Modular Organization of Residue-Level Contacts Shapes the Selection Pressure on Individual Amino Acid Sites of Ribosomal Proteins

Saurav Mallik[1,2] and Sudip Kundu[1,2,*]

[1]Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, Kolkata, India

[2]Center of Excellence in Systems Biology and Biomedical Engineering (TEQIP Phase-II), University of Calcutta, Kolkata, India

*Corresponding author: E-mail: skbmbg@caluniv.ac.in.

Accepted: February 21, 2017

Abstract

Understanding the molecular evolution of macromolecular complexes in the light of their structure, assembly, and stability is of central importance. Here, we address how the modular organization of native molecular contacts shapes the selection pressure on individual residue sites of ribosomal complexes. The bacterial ribosomal complex is represented as a residue contact network where nodes represent amino acid/nucleotide residues and edges represent their van der Waals interactions. We find statistically overrepresented native amino acid–nucleotide contacts (OaantC, one amino acid contacts one or multiple nucleotides, inter-nucleotide contacts are disregarded). Contact number is defined as the number of nucleotides contacted. Involvement of individual amino acids in OaantCs with smaller contact numbers is more random, whereas only a few amino acids significantly contribute to OaantCs with higher contact numbers. An investigation of structure, stability, and assembly of bacterial ribosome depicts the involvement of these OaantCs in diverse biophysical interactions stabilizing the complex, including high-affinity protein–RNA contacts, interprotein cooperativity, intersubunit bridge, packing of multiple ribosomal RNA domains, etc. Amino acid–nucleotide constituents of OaantCs with higher contact numbers are generally associated with significantly slower substitution rates compared with that of OaantCs with smaller contact numbers. This evolutionary rate heterogeneity emerges from the strong purifying selection pressure that conserves the respective amino acid physicochemical properties relevant to the stabilizing interaction with OaantC nucleotides. An analysis of relative molecular orientations of OaantC residues and their interaction energetics provides the biophysical ground of purifying selection conserving OaantC amino acid physicochemical properties.

Key words: contact network, network motif, purifying selection, ribosome, protein–RNA interaction, stability.

Introduction

Macromolecular complexes are the functional units of a myriad of core cellular processes, such as replication, transcription, translation, protein degradation, and signal transduction (Saiz and Vilar 2006). In cellular environment, these complexes are biosynthesized in the course of ordered assembly pathways that involve specific temporal orders of binding of different subunits. These subunits are building blocks of macromolecular complexes and these building blocks are held together by different biophysical constraints ranging from direct physical interactions (Dubin et al. 2012) to long-distance cooperative relationships (Williamson 2008). Combination of different subunits encoded by different genes encompasses a large sequence space where fixation of mutations is constrained by the maintenance of these different biophysical interactions (Osada and Akashi 2012; Mallik et al. 2015; Marsh and Teichmann 2015; Weng et al. 2016).

Numerous studies in the past decades have established that selection pressure on monomeric proteins can vary more than 1000-fold (from rapidly evolving viral proteins to highly conserved metabolic enzymes) and this variation refers to several factors such as protein function, mRNA expression, protein abundance, synonymous codon usage, etc. (McInerney 2006). Individual amino acid sites of any protein also exhibit evolutionary rate heterogeneity that again depends on several factors including solvent accessibility, packing density, folding kinetics, flexibility, and functional constraints (Echave et al. 2016). Macromolecular complexes are generally subjected to

strong purifying selection pressure (van Dam and Snel 2008; Chakraborty and Ghosh 2013; Weng et al. 2016) that disfavors fixation of deleterious mutations and thereby maintains critical interactions in the course of evolution (Shah et al. 2015). However, the magnitude of difference of selection pressure between complexes and monomeric proteins is not so dramatic (Krylov et al. 2003), which suggests that depending on the nature of biophysical constraints certain regions of complexes can also evolve rapidly. This is supported by the observable residue-level positive selection in many critical macromolecular complexes, such as ribosome (Swanson et al. 2004), proteosome (Takada et al. 2015), and spliceosome (Roy and Gilbert 2006) though the overall substitution rates of their constituent proteins are very slow.

Biological systems, from the scale of molecular nanostructures to ecological populations, exhibit an inherent modular organization. Modular organization dictates that regulation of the entire system is an outcome of the interactions among its different building blocks. Interestingly, if we represent the ensemble of interactions of a biomolecular system in terms of a network (nodes represent structural units, edges represent their interactions), its topological characteristics (namely contact number) are observed to have significant correlations with the selection pressure on structural units (Fraser et al. 2002; Jovelin and Phillips 2009; Toft and Fares 2010; Echave et al. 2016). For example, physical interactions among the myriad of protein molecules within a cell can be portrayed as a protein–protein interaction (PPI) network. Proteins and macromolecular complexes can be represented in terms of networks of physically interacting residues (David-Eden and Mandel-Gutfreund 2008; Zhang et al. 2013), known as residue contact networks (RCN). For both PPI and RCN, the rate of evolution of a node (whole protein or an amino acid site) negatively correlates with its contact number (Fraser et al. 2002; Toft and Fares 2010; Echave et al. 2016), defined as the count of other nodes with which the node of interest has a direct contact. Exceptions to this trend include proteins/amino acids associated with critical functional interactions, which constrains them to evolve slowly despite their low contact numbers (Toft and Fares 2010). Though a number of studies have analyzed how topological aspects of RCNs and functional constraints on residue sites shape the evolutionary rate heterogeneity of amino acid sites of globular proteins, a thorough investigation on macromolecular complexes from this perspective is still missing.

Here, taking advantage of the available high-resolution crystallographic data of bacterial ribosomal subunits, wide-ranging experimental data regarding ribosome biogenesis, stability and function, and the rich bacterial genomic information, we seek to address how the modular organization of native molecular contacts between protein–rRNA constituents shapes the selection pressure on individual amino acid sites of ribosomal proteins. Generally, networks are comprised over-represented distinct subgraph patterns (termed as motifs) that represent functional building blocks (Neduva et al. 2005; Shen-Orr et al. 2002; Wuchty et al. 2003, Milo et al. 2002). By definition, a motif consists of at least two nodes and two motifs having the same number of nodes are differentiated in terms of the subgraph connection patterns (Wuchty et al. 2003). For this investigation, we identify all the statistically overrepresented amino acid–nucleotide contacts (OaantCs) present in the native crystal structures. We only focus on OaantCs where one amino acid contacts one or multiple nucleotides (contact number is defined as the number of nucleotides contacted) while all the internucleotide contacts are disregarded. We find OaantCs with higher contact numbers are overrepresented in critical biophysical interactions, such as high-affinity protein–rRNA contacts, cooperativity phenomenon, protein extensions stabilizing multiple rRNA domains, exit-tunnel wall, and intersubunit bridging. These biophysical constraints are associated with significantly slower evolutionary rates of OaantCs ω with higher contact numbers, due to an extreme decrease in nonsynonymous substitution rates rather than a difference in synonymous substitution rates. In the course of evolution, accepted point-mutations in OaantCs with higher contact numbers are comparatively more conservative than substitutions at OaantCs with smaller contact numbers. The conservative nature of substitutions in OaantCs with higher contact numbers has a stronger tendency of maintaining amino acid chemical properties, such as charge, hydropathy, and polarity. There are two factors that determine whether purifying selection pressure would conserve a specific chemical property in an OaantC amino acid site. The first factor is the relative 3D orientation of amino acid–nucleotide components that determines which chemical property would be relevant to the stabilizing interaction. For example, an amino acid site that is surrounded by negatively charged nucleotide phosphates in 3D space generally experiences purifying selection pressure conserving positive charge at that residue site. The second factor is the magnitude of the net interaction energy and its stabilizing/destabilizing nature. In summary, OaantC analyses combined with advanced phylogenetic analysis and biophysical techniques suggest the molecular basis of purifying selection in ribosomal particle in the course of evolution.

Materials and Methods

Sequence and Structure Data Set

For the present work, r-protein gene sequences for 12 Gram-Negative Gammaproteobacteria species (*Escherichia coli*, *Buchnera aphidicola*, *Morganella morganii*, *Pantoea ananatis*, *Proteus mirabilis*, *Salmonella typhi*, *Citrobacter freundii*, *Cronobacter sakazakii*, *Xenorhabdus nematophila*, *Vibrio cholerae*, *Aeromonas hydrophila*, *Shewanella baltica*) are retrieved from GenBank (Benson et al. 2015) under the following criteria: 1) Variation of protein length averaged over all the

r-proteins is ≤2 amino acids; 2) number of synonymous substitutions per synonymous site, averaged over all the r-proteins at all branches of the phylogenetic tree is $d_S < 0.7$; and 3) similar habitat conditions (lives in animal/insect guts, optimum growth temperature ranges between 33 and 35 °C) of the respective organisms. The first two criteria ensure minimum unreliable regions in the alignment which ultimately leads to minimum false positive prediction of ancestral sequences and selection phenomena. The third criterion minimizes the effect of any episodic selection emerging from environmental adaptation. These sequences have an overall 85% sequence similarity. For structural analysis, the recently published X-ray crystallographic structure of the 70S ribosomal complex (PDB ID: 4YBB) resolved at 2.1 Å resolution (Noeske et al. 2015) is used along with ten additional crystal structures used in our previous work (PDB IDs 3R8S, 3OAS, 3OFC, 1VT2, 3ORB, 2QAM, 2QOZ, 2I2T, 1VS6, and 2AW4). These ten structures were previously chosen under the criterion that they exclude modified residues, full-length tRNA, mRNA, elongation, and release factors (Mallik and Kundu 2015).

Network Analysis

Network Construction

Undirected, unweighted residue-level amino acid–nucleotide contact network of the full 70S ribosome is designed where individual amino acid and nucleotide residues represent the nodes, and edges represent van der Waals interactions between residue pairs. The distance cutoff for van der Waals interactions differs for different atom types, depending on the atomic radii and protonation status. For simplicity, here we consider van der Waals interactions between any two residues (amino acid/nucleotide) if any two nonhydrogen atoms from two different residues are within cutoff distance $d_{ij} <$ 5.0 Å (Aftabuddin and Kundu 2007). A subnetwork of the full 70S contact network, which includes only amino acid–nucleotide contacts (contacts between 30S proteins and 16S rRNA, 50S proteins and 23S + 5S rRNA, 30S proteins and 23S rRNA and 50S proteins and 16S rRNA), is subjected to OaantCs analysis. The interamino acid and internucleotide contacts (both intrasubunit and intersubunit) are excluded. All the edges are considered equivalent. The consensus network (an edge is considered if it is present in at least 70% of the networks) for 11 unique crystal structures includes 2,279 amino acids having 6,132 contacts with 1,645 nucleotides.

