyword search (strong, lethal, severe and deleterious) excluding mutations with mild or moderate effect. Further, the mutations were manually filtered to include only single point mutations. In addition, those mutations which had their phenotypic effect tested with the presence of antibiotics were excluded. To avoid redundancy, we included mutations with a minimal space of 3 nt in the primary sequence. Additionally, we excluded mutations that appeared in *Datasets 3* and *4*. In total, *Dataset 1* included 44 nt: 25 nt in the 16S rRNA and 19 in the 23S rRNA. The 16S rRNA mutated nucleotides included positions: 13, 18, 517, 529, 571, 627, 643, 702, 770, 787, 792, 865, 914, 922, 967, 981, 1200, 1207, 1401, 1409, 1414, 1418, 1483, 1491 and 1498. The 23S rRNA mutated nucleotides included positions: 1832, 1836, 1849, 1896, 1916, 1926, 1932, 1940, 1946, 1955, 1960, 1972, 1979, 1984, 2252, 2504, 2507, 2580 and 2584 (Supplementary Table S5). *Dataset 2* includes mutations with mild effect of the bacteria function. In total, *Dataset 2* included 30 nt: 18 nt in the 16S rRNA and 12 in the 23S rRNA. The 16S rRNA mutated nucleotides included positions: 531, 534, 618, 624, 631, 634, 641, 645, 651, 912, 966, 1203, 1341, 1351, 1388, 1397, 1404 and 1518. The 23S rRNA mutated nucleotides included positions: 1067, 1098, 1914, 1921, 1940, 1951, 1979, 2249, 2254, 2477, 2561 and 2661.

Dataset 3, 4. These two sets were derived from two recent studies, which applied random mutagenesis procedure on *E. coli* 16S and 23S rRNA genes (44,45). In these studies, 53 and 77 mutations in the 16S and 23S rRNA, respectively, were classified according to their phenotypic severity. The 16S and 23S rRNA included 50 and 69 base substitutions, and 3 and 8 deletions, respectively. Among the 16S base substitutions, 13 mutations were classified as strong, 17 as mild and 20 as moderate. In the 23S base substitutions, 12 mutations were classified as strong, 34 as mild and 23 as moderate. Overall, the datasets included 25 positions (13 + 12) that were classified as strong and 51 positions (17 + 34) that were classified as mild.

The 16S rRNA mutated nucleotides with strong deleterious phenotype (*Dataset 3*) included the following mutations: Y516G, C518U, C519U, A520G, G521A, G973A, G1058A, G1068A, A1111U, C1208G, C1395U, U1406C and U1495C. The 16S rRNA mutated nucleotides with mild effect included: U49C, A51G, G57A, A161G, G299A, A373G, A389G, C536U, G568C, C614A, A622G, U684C, A1014G, C1054U, A1055G, U1073C and U1085C. The 23S rRNA mutated

Table 1. Topological properties of four rRNA structures

rRNA structure	Nodes	Edges	Average shortest path length			Clustering coefficient		
			L_{rand}	L_{observed}	L_{reg}	C_{rand}	C_{observed}	C_{reg}
<i>T. thermophilus</i> (30S)	1494	5389	3.7	12.5	104	0.0048	0.50	0.63
<i>D. radiodurans</i> (50S)	2766	10 673	3.9	11.6	179.6	0.0028	0.47	0.64
<i>H. marismortui</i> (50S)	2754	10 447	3.9	11.9	181.2	0.0028	0.48	0.64
<i>E. coli</i> (70S)	4371	16 049	4.2	15.5	198.2	0.0017	0.48	0.63

Average shortest path length and clustering coefficient were calculated for the small ribosomal subunit of *T. thermophilus* (PDB code 1J5E), the large ribosomal subunit of *D. radiodurans* (PDB code 1NKW), the large ribosomal subunit of *H. marismortui* (PDB code 1JJ2) and the structure of the 70S ribosome from *E. coli* (PDB codes 2AVY and 2AW4). The C and L of the random and the regular networks were calculated as described in text.

nucleotides with strong deleterious phenotype included: U860C, G864A, U2249A, G2250A, U2438C, G2446A, A2450G, A2451U/G, A2453G, C2499U, G2603A, and G2664A. The 23S rRNA mutated nucleotides with mild effect (*Dataset 4*) included: G46A, A49G, G55A, G58A, G123A, G215A, A324G, U328C, A330G, A371G, G380A, C420U, G425A, G481A, U558A, G578A, G656U, G778A, G1259A, C1297U, A1342G, C1345U, U1397G, G1421A, A1572C, U1602C, A1614G, U1683C, U1765G, C1800U, U1995C, G2027A, G2641A and U2898C.

Throughout this work, nucleotides were presented according to the *E. coli* sequence numbering.

RESULTS AND DISCUSSION

Ribosomal RNA exhibits small-world properties

Previously, it has been shown that amino acid network of a model protein structure have the unique properties of a small-world network (25). To study the properties of the ribosome network, we investigated the topological properties of four rRNA networks, including the small ribosomal subunit (PDB code 1J5E), two large ribosomal subunits (PDB codes 1NKW, 1JJ2) and the rRNA of the entire ribosome structure (PDB codes 2AVY, 2AW4). We presented the nucleotides and amino acids of the ribosome as nodes in the graph. Two nodes were connected by an edge when at least one atom from 1 nt/aa was in contact with at least one atom from a neighboring node. The inter-atomic contacts were defined based on surface complementarity, calculated with the CSU software (33).

The topological characteristics of a network are generally defined by two major parameters: the clustering coefficient (C) and the average shortest path length (L) (see Materials and Methods section). The clustering coefficient is the probability that any two neighbors of a given node in a graph are directly connected to each other, whereas the average path length is defined as the average of the minimal distance between all pairs of nodes in the graph. Different than random networks, small-world networks are characterized by a small number of highly central nodes (21). In comparison to random networks of the same size, small-world networks have a significantly higher clustering coefficient ($C_{\text{small-world}} \gg C_{\text{rand}}$) and an average shortest path length higher or in the same order of magnitude ($L_{\text{small-world}} \geq L_{\text{rand}}$ or $L_{\text{small-world}} \approx L_{\text{rand}}$) (24). To test whether the ribosome structures have

characteristics of small-world networks, we calculated the values of L and C of the ribosome networks, and compared to their equivalent values in regular and random networks with the same number of nodes and an average number of edges. The values of L and C of four rRNAs networks are presented in Table 1. Notably, the C values calculated for the ribosome structures were orders of magnitude larger than in randomly generated networks, while the L values were in the same order of magnitude as the theoretical random networks. In other words, rRNA structure can be described as a network of interacting nucleotides that have an intermediate value of L and C , a property which characterizes the class of small-world networks (21). In all four rRNA networks we analyzed, the clustering coefficient values as well as the average path lengths scaled with the values measured for protein structures (19,46).

