Evaluation of the bacterial ribosomal protein S1 gene (*rpsA*) from the position of structural repetition

Machulin A.[1]*, Deryusheva E.[2], Galzitskaya O.[3,4]*

[1] *Skryabin Institute of Biochemistry and Physiology of Microorganisms, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino , Russia*

[2] *Institute for Biological Instrumentation, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Russia*

[3] *Insitute of Protein Research, Pushchino, Russia*

[4] *Institute of Theoretical and Experimental Biophysics, Pushchino, Russia*

* *and.machul@gmail.com, ogalzit@vega.protres.ru*

Motivation and Aim: The family of bacterial ribosomal S1 proteins (*rpsA*) with repetitive S1 domains differs in a canonical sense from other similar proteins (**Figure 1**) [1]. The S1 protein is essential for cell viability because it interacts with both mRNA and proteins. Protein domain repeats are known to arise due to tandem duplications of internal genes. Analysis of the variability of base composition and tandem gene duplications in *rpsA* genes using a genomics-based approach can provide insight into the distinctive features and possible function and evolution of structural repetition.

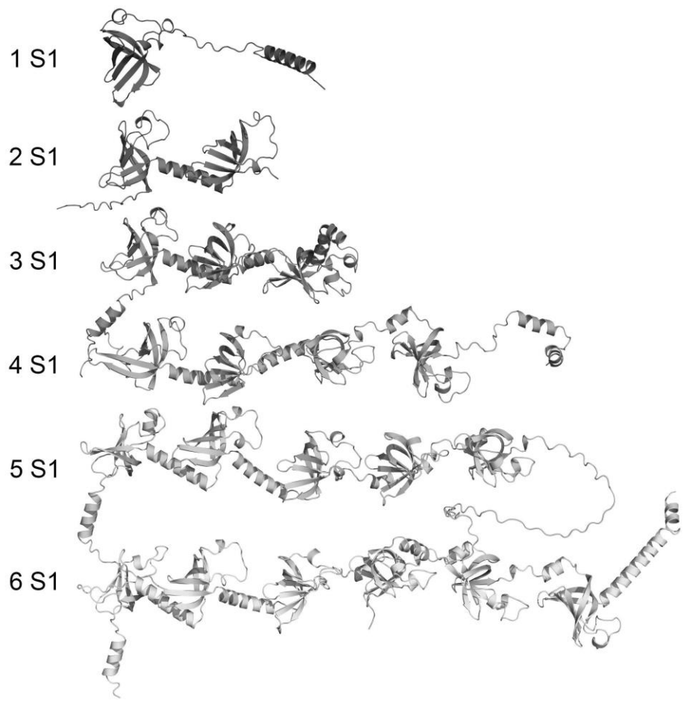

Figure 1. Representation of the different number of structural S1 domains in different bacteria.

