

An evolutionary conserved pattern of 18S rRNA sequence complementarity to mRNA 5' UTRs and its implications for eukaryotic gene translation regulation

Josef Pánek^{1,*}, Michal Kolář², Jiří Vohradský¹ and Leoš Shivaya Valášek³

¹Laboratory of Bioinformatics, Institute of Microbiology, Academy of Sciences of Czech Republic, v.v.i., 14220 Prague, Videnska 1083, Czech Republic, ²Laboratory of Genomics and Bioinformatics, Institute of Molecular Genetics, Academy of Sciences of Czech Republic, v.v.i., 14220 Prague, Videnska 1083, Czech Republic and ³Laboratory of Regulation of Gene Expression, Institute of Microbiology, Academy of Sciences of Czech Republic, v.v.i., 14220 Prague, Videnska 1083, Czech Republic

Received November 12, 2012; Revised May 23, 2013; Accepted May 24, 2013

ABSTRACT

There are several key mechanisms regulating eukaryotic gene expression at the level of protein synthesis. Interestingly, the least explored mechanisms of translational control are those that involve the translating ribosome *per se*, mediated for example via predicted interactions between the ribosomal RNAs (rRNAs) and mRNAs. Here, we took advantage of robustly growing large-scale data sets of mRNA sequences for numerous organisms, solved ribosomal structures and computational power to computationally explore the mRNA–rRNA complementarity that is statistically significant across the species. Our predictions reveal highly specific sequence complementarity of 18S rRNA sequences with mRNA 5' untranslated regions (UTRs) forming a well-defined 3D pattern on the rRNA sequence of the 40S subunit. Broader evolutionary conservation of this pattern may imply that 5' UTRs of eukaryotic mRNAs, which have already emerged from the mRNA-binding channel, may contact several complementary spots on 18S rRNA situated near the exit of the mRNA binding channel and on the middle-to-lower body of the solvent-exposed 40S ribosome including its left foot. We discuss physiological significance of this structurally conserved pattern and, in the context of previously published experimental results, propose that it modulates scanning of the 40S subunit through 5' UTRs of mRNAs.

INTRODUCTION

There are two general mechanisms regulating gene expression at the translational level in response to ever-changing environmental conditions, in addition to many others operating mainly in a gene-specific manner. Environmentally controlled cell signaling can regulate (i) the levels of the so-called ternary complexes comprising initiating Met-tRNA_i^{Met} bound to eukaryotic initiation factor (eIF) 2 in its GTP form and (ii) the phosphorylation status of several eIFs and their binding partners that are required to activate mRNAs to bind to pre-initiation complexes (PICs) [reviewed in (1)]. Significant reductions in ternary complexes levels or hyper-phosphorylation of the mRNA cap-binding protein eIF4E and its 'inhibitor', eIF4E-binding protein (eIF4E-BP), evoked by phenomena such as changes in the availability of nutrients, stress conditions or an absence of growth factors, rapidly shut down the translation of most mRNAs. The fact that, in contrast to what occurs in bacteria, the eukaryotic ribosome has to scan the mRNA 5' untranslated region (UTR), usually for the first AUG, to start producing the encoded protein has led to the development of many gene/mRNA-specific controls using specific RNA sequences, structures and features such as short open reading frames that either impede or enhance the ability of ribosomes to interact with the 5' UTR in a single-stranded form or prevent scanning PICs from reaching the authentic start site [reviewed in (2)]. Some of these sequences recruit sequence-specific RNA-binding proteins or microRNAs [reviewed in (3)], with the regulatory strategy of interfering with PIC assembly in a similar fashion to both general control mechanisms described earlier in the text. On the

*To whom correspondence should be addressed. Tel: +296442512; Fax: +296442544; Email: panek@biomed.cas.cz

other hand, the expression of many viral and a small subset of eukaryotic mRNAs is regulated by specialized sequences referred to as internal ribosome entry sites that recruit the PIC directly to the start codon in a manner that is relatively analogous to the Shine–Dalgarno mode of initiation in bacteria [reviewed in (4)]. There is clearly a colorful palette of mechanisms that manipulate the ribosome in specific ways to (i) influence the efficiency of the PIC assembly, (ii) prevent the PIC from moving along the 5' UTR, (iii) cause the PIC to avoid the AUG start site or (iv) block the formation of the functional 80S elongation-ready initiation complex (IC). What is missing in this overview is the contribution of the central hub of this process, the regulatory input of the translating ribosome *per se*, which is vastly underexplored.

Intuitively, translational regulation at the ribosomal level must be mediated by (i) ribosomal proteins and their prospective interactions with mRNA and various eIFs and (ii) those segments of ribosomal RNA (rRNA) that are exposed to the solvent and might therefore contact incoming mRNAs. Experimental evidence of interactions between mRNAs and rRNA segments that might be involved in the regulatory process started to emerge when the eukaryotic 18S and 28S RNAs were shown to be capable of forming stable hybrid structures with mRNAs in mice (5). The authors performed a computational search and identified numerous 18S RNA sequences that were complementary to oligonucleotide sequences that are frequently observed in mRNAs, which were designated potential hybridization regions (clinger fragments). They managed to experimentally demonstrate the hybridization properties of these RNA species and thus proposed that the regions showing mRNA–rRNA sequence complementarity might serve in translation as universal regions for mRNA binding.

The computational search for complementary sequences in mouse mRNAs and rRNAs performed at a larger scale showed a massive occurrence of rRNA-like sequences in either a sense or antisense orientation in mRNAs (6). Northern blot analysis of poly(A) RNA from different vertebrates revealed that a large number of discrete RNA molecules hybridize at high stringency to cloned probes prepared from 28S or 18S rRNA sequences that match those in mRNAs. The authors therefore hypothesized that functional rRNA-like sequences might occur in primary transcripts and differentially affect gene expression, possibly by helping to recruit mRNAs to the PIC in a selective manner. In a follow-up study (7), the same authors proposed that direct base pairing of particular mRNAs to rRNAs within ribosomes may function as a mechanism of differential translational control by (i) generally slowing or stalling the ribosome's run during elongation, depending on the strength of a particular base-pairing interaction and (ii) owing to binding of specific mRNAs to ribosomes to either block interactions with other molecules (e.g. with canonical translation factors) or to evoke conformational changes in the ribosome. Both of these processes would affect the ultimate translational output in a gene-specific manner. The existence of the former general effect was

supported experimentally by the observation that decreased mRNA–rRNA complementarity increases translational yields (7) and also by an independent study using an *in vivo* antisense RNA approach that identified a fragment complementary to the 18S rRNA 3' domain showing an inhibitory effect on the efficiency of translation of a reporter gene (8). Finally, Schneider *et al.* (9) suggested that the rRNA–mRNA hybrid interactions might also be important for alternative initiation mechanisms, including that using the internal ribosome entry sites.

