In silico analysis of bacterial translation factors reveal distinct translation event specific pI values

SOMA JANA

Indian Institute of Science Education and Research Kolkata

Partha Pratim Datta (✉ partha_datta@iiserkol.ac.in)

"Indian Institute of Science Education and Research Kolkata" <https://orcid.org/0000-0002-3478-2616>

Research article

Keywords: Ribosome, Translation, Translation factors, Isoelectric point, Molecular Weight, Phylogeny

Posted Date: September 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-75549/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on March 29th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07472-x>.

Abstract

Background

Protein synthesis is a cellular process that takes place through the successive translation events within the ribosome with the help of the event-specific protein factors, namely, initiation, elongation, release, and recycling factors. The translation process is fundamental to all organisms living in the wide variety of environments. In this regard, we asked the questions about how similar are those translation factors to each other from a wide variety of bacteria? Hence, we did a thorough *in silico* study of the translation factors from 495 bacterial sp., and 4262 amino acid sequences, wherein we theoretically measured their pI and MW values that are the two determining factors for distinguishing individual proteins in 2D gel electrophoresis. Then we analyzed the output from various angles.

Results

Our study revealed that, not all the pI values are same or random, but there is a distinct order, such that the pI values of translation factors are translation event specific. We found that the translation initiation factors are mainly basic, whereas, elongation and release factors that interact with the inter-subunit space of the intact 70S ribosome during translation are strictly acidic. Further analysis revealed that the acidic property of those factors is due to the higher frequencies of glutamic acids. However, two translation factors, the translation initiation factor 2 (IF2) and the ribosome recycling factor (RRF) showed variable pI values. Remarkably, the variability of the pI values of these two factors showed distinct lineage with the order of phylogeny.

Conclusion

From our results we conclude that, among all the bacterial translation factors, elongation and release factors are more conserved in terms of their pI values in comparison to initiation and recycling factors. Acidic properties of these factors are independent of habitat, nature, or the phylogeny of the bacterial species. Furthermore; irrespective of the different shapes, sizes, and functions of the elongation and release factors, possession of their strictly acidic pI values indicate that the acidic nature of these factors is a necessary criterion, perhaps to interact into the partially enclosed rRNA rich inter-subunit space of the translating 70S ribosome.

Background

The translation is a complex universal biological process that takes place in a large macromolecular machine called ribosome in all living organisms. It is an energy exhaustive cellular process. In *Escherichia coli*, 40% of the total cellular energy is utilized by the translation system [1]. With the help of specific protein factors and aminoacyl tRNAs, ribosomes carry out protein synthesis following the decoding of the

genetic information from mRNA in successive events, namely, initiation, elongation, and termination (release and recycling). The protein factors that are involved in the successive events are initiation factors (IF), elongation factors (EF), release factors (RF), and ribosome recycling factors (RRF). Here, the accurate coordination of every participant protein factors is necessary to perform the process successfully. Based on several years of biochemical and structural biological studies worldwide, a fairly detailed knowledge of the mechanisms of cellular protein synthesis is now known [2]. However, in the broad aspects, which characteristics of the translation factors i.e., IF, EF and RF are necessary to be conserved for the accuracy of the universal process of protein synthesis among the different kind's organisms needs to be investigated.

For this study, our concept was to comprehend the fundamental properties of the universally conserved process of protein synthesis by studying the acidic and basic properties of the translation factors involved in the reactions. For this, we needed to consider some common distinguishing property among these translation factors. Hence, we made use of the principle of the 2D gel electrophoresis [3], whereby we computationally calculated isoelectric point (pI) and molecular weight (MW) values of all the translation factors (this study), and used those calculated values to decipher how similar or dissimilar the translation factors are from each other's.

