How much can we learn about the function of bacterial rRNA modification by mining large-scale experimental datasets?

Petr V. Sergiev*, Anna Y. Golovina, Olga V. Sergeeva, Ilya A. Osterman, Mikhail V. Nesterchuk, Alexey A. Bogdanov and Olga A. Dontsova

Lomonosov Moscow State University, Department of Chemistry and A.N. Belozersky Institute of Physico-Chemical Biology, Moscow 119992, Russia

Received December 5, 2011; Revised February 16, 2012; Accepted February 21, 2012

ABSTRACT

Modification of ribosomal RNA is ubiquitous among living organisms. Its functional role is well established for only a limited number of modified nucleotides. There are examples of rRNA modification involvement in the gene expression regulation in the cell. There is a need for large data set analysis in the search for potential functional partners for rRNA modification. In this study, we extracted phylogenetic profile, genome neighbourhood, co-expression and phenotype profile and co-purification data regarding *Escherichia coli* rRNA modification enzymes from public databases. Results were visualized as graphs using Cytoscape and analysed. Majority linked genes/proteins belong to translation apparatus. Among co-purification partners of rRNA modification enzymes are several candidates for experimental validation. Phylogenetic profiling revealed links of pseudouridine synthetases with RF2, RsmH with translation factors IF2, RF1 and LepA and RlmM with RdgC. Genome neighbourhood connections revealed several putative functionally linked genes, e.g. *rlmH* with genes coding for cell wall biosynthetic proteins and others. Comparative analysis of expression profiles (*Gene Expression Omnibus*) revealed two main associations, a group of genes expressed during fast growth and association of *rrmJ* with heat shock genes. This study might be used as a roadmap for further experimental verification of predicted functional interactions.

INTRODUCTION

Modification of ribosomal RNA is ubiquitous among all living organisms. The proportion of modified residues roughly correlates with the complexity of an organism (1). Many efforts were invested in the determination of specific genes coding for rRNA modification enzymes and the mapping of modified residues in rRNA (2–4). A common opinion is that rRNA modification is needed for fine-tuning ribosome structure and function. All modifications of rRNA residues studied so far in *Escherichia coli* are dispensable. It is likely that modified nucleotides could be involved in the functioning of specific regulatory mechanisms of gene expression. In particular, modification of ribosomal RNA could be used for functional diversification of ribosomes. Hence, the existence of different ribosome species and the appearance of ‘altered’ ribosomes is well documented for eukaryotes (5). Plasmodium species express different rRNA genes when infecting either vertebrate or mosquito hosts (6). Ribosomal rpS6 phosphorylation alters translation efficiency (7). In yeast, ribosomal proteins encoded in the paralogue genes were shown to be specialized for mRNA-specific translation (8).

When speaking about possible ribosome specialization by rRNA modification, we should distinguish two possibilities. Firstly, ‘constitutive’ rRNA modification, independent of regulation by itself, could be necessary for the functioning of a specific regulatory mechanism. In this case, rRNA modification would be necessary for alteration of translation regulation but does not regulate it in a strict sense. This type of function could be revealed by a comparison of gene expression in the wild-type strain and the rRNA modification-deficient strain. Several laboratories, including our own, are involved in systematic analysis of gene expression differences caused by specific
rRNA modifications. Second, 'conditional' rRNA modification dependent on some other factor would in turn influence translation of some mRNAs. This would be regulation by rRNA modification in a strict sense. In this case, not only the influence of rRNA modification on gene expression, but also some conditions influencing the extent of particular rRNA modification should be demonstrated.

In this study, we used mining several large data sets in search for genes/proteins functionally related to rRNA modification. As queries for this analysis, we chose known genes coding for rRNA modification enzymes of the most studied bacteria, E. coli (Table 1). The results of this study could be used to guide hypothesis-driven studies on the specific functional role of rRNA modification. An equally probable result of this study might be to suggest a secondary function of rRNA modification enzymes.