Setting aside the number of proposed putative functions for the mRNA–rRNA duplexes in translational control and the debatable significance of the *in vitro* hybridization experiments supporting them, true demonstration of the existence and the physiological functionality of these interactions has been limited to a 9-nt element found in mouse Gtx homeodomain mRNA, which is the only clinger fragment that has been experimentally shown to promote translation initiation via its hybrid interaction with 18S rRNA (10,11). Considering these findings together, it is evident that there is a lack of both (i) large-scale experimental evidence for the universality and physiological significance of the role of mRNA–rRNA interactions in the translational control of native mRNAs and (ii) a robust, systematic and statistically valid computational search for mRNA–rRNA complementarity across the species.

In effort to address these concerns, we set out to conduct a systematic computational search for all segments showing statistically valid complementarity between rRNAs and mRNAs. We took advantage of availability of the large-scale data sets of sequences of mRNA 5' UTRs from multiple eukaryotic organisms, the solved high-resolution structure of the yeast ribosome and the most up-to-date computational tools. In our search, we used a large number of 5' UTR and coding region sequences from 14 eukaryotic species, an efficient computer model of sequence complementarity, and, most importantly, a statistical evaluation of the significance of the occurrence of mRNA–rRNA complementary regions in both mRNA and rRNA sequences separately. We show that statistically significant regions of rRNA–mRNA complementarity occur only in rRNA sequences toward specifically 5' UTRs of mRNAs but not in mRNA 5' UTRs or coding regions toward rRNAs. Moreover, the rRNA regions showing significant complementarity to 5' UTRs exist only in 18S RNAs, but not in 28S RNAs, and we found this phenomenon to be evolutionarily conserved. In addition, these 18S rRNA regions specifically cluster on the ribosomal surface. An extensive literature search revealed several lines of independent experimental evidence that not only validate our results but also extend them into a unified, biologically meaningful theory for function of the rRNA complementarity to mRNAs. Based on our results, we discuss how this 'complementarity mechanism' controls the efficiency of translation via putative interactions between mRNA 5' UTRs and defined segments of 18S rRNA.

MATERIALS AND METHODS

Data sets

The rRNAs and mRNA 5' UTRs of 14 species were used. In detail, the species were *Saccharomyces cerevisiae*, *Homo sapiens*, *Caenorhabditis elegans*, *Monosiga brevicollis*, *Gallus gallus*, *Danio rerio*, *Drosophila melanogaster*, *Bos taurus*, *Rattus norvegicus*, *Tetrahymena thermophila*, *Plasmodium vivax*, *Xenopus laevis*, *Trichomonas vaginalis* and *Theileria parva*. For each species, 750 unique 5' UTR sequences longer than 50 nt were used. The number of UTR sequences was determined according to the yeast data set, which represents the most comprehensive set of all organisms for which the 5' UTR sequences are available. The yeast 5' UTR sequences were manually collected from the literature (12). The 5' UTR sequences of the other 13 species were collected from UTRdb database (13). The number of sequences analyzed here ensures sufficient statistical robustness of the analysis. The 13 species used here represent all the species currently deposited in the UTRdb database that have >750 5' UTR sequences.

The rRNA sequences were collected from Silva (14) and GenBank databases. The 28S RNA for *G.gallus* was not available at the time of writing the manuscript; therefore, the analyses for 28S RNAs were restricted to 13 species.

Sequence complementarity

The complementarity of mRNAs to rRNAs and vice versa was defined by segments of mRNA and rRNA sequences that were reverse complementary to each other without gaps and were at least 5 nt long. The limit was set up to the shortest length of the complementary segments with respect to the computational capabilities (under 5 the sequences were too many to be computed in an acceptable time). A test for robustness of the method with respect to the minimal length of the complementary segments was carried out using the minimal length of 7 and 9 nt (see the 'Results' section section). An upper length limit was not set (the longest 18S RNA sequence complementary to a 5' UTR without a gap that we found in yeast was 13 nt long). The complementary segments must have opposite orientations in mRNAs and rRNAs, i.e. 5'-3' to 3'-5' and vice versa. If complementary segments with various lengths were found to overlap over their entire length, i.e. with a longer sequence encompassing a shorter one(s), only the longer sequence was used.

The segments were tested for significant enrichment of mRNA-rRNA sequence complementarity in pre-defined regions of the mRNA and rRNA sequences (see the next paragraph). If corresponding regions on mRNA and rRNA sequences were found to be statistically significantly enriched for complementary segments, they were likely to establish an intermolecular interaction. This model of complementarity was derived based on sequences mediating interactions of biological molecules that, while forming long interacting sequence segments (tens or even hundreds of nucleotides), interact through base pairing of relatively short (typically up to 10 nt) reverse complementary segments separated by gaps of a few nucleotides. These sequence segments were found to

be statistically significantly enriched in locations of intermolecular interactions (unpublished observation).

Statistical evaluation of the significance of the enrichment of complementary sequences

Sequences showing complementarity to the 5' UTRs of mRNAs were detected both within the expansion segments (ESs) of the rRNAs and outside of the ESs. When sampling rRNA sequences for sticky regions (i.e. the regions with high density of complementary sequences, for detailed definition see the 'Results' section), the ES_{*i*} was replaced with the 20 nt window in the following equations. The density of the complementary sequences in ES_{*i*} for 5' UTRs was calculated as follows:

$$d_j^{\text{within}(i)} = \frac{\# \text{ of sequences within ES}_i}{(\# \text{ of nts in ES}_i)(\# \text{ of nts in 5' UTRs of } j)},$$

and their density outside of ESs was estimated using the formula below:

$$d_j^{\text{off}} = \frac{\# \text{ of sequences out of the ESs}}{(\# \text{ of nts out of the ESs})(\# \text{ of nts in 5' UTRs of } j)},$$

for each tested mRNA, *j*, by normalizing their count based on the number of nucleotides in the ES and the number of nucleotides in the 5' UTRs to obtain intensive quantities.

To select the ESs showing a significantly enriched density of complementary sequences, we compared the densities within ES_{*i*} and outside of the ESs across all mRNAs using a one-sided Student's *t*-test (assuming equal variances) and corrected through multiple testing by calculating Storey's *q*-values (15) *q*(*i*) from the resulting *p*-values *p*(*i*). Statistical significance was estimated at a significance level 0.05.

The analysis was performed independently for the rRNAs of the small and large ribosomal subunits, with *q*-values evaluated based on the union of the two sets of *p*-values.

RESULTS

Sites of enriched complementarity to rRNAs were spread randomly within mRNA sequences and all mapped onto a few clearly discernible locations in rRNAs

We first computationally identified and stored sequences in both human and yeast 18S and 28S rRNAs complementary to the 5' UTRs and coding sequences of the collected mRNAs (see 'Materials and Methods' section). Next, we identified regions of rRNAs and mRNAs showing an increased density of the complementary sequences between them. The local density of complementary sequences was evaluated statistically, and the regions with significantly higher density were determined (see 'Materials and Methods' section).