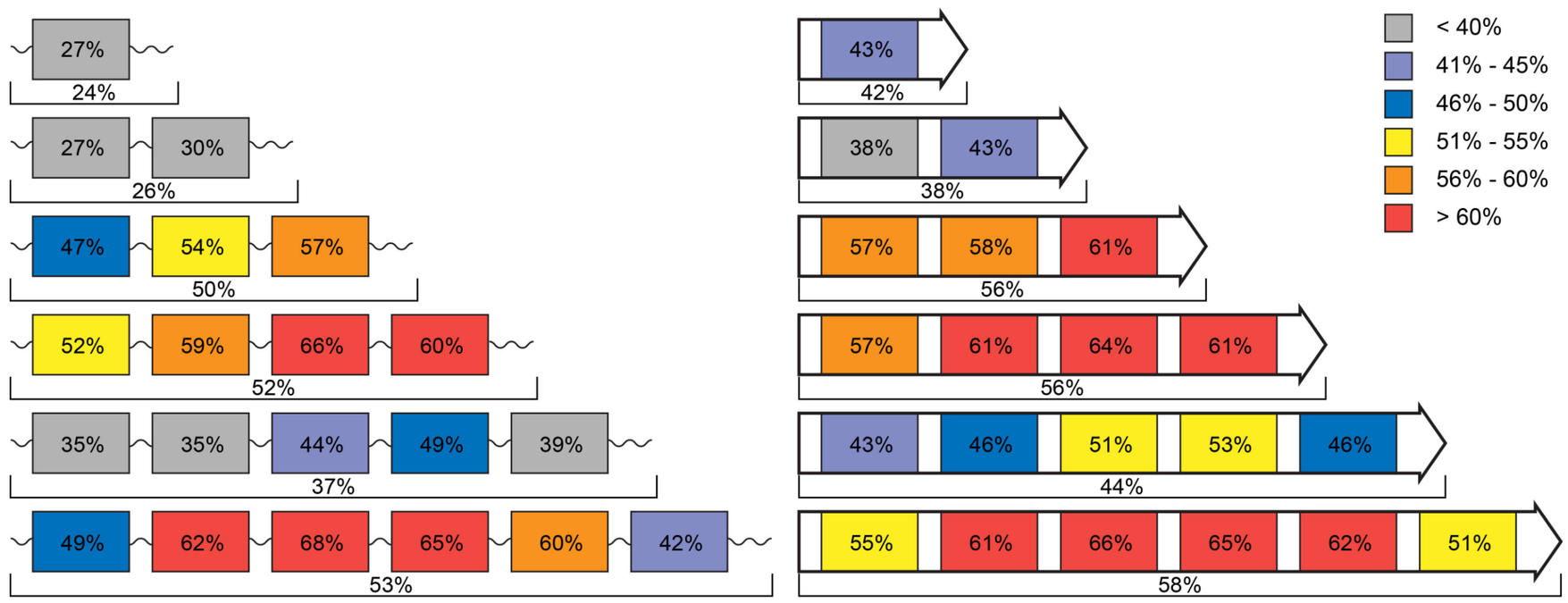

Figure 2. Conservatism of S1 domains within the family of 30S ribosomal protein (left) and conservatism of *rpsA* gene within the family of 30S riboso- mal protein S1 (right). Average percentage of identity within each domain as well as all domains in proteins containing different numbers of domains is given.

Methods and Algorithms: To study the mechanisms of repeat expansion, we studied 1324 *rpsA* sequences of S1 proteins with different number of S1 domains. Search, collection and analysis of the data were performed using a script implemented in PyCharm v.2018 software development environment, based on Python 3.3 language (https://www.python.org). The Multiple Sequence Alignment was implemented by the Clustal Omega service (https://www.ebi.ac.uk/Tools/msa/clustalo/). For codon nucleotide sequence alignment was done by MEGA Software using Clustal (codon) (https://www.megasoftware.net/). To evaluate sequence similarity MEGA pairwise distance analysis was used.

Results: Analysis of rpsA showed that the gene regions encoding individual S1 domains have no a strictly repetitive structure between groups containing different number of domains. The part of rpsA encoding the central domains in multidomain S1 proteins is more conserved than the terminal domains. The maximum value of rpsA identity for full-length proteins was found for S1 proteins containing six structural domains (58%) (**Figure 2**). The four- and six-domain S1 proteins predomi- nates. The regions of rpsA genes encoding adjacent domains are more identical. The regions of the gene encoding residues that form the RNA-binding site remain conserved (**Figure 3)**.

Conclusion: All the results obtained characterize the structural features of rpsA that determine the functioning of the S1 proteins. The portions of the rpsA gene encoding central domains in multidomain ribosomal S1 proteins are more conserved than the terminal domains what correlates with the assumption that duplication is found mostly in the central region of a protein chain between other repeats.

Figure 3. Consensus sequence of the *rpsA* gene in ribosomal S1 proteins. The consensus parts of the *rpsA* gene corresponding to the conserved residues on the surface of the S1 domain, which formed the RNA binding site, are highlighted in blue.

Also, analysis of datasets in [3] would suggest that simultaneous duplication of several domains underlies the formation of repetitive regions, while duplication of a single domain is rare. This fact may explain the predominance of four- and six-domain ribosomal S1 proteins. At the same time, the parts of rpsA genes encoding adjacent domains are more identical what correlates with the data suggesting that duplication occurs mainly in the middle of the protein chain between other repeats.

Acknowledgements: The study is supported by the Russian Science Foundation, grant number 18-14-00321 (Galzitskaya O.)

References

Deryusheva E.I. et al. Sequence and evolutionary analysis of bacterial ribosomal S1 proteins. Proteins. 2021;89:1111–1124.

Deryusheva E.I. et al. Structural, functional, and evolutionary characteristics of proteins with repeats. Mol. Biol. (Mosk.). 2021;55:748–775.

Björklund A.K. et al. Expansion of protein domain repeats. PLoS Comput. Biol. 2006; 2:e114.