Construction of Subnetworks Associated with Different Functional Aspects

We generate subnetworks associated with seven different functional aspects of ribosomal particles: 1) Critical and high-affinity protein–rRNA contacts, 2) interprotein cooperativity, 3) 5S rRNA assembly, 4) molecular packing of multiple rRNA domains, 5) intersubunit bridging, 6) stabilizing the exit-tunnel wall, and 7) peptidyltransferase center (PTC). For each of these seven functional aspects, we follow the following steps to generate the subnetwork: 1) From the available literature we identify the amino acid and nucleotide sites associated with the respective function, 2) we map these sites with the nodes present in the original consensus network, and 3) based on this mapping, we extract a subnetwork that includes contacts among the mapped amino acid–nucleotide residues only. For example, to construct the subnetwork associated with stabilizing the exit-tunnel wall, first we identify the 50 amino acid (refer to uL4, uL22, and uL24 proteins) and 80 nucleotide sites (refer to 23S rRNA) that constitute the inner surface of the exit-tunnel wall (supplementary data, Supplementary Material online) from the work of Voss et al. (2006). Mapping these residue sites with the consensus network, we extract a subnetwork consisting 130 nodes (50 amino acids and 80 nucleotides) and 391 edges (van der Waals interactions among residues). Similarly, we construct the subnetworks associated with other functional aspects (see supplementary data, Supplementary Material online).

Selection of Network Randomization Model

A crucial part of overrepresented pattern detection in any biological network is the choice of an appropriate random network model against which the overrepresentation is estimated. Therefore, first we estimate the statistical fit of different random network models to our subnetwork following Milenković et al. (2009) pipeline. Different random models include Erdös-Rényi model (Erdös-Rényi 1959), Erdös-Rényi with same degree distribution, Barabasi–Albert type scale free (Barabasi and Albert 1999), stickiness-index based random graph (Pržulj and Higham 2006), and 3D geometric random graphs (Penrose 2003). We use two highly constraining graphlet-based measures of structural similarity between the real and random networks—where graphlets are small substructures of large networks (Pržulj et al. 2004)—relative graphlet frequency distance (RGF-distance) and graphlet degree distribution agreement (GDD-agreement), estimated using GraphCrunch software package (Milenkovic et al. 2008). A high GDD-agreement and low RGF-distance denote a strong structural similarity between the original network and the random model.

GDD-agreement and RGF-distance-based random model fitting shows that Erdös-Rényi model has the strongest similarity with our amino acid–nucleotide contact network (supplementary table 1, Supplementary Material online). This is in contrast to the results obtained by Milenković et al. (2009) that geometric random graphs (GEO-3D) exhibit the highest similarity with residue-contact networks of globular proteins. This is in contrast to the results obtained by Milenković et al. (2009) that geometric random graphs (GEO-3D) exhibit the highest similarity with residue-contact networks of globular proteins. This discrepancy may arise due to several reasons,

including the one discussed below. The random network models used here are only topologically similar to the original residue contact network. But these random networks lack the preference of residue–residue contacts reflected in biological macromolecules based on their sequence separation, residue type, secondary structure, and even the neighborhood (Milenković et al. 2009). The preference of residue connectivity is entirely different at protein interiors, where interactions among hydrophobic amino acids are preferred, compared with that at protein–rRNA interfaces, where interactions among positively charged amino acids and nucleotide backbone are preferred (Nick Pace et al. 2014).

Identification of Overrepresented Contacts

To identify all the statistically overrepresented native amino acid–nucleotide contacts (network motifs), compared with random networks, we perform the following analysis. Our consensus network includes 2,279 amino acid and 1,645 nucleotide nodes and total 6,132 edges. In this network the number of nucleotides contacted by a single amino acid (k) varies between 1 and 12. We collate the entire network into isomorphic patterns (subgraphs), where one amino acid contacts $\geq$1 nucleotides, but internucleotide contacts are disregarded. FANMOD standalone tool (Wernicke and Rasche 2006) is used to identify these subgraphs. The number of occurrences of each subgraph is enumerated in the input network. For example, we find 35 subgraphs with $k = 7$ (fig. 1*A*). Next, we determine which of these subgraphs are topologically equivalent and they are grouped into unique subgraph classes accordingly. We further consider only those subgraphs that have $\geq$5 number of occurrences in the consensus network (Yeger-Lotem et al. 2004); this filtering excludes $k = 11$ and $k = 12$ subgraphs. Finally, statistical overrepresentation of each subgraph in the original network is estimated by computing their occurrences in 1 million Erdös-Rényi model random networks having the same number of nodes and edges as the original network. If the overrepresentation of a subgraph in the original network is associated with statistical significance $P < 0.05$ (one sample permutation Z-test) we consider it an OaantC (supplementary table 2, Supplementary Material online). This includes 2,276 OaantCs. For example, the OaantC, where a single amino acid contacts seven nucleotides, is associated with a statistical significance of $P < 10^{-16}$. The schematic representation and frequencies of occurrences of these OaantCs are shown in figure 1*A*.

Sequence Analysis

The r-protein gene sequences are codon-aligned using ClustalW (Larkin et al. 2007). Phylogenetic inference is made by Maximum Likelihood method (BioNJ initial tree, GTR model, 1,000 bootstrap replicates) implemented in MEGA version 6.0 (Tamura et al. 2013).

The ratio of nonsynonymous and synonymous substitution rates ($\omega = d_N/d_S$) represents the selection pressure on a protein or on individual amino acid sites. For each r-protein, the selective pressure is estimated using branch-site codon substitution models implemented in CodeML of the PAML software package (Yang 2007). Codon substitution site-models implemented in CodeML assume that selective pressures do not act equally on the entire protein sequence but vary between amino acid sites in a gene family. Because no a priori expectation exists for positive selection (all the taxa are habituated in nearly same environmental condition), we estimate and compare the log-likelihood values ($ln\,L$) between two pairs of models with different assumed ω distributions: 1) M1a (model of neutral evolution), where all the sites are assumed to be either negatively selected ($\omega < 1$) or undergoing neutral selection ($\omega = 1$) versus M2a (model of positive selection), which in addition to the site classes mentioned for M1a assumes a third category of sites; sites with $\omega > 1$. 2) M7 (β) versus M8 (β and ω).

The comparison between a null model that does not consider $\omega > 1$ (M1a and M7) and an alternate model that does consider $\omega > 1$ (M2a and M8) is a measure of positive selection. If the null (neutral) model is rejected in favor of the alternative (selection) model, positive selection is inferred (Yang 2006). The *LRT* value is estimated as $LRT = 2(ln\,L_{alternative} - ln\,L_{null})$, where $ln\,L$ is the maximum-likelihood estimate of the probabilities of selection. The LRT is asymptotically distributed as a χ^2_k function of k degrees of freedom. This k is given as the difference of the number of parameters (np) in the null and alternative models. The *LRT* values, for which the χ^2_k distribution shows a P-value < 0.05, indicate that some sites are significantly under positive selection. Finally, the Bayes Empirical Bayes (BEB) approach (Yang et al. 2005) is employed to identify sites under positive selection by calculating the posterior probabilities (PP) that a particular site belongs to a given selection class (neutral, negatively selected, or positively selected). Sites with $PP > 0.9$ are inferred to be positively selected. Additionally, we have used Fast Unconstrained Bayesian Approximation (FUBAR) (Murrell et al. 2013) implemented in HyPhy (Pond et al. 2005) to identify sites evolving under the influence of pervasive diversifying and purifying selection pressures.

Previous research on organelle ribosomal protein genes (Swanson et al. 2004) showed that they are generally subjected to strong purifying selection pressure, though in nearly 30% of r-proteins a few residue sites are found to be positively selected. Statistical analyses on bacterial ribosomal proteins identify signatures of residue-level positive selection in 25 out of 47 r-proteins included in the analysis. These positively selected sites include 44 amino acid sites only, supporting the previous reports that purifying selection is the major driving force of molecular evolution of r-proteins (Barreto and Burton 2013). We ask whether the strong purifying selection at individual amino acid sites maintains specific physicochemical properties in the course of evolution. We employ the PRoperty Informed Models of Evolution (PRIME), implemented

A

# OaantC	No. of OaantCs found (no. of selection event)	d_N/d_S	d_N/d_S random	d_N/d_S ratio	Shannon Entropy	Entropy (random)	Entropy ratio
1	552 (17)	0.143	0.138	1.036	0.678	0.521	1.301
2	765 (39)	0.091	0.124	0.734	0.657	0.504	1.304
3	380 (166)	0.086	0.176	0.489	0.644	0.476	1.353
4	292 (103)	0.078	0.184	0.424	0.627	0.485	1.293
5	164 (81)	0.069	0.217	0.318	0.619	0.519	1.193
6	63 (31)	0.057	0.161	0.354	0.607	0.527	1.152
7	35 (24)	0.052	0.155	0.335	0.571	0.495	1.154
8	14 (12)	0.051	0.234	0.217	0.432	0.531	0.814
9	6 (6)	0.050	0.109	0.459	0.269	0.661	0.407
10	5 (5)	0.050	0.191	0.262	0.064	0.698	0.092