Statistically significant well-defined regions within rRNAs that were complementary to the corresponding, not uniformly distributed segments within the 5' UTRs or coding sequences of mRNAs were observed in yeast and human rRNAs (Figure 1). These results were robust with respect to changes of the minimal length of

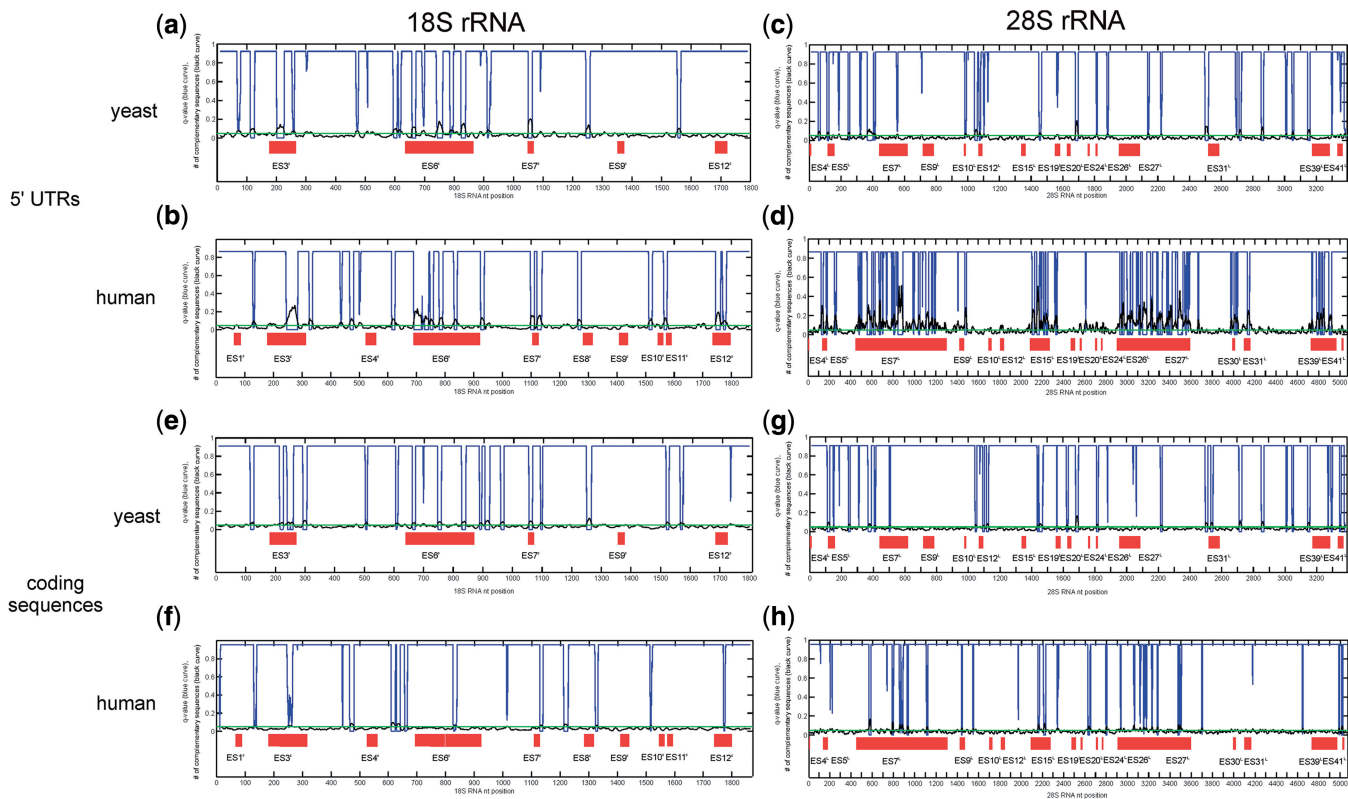


Figure 1. Graphical overview of the sticky regions of 18S and 28S rRNAs complementary to mRNA 5' UTRs and coding sequences in yeast and humans. The panels denoted (a–h) show sticky regions according to a column and a row in which they are placed, e.g. the panel (a) shows sticky regions of yeast 18S rRNA and 5' UTRs. In each panel, the horizontal axis shows the nucleotide positions in rRNAs, whereas the vertical axis indicates both the q -values (i.e. the level of statistical significance) and counts of complementary sequences within the 10 nt neighborhoods marked in the center of each neighborhood on the horizontal axis. In other words, the boundaries of each sequence region are -10 nt and $+10$ nt from the particular position on the horizontal axis. The counts are relative, weighted by the length of the sequences in which they occurred (see the 'Materials and Methods' section); the values of the counts are illustrated by the black curve, whereas q -values are shown by the blue curve. The horizontal black line indicates the statistical significance level, equal to 0.05. The counts of complementary sequences (black spikes) with q -values below this line were evaluated as statistically significant, forming the sticky regions. The red rectangles at the bottom of the plots represent ESs in the rRNAs. ESs are denoted by 'ES x^y ', respectively, where 'x' represents a number and 'y' indicates a subunit ('s' for small, 'L' for large).

complementary sequences (Supplementary Figure S1). We designated these rRNA regions 'sticky regions' owing to their theoretical potential to interact with mRNAs. Interestingly, as shown in Figure 1, most of the sticky regions were in general not similarly distributed between yeast and human rRNAs with the striking exception of the 18S rRNA sticky regions that were complementary to specific sequences in mRNA 5' UTRs. Similarity is indicated by the blue curve of statistical significance, with downward peaks marking the positions of sticky regions in the rRNA sequences that can be found in approximately the same regions in both yeast and human 18S rRNA sequences (cf. Figure 1a and b).

The latter observation, suggesting a potential evolutionary conservation, led us to extend our computational analysis of sequence complementarity between 18S and 28S rRNAs and specifically mRNA 5' UTRs by 12 more species by the same procedure as described earlier in the text. Indeed, statistically significant well-defined regions within rRNAs that were complementary to the corresponding segments within the 5' UTRs were found for all of these species (Supplementary Figure S2a and b). Thorough analysis of the evolutionary conservation of

the rRNA–mRNA complementarity of all 14 species is described in the next section.