Small-world and scale-free properties have been found in a large number of real world networks. However, the small-world property is less studied in networks that are not scale-free (47). To study the network properties of the ribosome, we selected the 2.4 Å high-resolution structure of the large ribosome subunit from *H. marismortui* (PDB code 1JJ2). We compared the 23S rRNA network extracted from 1JJ2 to five random graph models (48): Erdős-Rényi random graphs ('er') (49), random graphs with the same degree distribution as the data model network ('er-dd'), 3D geometric random graphs ('geo') (50), Barabási-Albert-type scale-free networks ('sf') (51) and stickiness-index-based networks ('sticky') (52). Based on this comparison (Supplementary Table S1), we found that the rRNA structure-derived network fits best with the geometric random graph model both by global and local properties. Overall, the rRNA structure-derived network fits the geometric random graph model but is highly clustered as typical for small-world networks.

Centrality measures and unique properties of the ribosome

Network analyses of protein structures have concluded that high centrality measures are associated with key (functional) positions, such as: enzyme catalytic sites, ligand-binding sites (17,27,28) as well as hot spots in protein-protein interactions (20,30). The major parameters examined in these studies included both local and global properties of the protein structures. In an attempt to identify prominent nucleotides that have a central role in the rRNA structure, we calculated three centrality measures

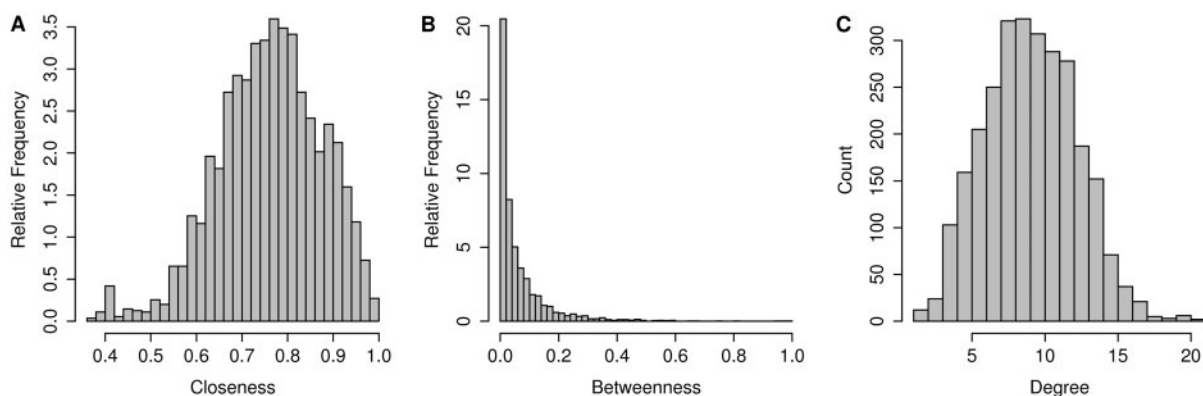


Figure 1. Closeness, betweenness and degree distributions calculated for the structure-based network of the large ribosomal subunit (PDB code 1JJ2). The presented distributions correspond to the 23S rRNA nucleotides. (A) Closeness, (B) betweenness and (C) degree.

for all nucleotides (nodes) in the network, i.e. degree, closeness and betweenness. Degree is a local measure that is defined as the number of edges directly connected to a node, whereas closeness and betweenness are based on the global centrality of the node. Closeness reflects the average length of the shortest path connecting a node to all other nodes in the network. Betweenness reflects the frequency in which a node appears on the shortest path between any two nodes in the network.

Closeness, betweenness and degree distributions have been utilized to compare network topologies. Among them, networks with small-world properties, such as protein-protein interaction network (53) and protein structure-based network (26). The distribution of the three centrality measures of all nucleotides in the 23S rRNA network (1JJ2) is presented in Figure 1. As previously observed in protein structures (17,26), the closeness parameter in the rRNA ribosome network showed a bell-shaped distribution (Figure 1A). Additionally, in accordance with previous observations in protein structures (19,25,54), the degree parameter of the rRNA structure follows a Poisson distribution (Figure 1C). The latter observation is different from other small-world networks such as regulatory networks that are scale-free (51), namely, a small number of nodes having very high degree values (51,55). Important to note that based on a visual investigation of wiring diagrams of the 16S rRNA, it has been suggested that the ribosome has characteristics of a scale-free network (32). Nevertheless, the observed distribution of the degree values in our nucleotide network (Figure 1C) indicates that the probability of highly connected nodes is relatively low. This result is expected based on the limited binding capacity of a given nucleotide, due to the excluded-volume effect in the RNA (56). Unlike the distribution of the degree parameter, the betweenness centrality parameter displays a power law distribution (Figure 1B), meaning that the majority of nucleotides demonstrate low values of betweenness, while only a few possess high values.

We further calculated the Pearson correlation coefficient for the different parameters in the 23S rRNA network (PDB code 1JJ2, 2754 nt). Our results indicate a weak positive correlation between the parameters. The correlation between closeness and betweenness is 0.42 ($P < e-15$),

while the correlation between degree and closeness, and between degree and betweenness is 0.54 and 0.50, respectively ($P < e-15$). The relative weak correlation between the network parameters suggest that each parameter can be applied independently to describe specific nodes (nucleotides) in the network, i.e. a node in the network can have a small number of neighbors (low degree) but can still be located in a central position in the structure (high closeness and betweenness). A notable example is nt 2602 in the 23S rRNA (*E. coli* numbering throughout), which has a high closeness value in *E. coli* 70S (z -score = 1.7, $P < 0.05$) but has a very low degree, i.e. in contact with only two neighbors. Moreover, the betweenness of this nucleotide was zero, meaning that there is no short path that passes through this position. Interestingly, nt 2602 is the only nucleotide in the PTC that has zero betweenness. The unique global and local properties of nt 2602 are consistent with the different conformations observed in various functional complexes of the large subunit with different antibiotics and substrate analogs (57–60).