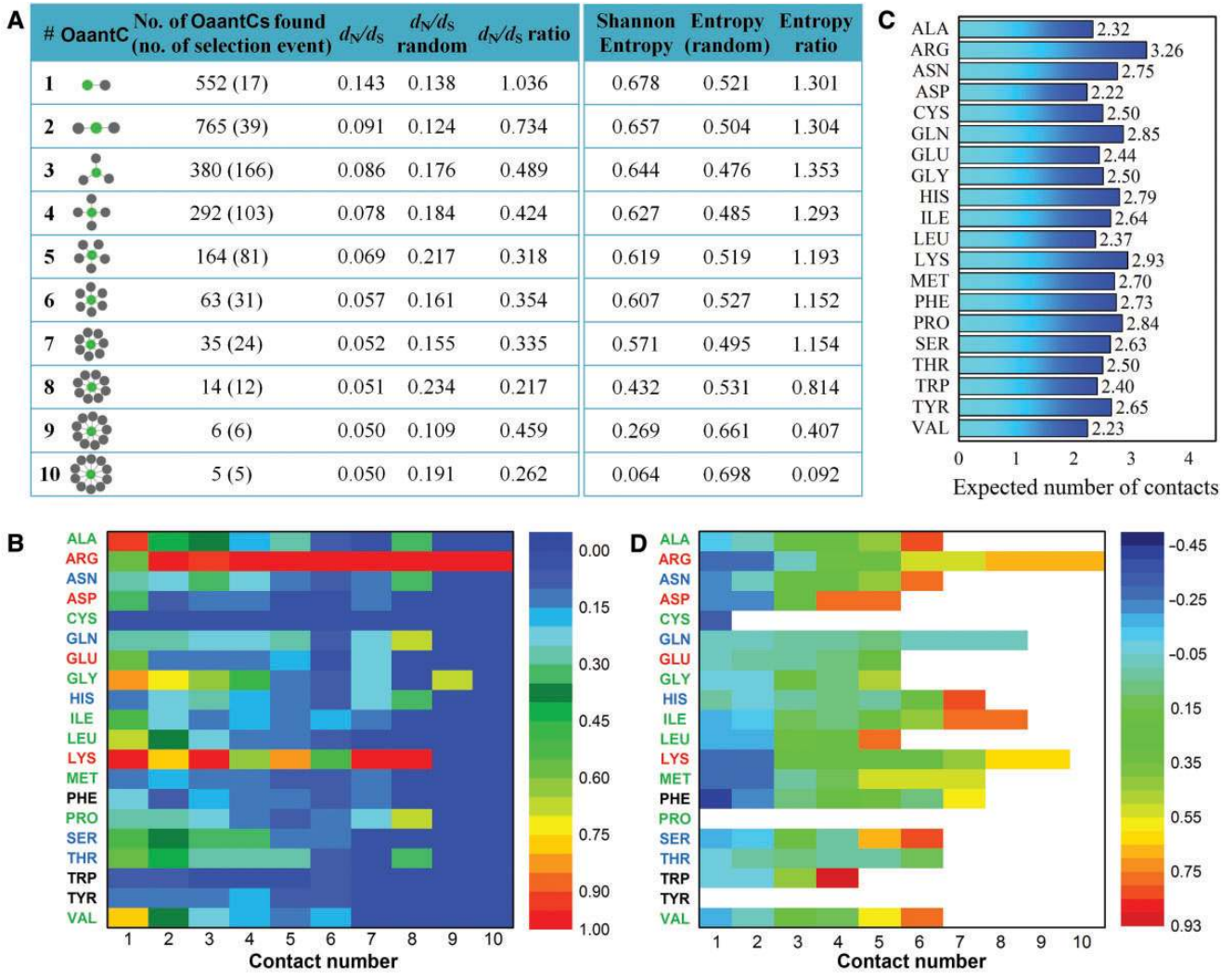

Fig. 1.—(*A*) Different OaantCs identified in the 70S ribosomal particle. In the second column, we show the different OaantCs. Left half of the panel enlists the amino acid data: 1) Column 3 represents the number of occurrences of each OaantCs, the number of cases where statistically significant signature of the respective amino acid chemical property being conserved is found is mentioned in the parenthesis, 2) average d_N/d_S ratio obtained at each OaantCs class (column 4), and 3) randomized evolutionary rates (column 5) and their ratios (column 6). Right half of the panel represents nucleic acid data: 1) Average Shannon entropy of nucleotides associated with each OaantCs class (column 7), 2) the randomized entropy values (column 8), and 3) their ratios (column 9). (*B*) Scaled occurrence bias of different types of amino acids in different contact OaantCs. The bias estimates are scales to the range 0–1 and are plotted as a colored heatmap. (*C*) The expected number of nucleotide contacts for individual amino acids is shown as bar-plots. (*D*) The heatmap represents the statistical bias of individual amino acids associated with different OaantCs classes being subjected to selection pressure conserving their chemical properties.

in the HyPhy package (Pond et al. 2005) for this analysis. The nonsynonymous rates (β) of replacing amino acid x with amino acid y is given as $\beta_{xy,x\neq y} = f^{(S)}\{\mathbf{d}(x,y)\}$, where $\mathbf{d}(x,y)$ is a vector that consists of property-specific distance measures, each of which indicates the degree of dissimilarity of amino acids x and y with respect to a particular biochemical property or set of properties. The function $f^{(S)}$ maps the distances into an exchangeability estimate for site S. This exchangeability depends on the relative importance of the various amino acid properties. Each individual prediction is compared with a null model by a likelihood ratio test using the χ^2 distribution to assess significance. The *LRT* values, for which the χ^2_k distribution shows a *P*-value < 0.05, indicate that some sites are significantly under purifying selection conserving the respective chemical property. We focus on three chemical properties (charge, polarity, hydropathy) in our analysis.

Amino Acid–Nucleotide Interaction Energetics

We estimate the electrostatic interaction energies of individual OaantCs (irrespective of the remaining structure)

using DelPhi electrostatic program (Sheinerman and Honig 2002), which assumes a minimally frustrated model of molecular recognition. For accurate energetic analysis, hydrogen atoms are added in the 70S structure and their positions are minimized (Donald et al. 2007) by 1,000 steepest descent steps and 200 conjugate gradient steps, using the AMBERff99SB force field (Ponder and Case 2003) implemented in UCSF Chimera (Pettersen et al. 2004). From thermodynamic point-of view, the total free energy of binding (ΔG^0_{bind}) can be represented as a sum of the following mutually independent components: Electrostatic energy of rigid binding (ΔG^0_{elec}), configurational entropy loss ($T\Delta S_{\text{conf}}$) and the structural deformation energy (ΔE_{def}). The native structure of ribosome is the result of a complex self-assembly mechanism including a series of structural reconstitutions (Shajani et al. 2011), which does not fit in the scope of a minimally frustrated model (Noel et al. 2016). Hence, for simplicity we exclude the configurational entropy and deformation energy components from the scope of our analysis and enumerate only the electrostatic energy of rigid binding (ΔG^0_{elec}). We use the CHARMM param22 charges and radii for our calculations (MacKerel et al. 1998), for 2 grids/Å scale with a grid size of 100. The electrostatic contribution to the free energy of binding in water with no salt is calculated as (Wang and Kollman 2000): $\Delta G^0_{\text{elec}} = \Delta G_{\text{Coul}} + (\Delta G^{4-80}_{\text{RFE}}{}_{a:b} - \Delta G^{4-80}_{\text{RFE}}{}_{a} - \Delta G^{4-80}_{\text{RFE}}{}_{b}) = \Delta G_{\text{Coul}} + \Delta\Delta G_{\text{sol}} = \Delta G_{\text{Coul}} + \Delta\Delta G_{\text{sol}}$, where $\Delta G^{4-80}_{\text{RFE}}{}_{a:b}$, $\Delta G^{4-80}_{\text{RFE}}{}_{a}$ and $\Delta G^{4-80}_{\text{RFE}}{}_{b}$ are the corrected reaction field energies of the whole OaantC (*a:b*), amino acid *a* in isolation, and nucleotides *b* in isolation as calculated by DelPhi with interior and exterior dielectric constants of 4.0 and 78.2, respectively. ΔG_{Coul} is defined as $\Delta G_{\text{Coul}} = E^{Coul}_{a:b} - E^{Coul}_{a} - E^{Coul}_{b}$, where $E^{Coul}_{a:b}$, E^{Coul}_{a} and E^{Coul}_{b} are the total Columbic energies of the complex, amino acid *a* and nucleotides *b*, respectively. The ionic contributions (ΔG_{ions}) are calculated by subtracting the total grid energy calculated in the 0.1 M salt condition by the total grid energy in no-salt condition. In salt solution, the electrostatic contribution to the rigid binding component of total free energy is $\Delta G^0_{\text{elec}} = \Delta G_{\text{Coul}} + \Delta\Delta G_{\text{sol}} + \Delta G_{\text{ions}}$.

Statistical Bias Estimation

The standard bias function for a parameter θ is defined as $B(\hat{\theta}) = \theta - E[\hat{\theta}_{\text{rand}}]$; where $E[\hat{\theta}_{\text{rand}}]$ is the expected value of $\hat{\theta}_{\text{rand}}$, obtained by 10,000 times random sampling of the data. To estimate the statistical significance of the bias, we compute the *Z*-score, defined as $Z = (\theta - E[\hat{\theta}_{\text{rand}}])/\sigma_E$, where σ_E is the standard deviation of $E[\hat{\theta}_{\text{rand}}]$. A positive/negative value of the bias function, associated with $P < 0.05$ statistical significance ensures the presence of positive/negative biasness in the data, with respect to the null statement that there is no biasness in the data.

Results and Discussions

Different OaantCs Identified in the Contact Network of 70S Particle

In different experiments ribosome structures are crystallized at different chemical conditions and conformational states according to the biological question being addressed, which gives rise to small-scale structural variations (Mallik and Kundu 2015). Here we seek a consensus of residue level interactions among 11 unique crystal structures of the 70S particle, where 2,279 amino acids make 6,132 contacts with 1,645 nucleotides. We identify all the OaantCs where an amino acid node contacts one or multiple nucleotide nodes, but internucleotide contacts are disregarded. The contact number of an OaantC (*k*) is defined as the number of nucleotides contacted. In 70S contact network contact number varies from $k = 1$ to 10. The schematic representations of OaantCs and their frequencies of occurrences are included in figure 1*A*.

Amino Acids Are More Randomly Distributed in OaantCs with Smaller Contact Numbers While Positively Charged Amino Acids Mostly Contribute OaantCs with Higher Contact Numbers

We investigate the statistical bias in amino acid abundance in various OaantCs with increasing contact number. For each OaantC class (one OaantC class means OaantCs having the same contact number), bias estimates (Materials and Methods) are scaled to the range 0 to +1 and are plotted in figure 1*B*. Depending on the size of the side chain, there must be a physical limit in how many nucleotides a specific amino acid can contact and this number should increase with the size of the side chain. Hence, we have estimated the expected number of nucleotides (k_e) contacted by each amino acid to determine whether the observed biased distribution of different amino acids in various OaantCs is purely due to their different side chain sizes or this biased distribution has some functional relevance. In a 10,000 times random resampling approach we pick up 100 amino acid sites (those that contact rRNA nucleotides) at each step and count the average number of nucleotides each amino acid is contacting. Averages of these estimates give us the expected number of nucleotides contacted by each amino acid and these values are plotted in figure 1*C*. For any random amino acid, the expected number of nucleotide contacted is 2.62. Except Arginine ($k_e = 3.26$), for all amino acids OaantCs with $k \geq 3$ represent an amino acid–nucleotide interaction where the number of nucleotides contacted by the amino acid is higher than that expected from random contacts.

Consistent with this result, distribution of different amino acids in OaantCs with smaller contact number ($k \leq k_e$) is observed to be more random, whereas in OaantCs with higher contact number ($k > k_e$) only a few amino acids occur in

significant proportions (fig. 1*B*). In other words, $k \leq k_e$ OaantCs are topologically quite similar to random contacts, and random graph generation models relying on small-world properties successfully reproduce the topological characteristics of these subgraphs (Estrada 2010). Small side-chain hydrophobic amino acids (Ala, Gly, Leu, Val), polar amino acids with smaller side chains (Ser, Thr), and positively charged amino acids (Arg, Lys) are found at high proportions in $k \leq 2$ OaantCs (fig. 1*B*). But as contact number increases ($k \geq 3$), polar amino acids become rare, biasness for hydrophobic amino acids with smaller side chains (Ala and Gly) gradually reduces, whereas Arg and Lys are found at the highest proportion.