Although the rRNA was in our analysis represented by two of its major forms, mRNAs were represented by multiple sequences (in particular by 750 sequences for each species). Hence, although the mRNA complementarity to rRNA occurs within a single rRNA sequence of a well-defined length, either 18S or 28S rRNA, the rRNA complementarity to mRNA has to be sought for on multiple sequences with naturally varying length. Here, the problem arises of how to determine the similarity of rRNA complementarity among multiple randomly long mRNAs. Arbitrarily setting the length of each mRNA 5' UTR sequence to 100% to enable the overall comparison of similarity patterns will inevitably lead to condensing or expanding the information in the peaks of similarity for longer or shorter sequences, respectively, to match the same percentage range. This may in turn produce both false positive and true negative results by matching the peaks that would not normally match if the sequences were not normalized for their length. Also, even though there is a standard procedure of aligning multiple sequences with different length, we could not use it as it is

biologically irrelevant for mRNAs without any functional relationship like the 5' UTRs. We therefore inspected the first 100 and 200 nt of those 5' UTR sequences in our data sets that were long enough (i.e. that were at least 100 and 200 nt long) for non-random locations of enriched complementarity (Figures 2a, Supplementary Figure S3a and

b) assuming that if the rRNA complementarity does exist and is specific, this length is long enough to display some unified pattern. The same analysis was also carried out for the last 100 and 200 nt of the available 5' UTRs (Figures 2b, Supplementary Figure S3c and d). As shown in Figures 2a and b and Supplementary Figure S3a–d, no

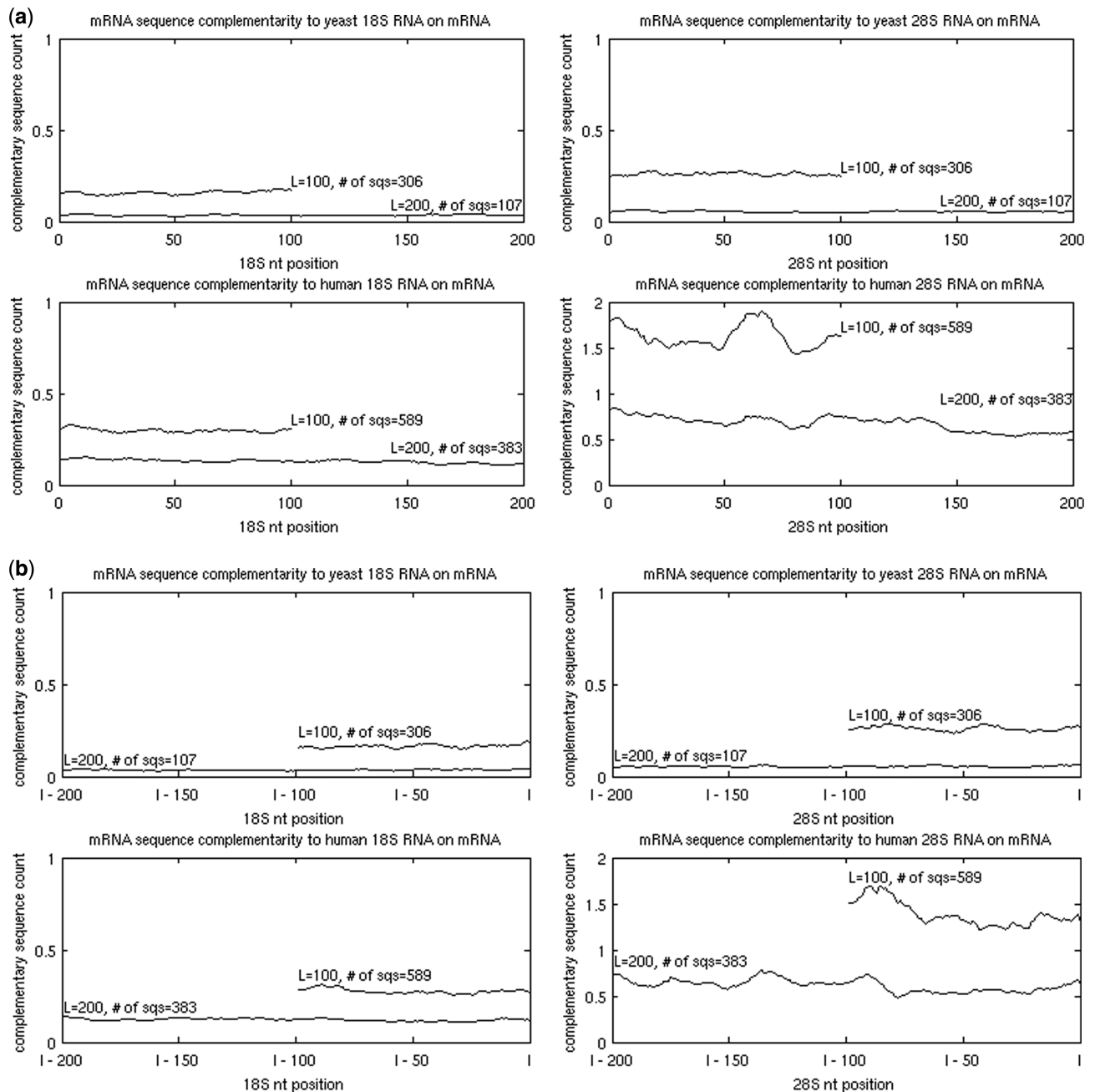


Figure 2. mRNA–rRNA sequence complementarity on mRNA 5' UTRs. For the space reasons, only data for yeast and human, the two most evolutionary distant organisms in the analysis, are shown. The remaining 12 organisms are shown in Supplementary Figure S3a–d. In each section, four panels show mRNA sequence complementarity to yeast 18S, human 18S, yeast 28S and human 28S rRNAs. Section 2a of the figure shows data for first 100 and 200 nt of 5' UTR sequences, whereas section 2b shows data for the last 100 and 200 nt of 5' UTR sequences. In each panel, the horizontal axis represents the nucleotide positions in mRNAs, shown in the opposite direction (3' to 5') in the section 2b ('1' stands for the length of the sequence). The vertical axis indicates counts of complementary sequences within 10 nt neighborhoods marked on the horizontal axis in the center of each neighborhood. In other words, the boundaries of each sequence region are ± 10 nucleotides from the particular position on the x-axis. The curve has the equivalent meaning as the black curve in Figure 1 (counts of complementary sequences). Complementarity for the first 100 and 200 nt of mRNAs is marked by 'L = 100' and 'L = 200', respectively. The numbers of the relevant mRNAs that were analyzed here; i.e. those that are longer than 100 or 200 are indicated as '# of sqs = x', where 'x' is the number of mRNAs.

statistically significant and discernible regions of enriched sequence complementarity of both the first and last 100 and 200 nt of mRNA 5' UTRs to rRNAs were observed. The mRNA sequences complementary to both 18S and 28S rRNAs were randomly spread throughout the 5' UTRs. The only exception among all 14 species was the sequence complementarity of 5' UTRs to 28S rRNA in human (Figure 2a and b, bottom right panels), for which we do not have a clear explanation.

In summary, we observed that within mRNA 5' UTRs and coding regions, segments with enriched complementarity to rRNAs do not exist; however, rRNAs sequences, especially 18S rRNA, do contain a few clearly discernible locations that are complementary to sequences specifically in mRNA 5' UTRs.