Though this study focuses on the unique properties of the rRNA, since the ribosome consists of both the rRNA and proteins, it was intriguing to examine the influence of the ribosomal proteins on the network properties of the rRNA. Thus, we analyzed the centrality measures separately for the 23S rRNA structure and for the 50S subunit, which comprises the 23S rRNA, the 5S rRNA and the ribosomal proteins. Both networks are based on the atomic coordinates of the 2.4 Å resolution structure from *H. marismortui* (PDB code 1JJ2). The first network included 2754 nodes and 10 447 edges (Table 1), while the second network, including the ribosomal proteins, was comprised of 6577 nodes and 33 013 edges. Interestingly, high correlations between the rRNA and the 50S subunit were obtained for closeness ($r = 0.96$, $P < e-15$) and betweenness ($r = 0.72$, $P < e-15$). While a lower correlation was measured for the degree parameter ($r = 0.63$, $P < e-15$). Similarly, we compared between the *T. thermophilus* (PDB code 1J5E) 16S rRNA network (Table 1) and the 30S structure that comprises the 16S rRNA and the ribosomal proteins (3850 edges and 18 929 edges). The Pearson correlation coefficient of the 16S rRNA nucleotides was 0.91, 0.86 and 0.67 ($P < e-15$) for closeness, betweenness and degree, respectively. The high correlations

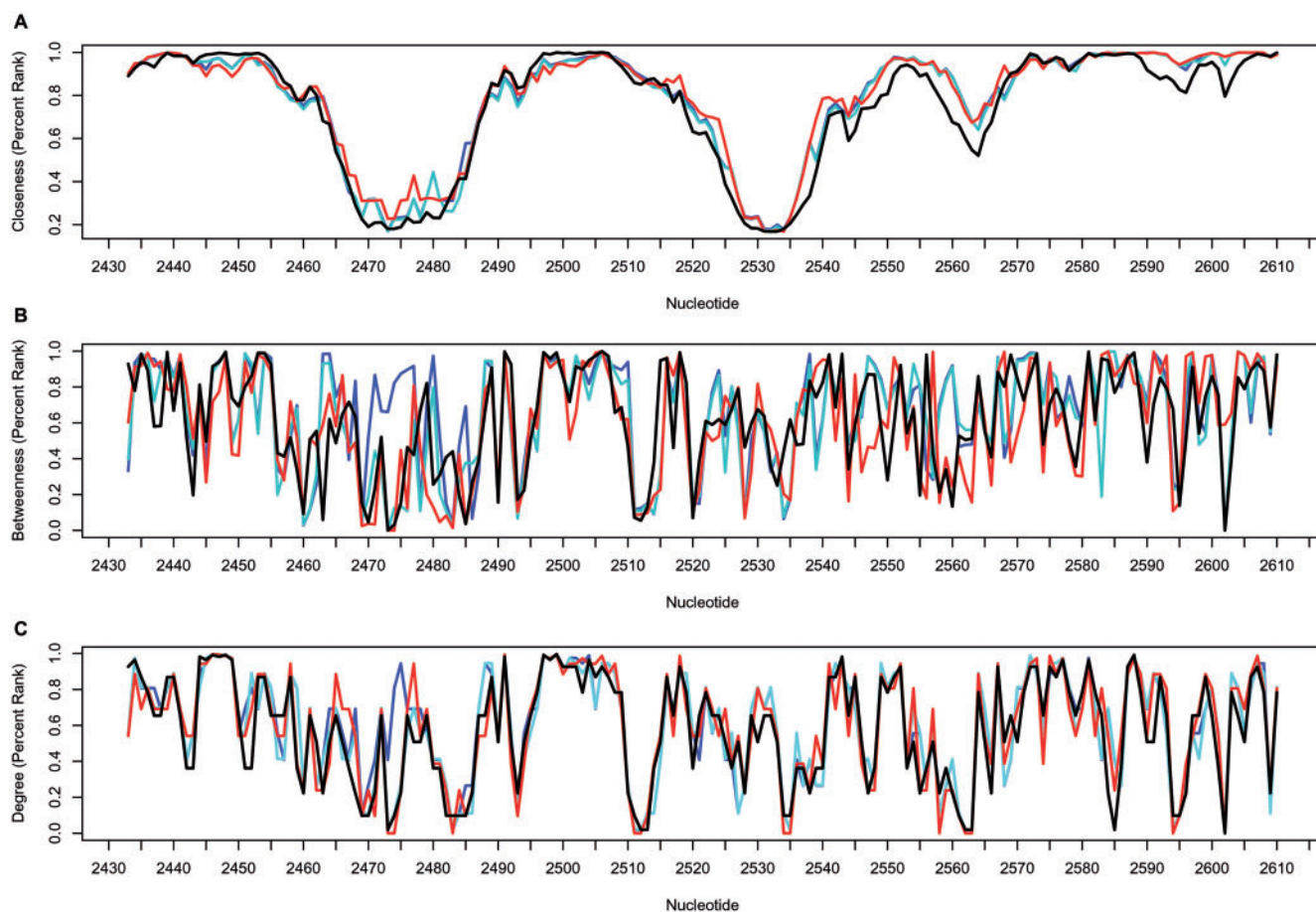


Figure 2. The correlation between the centrality measures in different organisms is presented for the central domain V, nt 2433–2610. (A) Closeness, (B) betweenness and (C) degree. Network parameters are based on the rRNA network calculated for the *E. coli* 70S (2AVY, 2AW4: blue line), (2AW7, 2AWB: cyan line), *T. thermophilus* 70S (2J00, 2J01: red line) and *H. marismortui* 50S (1JJ2: black line).

in the network parameters, specifically the closeness, between the rRNA and the entire ribosome networks of both the 30S and the 50S suggest that the proteins do not change the network properties. These results reinforce that the rRNA network can be analyzed independently.

Comparative analysis of the rRNA structure using centrality measures

We further compared the centrality values of ribosomal networks from different organisms. When comparing between the 70S rRNA of *E. coli* (8) (PDB codes 2AVY and 2AW4) and the 70S of *T. thermophilus* (9) (PDB codes 2J00 and 2J01) we observed extremely high correlations for both closeness ($r = 0.98$, $P < e-15$), and degree ($r = 0.87$, $P < e-15$), though a lower correlation was observed for betweenness ($r = 0.7$, $P < e-15$). Similar results were obtained when comparing the 23S rRNA network from bacteria to the 23S rRNA from archaea, as well as for the 16S rRNA from different bacteria (*E. coli* and *T. thermophilus*) (Supplementary Table S2). Here again, among the three centrality parameters closeness was the most conserved, while betweenness tended to be more variable.

A comparison between the network parameters from different organisms is shown for a stretch of consecutive nucleotides in the central domain V of the 23S rRNA (Figure 2). The fine-tuned correlation in the closeness values (Figure 2A) is notable when comparing the 70S structures from *E. coli* (8) and *T. thermophilus* (9) as well as to the *H. marismortui* 50S structure (6). Conversely, the betweenness (Figure 2B) and the degree (Figure 2C) parameters show notable variations among the different structures. For example, variations in the betweenness and degree parameters between the two *E. coli* structures were observed in the region between nucleotides U2473 and C2475 in helix H89. These variations may correspond to the movement of the L1 arm, as observed in the two ribosome structures (8).