MATERIALS AND METHODS

Database mining

Analysis of phylogenetic profiling was done using mutual information criteria using the STRING database search system (9), utilizing the probability assessment algorithm provided in the database engine. Significance criteria used was a probability of the functional linkage being more than or equal to 0.4. Data of gene co-occurrence in the bacterial genomes and all other data on functional association of genes were visualized by Cytoscape (10). Genome neighbourhood was analysed using the STRING database using significance criteria ≥0.7.

Operone structure of rRNA modification genes was checked manually on the basis of published results of a deep sequencing analysis of the complete E. coli transcriptome (11). Co-expression profiling was analysed by the microbesonline database (12) on the basis of gene expression omnibus (GEO) database (13) entries related to E. coli. Only genes whose co-expression correlation coefficients with rRNA modification genes were ≥0.6 were selected. Protein co-purification data were extracted from the microbesonline database (12) based on the original data (14,15). Phenotype profile similarity data were extracted from http://ecoliwiki.net/tools/chemgen/ based on the original work (16). The phenotype landscape of rRNA modification gene knockout strains was extracted from an original published data set (16) and visualized by Microsoft Excel.

RESULTS AND DISCUSSION

Co-occurrence with rRNA modification genes in bacterial genomes

If enzymes are used in the same metabolic pathway, their genes’ presence across genomes is strictly correlated. A mutual information criteria were used as implemented into the STRING database (9) search engine. It is a more flexible criterion that accommodates not only simultaneous presence or absence of a pair of genes in a number of genomes but also similar degrees of conservation between pairs in a wide range of species. A similar degree of conservation reflects a co-evolution of genes. The confidence cut-off was set at 0.4 probability of functional relation. Mutual information analysis of genes

Figure 1. Genes whose presence in bacterial genomes coordinated with the presence of rRNA modification genes. Shown are genes whose presence in bacterial genomes correlated with the presence of rRNA modification genes (dark grey) at the confidence level 0.4 and higher. Confidence measure was based on mutual information. Edge width corresponds to the confidence of interaction. Figure was created with the help of Cytoscape.
Table 1. Known E. coli tRNA modification enzymes and modifications they make

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Nucleotide</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pseudouridine synthases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RsuA (YcgD)</td>
<td>16S rRNA m16Ψ</td>
<td>(25)</td>
</tr>
<tr>
<td>RluA (YabO)</td>
<td>23S rRNA m746Ψ</td>
<td>(26)</td>
</tr>
<tr>
<td>RluB (YciL)</td>
<td>23S rRNA 2605Ψ</td>
<td>(22)</td>
</tr>
<tr>
<td>RluC (YccC)</td>
<td>23S rRNA 955Ψ, 2504Ψ, 2580Ψ</td>
<td>(17,27)</td>
</tr>
<tr>
<td>RluD (YfiH)</td>
<td>23S rRNA 1919Ψ, 1917Ψ</td>
<td>(17,18)</td>
</tr>
<tr>
<td>RluE (YmiC)</td>
<td>23S rRNA 2457Ψ</td>
<td>(22)</td>
</tr>
<tr>
<td>RluF (YjiC)</td>
<td>23S rRNA 2604Ψ</td>
<td>(22)</td>
</tr>
<tr>
<td><strong>Methyltransferases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RsmG (GidB)</td>
<td>16S rRNA mG527</td>
<td>(28)</td>
</tr>
<tr>
<td>RsmD (YhhF)</td>
<td>16S rRNA mG966</td>
<td>(3)</td>
</tr>
<tr>
<td>RsmB (YjdB)</td>
<td>16S rRNA mG907</td>
<td>(29,30)</td>
</tr>
<tr>
<td>RsmC (YjT)</td>
<td>16S rRNA mG1207</td>
<td>(31)</td>
</tr>
<tr>
<td>RsmI (YraL)</td>
<td>16S rRNA mC1402</td>
<td>(32)</td>
</tr>
<tr>
<td>RsmH (MraW)</td>
<td>16S rRNA mC1402</td>
<td>(32)</td>
</tr>
<tr>
<td>RsmF (YebU)</td>
<td>16S rRNA mC1407</td>
<td>(33)</td>
</tr>
<tr>
<td>RsmE (YggJ)</td>
<td>16S rRNA mU1498</td>
<td>(34)</td>
</tr>
<tr>
<td>RsmA (KsgA)</td>
<td>16S rRNA mU1518, m6A1519</td>
<td>(35,36)</td>
</tr>
<tr>
<td>RlmC (YbjF, RumB)</td>
<td>23S rRNA mU747</td>
<td>(38)</td>
</tr>
<tr>
<td>RlmF (YbiN)</td>
<td>23S rRNA m6A1618</td>
<td>(2)</td>
</tr>
<tr>
<td>RlmG (YgO)</td>
<td>23S rRNA mG1835</td>
<td>(39)</td>
</tr>
<tr>
<td>RlmH (YbeA)</td>
<td>23S rRNA mW9195</td>
<td>(40,41)</td>
</tr>
<tr>
<td>RlmD (YgeA, RumA)</td>
<td>23S rRNA mU1939</td>
<td>(38,42)</td>
</tr>
<tr>
<td>RlmI (YceW)</td>
<td>23S rRNA mG1962</td>
<td>(43)</td>
</tr>
<tr>
<td>RlmB (YjiH)</td>
<td>23S rRNA mG2251</td>
<td>(44)</td>
</tr>
<tr>
<td>RlmL (YebY)</td>
<td>23S rRNA mG2445</td>
<td>(45)</td>
</tr>
<tr>
<td>RlmM (YggE)</td>
<td>23S rRNA mC2498cm</td>
<td>(4)</td>
</tr>
<tr>
<td>RlmN (YggB)</td>
<td>23S rRNA mA2503</td>
<td>(46)</td>
</tr>
<tr>
<td>RlmE (Fsd, RumJ)</td>
<td>23S rRNA Um2552</td>
<td>(47,48)</td>
</tr>
</tbody>
</table>