Evolutionary conservation of the occurrence of rRNA sticky regions

Evolutionary conservation of rRNA sticky regions could indicate that both the occurrence and distribution of statistically significant rRNA–mRNA sequence complementarity are not random and thus that the complementarity does play some regulatory role in translation. It may also provide some clues regarding the molecular mechanism underlying this prospective regulatory role.

The evolutionary conservation of the sticky regions is indicated by similarity of both occurrence (q -values < 0.05) and distribution (position on rRNA sequences) of the sticky regions in the rRNA sequences of individual species. Both the occurrence and the distribution of sticky regions in rRNAs are shown in Figure 1a and b and Supplementary Figure S2a for 18S rRNA and in Figure 1c and d and Supplementary Figure S2b for 28S rRNA. Again, as in case of mRNAs (as described in the previous section), determination of the similarity pattern among all tested species is hindered by different lengths of their respective rRNA sequences. Unlike mRNAs, however, rRNAs have the same biological function in all species, and thus their direct sequence alignment is possible. Therefore, we first aligned 18S rRNA sequences of all 14 species using a multiple sequence alignment using Clustal Omega (16), and the resulting alignment was subsequently refined using the SINA model of the 18S rRNA multiple alignment (17). Next, we identified positions of all nucleotides that form sticky regions in the individual species (i.e. those with q -values < 0.05 ; where the blue curve fell below the green line in the individual panels in Figure 1 and Supplementary Figure S2a and b) and mapped them onto the multiple sequence alignment of the rRNA sequences. Because the nucleotide positions in the sequence alignment are aligned, the positions of the nucleotides forming the sticky regions are aligned as well. Subsequently, we counted numbers of species having the sticky regions at the same position in aligned 18S and 28S rRNAs and plotted these numbers against the corresponding aligned positions in the form of histograms (Figure 3a and b). In other words, individual positions of the bars (horizontal axis) in our histograms indicate locations of the 'sticky' nucleotides in the rRNA sequence. The height of the bars (vertical axis) then indicates how many of the

14 species do contain a statistically significant sticky region in a corresponding location in the rRNA sequence. Importantly, the histograms show clear and distinctive peaks formed by uninterrupted streaks of bars unambiguously pointing at specific regions located specifically in 18S rRNA with statistically significant complementarity to mRNA 5' UTR sequences. The color of the peaks then indicates a number of species containing a given sticky region falling in a particular bar streak.

To demonstrate that the peaks shown in Figure 3a and b are specific and not accidental, we randomized nucleotide sequences of rRNAs by random shuffling of nucleotide positions in the sequences without changing the overall nucleotide composition and computed their complementarity to the same mRNAs by the same procedure as used earlier in the text. As in case of native rRNA sequences (Figure 3a and b), we then plotted the counts of species with sticky regions mapping at the same positions in aligned randomized 18S and 28S rRNAs (see Figure 3c and d). The highest number of species with mRNA complementarity to randomized rRNAs in a single position (i.e. the highest bar in Figure 3c and d) determined a baseline (i.e. 4 for both rRNAs), above which only the counts are considered to be statistically significant with respect to mRNA complementarity between mRNA and rRNA in a certain position on a given rRNA sequence. Importantly, the same baseline value for both 18S and 28S rRNAs reflects a similar nucleotide composition of 18S and 28S rRNAs in individual species, which was not affected by randomization, and which therefore retains the same level of statistically significant complementarity to mRNAs that can be formed by chance.

We thus propose that all bars higher than four suggest that a given sticky region is evolutionary conserved. In particular, there are 18 bars in six peaks for 18S rRNA (Figure 3a) and one bar in a single peak for 28S rRNA (Figure 3b) that comply with this rule. This striking difference reflected in the size and colors of the peaks (18S rRNA peaks are larger and contain more unique species than those in 28S rRNA) clearly suggests that mRNA 5' UTR sequence complementarity is evolutionary conserved only in 18S rRNA, perhaps with the only exception occurring between ESs ES20^L and ES24^L in the 28S rRNA (Figure 3b). Otherwise, the peaks in 28S rRNA do not differ from those of randomized rRNAs (Figures 3c and d). This may suggest that the 18S rRNA-5' UTR complementarity might have a regulatory role(s) in translation, which could come into play most probably during scanning of the 48S PIC for the AUG start codon (see 'Discussion' section for further details).

Evolutionary conserved sticky regions predominantly map to the ESs and cluster on the ribosomal surface

In addition to mRNA–rRNA complementarity, Figure 3 also illustrates the positions of individual ESs (by horizontal thick red lines) and shows consensus of multiple sequence alignment of the analyzed rRNAs (black histogram at the bottom of individual panels). The positions of ESs reflect those predicted in human rRNAs based on