Centrality and functionality

The most intriguing question when studying the structural properties of the ribosome is to what extent these properties correspond to functionality. Given that evolutionary conservation is indicative of functionality, we computed the correlation between the network parameters and the evolutionary rate of the 23S rRNA sequence. Evolutionary rate in bacteria was calculated using the ARB software,

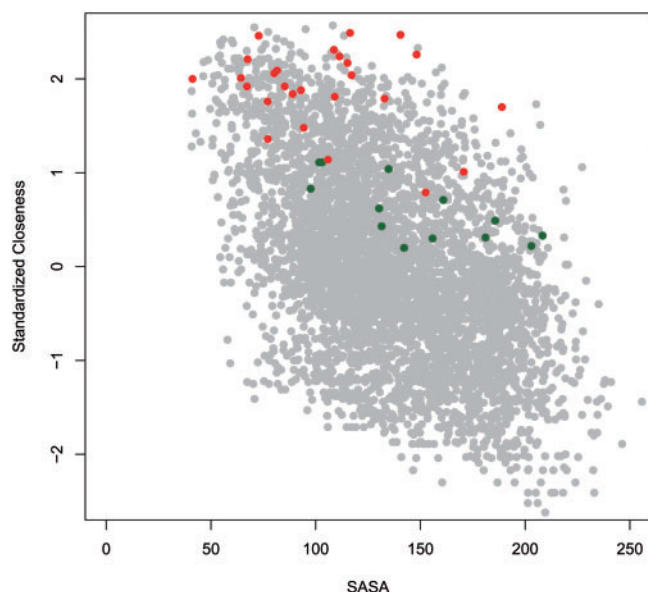


Figure 3. Closeness versus solvent accessible surface area (SASA) calculated for the *E. coli* 70S rRNA. The Pearson coefficient shows a weak negative correlation ($r = 0.51$, $P < e-15$). The A-site nucleotides of the 16S rRNA are colored green. The PTC nucleotides are colored red. Notably, the PTC nucleotides demonstrate high closeness values independent of the solvent accessibility.

based on aligned sequences and a guide tree downloaded from the SILVA database (38) (see Materials and Methods section). Our results show a weak negative correlation between the evolutionary rate and closeness (PDB codes 2AVY and 2AW4) in the 16S and 23S rRNA ($r = 0.44$, $P < e-15$). These results are consistent with earlier studies in proteins in which the closeness was found to be in correlation with evolutionary conservation, with values ranging between $r = 0.24$ and $r = 0.62$ (17).

To test the conjecture that the network parameters can give new insight into the ribosome function, we focused on two major functional sites, the PTC and the nascent peptide exit tunnel. The PTC is the catalytic center, located on the interface side of the large subunit at a cavity that opens into a roughly cylindrical tunnel, which is ~ 100 Å long and spans the entire body of the large ribosomal subunit. Quite expected, we found that the PTC nucleotides (defined in the Materials and Methods section) demonstrated higher than average closeness values (the average z -score for all PTC nucleotides was 1.47). Moreover, when compared to all nucleotides in the entire ribosome complex, the closeness values of the PTC nucleotides were significantly higher (Figure 3). This result is consistent with the observation that the catalytic sites in enzymes are located in regions with high closeness (17,27,61).

In order to examine whether high closeness of the ribosomal nucleotides simply reflect their low accessibility to the solvent, we calculated the Pearson correlation between the solvent accessible surface area of each nucleotide in the 23S rRNA (36,37) and the network parameters. The correlations for closeness, betweenness and degree were -0.57 , -0.37 and -0.65 , respectively (in all cases the P -value was highly significant $P < e-15$). As shown in

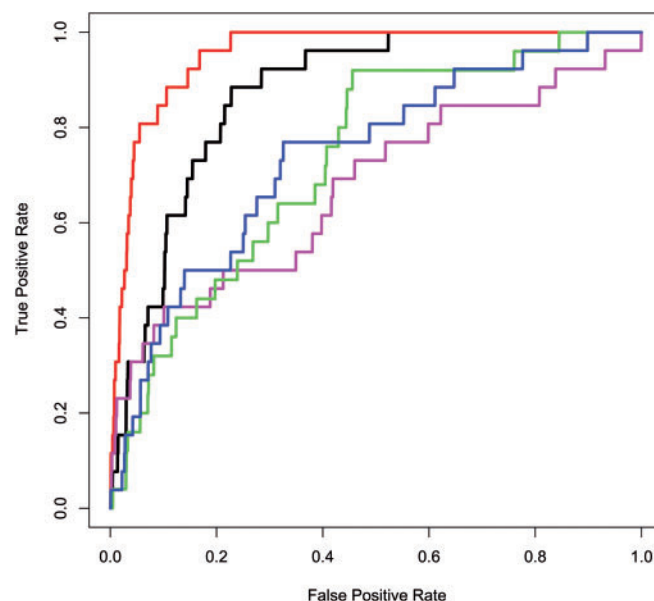


Figure 4. ROC curves classifying the PTC nucleotides. The structural parameters are based on the *E. coli* 70S network (2AVY, 2AW4). As shown, the PTC nucleotides were better classified using the closeness parameter (red line, AUC = 0.95) rather than the evolutionary rate (black line, AUC = 0.87), solvent accessibility (blue line, AUC = 0.74), degree (green line, AUC = 0.71) and betweenness (magenta line, AUC = 0.67).

Figure 3, overall the closeness and the surface accessibility demonstrated a weak negative correlation. Nevertheless, it can be noticed that the high closeness of the PTC is independent of the surface accessibility.

The outstanding high closeness centrality values observed here for the PTC is consistent with the high evolutionary conservation previously shown for this region (57,58,62–64). To test the ability of each one of the parameters: evolutionary rate, solvent accessibility, closeness, betweenness and degree to exclusively characterize the PTC nucleotides, we applied a ROC curve analysis (65) for each property independently (Figure 4). As described in the Materials and Methods section, the PTC nucleotides were labeled as ‘positives’ and the remaining nucleotides in the 70S ribosome (16S and 23S) were labeled as ‘negatives’. The ROC curves show the true positive rate versus the false positive rate. To quantify the classification performance, we compared the area under the curve (AUC) of the different ROC plots. As can be noticed (Figure 4) the closeness centrality better classifies the PTC nucleotides (AUC = 0.95) compared to all other structural parameters, as well as to the evolutionary conservation rate (AUC = 0.87). Remarkably, for 100% sensitivity, i.e. all 26 nt comprises the PTC nucleotides being classified correctly, the closeness achieved 78% accuracy while the evolutionary rate achieved only 48% accuracy. Additionally, our results show that the solvent accessibility parameter did not improve the PTC nucleotides’ identification. This latter result differs from observations in proteins showing that solvent accessibility improves the identification of enzyme catalytic residue significantly (17,26). Taken together, our results indicate that the ribosome major catalytic site can be accurately

identified by a global topological measure (closeness) with no requirement for evolutionary information.