Mutual information analysis of nearly universal among bacteria methyltransferase RsmH responsible for 16S rRNA C1402 N6-modification (32) revealed a link with several translation-related proteins of high conservation, such as ribosomal proteins S9, L9 and L20 and translation factors EF-G, IF2, RRF, LepA, RF1 and ribosome-related GTPase YchF. Another cluster is formed by RsmD, RlmB and RlmN methyltransferases connected with a set of genes involved in a number of unrelated processes, namely cell wall biosynthesis, transcription, DNA repair and amino acid biosynthesis.

The RsmG gene responsible for N7-methylation of the 16S rRNA nucleotide G527 (28,49) is linked with rsmL, which is responsible for ribose methylation of the 16S rRNA nucleotide C1402 (28). The RelA gene links both modification enzyme genes by an additional connection. Putative involvement of rsmG and rsmL methyltransferases in relA-dependent stringent response (50) could be suggested. RunB and RsmF methyltransferase genes form a poorly defined cluster with a number of genes with an unknown function. RlmL, responsible for N2-methylation of the 23S rRNA nucleotide G2445 (45), connects with the isoleucyl-tRNA synthetase gene, the ribosomal protein L5 gene and genes involved in 4-thiouridine biosynthesis. The KsgA methyltransferase gene forms a cluster with the rimN, gene coding for the ribosome maturation protein histidyl-tRNA synthetase gene, and several genes unrelated to translation. RimJ forms a cluster with glutaredoxin 4 and the dnaK chaperone. The RimH gene responsible for mΨ1915 formation in the 23S rRNA (40,41) forms a cluster with formyltransferase, tRNA (mG46) methyltransferase, ribosomal protein L11-specific methyltransferase (prnA) and several genes unrelated to translation.

**Neighbourhood in bacterial genomes**

Knowing the operon structure of rRNA modification enzyme genes can be informative for formulating an ‘educated guess’ of regulatory processes involving rRNA modification (Figure 2 and Supplementary File S2). In Figure 2, a genomic neighbourhood analysis of rRNA modification genes among bacterial genomes is presented. The operon structure of rRNA modification genes in E. coli (11) is marked by bold edges connecting the genes on the same operon. Among rRNA modification genes, only few are coded as stand-alone cistrons. Several rRNA modification enzymes are co-expressed with tRNA modification enzymes. RsmG, responsible for 16S rRNA mG527 formation (28), is located in the operon with mnmG6, the gene coding for the tRNA modification enzyme, necessary for carboxymethylaminomethyl (cmnm) modification of U34 residue of certain tRNAs (51). This arrangement is conserved according to the large-scale genomes analysis. Although there is no direct contact between mG527 and U34 of A-site-bound tRNA (52), modified nucleotides might ‘communicate’ via G530 undergoing syn-anti flip upon decoding (52).