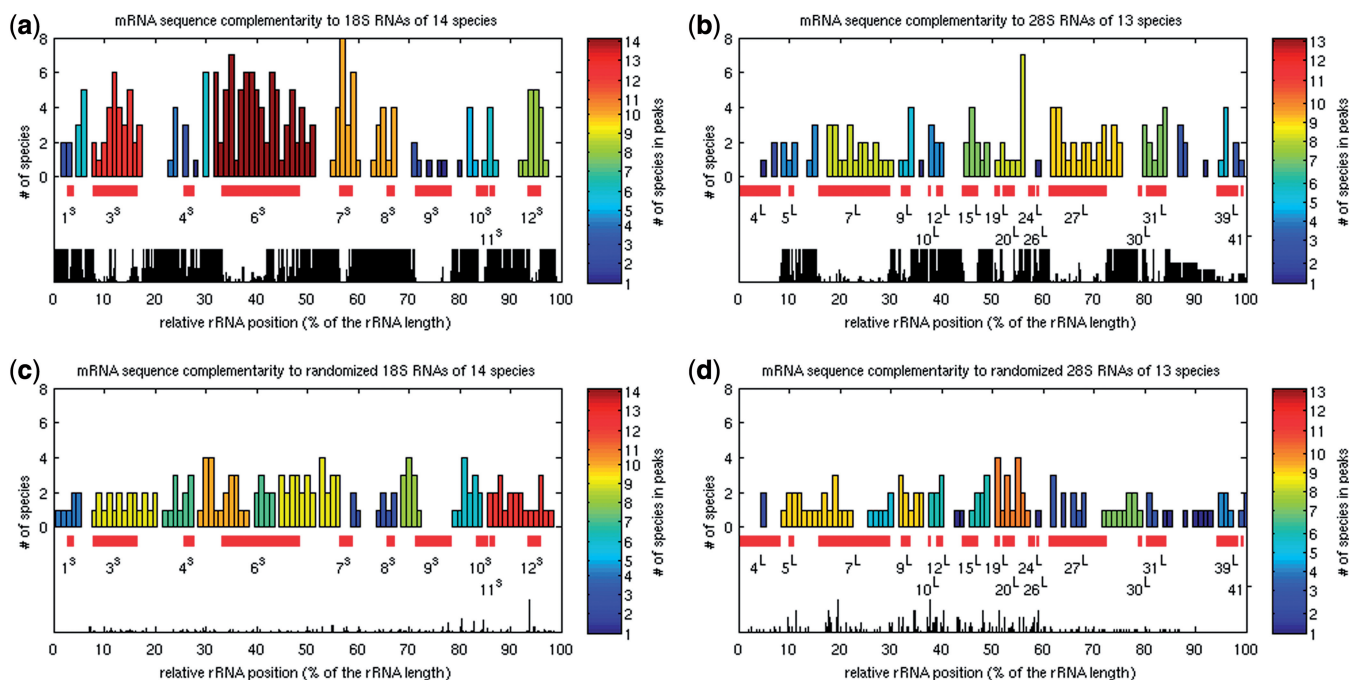


Figure 3. Evolutionary conservation of the sticky regions. The horizontal axis shows relative positions (in percentage of the original length) of the rRNA sequences of 14 (a and c) and 13 (b and d) analyzed species. The vertical axis shows number of the species. Bar colors indicate number of the species with significant mRNA–rRNA complementarity in defined locations on rRNAs indicated by peaks (for explanation, see the ‘Results’ section). Color coding is indicated by the rainbow bar to the right of the panels. The red thick horizontal lines below the histograms mark positions of ESs labeled by ‘x^y’, where ‘x’ represents a number and ‘y’ indicates the subunit (^s small, ^L large). Black profiles at the bottom of the panels represent the consensus sequence of the multiple sequence alignments of the analyzed rRNAs for either 18S (a and c) or 28S (b and d) rRNAs. The height of the peaks in the profile reflects the degree of the consensus at a given nucleotide position of the analyzed rRNAs.

yeast, canine, *T. thermophila* and *Triticum aestivum* ribosomal structures (18–21).

Careful inspection of the conserved occurrence of sticky regions in 18S rRNAs associated with mRNA 5′ UTRs (represented by colored peaks in Figure 3) revealed that they predominantly map to ribosomal ESs (Figure 3a). It is well known that ESs occur on those segments of rRNA sequences that show the lowest degree of conservation among all rRNAs (compare locations of ESs with the strength of consensus of multiple sequence alignment in Figure 3). In addition, the ribosomal structure around the ESs is the least conserved as well. It is thus truly fascinating to find out that the highest degree of conservation of the mRNA–18S rRNA complementarity maps predominantly to the least conserved parts of rRNAs.

In detail, there are four major sticky regions encompassing four ESs (ESs 3^s, 6^s, 7^s and 12^s)—therefore, we designate these ESs ‘sticky ESs’—and two minor sticky regions outside ESs that immediately precede ESs 3^s and 6^s in the 18S rRNA. An important indication of a potential functionality of the sequence complementarity is the fact that all of these sticky ESs exist in the two most evolutionary distant species analyzed here; i.e. in yeast and human. On the other hand, none of the ESs found exclusively in human 18S rRNA (namely, ESs 1^s, 4^s, 8^s, 10^s and 11^s) belong to the family of ESs with evolutionary conserved sticky regions.

Additional indication of a potential biological importance of the mRNA–rRNA complementarity is another

fact that three of the sticky ESs, namely, ES6^s, ES3^s and ES12^s, form a specific cluster at the bottom of the 40S ribosome, on its solvent exposed side, in which rRNA constitutes a dense network of interconnected RNA on the surface of the 40S ribosome (21,22).

The sticky ESs would be good candidates for establishing the rRNA–mRNA interactions for two reasons: (i) they are exposed to the solvent and (ii) they contain flexible structures with a high probability of dynamic production of single-stranded rRNA segments during often occurring structural rearrangements (21), which can readily interact with other RNA molecules including mRNAs in the 48S PICs. Such a specific distribution of the evolutionary conserved sequence complementarity to mRNA in 18S rRNA ESs suggests that the sticky regions form evolutionary conserved patterns not only in the primary sequences of 18S rRNA but also in their structures.

Do sticky regions constitute a post-exit mRNA threading path on the surface of the 40S subunit?

Finally, we took advantage of the recently solved high-resolution X-ray structure of 18S Rrna, as it appears in the mature yeast 40S subunit (22). We used it as a model for determination of the structural environment of the evolutionary conserved sticky regions. Toward this end, the yeast 18S rRNA and the evolutionary conserved 5′ UTR sticky regions (Figure 3a, Table 1) were mapped onto the yeast 40S structure (Figure 4). In accord with

Table 1. Positions of yeast 5' UTRs—18S rRNA sticky regions and structural elements that they fold into

Position of 5' UTR sticky regions on yeast 18S rRNA sequence	18S rRNA structural elements encompassing sticky regions ^a
103–135	h7
192–235, 249–269	ES3 ^s
649–679, 734–771, 777–800, 786–806, 815–848	ES6 ^s
902–927	h23
1042–1073	ES7 ^s

^aHelices and ESs are denoted by 'hx' and 'ESx^y', respectively, where 'x' stands for a number, 'y' indicates subunit ('s' small, 'L' large).

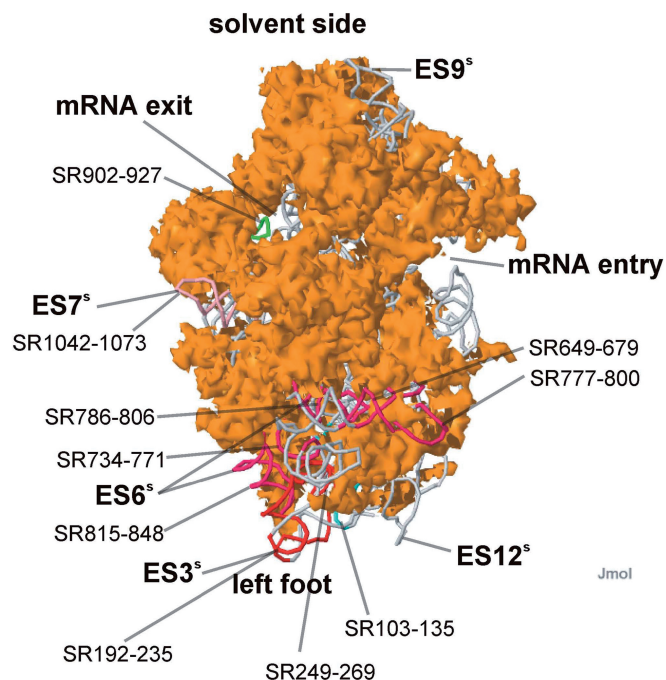


Figure 4. The yeast 18S rRNA sticky regions set in the context of the crystal structure of the 40S subunit. The structure was adopted from 3U5B PDB structure {PDB, #31} {Ben-Shem, 2011 #29}. The 18S rRNA structure is shown using ribbons overlaid with the surface representations of the small ribosomal proteins to demonstrate the accessibility of the 18S rRNA sticky regions on the ribosomal surface. For the color coding of the sticky regions denoted by 'SR' followed by the coordinates of their positions in the 18S RNA sequence, please see the last chapter of the 'Results' section. The ESs are labeled as 'ESx^y', where 'x' represents a number and 'y' indicates the subunit ('s' small, 'L' large).

the canine ribosomal structure, all sticky regions of yeast 18S rRNA occur within the solvent-exposed segments and spatially clustered near the exit of the 40S mRNA-binding channel and on the solvent-exposed middle-to-bottom body of the 40S ribosome including its left foot.