A major fraction of rRNA contact is mediated through the disordered extensions of r-protein spanning the ribosome core. These extensions feature a distinct abundance of hydrophobic amino acids with smaller side chains (Gly, Ala, Ser), along with Arg and Lys (Klein et al. 2004; Mallik and Kundu 2015). Small residues exhibit the necessary backbone torsion angles (not accessible to other amino acids) and smaller sizes to pack in an around rRNA helices (Klein et al. 2004). However, small amino acids can hardly contact multiple nucleotides, for which 95% of Gly, Ala, and Ser amino acids appear at $k < 4$ OaantCs (fig. 1*B*). Stabilizing the rRNA core is the principle biological role of r-proteins, which is achieved by their positively charged amino acids neutralizing the rRNA backbone phosphates (Klein et al. 2004). Positively charged amino acids Arg and Lys also exhibit large side chains that can contact multiple backbone phosphates in the structure. These factors explain the observable high occurrence bias of Arg and Lys in all $k > 2$ OaantCs. Arg and Lys are overrepresented in r-proteins and these amino acids play a critical role in neutralizing the negatively charged backbone phosphates of rRNA. This may be the reason that these amino acids are overrepresented in $k \leq k_e$ OaantCs as well. Ribosomal proteins feature a general deficiency of Cys and negatively charged amino acids Glu and Asp, which explains their lower abundance in various OaantCs.

Involvement of Various OaantCs in Critical Biophysical Interactions

For PPI networks, it has been shown that different biological functions are associated with characteristic conserved topological motifs (Wuchty et al. 2003). Motivated by this observation, we investigate the overrepresentation of amino acid–nucleotide contacts in various critical biophysical interactions associated with maintaining ribosome assembly and function. We focus on seven functional aspects of 70S particle—1) critical and high-affinity protein–rRNA contacts, 2) interprotein cooperativity, 3) 5S rRNA assembly, 4) molecular packing of multiple rRNA domains, 5) intersubunit bridging, 6) stabilizing the exit-tunnel wall, and 7) PTC—and investigate OaantC overrepresentation in different subnetworks generated for each case (Materials and Methods). The involvement of various OaantCs in these seven critical biophysical interactions is demonstrated in figure 2. For the seven functional classes, each of the ten studied OaantCs is significantly overrepresented. As $1 \leq k \leq 2$ OaantCs are the most abundant, they are present in nearly equal proportions in each functional class, with the exception of OaantCs associated with molecular packing of multiple rRNA domains, where $1 \leq k \leq 2$ OaantCs are predominantly absent.

Experimentally Identified Critical and High-Affinity Protein–rRNA Contacts

Here we observe that molecular contacts between a protein and the nucleotides critical for its high-affinity binding generally involve $k > k_e$ OaantCs. For example, we observe eleven $k \geq 4$ OaantCs associated with the 410G–432A segment of helix h16 (*Escherichia coli* numbering) of 16S rRNA including Lys10 ($k = 4$), Arg13 ($k = 7$), Lys22 ($k = 4$), Lys31 ($k = 4$), and Gln40 ($k = 8$). These OaantCs play critical roles in stabilizing the intrahelix loop of h16 and induce intimate molecular packing of h16 with h4 and h17 helices (Brodersen et al. 2002). Deletion experiments have earlier confirmed that these contacts are critical for initial recognition of uS4 protein and its high-affinity binding (Bellur and Woodson 2009).

Another interesting example is a positively charged pocket created by Arg5 ($k = 3$) and Arg10 ($k = 3$) of uS7 and Arg109 ($k = 3$) and Arg122 ($k = 9$) of uS9 at SSU head region that neutralizes the negative charge cluster of the S-turn motif at the junctions of helices h29 and h43. Stabilization of this critical helix junction ensures native molecular packing of the entire SSU head region and supports the native orientations of two complex S-turn motifs at h29–h43 junction and h41–h43 participate in these protein–rRNA contacts (Brodersen et al. 2002).

Maintenance of Inter-protein Cooperativity

The thermodynamic principle of cooperativity is manifested as energetic coupling among different subunits of the whole structure that ensures the fact that the whole is more stable than the sum of its parts, which drives the complex self-assembly forward (Williamson 2008). We identify 74 nucleotides in SSU and 61 nucleotides in LSU that are shared by at least two $k \leq k_e$ OaantCs in two cooperative proteins. There are 537 such $k \leq k_e$ OaantCs (fig. 2) in the entire 70S complex. For example, 597G (contacts Tyr86, Glu124 from uS8 and Phe28, Phe37 from uS17, $\langle k \rangle = 4.1$) and 598U (contacts Tyr86, Gly123, Glu124 from uS8 and Lys30, Phe37 from uS17, $\langle k \rangle = 3.5$) maintain a critical interdomain cooperativity in SSU between uS8 and uS17. Experiments by Moine et al. (1997) and Allmang et al. (1994) demonstrate that point mutations at these two nucleotide sites completely block uS8 binding and successive ribosome assembly.

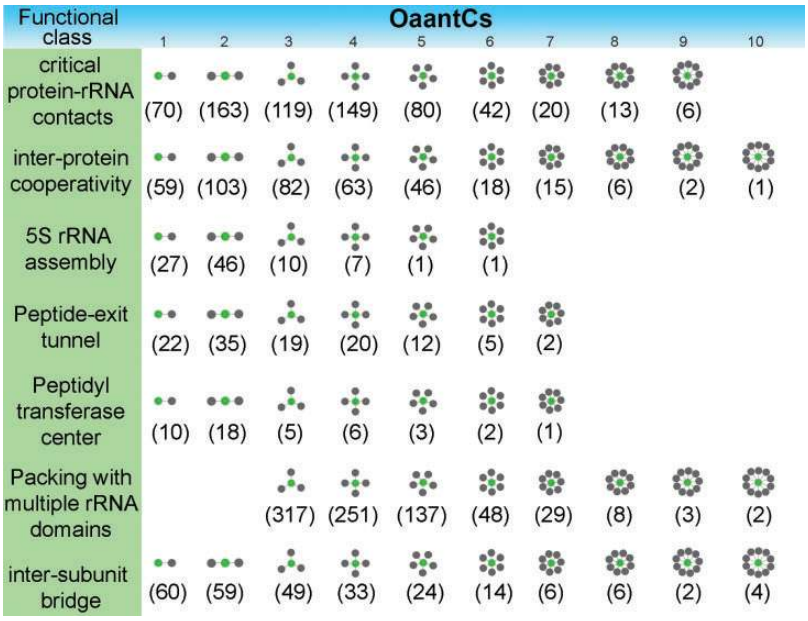

Fig. 2.—OaantCs associated with various functional aspects of the 70S particle. Complex interactions such as stabilizing multiple ribosomal RNA domain interfaces, cooperativity phenomenon, and intersubunit bridging OaantCs with higher contact number. One OaantC can be associated with multiple types of interactions. For example, a major fraction of the OaantCs maintaining interprotein cooperativity also maintain the molecular packing among multiple rRNA domains. The frequency of occurrences of different OaantCs in the respective functional classes is mentioned in the parenthesis.

1202U nucleotide is another fascinating example, that maintains a critical cooperative interaction between uS10 (Pro55, His56, $\langle k \rangle = 6.5$) and uS14 (Ala2, Thr67, Arg69, Ile82, Lys83, and Glu86, $\langle k \rangle = 3.4$). Göringer et al. (1991) showed that a U→C mutation at this site is dominant lethal. Nucleotide 1351U contacts Asp33 ($k = 3$) and Lys35 ($k = 5$) from uS7 and Arg119 ($k = 4$) and Lys120 ($k = 3$) from uS9. This nucleotide is an essential part of uS7:uS9 cooperativity and an U→C mutation at this site reduces uS7 binding affinity and binding rate (Dragon et al. 1996).

OaantCs Involved in 5S rRNA Assembly

We identify 92 OaantCs associated with specific and high-affinity binding of 5S rRNA with uL5, uL18, and bL25 proteins. These OaantCs involve Lys14 ($k = 6$), Met25 ($k = 5$) and Tyr31 ($k = 4$) of bL25, Ser24 ($k = 4$) and Met26 ($k = 6$) of uL5, and Ser45 ($k = 6$), Val54 ($k = 6$) and Glu46 ($k = 5$) of uL18. Iodine cleavage experiments demonstrate these OaantC amino acid sites protect the nucleotides (G6, G7, G10, G21, A34, and A73) critical for 5S rRNA binding (Shpanchenko et al. 1996).

OaantCs Involved in Maintaining Molecular Packing among Multiple rRNA Domains

There are 795 OaantCs associated with protein segments stabilizing the packing among multiple rRNA domains. The Ile85–Arg99 extension of uL22 is one interesting example that maintains contacts with four 23S rRNA domains (domain II, III, IV, and V) near the exit-tunnel wall (Klein et al. 2004). Nine $k > k_e$ OaantCs ($\langle k \rangle = 4.7$) are identified at this region including Arg88 ($k = 9$), Arg95 ($k = 5$), and Arg99 ($k = 5$). Experimental deletion of this segment results very weak incorporation of uL22 into the LSU during the biogenesis, though ribosomes remain functional except for a widened exit-tunnel structure (Zengel et al. 2003). The Lys88 of uS12 ($k = 5$) is another fascinating example (maintaining contacts with 5′ and Central domain of 16S rRNA), K→Q mutation at which results 15% reduction of growth rate (Toivonen et al. 1999).

A significant fraction of these OaantCs involve long stretches of idiosyncratically folded disordered segments that tightly hold multiple rRNA domains together. These extensions exhibit a distinct amino acid composition (rich in Ala, Gly, Arg, and Lys) and lack the hydrophobic cores typical of globular proteins. Segregation of positively charged residues at the disordered extensions enables them to neutralize the negatively charged backbone phosphates in the core of the subunit, which is critical for the proper rRNA folding (Brodersen et al. 2002; Klein et al. 2004).