RluE, the gene coding for pseudouridine synthase, required for U to Ψ conversion at 23S rRNA position...
2457 (22), is located adjacent to mnmA, a gene responsible for 2-thiouridine formation at tRNA nucleotide 34 (53). However, analysis of E. coli transcriptome does not reveal existence of the joint operon (11). The s'U residue could be subsequently converted to s'U residue, modified at position 5, depending on tRNA identity. It is difficult to imagine any communication between position 34 of tRNA and /C9 2457 of the 23S rRNA, except for communication through the tRNA body during the process of accommodation (54), which might be relevant to translation fidelity (55). A gene located in the same operon with rluE is nudJ, coding for a member of Nudix hydrolases family, responsible for catabolism of damaged and modified nucleoside triphosphates (56). Similarly, rsuA coding for the 16S rRNA Ps516 synthetase (25) could be found on the same operon with a number of genes coding for pseudouridine metabolism.

HepA, located in the same transcript with rluA, is a protein involved in transcription termination (57), whereas RluD forms an operon with yfiH as evidenced from complete transcriptome sequencing (11). RsmB methyltransferase is coded in an operon (11) after protein deformylase and formyltransferase, involved in initiator fMet residue cleavage (58) and synthesis (59). After these two genes a strong Rho-independent terminator is encoded, followed by the rsmB gene that is not preceded by any Shine–Dalgarno-like sequence (60). RsmB works early in assembly of the small ribosomal subunit to form nucleotide m5C967 of the 16S rRNA (61). It is interesting how the RsmB enzyme could act in time for the modification of 16S rRNA if its expression has so many obstacles leading apparently to low enzyme concentration. RsmB has a NusB-like RNA-binding domain that could potentially serve as an assembly chaperone. A functional link between the initiator fMet-tRNAfMet tRNA and the 16S rRNA modified nucleotide m7C967 is quite logical. This modified residue is in direct contact with the initiator fMet-tRNAfMet tRNA during the process of translation initiation (62).

A surprisingly high proportion of rRNA modification genes are co-expressed with various genes coding for transmembrane proteins, proteins involved in cell wall synthesis, transmembrane transport, etc. The YhhL gene that partially overlaps with the 16S rRNA G966 N2-methyltransferase gene rsmD (3,60) and YbjO, located in the same operon (11) with the 23S rRNA U747 C5-methyltransferase gene rumB (38), are considered to be putative membrane proteins (63). RsmH is located in a large operon coding for many cell wall biosynthetic genes (11,63). The operon structure of rsmH is conserved and shows high significance by large-scale genome analysis (9). This arrangement suggests an experimental verification of a mutual influence of the 16S rRNA N4-C1402 methylation and cell wall biosynthesis. RlmH, which is involved in methylation of 23S rRNA nucleotide ?1915 (40,41), is located in an operon with several cell wall biosynthesis genes (11,63), but its nearest operon neighbour ybeB codes for a protein physically associated with 50S
ribosomal subunits (64,65). The YbeB protein resembles plant Ioja protein associated with chloroplast ribosomes (66); however, its function remains unknown and is certainly worth experimental investigation.