More specifically, one sticky region occurred at the exit of the mRNA-binding channel at position 902–927 (SR902–927, colored in green in Figure 4), followed by SR1042–1073, mapping directly to ES7^s below the exit of the mRNA binding channel at position 1042–1073 (colored in pink). These sites were followed downwards

along the left foot by three hits in ES6^s (SR815–848, SR734–771 and SR786–806, in magenta) and ended at the bottom of the left foot with two hits in ES3^s (SR192–235 and SR249–269, in red) and in helix 7 (SR103–135, in cyan). It is tempting to speculate that this particular distribution may reflect the threading path of an mRNA not only through the mRNA-binding channel during the search for the AUG start codon but also after it has left the immediate intersubunit space; yet, it still remains in the vicinity of the small subunit, where it could perhaps influence ribosomal scanning (see 'Discussion' section). Finally, to the right of this path, two hits in the horizontal arm of ES6^s (SR649–679 and SR777–800, in magenta) were found. Another version of the Figure 4 rotating in space is provided in Supplementary Figure S4. Given the high evolutionary conservation of the sticky regions demonstrated earlier in the text, it is likely that an analogous arrangement of them into the proposed post-exit mRNA threading path will be seen in other eukaryotic species, when the structure of their 40S ribosomal subunits is solved.

DISCUSSION

Given the fact that the universality and physiological significance of the role of mRNA–rRNA base pairing in translational control has never been investigated using a large-scale computational approach, in this study, we embarked on a robust, systematic and statistically valid computational search for mRNA–rRNA complementarity across eukaryotic species. Our study purposely focused on mRNA–rRNA complementarity, leaving aside another interesting aspect of this scenario, which is the mRNA sequences that exactly match the rRNA sequences in sense orientation. As wet-lab experiments aimed at addressing these questions are extremely resource-demanding or even technically unfeasible, we took advantage of statistically sufficient amounts of reliable biological data, which are available in various sequential and structural databases, to produce biologically plausible knowledge through systematic hypothesis-driven computing. Our results demonstrate, for the first time, the non-random and systematic nature of the complementarity of specific segments (sticky regions) of eukaryotic rRNAs to mRNA 5' UTRs with indications of evolutionary conservation.

Our model of sequential complementarity between mRNA and rRNA was designed for computational and quantification purposes of large-scale data sets. Beginning with the inspection of the nature of how complementary sequences mediate intermolecular interactions, we took advantage of the fact that long DNA and RNA sequences usually interact with each other through base pairing between shorter segments of several nucleotides in length, often separated by gaps of a few nucleotides in their sequences. These sequence-specific segments can then be found concentrated in space, in the secondary and tertiary structures of those sections of mutually interacting macromolecules, which directly mediate their contact. We aimed to identify these relatively short continuous segments that are spatially condensed in defined

locations, rather than detecting long interacting sequences through the prediction of hybridization energies, as is routinely done in classical studies. The latter approach is rather slow and often inefficient, especially when a large number of complementary sequences need to be inspected for potential interactions.

All of the previous studies on this topic have certainly made a valuable progress in our common effort to elucidate the existence of mRNA–rRNA complementarity and its function in translational control in eukaryotes. However, none of them included statistical evaluation of the significance of rRNA–mRNA complementarity in their preliminary computational searches, which would be based on sufficient number of sample sequences from multiple organisms, to examine evolutionary conservation and functionality of the observed complementarity. Using large-scale data sets of mRNA sequences from 14 considerably evolutionary distant organisms, between which significant differences could be expected, we identified several regions in 18S rRNA but not in 28S rRNA showing statistically significant complementarity to mRNA 5' UTRs and a non-random spatial distribution over the ribosomal surface, suggesting an unambiguous potential to be biologically relevant (Figures 1–3). In accord, the ability of the 5' UTRs of mRNAs to interact with 18S rRNA was already demonstrated in the mice pilot studies on this topic (5). In contrast to rRNAs, the examined mRNA sequences showed no statistically significant complementarity to rRNA sequences that would be common in mRNAs. These findings are further supported by our demonstration of the striking similarity of sequential complementarity among the 18S rRNA sticky regions and mRNA's 5' UTRs between all 14 species.

As suggested by Matveeva *et al.* (5), the complementarity between rRNA and mRNA forms universal regions of mRNA binding in translation processes. Mauro *et al.* (23) then proposed that the mRNA–rRNA complementarity provides the eukaryotic ribosome with a regulatory function that could specifically 'filter' what mRNAs and with what frequency will be recruited to the 40S PICs and thus translated. In other words, increased complementarity should increase the odds of a given mRNA being recruited to the 40S ribosome to facilitate the initiation of its translation, as demonstrated experimentally (11). At the same time, however, it could also decrease its odds of being translated by sequestering it in a non-productive manner, possibly owing to the establishment of a strong base-pairing interaction that would be difficult to disrupt, which has also been demonstrated experimentally (7,24). In any case, the unified hypothesis of all previous reports was that rRNA–mRNA complementarity primarily promotes the recruitment of diverse mRNAs in a selective manner, followed by differential effects on the efficiency of their translation. Based on the clustering pattern of the 18S rRNA sticky regions that we observed, specifically concentrated on the lower solvent-exposed body of the 40S ribosome (Figure 4), we think that the likelihood of these regions serving primarily as mRNA recruiters is low.

The recruitment of the capped mRNAs occurs with the 5' cap situated near the mRNA exit channel to fully enter

the interface-based mRNA-binding channel, with the latch formed by helices 18 (h18) and 34 (h34) of the 40S beak and shoulder temporarily dissolved to enable the small ribosomal subunit to begin inspecting mRNA bases from the farthest 5' end (25,26). This simple fact implies that mRNA, on initial contact, must approach the ribosome from its interface side. Hence, it is improbable that the sticky regions situated on the far lower back side of the ribosome could actively promote the recruitment step *per se*, as they would contact mRNAs at regions that are not uniformly distributed within 5' UTRs with respect to their positions relative to the 5' cap. This situation would impose a rather challenging task on the ribosome to enforce the 5' end of these 'diverse' mRNAs wrapped around the ribosomal body through interactions with rRNA to reach the mRNA-binding channel in a way that ensures the exact placement of the 5' cap immediately behind the mRNA exit channel.