In addition to the PTC, other functional sites in the ribosome such as the A-site in the 30S (Figure 3) and the exit tunnel in the 50S (Supplementary Table S4) demonstrated high closeness values. Generally, the 23S rRNA nts were found to have higher closeness values compared to the 16S rRNA nucleotides, with a maximum z -score of 2.37 and 1.85, respectively. Nevertheless, it can be noticed that both the PTC and the A-site nucleotides encompass relatively high closeness values compared to the rest of the 70S nucleotides (Figure 3). Additionally, the exit tunnel nucleotides (50S) were found to possess high closeness values with the highest values for the tunnel nucleotides facing the PTC (Supplementary Table S4). Further, we examined the number of contacts (degree) of the tunnel nucleotides (PDB code 1JJ2). Interestingly, the high degree values, like the closeness values, were independent of the solvent accessibility. Overall, the average degree of the tunnel nucleotides (123 nt) was found to be significantly higher (8.3) than the average degree of all 23S rRNA nucleotides (7.6) ($P = 0.001$, t -test).

When examining the distribution of degree values along the tunnel by dissecting it into 10 Å wide slices (5 Å overlap) starting from the PTC (nt 2451), we found that the highest degree values in the tunnel were within a distance of 20–30 Å from the PTC (Supplementary Table S3). The distribution of the degree values along the ribosome tunnel could be indicative of the tunnel dynamics. To better understand the correspondence between the tunnel dynamics and the degree we compared the profile of the degree in the tunnel with a theoretical B-factor calculated with the Gaussian Network Model (GNM) method (66,67). GNM has been introduced as an efficient tool to explore the dynamics of proteins and large biomolecular structures such as the ribosome (12,54,68). In the GNM, a structure is represented as a network of N nodes corresponding to the α -carbon atoms of proteins and selected atoms of the nucleotides (see Materials and methods section). In accordance with the degree distribution, the theoretical B-factor confirmed that the center of the tunnel (15–25 Å from nt 2451) is more rigid, but its flexibility increases as it reaches the tunnel exit. Based on the crystallographic B-factor values for the RNA atoms in the *H. marismortui* large ribosomal subunit, an upper bound estimate of nucleotide thermal fluctuations of 0.63 Å was suggested (6). It has been argued that random structural fluctuations of this magnitude will not allow drastic changes in the tunnel dimensions, which would be required in order for the nascent polypeptide to form tertiary structures (69). Our network analysis, demonstrating high closeness of the tunnel nucleotides and a relatively high degree of nucleotides that construct the tunnel wall (Supplementary Tables S3 and S4, Figure S1) supports this hypothesis (69). The relatively high degree in a central region within a macromolecule is quite expected and is consistent with other observations in globular proteins showing that the residues in the protein core are characterized by a higher mean degree than the surface residues (18).

As shown, the PTC nucleotides have significantly high closeness values among all the ribosome nucleotides. Notably, the top ranked nucleotides in the 70S rRNA only partially overlap with the PTC nucleotides. This can also be observed in the exit tunnel profile (Supplementary Table S4), which shows that the highest average closeness is within a distance of 5–15 Å from nt 2451. To identify the nucleotides with the highest closeness in the rRNA, we ranked the closeness values of nucleotides in the *E. coli* 70S rRNA (8) (PDB codes 2AVY and 2AW4) and the *T. thermophilus* 70S rRNA (9) (PDB codes 2J00 and 2J01). Interestingly, among the top ten nucleotides with the highest closeness values in each of the ribosome networks (from different bacteria), 7 nt were common (1938, 2439, 2584, 2591, 2604, 2605 and 2608). Interestingly, these nucleotides only partially overlap the PTC nucleotides. This result is consistent with previous observations in proteins (26) showing that the most central residues fall near the center of the structure, which often does not overlap with the catalytic residues. Notably, in the ribosome networks, among the top ranked nucleotides there were no nucleotides that directly interact with the ribosome substrate, such as nt 2062, 2573, 2585, 2602 and 2609 (62). This could be partially explained by the conformation variations observed for these nucleotides in different ribosome structures (57,62,70). Interestingly, among the ten top ranked nucleotides in the two bacteria (70S) and in the archaeon (50S) was nt A2439 with extremely high closeness. This nt (A2439, H74) is located in the symmetrical region of the ribosome (62), and forms an A- to P-region contact through its ribose (A2439 is bulged with U2586, H93). The high closeness values in the vicinity of the PTC both in the bacteria (70S), and in archaea (50S), may suggest that the spatial organization of the rRNA nucleotides is required to enable signal transmission between remote locations in the ribosome. This hypothesis is supported by the observation of a 2-fold symmetry axis within domain V, which suggests that the PTC resulted from an early gene duplication event and may indicate that domain V is the most ancient part of the primitive ribosome (62,70).

Network properties of critical positions in the ribosome

As demonstrated in Table 1, the network of interacting nucleotides in the 70S rRNA is highly clustered (Table 1). Consequently, there is a relatively small number of nucleotides that are highly central and serve as interconnection hubs between all nucleotides in the structure. A fascinating conjecture is that the latter nucleotides have a central role in the function or stability of the ribosome complex. Here, we focused on functional information derived from mutagenesis studies, including directed mutations and random mutations, which have been reported to specifically affect ribosome function or subunits association (39–45,71).