KsgA, one of the most universal rRNA modification genes common in all living organisms, is responsible for formation of modified nucleotides m5A1518 and m2A1519 in the 16S rRNA (35,36). According to a recent finding, KsgA is not simply a modification enzyme but rather a switch protein sequestering incompletely assembled 30S subunits from being involved in translation (67–69). KsgA is encoded in an operon with a periplasmic peptidyl-prolyl isomerase (surA) (63) and apaH, coding for AppppA hydrolase (70). Diadenosine tetraphosphate (AppppA) is an ubiquitous alarmone molecule produced by lysyl-tRNA synthetase (71). Little is known about its function in bacteria, but its possible involvement in stress response was suggested in scientific literature (72). Several other stress–response genes are located on the same operons with rRNA modification genes. The GshB, enzyme necessary for the last step in glutathione biosynthesis (73), is located in an operon with rsmE (11), whereas the RelA protein involved in ppGpp alarmone biosynthesis in response to amino acid starvation (28,74) is located next to the rrmA gene. The RrmJ gene is located in the same operon with fisH, a membrane-bound protease (47,48). Both genes are induced upon heat shock (75). Involvement of RrmJ, responsible for formation of Um2552 of the 23S rRNA (47,48), is certainly possible in a heat shock response. What is mysterious, however, is the fact that Um2552 is methylated constitutively, independent of heat shock (47).

The observed genomic neighbourhood of the 23S rRNA nucleotide G2251 ribose methyltransferase gene rlmB (44) with rnr gene, coding for RNase R involved in rRNA maturation and RNA decay under cold shock conditions and in a stationary phase, seems logical (76). Both proteins are involved in rRNA maturation. RNase R co-purifies with several ribosome maturation proteins, such as RlmN, RluB, RluC, RlmI and RlmL (14,15).

The Uup gene is linked with the rRNA methyltransferase gene rlmL, responsible for formation of the m5G2445 modified nucleotide in 23S rRNA (45). Uup codes for a soluble ATP-binding cassette (ABC) protein (77). Such proteins, exemplified with ABCE1 (78), could be involved in regulation of translation. Deletion of uup makes cells susceptible to contact-dependent inhibition (CDI) by the parent strain (77). Although the mechanism of such inhibition is unknown and involvement of rRNA modification in this process is questionable, it opens a possibility for further studies. It might be that this mechanism involves inactivation of rRNA either by a small molecule or enzyme and rRNA modification is relevant to this process.

**Similar patterns of expression**

Expression pattern similarity data were extracted from the GEO database (13) with the help of www.microbesonline.org website (12). Co-expression in a number of growth conditions might reflect similar mechanisms of gene expression regulation, which, in turn speaks in favour of co-involvement in the same functional pathway. Using a confidence cut-off value of 0.6, one can observe that expression profiles for rRNA modification genes are separated into several uneven clusters (Figure 3, Supplementary Table S1 and Supplementary File S3). Two stand-alone binary clusters were visible reflecting co-expression based on the operon structure: ksgA with apaG and rluA with hepA. A larger cluster contains the rrmJ methyltransferase gene and several heat shock-related chaperone and protease genes. This link is expected one since rrmJ is located in the heat shock-induced operon and its expression increases with the temperature (75). All other rRNA modification genes form a huge cluster of a total 235 gene associated with exponential growth. This clusterization seems absolutely logical because exponential growth requires massive ribosome assembly and thus ribosome modification factors.

At the co-expression profile confidence level cut-off of 0.7, several ribosome modification genes fall off the main cluster. RsmH forms a separate cluster with its operon neighbours mwaZ and fisL (11,63), and similarly rsmE clusters with gshB (11,63). Ribosomal RNA modification genes that remained associated with other genes induced in an exponential growth phase are rluC, rlmH, rlmI, rsmB, rsmG, rlmN and rlmL.

**Phenotypes of knockout strains**

The availability of a comprehensive *E. coli* gene knockout collection (Keio) (79) made it possible to check gene influence on cell fitness under various stress conditions in high-throughput experiments (16). Phenotype profiles of knockout strains could be compared and this new measure could be used to find a functional relation between the genes. Analysis of gene knockout phenotype similarity revealed several genes whose knockout phenotypes resemble that for rRNA modification gene knockouts (Figure 4 and Supplementary File S4). Analysis of genes associated with rRNA modification genes according to phenotype profiling does not reveal any correlation with other data sets. Attention can be given to genes *potA*, *potI* and *mgtA* involved in transport of polyamines and magnesium ions (46). Both magnesium ions and polyamines associate with the ribosome and are required for the folding of rRNA (80). Genetic interaction between polyamine biosynthesis and rRNA modification was previously reported for KsgA rRNA methyltransferase and SpeD S-adenosylmethionine decarboxylase essential for spermidine biosynthesis (81). Double *ksgA/speD* knockout had a synergistic effect on kasugamicin resistance (81). Since KsgA is a switch protein involved in 30S subunit assembly, its functional link with the synthesis of essential rRNA counterions is logical. It should be noted, however, that phenotype profiling analysis does not link *ksgA* and *speD* (16), but other rRNA modification and polyamine-related genes (Figure 4).