OaantCs Involved in Intersubunit Bridges

The two subunits of bacterial ribosome join together via a series of 12 distinct, conserved intersubunit bridges to commence the translation process (Liiv and O'Connor 2006). Five of these bridges (B1a, B1b, B4, B5, B6, B7b, B8) involve

ribosomal protein segments and we identify 183 OaantCs with average contact number $\langle k \rangle = 4.1$ associated with these bridges. For example, in B4 bridge contacts between uS15 protein and h34 helix of 23S rRNA involve 28 OaantCs. Examples include Ser52 ($k = 4$), Arg54 ($k = 5$), Lys65 ($k = 6$), and Lys48 ($k = 4$). Komoda et al. (2006) showed that deletion of entire h34, which disrupts these key OaantCs, moderately inhibits bacterial growth. Liiv and O'Connor (2006) systematically performed UAAC→AUCU, UAAC→CCUC, UAAC→UUGU and UAAC→GCAC mutations at 714U–717C region of h34, which again disrupts the above-mentioned key OaantCs, resulted reduction of bacterial growth rate at 30 °C.

Bridge B6 and B8 are two other fascinating examples where bL19 interacts with 16S rRNA 1431A loop of h44 helix and h14 helix, respectively. Deletion of bridge B6 and B8 nucleotides disrupts 11 ($\langle k \rangle = 6.8$) and 9 OaantCs ($\langle k \rangle = 6.2$) associated with bL19, both severely inhibiting cell growth (McClory et al. 2010).

Interactions at Exit-Tunnel and PTC

Protein–rRNA interactions at the exit-tunnel wall involve 119 OaantCs (fig. 2) majority of which are contributed by uL4, uL22, and uL23. For example, a number of large, conserved OaantCs are found at Met82–Arg99 segment of uL22 (Arg88, Arg95 and Arg99, $\langle k \rangle = 6.4$). Zengel et al. (2003) showed that deletion of this segment drastically reduces uL22 binding affinity.

The PTC is the functional center of ribosome and this region is poor in protein presence (Klein et al. 2004). However, 48 OaantCs are identified in this region (fig. 2) stabilizing the distantly extensions of helices constituting the functional center (Klein et al. 2004).

Thus, our analysis demonstrates that different OaantCs located at different regions of the 70S ribosomal particle are associated with diverse biological functions. We now ask how these biophysical constraints curb the selection pressure on individual residue sites.

Both Amino Acid and Nucleotide Sites Associated with $k > K_e$ OaantCs Generally Evolve Slowly

Mapping the results of evolutionary selection analyses on the sequence alignment of ribosomal proteins revealed the greater influence of purifying selection on regions responsible for rRNA binding. A total of 6,251 amino acid sites are present in the 47 SSU and LSU r-protein genes included in the study, and out of them, 2,279 amino acids contact the rRNA. The FUBAR analysis, which detects sites evolving via pervasive diversifying and purifying selection, identifies only 17 amino acid sites constrained by pervasive positive selection pressure (posterior probability > 0.9). Similarly, the M7 versus M8 model comparison followed by BEB approach implemented in CodeML identifies only 30 positively selected sites with posterior probability > 0.9. There are three common positively selected sites in the two methods, resulting 44 total positively selected sites. We identify only 14 out of these 44 amino acid sites are located at protein–rRNA interfaces. Posterior mean d_N/d_S at the remaining 2,211 rRNA-contacting amino acid sites is very low (average $d_N/d_S < 0.25$), demonstrating that they are subjected to strong purifying selection pressure.

In the previous section, we have shown how the critical functional constraints mostly involve $k > k_e$ OaantCs and how the presence of these OaantCs is associated with specific and high-affinity contacts. In the light of these discussions, the presence of positive selection only in $k \leq k_e$ OaantCs may be a reflection of the relaxed biophysical constraints they are subjected to. To investigate this in further detail, we perform the following analysis. First, we use the posterior mean d_N/d_S estimated at each amino acid site by CodeML to enumerate the average d_N/d_S for each OaantC with a given k. We find strong negative Pearson correlations between contact number and mean d_N/d_S for both CodeML (−0.88) and FUBAR estimations (−0.85). If we take individual contact numbers and their respective d_N/d_S values, these two correlations are −0.36 and −0.31, respectively. To test whether this correlation between evolutionary rates and contact number is a random correlation, we perform a network randomization according to Wuchty et al. (2003). In this randomization approach, amino acid positions are randomly shuffled 100,000 times conserving the total number of contacts in the network and at each iteration we re-estimate the average d_N/d_S for each OaantC class with a given k. If the correlation between network topology and average d_N/d_S for individual OaantC class is random, the above-mentioned correlation should disappear in this randomization.

Performing 100,000 cycles of edge-swapping randomization (conserving the network density) we find no significant correlation between the randomized average d_N/d_S for different OaantC class and the respective contact numbers. For example, there are 35 OaantCs having $k = 7$ and the average d_N/d_S of these 35 OaantCs estimated in the original network is 0.05, which elevates to 0.21 while averaging over 100,000 randomization replicates. A measure of this elevation is presented in figure 1*A* as the ratio of original and random d_N/d_S estimates. One tailed *t*-tests are performed under a null-hypothesis of equal mean to test whether the original d_N/d_S estimates significantly differ from the random distribution. Except for OaantCs with $k \leq 2$, the null hypothesis is rejected in all other cases with at least $P < 0.001$. This result clearly indicates that participation in $k > k_e$ OaantCs causes the respective amino acids being subjected to stronger biophysical constraints which in turn drive them to evolve slowly. This result is consistent with previous observations for both PPI network and RCNs, that the rate of evolution a node (whole protein/amino acid site) negatively correlates with its number of contacts (Fraser et al. 2002; Toft and Fares 2010; Echave et al. 2016).

To test whether such relationships exist for nucleotide sites associated with the respective OaantCs as well, we estimate the sequence diversity at each nucleotide site of 16S and 23S rRNA in terms of the Shannon entropy score enumerated for our previously used large alignment of 280 species (Mallik et al. 2015). In this case, for 1,645 nucleotides (647 16S and 998 23S rRNA nucleotides), we find a negative correlation (−0.86) between contact number and sequence entropy (which means nucleotides associated with OaantCs with higher contact numbers evolve slowly) which completely disappears when we randomize the contact network as before (fig. 1*A*). One tailed *t*-tests are performed under a null-hypothesis of equal mean to test whether the original entropy scores significantly differ from the random distribution. Exactly as the previous case with amino acid sites, the null hypothesis is rejected in all the other cases except for OaantCs with $k \leq 2$, with at least $P < 0.001$. Both these results are consistent with two biological facts: 1) OaantCs constituents do not emerge within a contact network due to random interactions among residues, rather they are evolutionary conserved building blocks of the 70S particle; and 2) critical functional constraints associated with $k > k_e$ OaantCs probably constraints the rate of evolution of the constituent amino acid–nucleotide sites.

Slower Evolutionary Rates at $k > k_e$ OaantCs Emerge from Slower Substitution Rates at Nonsynonymous Sites

We ask whether the smaller d_N/d_S ratio associated with $k > k_e$ OaantCs is an outcome of strong functional constraints at the nonsynonymous sites (slow nonsynonymous substitution rates) or relaxed functional constraints at synonymous sites (accelerated synonymous substitution rates) (Barreto and Burton 2013). To investigate this, we compute the rates of synonymous substitutions (d_S) and the rates of nonsynonymous substitutions (d_N) at each amino acid site and compare the respective distributions for different OaantCs classes (fig. 3). Permutation Mann–Whitney *U* tests under the null hypothesis of equal mean are performed to test whether the d_N and d_S distributions associated with various OaantC classes differ significantly from one another. Results show that different OaantCs show no difference in d_S distribution (fig. 3*A*), but a gradual and significant fall of average d_N is observed with increasing contact number (fig. 3*B*). These results can be compared with the previous work of Barreto and Burton (2013) where it was reported that in eukaryotic mitochondrial ribosome mitochondrial gene-encoded r-proteins evolve faster than the nuclear gene-encoded r-proteins. The authors demonstrated that this difference is inherent to the slower substitution rates at nonsynonymous sites of the latter, as d_S distributions between the two groups of proteins do not differ significantly.

The inherent message of this result is that for $k > k_e$ OaantCs the evolutionary pressure to keep the respective amino acid unaltered is much higher compared with that at $k \leq k_e$ OaantCs. Based on the analysis so far and in the light of our previous discussions, one can argue that the biological origin of this evolutionary pressure is likely inherent to the association of $k > k_e$ OaantCs with biologically essential functional interactions and specific and high-affinity protein–rRNA contacts. However, just substitution rates do not provide a clear relationship between the nature of biophysical interactions and selection pressure. Hence, to understand how the evolutionary pattern at different OaantC classes is coupled with the biophysical constraints they are subjected to, we look into their amino acid substitution patterns in the course of evolution.

Evolutionary Substitutions at $k > k_e$ OaantCs Exhibit Stronger Tendencies to Maintain Respective Amino Acid Physicochemical Properties

Maximum Likelihood methods implemented in CodeML allows a probabilistic reconstruction of the ancestral codons given a codon alignment and a phylogenetic tree representing the evolutionary divergence (Yang 2007). We take advantage of this method to predict the posterior probabilities of different amino acid substitutions at various branches of the phylogenetic trees. Only the branch-specific substitutions predicted with > 0.7 posterior probability are included in further analysis (Bhattacherjee et al. 2015). Our aim here is to understand how different physicochemical properties alter upon individual evolutionary substitutions, irrespective of their temporal order. We focus on two physical properties: Flexibility (based on Bhaskaran and Ponnuswamy [1988] scale) and volume (solvent accessible volume, estimated for 1.4 Å probe) and two chemical properties: Hydrophobicity (based on Kyte and Doolittle [1982] scale) and polarity (based on Zimmerman et al. [1968] scale) for our analysis. The flexibility inherent to a protein molecule is generally correlated with an elevated propensity of molecular recognition and binding (Wright and Dyson, 1999; Smith et al. 2003). Ribosomal proteins generally include a globular domain that rests on ribosome surface, and some segments that pierce into the subunit core. These segments, densely packed with the surrounding rRNA helices, are devoid of secondary structural elements and they are rich in amino acids with smaller side chains and higher flexibility (Brodersen et al. 2002). On the other hand, the globular domains are generally rich in amino acids with larger side chains and lower flexibility (Brodersen et al. 2002). Chemical properties of amino acids are relevant to their biophysical interactions with rRNA nucleotides. For example, hydrophobic amino acids Ala and Val often pack with the hydrophobic crevices of rRNA base pairs and these interactions are stabilized by hydrophobic forces (Brodersen et al. 2002). We ask whether and how the involvement in different OaantCs is associated with evolutionary conservation of these physical and chemical properties. At a given amino