What is then the possible role of the ES-based solvent-side-situated sticky regions? Having noted that mRNA contacts the 40S ribosome from the interface side, the time frame for the potential regulatory action of these sticky regions must come after at least a part of the 5' UTR has emerged from the mRNA exit channel. Interestingly, hydroxyl radical cleavages directed from eIF4G HEAT-1 in reconstituted mammalian 48S PICs were found to occur primarily in ES6 of 18S rRNA (27). These results suggest that at least the HEAT-1 domain of eIF4G, but possibly also other eIF4G-binding partners, such as eIF4A, is situated below the platform, close to the 'left foot' of the 40S subunit. Moreover, it was also proposed that eIF4G, in complex with eIF4A and the cap-binding protein eIF4E (the so-called eIF4F complex), remains close to the cap even during scanning, and thus the already-inspected nucleotides do not purposelessly move about in space but form a loop between the cap and the traversing ribosome (28). The mRNA path following the localization of these factors on the ribosome matches the proposed post-exit mRNA-threading path formed by sites of sticky regions spatially arranged on the 40S solvent side starting at the exit of the mRNA-binding channel and going down toward the left foot of the 40S ribosome. Hence, it could be proposed that the solvent-side-situated sticky regions occurring in the vicinity of eIF4F could contact complementary regions within these loops and modulate the eIF4F activity during scanning. This would be consistent with the exclusive occurrence of the sticky regions in 18S but not 28S rRNA, as the scanning takes place without the 60S subunit. Sticky regions, and especially those located in the ESs, could for example help to organize the spatial arrangement of the loop so that the already scanned mRNA sequences do not sterically impede the progression of scanning, or, more probably, they could impact the rate of scanning by holding onto complementary sections of 5' UTRs. Both of these effects would have differential effects on the translatability of individual mRNAs depending on the number of the 5' UTR sequences complementary to the sticky regions and the degree of their mutual complementarity.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–3 and Supplementary Movies 4 and 5.

ACKNOWLEDGEMENTS

The authors are thankful to C. W. Akey for providing us with the figure of the canine 80S ribosome. JP: conceived and designed the study, carried out the computational experiments. MK: conceived and designed the statistics. JP and LSV: biological interpretation of the computational results. JP, MK, JV and LSV: wrote the manuscript.

FUNDING

Centrum of Excellence of the Czech Science Foundation [P305/12/G034 to L.S.V.]; Wellcome Trust [090812/Z/09/Z to L.S.V.]; Academy of Sciences of the Czech Republic [RVO: 68378050 to M.K.]. Funding for open access charge: Academy of Sciences of the Czech Republic.

Conflict of interest statement. None declared.

REFERENCES

1. Sonenberg, N. and Hinnebusch, A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, **136**, 731–745.
2. Kong, J. and Lasko, P. (2012) Translational control in cellular and developmental processes. *Nat. Rev. Genet.*, **13**, 383–394.
3. Gebauer, F., Preiss, T. and Hentze, M.W. (2012) From cis-regulatory elements to complex RNPs and back. *Cold Spring Harb. Perspect. Biol.*, **4**, a012245.
4. Hellen, C.U. (2009) IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. *Biochim. Biophys. Acta*, **1789**, 558–570.
5. Matveeva, O.V. and Shabalina, S.A. (1993) Intermolecular mRNA-rRNA hybridization and the distribution of potential interaction regions in murine 18S rRNA. *Nucleic Acids Res.*, **21**, 1007–1011.
6. Mauro, V.P. and Edelman, G.M. (1997) rRNA-like sequences occur in diverse primary transcripts: implications for the control of gene expression. *Proc. Natl Acad. Sci. USA*, **94**, 422–427.
7. Tranque, P., Hu, M.C., Edelman, G.M. and Mauro, V.P. (1998) rRNA complementarity within mRNAs: a possible basis for mRNA-ribosome interactions and translational control. *Proc. Natl Acad. Sci. USA*, **95**, 12238–12243.
8. Verrier, S.B. and Jean-Jean, O. (2000) Complementarity between the mRNA 5' untranslated region and 18S ribosomal RNA can inhibit translation. *RNA*, **6**, 584–597.
9. Schneider, R., Agol, V.I., Andino, R., Bayard, F., Cavener, D.R., Chappell, S.A., Chen, J.J., Darlix, J.L., Dasgupta, A., Donze, O. *et al.* (2001) New ways of initiating translation in eukaryotes. *Mol. Cell. Biol.*, **21**, 8238–8246.
10. Mauro, V.P. and Edelman, G.M. (2007) The ribosome filter redux. *Cell Cycle*, **6**, 2246–2251.
11. Dresios, J., Chappell, S.A., Zhou, W. and Mauro, V.P. (2006) An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.*, **13**, 30–34.
12. Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, **320**, 1344–1349.
13. Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E. *et al.* (2009) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.*, **38**, D75–D80.
14. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glockner, F.O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, **41**, D590–D596.
15. Storey, J.D. (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Stat.*, **31**, 2013–2035.
16. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
17. Pruesse, E., Peplies, J. and Glockner, F.O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, **28**, 1823–1829.
18. Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Blushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S. *et al.* (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc. Natl Acad. Sci. USA*, **107**, 19748–19753.
19. Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G. and Frank, J. (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions. *Cell*, **107**, 373–386.
20. Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A. and Ban, N. (2011) Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science*, **331**, 730–736.
21. Chandramouli, P., Topf, M., Menetret, J.F., Eswar, N., Cannone, J.J., Gutell, R.R., Sali, A. and Akey, C.W. (2008) Structure of the mammalian 80S ribosome at 8.7 Å resolution. *Structure*, **16**, 535–548.
22. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G. and Yusupov, M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
23. Mauro, V.P. and Edelman, G.M. (2002) The ribosome filter hypothesis. *Proc. Natl Acad. Sci. USA*, **99**, 12031–12036.
24. Hu, M.C., Tranque, P., Edelman, G.M. and Mauro, V.P. (1999) rRNA-complementarity in the 5' untranslated region of mRNA specifying the Gtx homeodomain protein: evidence that base-pairing to 18S rRNA affects translational efficiency. *Proc. Natl Acad. Sci. USA*, **96**, 1339–1344.
25. Pestova, T.V. and Kolupaeva, V.G. (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.*, **16**, 2906–2922.
26. Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R. and Ramakrishnan, V. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell*, **26**, 41–50.
27. Yu, Y., Abaeva, I.S., Marintchev, A., Pestova, T.V. and Hellen, C.U. (2011) Common conformational changes induced in type 2 picornavirus IRESs by cognate trans-acting factors. *Nucleic Acids Res.*, **39**, 4851–4865.
28. Marintchev, A., Edmonds, K.A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C., Herdy, B., Sonenberg, N. and Wagner, G. (2009) Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell*, **136**, 447–460.