Despite the low efficiency of phenotype profile comparison for mining of functional relations between rRNA
Modification genes and other genes, phenotype profiles themselves give some insight into the functional role of rRNA modification. Figure 5 and Supplementary File S5 presents a phenotype landscape of rRNA modification gene knockouts in a subset of conditions relevant for translation. Several results were expected and match previously published experimental data, namely increased susceptibility of the *rluC* knockout strain to peptidyltransferase-associated antibiotics (82), streptomycin resistance of the *rsmG* knockout strain (28) and extreme cold sensitivity of the *ksgA* knockout strain (68).

Modification of rRNA in some cases increases susceptibility in some other cases resistance to antibiotics. Not only the *rsmG* knockout strain, but also the *rsmC*, and to some extent, the *rsmF* knockout strain, were found to be resistant to streptomycin (16). Knockouts of *rsmD*, *rsmH* and *ksgA* make cells hypersensitive to tobramycin, whereas knockouts of *rluA* and *rlmI* lead to moderate tobramycin resistance (16). Extreme and moderate spectinomycin sensitivity were found for the *ksgA* and *rsmG* knockout strains, whereas *rsmH* and to some extent *rluC* and *rluB* knockouts lead to spectinomycin resistance (16).
High tetracycline sensitivity could be ascribed to the rsmD knockout strain, which is not surprising, because the tetracycline binding site is juxtaposed with a methylated nucleotide m^2G966 of 16S rRNA that is formed by RsmD methyltransferase (83, 84). Several strains defective in rRNA modification genes have altered sensitivity towards erythromycin. Strains devoid of rsmH, rlulA, rumA and rlmlI are hypersensitive, whereas knockouts of rsmE and rlue are moderately resistant to erythromycin. Susceptibility patterns for azitromycin and claritromycin resemble, but not perfectly match those for erythromycin (16).

Curiously, the rsmG knockout strain gains advantage in a number of conditions, such as the presence of EDTA, bicyclomycin and absence of a nitrogen source except ammonium chloride. Only in the presence of spectinomycin and a low concentration of chloramphenicol does the rsmG knockout have a disadvantage in comparison with the parental strain. It seems that the reason behind selective pressure to retain rsmG in bacterial genomes is not yet known.

Proteins co-purified with rRNA modification enzymes

Several massive co-purification studies to investigate E. coli interactome were done in recent years (14, 15). Data regarding rRNA modification enzymes were extracted from databases using the www.microbesonline.org search engine (12). Most likely, interactions of the rRNA modification enzymes are formed in the context of the small and large ribosomal subunit assembly intermediates. Modification enzymes acting on rRNA utilize a complete range of assembly intermediates starting with deproteinized rRNA (39, 45) and ending with associated 70S ribosomes (40, 41). Apart from serving solely for rRNA modification, these enzymes might serve as switch proteins for assembly checkpoints [see ref. (85) for discussion]. The best documented case of rRNA methyltransferase involvement in assembly is KsgA, serving to prevent immature 30S subunits to start translation (67–69).