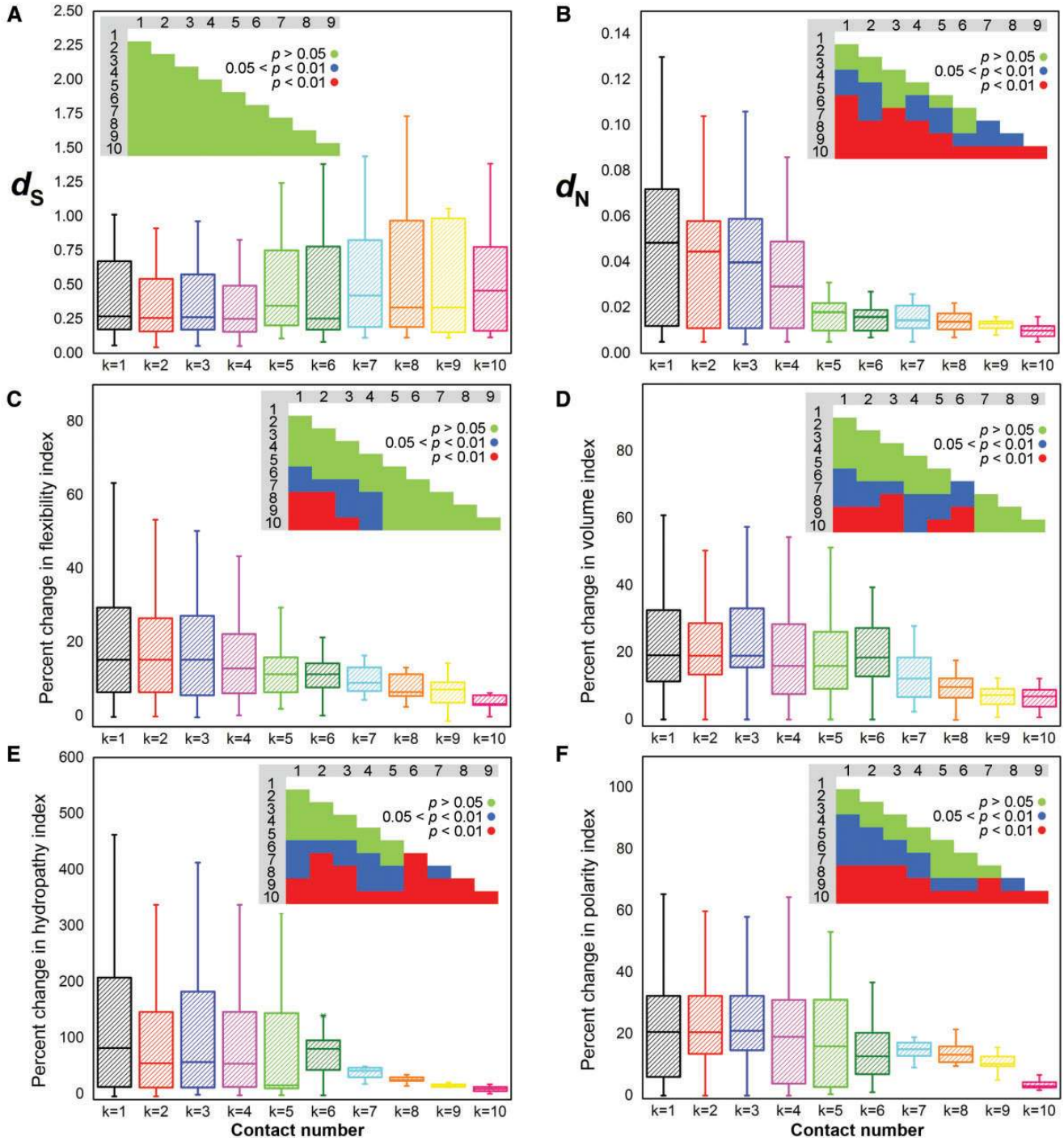

Fig. 3.—Box-plot distributions of (*A*) synonymous substitution rate (d_S) and (*B*) the nonsynonymous substitution rate (d_N) estimated at amino acid sites associated with different OaantC class. Permutation Mann–Whitney *U* tests under the null hypothesis of equal mean are performed for all possible OaantC pair to investigate whether the distributions significantly differ from one another. The respective *P* values of the *U* tests are shown in terms of a colormap matrix (significance of each color is mentioned in the figure). (*C*–*F*) The percent changes of two physical (flexibility and volume) and two chemical properties (hydropathy and polarity) upon individual point-mutations are plotted as box-plot distributions for each OaantC class. Distribution plots for the four properties are presented in panel (*C*), (*D*), (*E*), and (*F*), respectively. Permutation Mann–Whitney *U* tests under the null hypothesis of equal mean are performed to investigate whether the distributions associated with various OaantCs significantly differ from one another. *P* values are represented as colormap matrices and significance of each color is mentioned in the figure.

acid site, for an $i \rightarrow j$ substitution, if the values of the respective physical/chemical property for ancestor and daughter amino acids are p_i and p_j, respectively, we estimate the percent change of that property as $|p_i - p_j| \times 100/p_i$. The distributions of this parameter for the two physical and two chemical properties for different OaantC classes are plotted in figure 3C–F. Permutation Mann–Whitney *U* tests (null hypothesis of equal mean) are performed in each case to test whether the distributions associated with each OaantC class significantly differ from all the other classes.

Results show a general tendency that at $k > k_e$ OaantCs the observable substitutions are generally highly conservative, which means substantial fluctuations of any of the physicochemical properties are strongly disfavored. Permutation Mann–Whitney *U* tests are performed to investigate whether this trend (mutations are generally more conservative in $k > k_e$ OaantCs compared with $k \leq k_e$ OaantCs) is statistically significant. For the two chemical properties, hydropathy (fig. 3*E*) and polarity (fig. 3*F*), the above-mentioned trend is observed to be highly prominent as reflected from their respective *U* test matrices (each OaantC class significantly differs from all the other classes). However, for the two physical properties, flexibility index (fig. 3*C*) and volume (fig. 3*D*), this trend is comparatively weaker (substitutions are significantly more conservative in $k \geq 6$ OaantCs compared with the other OaantCs). Variations of amino acid physical properties such as flexibility and residue volume remain statistically indifferent for a wide range of contact number. Significant differences of amino acid physical property conservation are identified only when we compare very small (represent interactions generally made from globular regions, large variations of flexibility and volume is permitted) and very large contact number OaantC (represent interactions generally made from disordered extensions densely packed against rRNA helices, large variations of flexibility and volume is not permitted).

This analysis depicts an intimate relationship between biophysical constraints associated with OaantCs and patterns of their sequence evolution. Evolutionary rates are slower in OaantCs with higher *k* and evolutionary substitutions are also more conservative in terms of amino acid chemical properties. As these OaantCs are fundamental building blocks of protein–rRNA physical interactions, one can expect that the purifying selection pressure conserves the amino acid chemical property required for local stabilizing interaction with the OaantC nucleotides. In the next step, we investigate this in further detail.

Amino Acid Sites Associated with Higher Contact Number More Frequently Experience Purifying Selection Pressure Conserving Their Chemical Properties

From figure 3, it reflects that OaantCs with higher *k* are more likely to experience conservative substitutions compared with OaantCs with smaller *k*. But in some cases highly conservative substitutions occur in OaantCs with smaller *k* as well. To gain a clearer insight, we classify OaantC amino acids into three classes, positively charged, hydrophobic and polar. Maximum Likelihood methods implemented in PRIME package (Pond et al. 2005) are employed to identify residue sites associated with a statistically significant signature that purifying selection is conserving one of the three chemical properties. Likelihood ratio tests followed by significance test of individual predictions under χ^2_k distribution (95% confidence interval) identify 484 sites among the 47 proteins where purifying selection conserves one chemical property at the given amino acid site. We have enlisted the frequency of occurrence of these "selected" amino acid sites for each OaantC class in figure 1*A*. Positively charged, hydrophobic and polar amino acids contribute 49%, 24% and 27% of these "selected" sites.

First we ask whether the biophysical origin of this purifying selection pressure is inherent to amino acid–nucleotide contacts only or whether amino acid–amino acid contacts also play significant roles. Two different statistical tests are performed to address this issue. We compare d_N/d_S distributions between sites with the same number of contacts, but 1) only with the RNA (sample A) and 2) only with other amino acids (sample B). The differences are tested using two-sample Mann–Whitney *U* tests with the null hypothesis of equal median (supplementary table 3, Supplementary Material online). Results show that only for $k > k_e$, the d_N/d_S distributions between sample A and B differ significantly. This result depicts that nucleotide contacts have comparatively stronger impact in shaping the selection pressure on OaantC amino acid sites than amino acid contacts. In the second test, for each OaantC class, we classify OaantC amino acid sites into two classes: 1) Those that contact nucleotides only and 2) those that contact other amino acids as well. Permutation Mann–Whitney *U* tests confirm that for any OaantC class, d_N/d_S distributions of these two sets are not significantly different (supplementary table 4, Supplementary Material online). These two tests confirm that amino acid contacts do not have significant effect in regulating the selection pressure on $k > k_e$ OaantC amino acid sites. Hence, there is a clear statistical tendency that nucleotide contacts make the major contribution to shape the selection pressure on all $k \geq 3$ OaantCs (there may be a few special cases contradicting this tendency). We have further explored this by estimating the statistical bias of individual amino acids (associated with different OaantC classes) being "selected" (fig. 1*D*). The results show that for $k \leq k_e$ OaantCs (i.e., the number of nucleotides contacted is less than or equal to that expected from random distribution) the respective amino acid sites exhibit negative bias of being "selected" (i.e., a "selection" event is rare compared with a null distribution), whereas amino acid sites associated with $k > k_e$ OaantCs generally exhibit positive bias of being "selected" (i.e., a "selection" event is more frequent compared with a null distribution).

Though we find a clear statistical tendency that amino acid–nucleotide contacts play critical roles in shaping the selection pressure, participation in OaantC with higher contact number does not necessary mean the respective amino acid physicochemical property would be conserved by purifying selection pressure. Hence, network representation of molecular contacts is a powerful as yet incomplete representation of the actual biophysical interactions the residue sites are subjected to. To investigate what actually determines the nature of selection pressure at individual amino acid sites, for each OaantC we investigate the 3D structural orientations of its constituent residues. These orientations further determine the stabilizing/destabilizing nature of the interactions among OaantC constituents. Our hypothesis is, more stabilizing interactions may have stronger impacts in driving the selection pressure for the respective chemical property.