The first dataset we tested included 44 mutations, which have been reported to have strong, lethal, severe or deleterious effect on ribosome function, activity, growth, association or translation fidelity (see Materials and Methods section). We defined the mutated nucleotides and their structural neighbors as *experimental set 1*. Overall, the

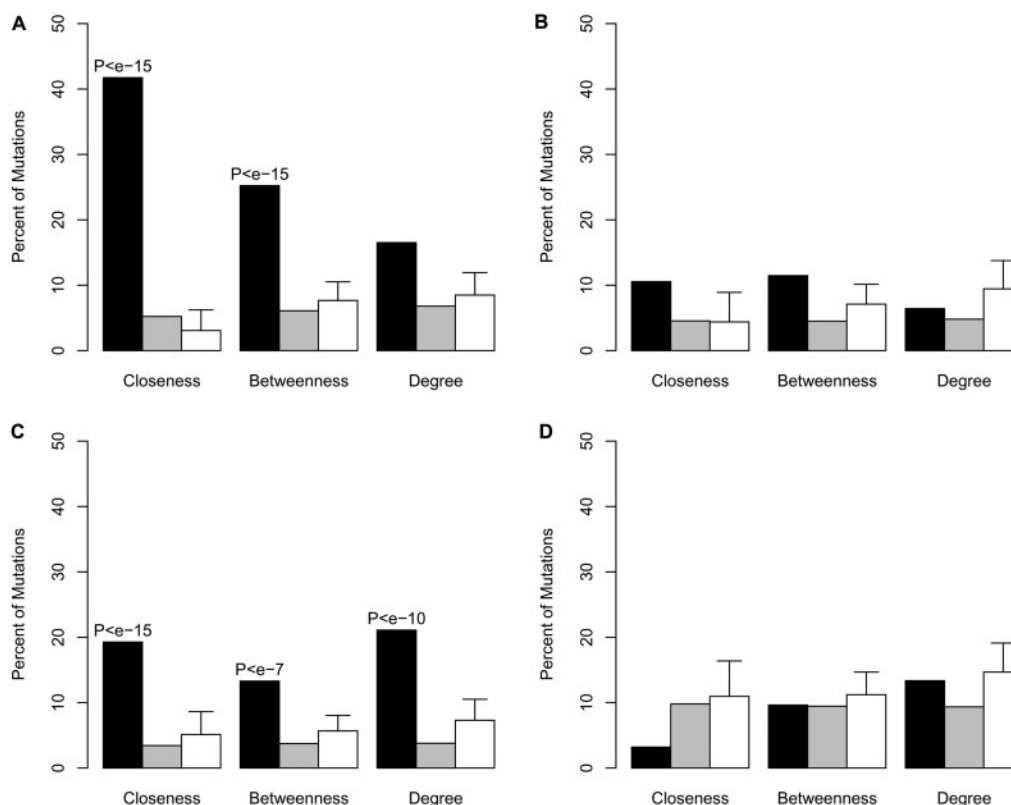


Figure 5. A histogram demonstrating the enrichment of mutated nucleotides among three sets of nucleotides: (A) *experimental set 1*, (B) *experimental set 2*, (C) *experimental set 3* and (D) *experimental set 4*. High centrality refers to nucleotides with significantly high values (top 5%) for closeness, betweenness and degree. Black bars represent nucleotides with high centrality values (top 5%). The gray bars represent subset of ribosomal nucleotides that do not possess high centrality. White bars represent randomly selected equal sized subsets of nucleotides. The enrichment of mutations that cause strong deleterious effect on ribosome function is statistically significant in the strong deleterious mutations (A and C) but not in the mild mutations (B and D).

nucleotides of the *experimental set 1* represented 7% of the ribosomal nt (308/4371). As a negative control, we extracted from various studies a dataset of 30 mutations that have mild effect of the bacteria function (see Materials and Methods section, Supplementary Table S5). The mild mutations and their structural neighbors were defined as *experimental set 2*. Overall, the nucleotides of the *experimental set 2* represented 5% of the ribosomal nt (212/4371).

The two additional datasets were based on a random mutants library of rRNA gene (44,45). The first dataset included deleterious mutation with strong (25 nt) effect on *E. coli* growth. The second dataset included mutations with mild (51 nt) effect on *E. coli* growth (44,45). Further, we defined the mutated nucleotides and their structural neighbors as *experimental sets*. Overall, the nucleotides of *experimental set 3*, derived from the strong deleterious mutations represented 4% of the ribosomal nt (185/4371), and the nucleotides of *experimental set 4* derived from the mutations with mild phenotype represented 10% of the ribosomal nt (422/4371).

Next, we defined three subsets of nucleotides that encompass high centrality, i.e. nucleotides with the top 5% values of closeness, betweenness and degree, in the *E. coli* 70S rRNA (PDB codes 2AVY and 2AW4). Their exact position is presented on the ribosome secondary

structure diagram (Supplementary Figures S4 and S5). In order to calculate the correspondence of the *experimental sets* to the network parameters, we calculated the fraction of the *experimental sets* nucleotides among the nucleotides with high values (top 5%) of closeness, betweenness and degree (Figure 5). To evaluate the enrichment of the mutations among nucleotides with the high centrality values, we applied the Hyper Geometric Distribution analysis using the Fisher's exact test. Additionally, to test the significance of the results, we selected 100 random sets with equal number of nucleotides as in each one of the *experimental sets* (see Materials and Methods section) and tested their overlap with the subsets of nucleotides with high centrality values. As shown in Figure 5A, the set of mutations with strong deleterious effects on bacterial growth (*experimental set 1*), was significantly enriched with nucleotides with high values of closeness and betweenness (Figure 5A, black bars). An enrichment (relatively to randomly selected nucleotides) was also observed in *experimental set 3* (Figure 5C) but was not observed in the control sets (*experimental set 2, 4*), which comprises mutations with mild effect on bacterial growth (Figure 5B and D). Interestingly, the significant enrichment of deleterious mutations among highly central nucleotides was found in two independent sets of data, which were obtained from different sources. By examining

different thresholds for high centrality definition, we obtained that in all thresholds (Supplementary Figure S2) there was an enrichment of nucleotides with strong deleterious mutation (*experimental set 1, 3*) among nucleotides with high closeness centrality. Here again, we did not observe a significant enrichment for the nucleotides with the mild phenotype (*experimental set 2, 4*), relatively to randomly selected nucleotides (Figure 5, Figure S2).

It has been shown (44,45) that evolutionary conservation (72) is high for both strong deleterious mutations and for a large fraction of the mild mutations. In order to quantify the ability of evolutionary rate (38), and the closeness centrality to classify strong deleterious mutations we applied a ROC curve analysis (65). Strong deleterious mutations comprising nucleotides in *experimental set 1 and 3* were labeled as 'positives' (433) and the remaining nucleotides in the 16S and 23S rRNA were labeled as 'negatives'. Our results showed (Supplementary Figure S3) that in the 16S rRNA, the performance of the closeness-based classifier was better (AUC = 0.77) than the evolutionary rate classifier (AUC = 0.67). An overall higher performance was observed in the 23S rRNA for both the closeness (AUC = 0.85) and the evolutionary rate classifiers (AUC = 0.84).