Firstly, interactions between rRNA modification enzymes and ribosomal proteins were analysed (Figure 6A and Supplementary File S6). Ribosomal proteins were sorted on the basis of small and large ribosomal subunit assembly maps (86, 87). It was interesting to see if one can determine the protein composition of an assembly intermediate that is utilized by certain rRNA modification enzymes. The result was somewhat puzzling. Proteins that are exclusively co-purified with small subunit proteins are RsuA, RlmF and RlmH. RsuA is involved in the 30S subunit maturation (25), RlmH interacts with 70S ribosomes (41), whereas RlmF acts on the assembly intermediate of the large subunit (2). RlmL, 23S rRNA G2445 N^2- methyltransferase (45) is exclusively associated with the 50S protein L1 protein, which binds independently with other proteins and away from the RlmL modification site (88). A number of 16S rRNA modification enzymes (RsmG and RsmH) and 23S rRNA modification enzymes (RluB, RluC, RlmB, RlmN, RlmI, RrmJ) co-purify with ribosomal protein components of both subunits (14, 15). RluB, RluC and RluE, pseudouridine synthetases that modify residues in the peptidyltransferase centre form single complex.

Proteins that co-purify with ribosomal proteins and rRNA modification enzymes (Figure 6B, Supplementary Table S2 and Supplementary File S7) are likely candidates for assembly chaperones. Most probably, co-purification results reflect not binary protein–protein interactions but a separation of a complete assembly intermediate. This category is populated by ribonuclease Rnr (76), helicases DeaD (89) and SrmB (90), chaperones DnaJ and DnaK (91, 92), GTPase ObgE (93), GTPase activating protein YihI (94) and the RNA-binding proteins YhbY (95) and YibL (64) of yet unknown function. Somewhat
surprisingly, factors involved in interaction with mature ribosomes, such as IF2, IF3, EF-Tu, SelB and EF-G could also be found among interaction partners. In addition, aminoacyl-tRNA synthetases and proteins involved in protein export could be found as interacting with rRNA modification enzymes and ribosomal proteins.

It is possible that rRNA modification enzymes could still bind to completely assembled and functionally active ribosomes. Hypothetically, this could indicate possible involvement of rRNA modification enzymes not only in...
the assembly of the ribosome but also in post-assembly regulation of ribosomal function. Several DNA binding proteins could be found in the list. Most prominently, the HU protein was shown to interact with ribosomal proteins and as much as four rRNA modification enzymes. Nearly the entire set of cold shock protein family members co-purifies with rRNA modification enzymes and ribosomal proteins. Proteases and chaperones are also present in the set of co-purified proteins as well as a number of proteins with RNA degradation/processing and other RNA-related activities. NudH pyrophosphohydrolase initiates mRNA decay by removal of the triphosphate ‘cap’ from mRNA 5’-end (96). Co-purification of NudH (RppH) with ribosomal proteins and three rRNA modification enzymes indicates that it could act on rRNA as well as on mRNA. Similar conclusion could be made for HrpA helicase (97). A set of uncharacterized proteins were co-purified with high confidence with ribosomal proteins and several rRNA modification enzymes, putative methyltransferase YfIF, the RNA binding protein YbcJ, ATPase YjeE that was later shown to regulate tRNA N6-threonylcarbamoylation at adenosine residue (98), the predicted adenylate cyclase YgiF, conserved proteins YeeX and the putative DNA-binding protein YncE (99). These proteins are potential ribosome biogenesis factors.

From the set of proteins that co-purified solely with rRNA modification enzymes, one should also mention a complete set of subunits of the pyruvate dehydrogenase complex. This result might be considered artificial if not for the observation that three–five different rRNA modification enzymes independently co-purified with the pyruvate dehydrogenase complex. Jiang et al. (64) noticed large quantities of pyruvate dehydrogenase complex in their preparations of 50S subunit precursors, but they ascribed it to a co-migration of 50S subunits and multimeric pyruvate dehydrogenase on sucrose density centrifugation. Affinity purification (14,15) should not allow co-purification solely on the basis of physical dimensions. This interaction certainly deserves more detailed experimental investigation. After enolase was shown to be an integral part of the RNA degradosome (100), pyruvate dehydrogenase association with rRNA modification enzymes does not seem so strange.