Amino Acid–Nucleotide Interaction Energetics Provide the Molecular Basis of Purifying Selection for Conserved Chemical Properties

Relative 3D Orientations of Amino Acid–Nucleotide Constituents Differ between "Selected" and "Nonselected" OaantCs

For individual amino acids, all the respective "selected" (statistically significant signature of purifying selection conserving a chemical property) and "nonselected" OaantCs (there is no statistically significant signature of purifying selection conserving a chemical property) are separately superimposed on one of their representative structures. Only the amino acid residues are aligned in this process, using the all-atom cealign structural alignment method implemented in PyMOL (Version 1.8 Schrödinger, LLC). So, for a given amino acid (say Arginine) and for a given OaantC class (say all "selected" $k = 5$ OaantCs) this superposition approach allows us to visualize the all-possible orientations of OaantC nucleotides around the central Arginine. Once we understand these orientation patterns, we can compare the differences between "selected" or "nonselected" OaantCs. Surface representation of these superposed OaantCs for positively charged, polar, and hydrophobic amino acids is shown in figure 4*A*. For "selected" positively charged amino acids, all possible combination of surrounding nucleotides creates a negatively charged pocket where backbone phosphates face the Arg/Lys residue whereas nucleotide bases are distantly placed (fig. 4*A*). This indicates that the major contribution of stabilizing interaction between negatively charged phosphates and positively charged amino groups comes from the Coulomb electrostatic interaction energy. This charged pocket surface is enriched in hydrogen bond acceptor groups (fig. 4*A*) and for ~90% of the individual OaantCs, we find 1–2 H-bonds between the respective amino acid–nucleotide components. One example of a "selected" OaantC with $k = 5$ including Arg12 of uL20 protein and five 23S rRNA nucleotides is shown in figure 4*B*. This specific orientation of backbone phosphates facing the amino acid is absent in "nonselected" OaantCs and the negative charge-rich pocket is absent as well (fig. 4*A*).

"Selected" hydrophobic OaantCs are characterized by the amino acids facing the hydrophobic bases, whereas charged phosphate backbones are distantly placed (fig. 4*A*). This suggests that the major contribution of stabilizing interaction between hydrophobic nucleotide bases and hydrophobic amino acids comes from their hydrophobic interactions. These specific interactions can occur in at least four ways: 1) Interactions between the amino acid and minor groove edges of base pairs, 2) interactions of amino acids with widened major grooves of RNA helices, 3) amino acid packing with the flipped out bases of bulged nucleotides, and 4) insertion of amino acids into hydrophobic crevices between exposed nucleotide bases. Superposition of such OaantCs creates a hydrophobic pocket enclosing the amino acid (fig. 4*A*), and hydrophobic interactions likely play the stabilizing role for these OaantCs. In case of "nonselected" OaantCs, nucleotide orientations are more random such that specific hydrophobic pockets are not observed in the superposed ensemble. The accepted value for the strength of the hydrophobic effect for protein folding and oligomerization is -25 cal mol^{-1} per Å^2 of buried hydrophobic surface area (Chothia and Janin 1975), though for nonselected OaantCs the electrostatic and polar contributions will affect this value. Energetic estimations have been performed in the next section to investigate whether interaction strength of OaantC constituents has any role in determining whether selection pressure will conserve the respective amino acid chemical property.

For "selected" polar amino acids, the preference of amino acid–nucleotide relative orientation is not very distinctive. Hydrophobic bases and ribose sugars appear to be equally preferred close to the amino acids, whereas backbone phosphates are generally distantly placed. An example of $k = 4$ OaantC involving Asn23 of bL17 protein is shown in figure 3*B*.

"Selected" OaantCs Are Generally Associated with Higher Stabilizing Interaction Energy Compared with the "Nonselected" OaantCs

Relative orientations of protein–rRNA constituents of "selected" OaantCs suggests specific stabilizing interactions may be the molecular basis of observed selection phenomena that conserves the respective chemical properties of the amino acids. We compute the native interaction energetics of all individual OaantCs with $k \geq 3$ assuming a minimally frustrated model of molecular recognition (Sheinerman and Honig 2002) that ignores the structural reconstitution processes through which the final structure is generated. Each OaantCs is considered individually, apart from the rest of the structure, and electrostatic interaction energy (ΔG^0_{elec}) is predicted assuming a rigid binding between the amino acid and the nucleotide components. In these calculations, the entire nucleotide

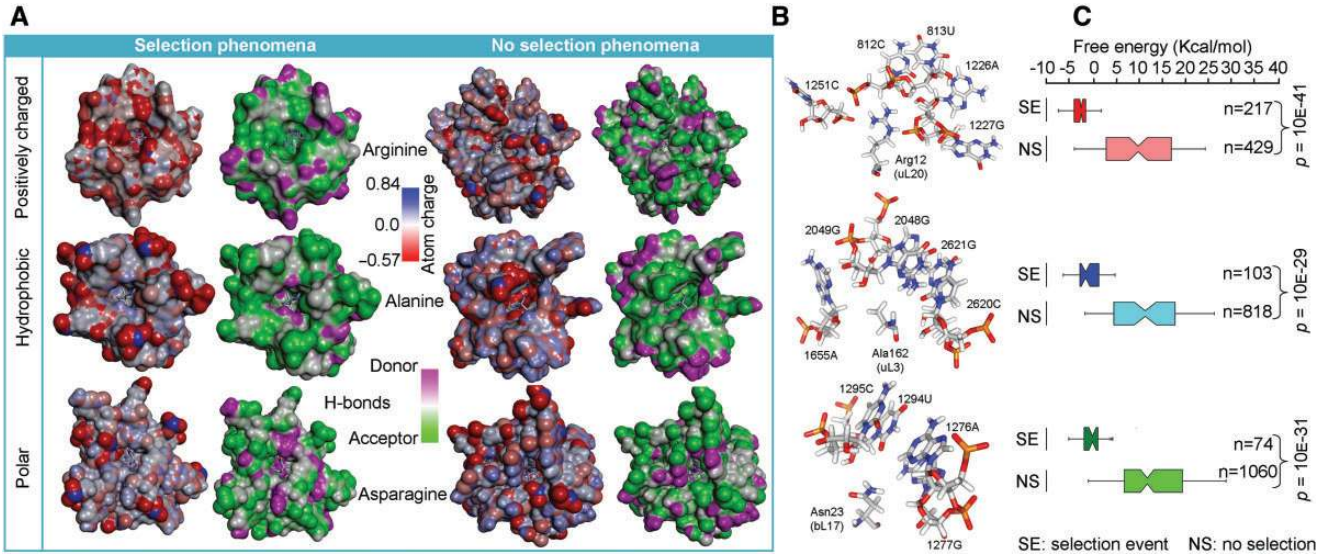

Fig. 4.—(*A*) Superimposed OaantCs with $k = 5$ are shown here as surface representation for positively charged, hydrophobic and polar amino acids. Surfaces are colored according to atom partial charges. (*B*) Molecular diagrams (stick representation) of "selected" OaantCs are shown for the three chemical categories. Stick colors are according to the amino acid partial charge scale provided in panel (*A*). At the top, the Arg12 of uL20 protein represents a positively charged "selected" OaantC that contacts five nucleotides. In the middle, Ala162 of uL3 represents a hydrophobic "selected" OaantC contacting five nucleotides. At the bottom, Asn23 of bL17 represents a polar OaantC. (*C*) Comparison of computed amino acid–nucleotide interaction free energies for selected and nonselected OaantCs (top: positively charged, middle: hydrophobic, bottom: polar). SE, selected; NS, not selected.

(base + sugar + phosphate) is subjected to energetic analysis. The predicted energy estimates vary in the range between −17.3 and 79.8 kcal/mol. Such overestimation of numerical energy values is quite common for computational analysis (with additional limitations of different assumptions), for which Zhang et al. (2013) mentioned that these theoretical estimations should not be assumed to be confirmed experimentally, but rather should be considered as a tendency.

Box-plot distributions of all the energy estimates are shown in figure 4C for "selected" and "nonselected" OaantCs of all three chemical categories. These plots predict a clear tendency of "selected" OaantCs being associated with more negative interaction free energies compared with the "nonselected" ones. For example, for "selected" positively charged amino acids, the average interaction energy is −3.3(±3.6) kcal/mol, whereas that for "nonselected" OaantCs is 8.5(±8.2) kcal/mol. Again, for "selected" hydrophobic amino acids, the average interaction energy is −1.9(±2.9) kcal/mol, whereas that for "nonselected" OaantCs is 11.2(±9.1) kcal/mol. In majority of the "selected" OaantCs the estimated ΔG^0_{bind} is close to or less than zero. Hence, computationally predicted $\Delta G^0_{bind} < 0$ values strongly associate with the occurrence of purifying selection conserving the amino acid chemical property. For 2,276 OaantCs with $k \geq 3$, energetic calculations statistically associate with such purifying selection phenomena with only 19% false negative and 4% false positive outcomes.

One interesting fact is, stabilizing interactions in some small OaantCs is also associated with purifying selection pressure conserving amino acid chemical property. Ser70 of uL4 ($\Delta G^0_{bind} = -1.57$ kcal/mol) that neutralizes the backbone phosphates of 674G and 675A located at peptide-exit tunnel wall is an interesting example of this trend. Another example is Lys54 of uL14, located at the B5 intersubunit bridge of bacterial ribosome that maintains an important contact with three nucleotides (1420U, 1421G, and 1422G) located at helix h44 of 16S rRNA. These examples suggest that though there is a clear statistical tendency that smaller contact OaantCs ($k < k_e$) are topologically quite similar to random contacts, depending on their orientations in 3D space they may have important functional roles including stabilizing the protein–rRNA interfaces.

Ribosome assembly is cooperative in nature (Shajani et al. 2011), which means a major fraction of protein binding energy is included as cooperative coupling energy that cannot be estimated from the native structure. Hence, it is difficult to relate our energy estimates with experimental estimations or with the exact binding affinity of the protein. Regardless of these limitations, our analyses indicate that participation in OaantCs associated with stabilizing interactions has a higher tendency to be subjected to strong purifying selection conserving the specific chemical property responsible for the interaction.

Summary

Thus, using a proximity-based network of amino acid–nucleotide residue pairs having van der Waals interactions with

each other, we perform a series of analysis to demonstrate that modular organization of protein–rRNA constituents of ribosomal particles has a strong impact in regulating the evolutionary rate heterogeneity of ribosomal proteins. The residue contact network of ribosome includes different overrepresented OaantCs (one amino acid contacts ≥ 1 nucleotides) distributed throughout the structure. Amino acids are more randomly distributed in OaantCs that comprise smaller number of nucleotides, whereas only a few specific amino acids are preferred in OaantCs that maintain contacts with multiple nucleotides. These OaantCs are associated with diverse functional aspects of ribosomal particles, ranging from biogenesis and structural stability to translation process. The biophysical constraints to maintain contacts among multiple residue sites likely impose stronger purifying selection pressure on larger OaantCs, resulting both the amino acid and nucleotide constituents to evolve slowly. The slower evolutionary rates of larger OaantCs amino acid sites refer to slower substitution rates at nonsynonymous sites that tend to preserve the respective amino acid physicochemical properties responsible for local molecular interaction. Cases in which amino acid–nucleotide constituents of OaantCs are associated with stabilizing interactions, we find statistically significant signatures of purifying selection preserving the respective amino acid physicochemical property.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Author Contributions

S.M. and S.K. designed research; S.M. implemented computational methodologies, performed research and analyzed data; S.M. and S.K. discussed and commented on the results and wrote the paper.