In summary, we have shown that deleterious mutations, in contrast to mild mutations, tend to have high closeness values in the rRNA. Furthermore, with the closeness parameter derived from a single ribosome structure we were able to distinguish the strong deleterious mutations from mutations which do not affect the ribosome function independent of the evolutionary conservation of the mutated nucleotides. Quite strikingly, in the small subunit the closeness parameter achieved even better performance compared to the evolution-based parameter, which is traditionally used for defining functionality.

CONCLUSION

Here, we have applied a network approach to study the unique properties of the bacterial and the archaeal rRNA. We show that the different ribosome structures have properties of a small-world network and that the network properties correlate weakly with structural properties of the ribosome as well as with evolutionary conservation. Interestingly, the active sites of the ribosome demonstrated high centrality measures independent of the surface accessibility and evolutionary conservation, reinforcing their unique structural properties. Most prominently, we show that the network parameters can point out nucleotides with critical functional roles. Specifically, we found that the major functional nucleotides in the ribosome demonstrate significantly high centrality values in both global and local properties. In addition, we were able to distinguish mutations in nucleotides, which have deleterious effect on the ribosome function or assembly from those which cause mild effect, based on their network centrality. We suggest that the closeness centrality measures can be used to classify functional and critical positions in the ribosome, independent of their evolutionary conservation. In conclusion, we present a new approach to

characterize essential positions in the bacterial ribosome, which can also be exploited as a tool for structural comparative analysis of large RNA structures.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Ilana Agmon and Anat Bashan for their help with ribosome structure analysis. Thanks to Alexander Mankin for the support and helpful suggestions during the course of the project and for valuable comments on the article. Many thanks to Ron Pinter for the helpful advice on the network analysis and for critical remarks on the article. This work was funded by the by the Israeli Science Foundation ISF grant number 923/05 granted to Y.M.G. Funding to pay the Open Access publication charges for this article was provided by the Israel Science Foundation Grant number 923/05.

Conflict of interest statement. None declared.

REFERENCES

- Steitz,T.A. (2008) A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.*, **9**, 242–253.
- Ramakrishnan,V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557–572.
- Maguire,B.A., Beniaminov,A.D., Ramu,H., Mankin,A.S. and Zimmermann,R.A. (2005) A protein component at the heart of an RNA machine: the importance of protein 127 for the function of the bacterial ribosome. *Mol. Cell*, **20**, 427–435.
- Schlunzen,F., Tocilj,A., Zarivach,R., Harms,J., Gluehmann,M., Janell,D., Bashan,A., Bartels,H., Agmon,I., Franceschi,F. *et al.* (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615–623.
- Wimberly,B.T., Brodersen,D.E., Clemons,W.M. Jr, Morgan-Warren,R.J., Carter,A.P., Vornrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
- Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
- Pioletti,M., Schlunzen,F., Harms,J., Zarivach,R., Gluehmann,M., Avila,H., Bashan,A., Bartels,H., Auerbach,T., Jacobi,C. *et al.* (2001) Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO J.*, **20**, 1829–1839.
- Schuwirth,B.S., Borovinskaya,M.A., Hau,C.W., Zhang,W., Vila-Sanjurjo,A., Holton,J.M. and Cate,J.H. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science*, **310**, 827–834.
- Selmer,M., Dunham,C.M., Murphy,F.V.4th., Weixlbaumer,A., Petry,S., Kelley,A.C., Weir,J.R. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
- Korostelev,A., Trakhanov,S., Laurberg,M. and Noller,H.F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*, **126**, 1065–1077.
- Cheatham,T.E. III (2004) Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.*, **14**, 360–367.
- Wang,Y., Rader,A.J., Bahar,I. and Jernigan,R.L. (2004) Global ribosome motions revealed with elastic network model. *J. Struct. Biol.*, **147**, 302–314.
- Trylska,J., Tozzini,V. and McCammon,J.A. (2005) Exploring global motions and correlations in the ribosome. *Biophys. J.*, **89**, 1455–1463.