More informative could be a comparison of proteins co-purified with rRNA modification enzymes and proteins whose genes are located close to the correspondent genes in the genomes. A number of putative functionally linked partners of rRNA modification enzymes could be identified in such a way: FtsY, FtsL, Rnr, DnaJ, YhbY and PrsA. FtsY is a signal recognition particle receptor, the protein involved in co-translational protein export (101). Ribonuclease R (Rnr) is involved in rRNA maturation (76), whereas chaperone DnaJ is involved in ribosome assembly (91,92). The role of protein YhbY in ribosome assembly is not clear yet, but its involvement in this process was demonstrated (64). Two proteins that were not yet shown to be involved in translation machinery function or assembly could be extracted from the list—the membrane-bound cell division protein FtsL and the phosphoribosylpyrophosphate synthase PrsA. Specific experiments are required to determine whether ribosome assembly/function is a secondary activity of these proteins or if it is an artefact of the performed experiments.

CONCLUSIONS

Analysis of genes/proteins associated with rRNA modification enzymes by phylogenetic profiling, genome neighbourhood, co-expression profiling, similarity in the phenotype profile and co-purification revealed a substantial number of putative partners. Many of them are undoubtedly false positives, i.e. proteins or genes accidentally linked to rRNA modification. The majority of linked genes/proteins belongs to the translation apparatus, but does not have any specific relations to the process of rRNA modification. However, in the list of putative functional interactions that are listed in this work, there are undoubtedly proteins functionally related to the modification of rRNA. Among the co-purification partners of rRNA modification enzymes, there are many proteins known to participate in rRNA maturation and ribosome assembly. Several proteins whose involvement in assembly or maturation of the ribosome was not demonstrated are still good candidates for experimental validation. In this category one would list NudH, HrpA, YfIF, YbcJ, YjeE, YgiF, YeeX and YncE. Association of the HU protein and the pyruvate dehydrogenase complex with assembled ribosomes might reflect some unknown function of these well-characterized proteins.

Phylogenetic profiling (co-occurrence) of rRNA modification enzymes revealed several interesting links: pseudo-uridine synthetases with RF2, RsrM with translation factors IF2, RF1 and LepA and RlmM with DNA-binding protein RdgC. The latter is most puzzling and deserves experimental examination.

Genome neighbourhood analysis revealed a number of genes predicted to be functionally linked to rRNA modification. Several genes of this subset belong to other rRNA maturation processes, like ybeB and rnr. Plenty of genome neighbourhood connections deserve experimental verification of their functional linkage. Specifically, the conserved involvement of rlmH gene into an operone coding for cell wall biosynthetic proteins and ksgA involvement into a
complex operon with a periplasmic peptidyl-prolyl isomerase (surA), AppppA hydrolase (apaH), the enzyme involved in vitamin B6 synthesis (pdxA) and a protein of unknown function (apaG) can bear further research. Several binary associations are also suggested to study experimentally, such as rsmD–yihL, rrmJ–ftsH, rluE–nudJ, rluD–yfiH, rlmL–upa and rlmM–ygD.

Comparative analysis of expression profiles through GEO revealed two main associations, the group of genes expressed during the fast growth, including rluC, rlmH, rlmI, rsmB, rsmG, rlmN and rlmL and the association of rrmJ with heat shock genes. The first association seems logical, since ribosome maturation should be especially effective during the logarithmic growth phase. The second association deserves special attention: are there any functional links between 23S rRNA U2552 ribose methylation and heat shock response?

The application of database mining in the search for potential functional links of rRNA modification enzymes allowed us to reveal many potential functional interaction partners. The results of this study might be used to plan further experimental verification of predicted functional interactions.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online: Supplementary Tables S1 and S2, Supplementary Cytoscape Files S1–S6.

ACKNOWLEDGEMENTS
The authors are very thankful to M. Gelfand for his efforts on improving the language of the article. His advice on database selection and to Alexander Lebedeff for his efforts on improving the language of the article.

FUNDING
Russian Foundation for Basic Research (grant numbers 10-04-01345-a, 11-04-01314-a, 11-04-01018-a and 11-04-12060-ofi); Russian Ministry of Science (grant 16.512.11.2108), Federal Agency for Science and Innovations (grant 02.740.11.0706); Moscow University Development Program (grant PNR 5.13). Funding for open access charge: Lomonosov Moscow State University.

Conflict of interest statement. None declared.

REFERENCES