Acknowledgments

The authors sincerely acknowledge the anonymous reviewers for their many constructive suggestions. This work is supported by the Center of Excellence in Systems Biology and Biomedical Engineering (TEQIP Phase II), University of Calcutta, India. The authors declare that no conflicting interests exist.

Literature Cited

Aftabuddin M, Kundu S. 2007. Hydrophobic, hydrophilic, and charged amino acid networks within protein. Biophys J. 93:225–231.

Allmang C, Mougel M, Westhof E, Ehresmann B, Ehresmann C. 1994. Role of conserved nucleotides in building the 16S rRNA binding site of *E. coli* ribosomal protein S8. Nucleic Acids Res. 22:3708–3714.

Barabasi AL, Albert R. 1999. Emergence of scaling in random networks. Science 286:509–512.

Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. Mol Biol Evol. 30:310–314.

Bellur DL, Woodson SA. 2009. A minimized rRNA-binding site for ribosomal protein S4 & its implications for 30S assembly. Nucleic Acids Res. 37:1886–1896.

Benson DA, et al. 2015. GenBank. Nucleic Acids Res. 43(Database issue):D30.

Bhaskaran R, Ponnuswamy PK. 1988. Positional flexibilities of amino acid residues in globular proteins. Int J Pep Prot Res. 32:241–255.

Bhattacherjee A, Mallik S, Kundu S. 2015. Compensatory mutations occur within the electrostatic interaction range of deleterious mutations in protein structure. J Mol Evol. 80:10–12.

Brodersen DE, Clemons WM Jr, Carter AP, Wimberly BT, Ramakrishnan V. 2002. Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. J Mol Biol. 316:725–768.

Chakraborty S, Ghosh TC. 2013. Evolutionary rate heterogeneity of core and attachment proteins in yeast protein complexes. Genome Biol Evol. 5:1366–1375.

Chothia C, Janin J. 1975. Principles of protein-protein recognition. Nature 256:705–708.

David-Eden H, Mandel-Gutfreund Y. 2008. Revealing unique properties of the ribosome using a network based analysis. Nucleic Acids Res. 36:4641–4652.

Donald JE, Chen WW, Shakhnovich EI. 2007. Energetics of protein-DNA interactions. Nucleic Acids Res. 35:1039–1047.

Dragon F, Spickler C, Pinard R, Carrière J, Brakier-Gringas L. 1996. Mutations of non-canonical base-pairs in the 3′ major domain of Escherichia coli 16 S ribosomal RNA affect the initiation & elongation of protein synthesis. J Mol Biol. 259:207–215.

Dubin P, Bock J, Davis R, Schulz DN, Thies C, editors. 2012. Macromolecular complexes in chemistry and biology. Springer Science & Business Media. ISBN 9783642784712.

Echave J, Spielman SJ, Wilke CO. 2016. Causes of evolutionary rate variation among protein sites. Nat Rev Genet. 17:109–121.

Erdös P, Rényi A. 1959. On random graphs. Publ Math. 6:290–297.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science 296:750–752.

Göringer HU, Hijazi KA, Murgola EJ, Dahlberg AE. 1991. Mutations in 16S rRNA that affect UGA (stop codon)-directed translation termination. Proc Natl Acad Sci U S A. 88:6603–6607.

Jovelin R, Phillips PC. 2009. Evolutionary rates and centrality in the yeast gene regulatory network. Genome Biol. 10:R35.

Klein DJ, Moore PB, Steitz TA. 2004. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. J Mol Biol. 340:141–177.

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

Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 13:2229–2235.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 157, 105–132.

Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Liiv A, O'Connor M. 2006. Mutations in the intersubunit bridge regions of 23 S rRNA. J Biol Chem. 281:29850–29862.

Mallik S, Akashi H, Kundu S. 2015. Assembly constraints drive co-evolution among ribosomal constituents. Nucleic Acids Res. 43:5352–5363.

Mallik S, Kundu S. 2015. Molecular interactions within the halophilic, thermophilic, and mesophilic prokaryotic ribosomal complexes: clues to environmental adaptation. J Biomol Struct Dyn. 33:639–656.
Marsh JA, Teichmann SA. 2015. Structure, dynamics, assembly, and evolution of protein complexes. Annu Rev Biochem. 84:551–575.
MacKerel AD Jr, Brooks III CL, Nilsson L, Roux B, Won Y, Karplus M. 1998. Charmm: The energy function and its parameterization with an overview of the program. In Schleyer v. R. et al, P. editor, The Encyclopedia of Computational Chemistry.
McClory SP, Leisring JM, Qin D, Fredrick K. 2010. Missense suppressor mutations in 16S rRNA reveal the importance of helices h8 and h14 in aminoacyl-tRNA selection. RNA 16:1925–1934.
McInerney JO. 2006. The causes of protein evolutionary rate variation. Trends Ecol Evol. 21:230–232.
Milenković T, Filippis I, Lappe M, Przulj N. 2009. Optimized null model for protein structure networks. PLoS One 4:e5967.
Milenkovic T, Lai J, Przulj N. 2008. Graphcrunch: a tool for large network analyses. BMC Bioinformatics 9:70.
Milo R, et al. 2002. Network motifs: simple building blocks of complex networks. Science 298:824–827.
Moine H, Cachia C, Westhof E, Ehresmann B, Ehresmann C. 1997. The RNA binding site of S8 ribosomal protein of *Escherichia coli*: selex and hydroxyl radical probing studies. RNA 3:255–268.
Murrell B, et al. 2013. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. Mol Biol Evol. 30:1196–1205.
Neduva V, et al. 2005. Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol. 3:e405.
Nick Pace C, Scholtz JM, Grimsley GR. 2014. Forces stabilizing proteins. FEBS Lett. 588:2177–2184.
Noel JK, Morcos F, Onuchic JN. 2016. Sequence co-evolutionary information is a natural partner to minimally-frustrated models of biomolecular dynamics. F1000Res. 5:Rev-106.
Noeske J, et al. 2015. High-resolution structure of the *Escherichia coli* ribosome. Nat Struct Mol Biol. 22:336–341.
Osada N, Akashi H. 2012. Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. Mol Biol Evol. 29:337–346.
Penrose M. 2003. Random geometric graphs. Oxford University Press. ISBN 0198506260.
Pettersen EF, et al. 2004. UCSF Chimera—a visualization system for exploratory research & analysis. J Comp Chem. 25:1605–1612.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.
Ponder JW, Case DA. 2003. Force fields for protein simulations. Adv Protein Chem. 66:27–85.
Pržulj N, Corneil DG, Jurisica I. 2004. Modeling interactome: scale-free or geometric? Bioinformatics 20:3508–3515.
Pržulj N, Higham D. 2006. Modelling protein-protein interaction networks via a stickiness index. J R Soc Interface. 3:711–716.
Roy SW, Gilbert W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet. 7:211–221.
Saiz L, Vilar JM. 2006. Stochastic dynamics of macromolecular-assembly networks. Mol Syst Biol. 2:0024.
Shah P, McCandlish DM, Plotkin JB. 2015. Contingency and entrenchment in protein evolution under purifying selection. Proc Natl Acad Sci U S A. 112:E3226–E3235.
Shajani Z, Sykes MT, Williamson JR. 2011. Assembly of bacterial ribosomes. Annu Rev Biochem. 80:501–526.
Sheinerman FB, Honig B. 2002. On the role of electrostatic interactions in the design of protein-protein interfaces. J Mol Biol. 318:161–177.
Shen-Orr SS, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet. 31:64–68.
Shpanchenko OV, Zvereva MI, Dontsova OA, Nierhaus KH, Bogdanov AA. 1996. 5S rRNA sugar-phosphate backbone protection in complexes with specific ribosomal proteins. FEBS Lett. 394:71–75.
Smith DK, Radivojac P, Obradovic Z, Dunker AK, Zhu G. 2003. Improved amino acid flexibility parameters. Protein Sci. 12:1060–1072.
Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168:1457–1465.
Takada K, et al. 2015. TCR affinity for thymoproteasome-dependent positively selecting peptides conditions antigen responsiveness in CD8(+) T cells. Nat Immunol. 16:1069–1076.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.
Toft C, Fares MA. 2010. Structural calibration of the rates of amino acid evolution in a search for Darwin in drifting biological systems. Mol Biol Evol. 27:2375–2385.
Toivonen JM, Boocock MR, Jacobs HT. 1999. Modelling in *Escherichia coli* of mutations in mitoribosomal protein S12: novel mutant phenotypes of rpsL. Mol Microbiol. 31:1735–1746.
van Dam TJ, Snel B. 2008. Protein complex evolution does not involve extensive network rewiring. PLoS Comput Biol. 4:e1000132.
Voss NR, Gerstein M, Steitz TA, Moore PB. 2006. The geometry of the ribosomal polypeptide exit tunnel. J Mol Biol. 360:893–906.
Wang W, Kollman PA. 2000. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. J Mol Biol. 303:567–582.
Weng ML, Ruhlman TA, Jansen RK. 2016. Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol Evol. 8:1824–1838.
Wernicke S, Rasche F. 2006. FANMOD: a tool for fast network motif detection. Bioinformatics 22:1152–1153.
Williamson JR. 2008. Cooperativity in macromolecular assembly. Nat Chem Biol. 4:458–465.
Wright PE, Dyson JH. 1999. Intrinsically unstructured proteins: re-assessing the protein-structure paradigm. J Mol Biol. 293:321–331.
Wuchty S, Oltvai ZN, Barabási AL. 2003. Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat Genet. 35:176–179.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.
Yang Z. 2006. Computational molecular evolution. Oxford: Oxford University Press.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yeger-Lotem E, et al. 2004. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc Natl Acad Sci U S A. 101:5934–5939.
Zengel JM, Jerauld A, Walker A, Wahl MC, Lindahl L. 2003. The extended loops of ribosomal proteins L4 and L22 are not required for ribosome assembly or L4-mediated autogenous control. RNA 9:1188–1197.
Zhang X, Perica T, Teichmann SA. 2013. Evolution of protein structures and interactions from the perspective of residue contact networks. Curr Opin Struct Biol. 23:954–963.
Zhang Z, et al. 2013. Enhancing human spermine synthase activity by engineered mutations. PLoS Comput Biol. 9:e1002924.
Zimmerman JM, Eliezer N, Simha R. 1968. The characterization of amino acid sequences in proteins by statistical methods. J Theor Biol. 21:170–201.

Associate editor: Eric Bapteste