14. Vishveshwara, S., Brinda, K.V. and Kannan, N. (2002) Protein structure: insights from graph theory. *J. Theor. Comp. Chem.*, **1**, 187–211.
15. Grindley, H.M., Artymiuk, P.J., Rice, D.W. and Willett, P. (1993) Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol.*, **229**, 707–721.
16. Jacobs, D.J., Rader, A.J., Kuhn, L.A. and Thorpe, M.F. (2001) Protein flexibility predictions using graph theory. *Proteins*, **44**, 150–165.
17. Amitai, G., Shemesh, A., Sitbon, E., Shklar, M., Netanel, D., Venger, I. and Pietrokovski, S. (2004) Network analysis of protein structures identifies functional residues. *J. Mol. Biol.*, **344**, 1135–1146.
18. Atilgan, A.R., Akan, P. and Baysal, C. (2004) Small-world communication of residues and significance for protein dynamics. *Biophys. J.*, **86**, 85–91.
19. Bagler, G. and Sinha, S. (2005) Network properties of protein structures. *Physica A Stat. Mech. Appl.*, **346**, 27–33.
20. del Sol, A. and O'Meara, P. (2005) Small-world network approach to identify key residues in protein-protein interaction. *Proteins*, **58**, 672–682.
21. Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.
22. Taylor, T.J. and Vaisman, I.I. (2006) Graph theoretic properties of networks formed by Delaunay tessellation of protein structures. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **73**, 041925.
23. Dokholyan, N.V., Li, L., Ding, F. and Shakhnovich, E.I. (2002) Topological determinants of protein folding. *Proc. Natl Acad. Sci. USA*, **99**, 8637–8641.
24. Vendruscolo, M., Dokholyan, N.V., Paci, E. and Karplus, M. (2002) Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **65**, 061910.
25. Bode, C., Kovacs, I.A., Szalay, M.S., Palotai, R., Korcsmaros, T. and Csermely, P. (2007) Network analysis of protein dynamics. *FEBS Lett.*, **581**, 2776–2782.
26. Chea, E. and Livesay, D.R. (2007) How accurate and statistically robust are catalytic site predictions based on closeness centrality? *BMC Bioinform.*, **8**, 153.
27. del Sol, A., Fujihashi, H., Amoros, D. and Nussinov, R. (2006) Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. *Protein Sci.*, **15**, 2120–2128.
28. Thibert, B., Bredesen, D.E. and del Rio, G. (2005) Improved prediction of critical residues for protein function based on network and phylogenetic analyses. *BMC Bioinform.*, **6**, 213.
29. Hu, Z., Bowen, D., Southerland, W.M., Del Sol, A., Pan, Y., Nussinov, R. and Ma, B. (2007) Ligand binding and circular permutation modify residue interaction network in DHFR. *PLoS Comput. Biol.*, **3**, e117.
30. del Sol, A., Fujihashi, H. and O'Meara, P. (2005) Topology of small-world networks of protein-protein complex structures. *Bioinformatics*, **21**, 1311–1315.
31. Wuchty, S. (2003) Small worlds in RNA structures. *Nucleic Acids Res.*, **31**, 1108–1117.
32. Lescoute, A. and Westhof, E. (2006) The interaction networks of structured RNAs. *Nucleic Acids Res.*, **34**, 6587–6604.
33. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E.E. and Edelman, M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
34. Freeman, L.C. (1979) Centrality in social networks I: conceptual clarification. *Soc. Networks*, **1**, 215–239.
35. Brandes, U. (2001) A faster algorithm for betweenness centrality. *J. Math. Sociol.*, **25**, 163–177.
36. Cavallo, L., Kleinjung, J. and Fraternali, F. (2003) POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res.*, **31**, 3364–3366.
37. Fraternali, F. and Cavallo, L. (2002) Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. *Nucleic Acids Res.*, **30**, 2950–2960.
38. Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.*, **35**, 7188–7196.
39. Triman, K.L. (2007) Mutational analysis of the ribosome. *Adv. Genet.*, **58**, 89–119.
40. Kim, H.M., Yeom, J.H., Ha, H.J., Kim, J.M. and Lee, K. (2007) Functional analysis of the residues C770 and G771 of *E. coli* 16S rRNA implicated in forming the intersubunit bridge B2c of the ribosome. *J. Microbiol. Biotechnol.*, **17**, 1204–1207.
41. Liiv, A., Karitkina, D., Maivali, U. and Remme, J. (2005) Analysis of the function of *E. coli* 23S rRNA helix-loop 69 by mutagenesis. *BMC Mol. Biol.*, **6**, 18.
42. Liiv, A. and O'Connor, M. (2006) Mutations in the intersubunit bridge regions of 23S rRNA. *J. Biol. Chem.*, **281**, 29850–29862.
43. Pulk, A., Maivali, U. and Remme, J. (2006) Identification of nucleotides in *E. coli* 16S rRNA essential for ribosome subunit association. *RNA*, **12**, 790–796.
44. Yassin, A., Fredrick, K. and Mankin, A.S. (2005) Deleterious mutations in small subunit ribosomal RNA identify functional sites and potential targets for antibiotics. *Proc. Natl Acad. Sci. USA*, **102**, 16620–16625.
45. Yassin, A. and Mankin, A.S. (2007) Potential new antibiotic sites in the ribosome revealed by deleterious mutations in RNA of the large ribosomal subunit. *J. Biol. Chem.*, **282**, 24329–24342.
46. Greene, L.H. and Higman, V.A. (2003) Uncovering network systems within protein structures. *J. Mol. Biol.*, **334**, 781–791.
47. Barnett, L. and Paolo, E.D. (2007) Spatially embedded random networks. *Phys. Rev.*, **76**, 056115.
48. Milenkovic, T., Lai, J. and Przulj, N. (2008) GraphCrunch: a tool for large network analyses. *BMC Bioinform.*, **9**, 70.
49. Erdos, P. and Renyi, A. (1959) On random graphs. *Publ. Math.*, **6**, 290–279.
50. Penrose, M. (2003) *Random Geometric Graphs*. Oxford University Press, Oxford.
51. Barabasi, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
52. Przulj, N. and Higham, D. (2006) Modelling protein-protein interaction networks via a stickiness index. *J. Roy. Soc. Inter.*, **3**, 711–716.
53. Hormozdiari, F., Berenbrink, P., Przulj, N. and Sahinalp, S.C. (2007) Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution. *PLoS Comput. Biol.*, **3**, e118.
54. Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O. and Bahar, I. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, **80**, 505–515.
55. Newman, M.E., Girvan, M. and Farmer, J.D. (2002) Optimal design, robustness, and risk aversion. *Phys. Rev. Lett.*, **89**, 028301.
56. Fernandez, A. (1991) Excluded-volume effects on the stacking of RNA base pairs. *Phys. Rev. A*, **44**, R7910–R7912.
57. Bashan, A., Agmon, I., Zarivach, R., Schlunzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H.A. et al. (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol. Cell*, **11**, 91–102.
58. Nissen, P., Hansen, J., Ban, N., Moore, P.B. and Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–930.
59. Schlunzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A. and Franceschi, F. (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, **413**, 814–821.
60. Schmeing, T.M., Seila, A.C., Hansen, J.L., Freeborn, B., Soukup, J.K., Scaringe, S.A., Strobel, S.A., Moore, P.B. and Steitz, T.A. (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat. Struct. Biol.*, **9**, 225–230.
61. Tang, Y.R., Sheng, Z.Y., Chen, Y.Z. and Zhang, Z. (2008) An improved prediction of catalytic residues in enzyme structures. *Protein Eng. Des. Sel.*, **21**, 295–302.
62. Agmon, I., Bashan, A., Zarivach, R. and Yonath, A. (2005) Symmetry at the active site of the ribosome: structural and functional implications. *Biol. Chem.*, **386**, 833–844.
63. Harms, J., Schlunzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001)

- High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679–688.
64. Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.
65. Metz, C.E. (1978) Basic principles of ROC analysis. *Semin. Nucl. Med.*, **8**, 283–298.
66. Bahar, I., Atilgan, A.R. and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des.*, **2**, 173–181.
67. Yang, L.W., Liu, X., Jursa, C.J., Holliman, M., Rader, A.J., Karimi, H.A. and Bahar, I. (2005) iGNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics*, **21**, 2978–2987.
68. Chennubhotla, C., Rader, A.J., Yang, L.W. and Bahar, I. (2005) Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.*, **2**, S173–S180.
69. Voss, N.R., Gerstein, M., Steitz, T.A. and Moore, P.B. (2006) The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.*, **360**, 893–906.
70. Agmon, I., Auerbach, T., Baram, D., Bartels, H., Bashan, A., Berisio, R., Fucini, P., Hansen, H.A., Harms, J., Kessler, M. *et al.* (2003) On peptide bond formation, translocation, nascent protein progression and the regulatory properties of ribosomes. Derived on 20 October 2002 at the 28th FEBS Meeting in Istanbul. *Eur. J. Biochem.*, **270**, 2543–2556.
71. Triman, K.L., Peister, A. and Goel, R.A. (1998) Expanded versions of the 16S and 23S ribosomal RNA mutation databases (16SMDBexp and 23SMDBexp). *Nucleic Acids Res.*, **26**, 280–284.
72. Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinform.*, **3**, 2.