Results And Discussion

We studied 495 bacterial species throughout the domain of bacteria. The habitat of these bacteria is very different from each other. The nature of these bacteria in terms of cell shape and size (coccus or bacillus), intracellular metabolic reactions (aerobic or anaerobic), and even the way they respond to the external environments (mesophilic or thermophilic or psychrophilic) are distinct [4]. Here, we considered the following bacterial Phyla: Deinococcus-Thermus, Chlorobi, Actinobacteria, Firmicutes, Chlamydiae, Fusobacteria, Spirochaetes, Chloroflexi, Tenericutes, Cyanobacteria, Bacteroidetes, Thermotogae, Acidobacteria, Aquificae, Caldiseae, Chrysiogenetes, Deferribacteres, Elusimicrobia, Fibrobacteres, Gemmatimonadetes, Lentisphaerae, Nitrospirae, Planctomycetes, Thermodesulfobacteria, Verrucomicrobia, and Proteobacteria for our study [5].

pI and molecular weight value distributions of translation protein factors

In the case of translation (Additional file 1), we found a unique pattern of pI value distribution as depicted in Figure 1A. Here, except for some bacterial species in the case of the initiation factor 2 (IF2), the other bacterial initiation factors were basic. On the other hand, the elongation and release factors were highly acidic except the ribosome-recycling factor (RRF). All the four quartiles of initiation factor 1 (IF1) and initiation factor 3 (IF3) were above pI 7. On the other hand, IF2 had more than one quartile of basic pI values. The elongation factor Tu (EF-Tu), elongation factor G (EF-G), elongation factor 4 (EF-4), & elongation factor P (EF-P) and the release factor 1 (RF1), release factor 2 (RF2), & release factor 3 (RF3) had all the four quartiles in the acidic range. On the other hand, IF2, and RRF had more than one quartile in basic pI ranges. For the comprehensive *in silico* study, along with the pI values, we also studied the

molecular-weight (MW) value distributions of these translation protein factors (Figure 1B). Like pI value distribution, the protein IF2 showed a wide range of variations in MW value distributions also (Figure 1B). All the other proteins showed precise MW value distributions. A surprising observation is to be noted here that although RRF proteins showed highly variable pI value distributions, their MW values were quite fixed.

Statistical analysis of pI values of translation proteins

We further performed asymptotic tests [6] for 5% quantile and 95% quantile (Table 1) of these translation factors. We found that the p values corresponding to the null hypotheses ($H_0: q_{05} \geq 7$, and $H_0: q_{95} \leq 9.95$) for the 5% and 95% quantiles, respectively, for both the initiation factors, IF1 and IF3 to be more than 0.05, from which we inferred that 90% data lies in basic pI values, i.e., between 7 to 9.95. On the contrary, in the case of elongation (EF-Tu, EF-G, EF-4, and EF-P) and release factors (RF1, RF2, and RF3), 90% data lies in completely acidic pI values i.e., between 4.635 and 6.225 (p values corresponding to $H_0: q_{05} \geq 4.635$ and $H_0: q_{95} \leq 6.225$ turned out to be more than 0.05, respectively). But we found a different scenario in the case of initiation factor, IF2, and ribosome recycling factor, RRF. In both these cases, 90% of data stretches in between acidic 5.1 to basic 9.25 (p values are more than 0.05 for $H_0: q_{05} \geq 5.1$ and $H_0: q_{95} \leq 9.25$, respectively).

Table 1. Asymptotic tests for 5% quantile and 95% quantile for translation factors.

Translation Factors (IF1 & IF3)	5% sample quantiles	p-values ($H_0: q_{05} \geq 7$)	95% sample quantiles	p-values ($H_0: q_{95} \leq 9.95$)
IF1	6.82	0.2839	9.98	0.1684
IF3	8.87	0.5	9.94	0.6444
Translational Factors (IF2)	5% sample quantiles	p-values ($H_0: q_{05} \geq 5.1$)	95% sample quantiles	p-values ($H_0: q_{95} \leq 9.25$)
IF2	5.09	0.4173	9.31	0.1198
Translational Factors (Elongation and Release Factors)	5% sample quantiles	p-values ($H_0: q_{05} \geq 4.635$)	95% sample quantiles	p-values ($H_0: q_{95} \leq 6.225$)
EF-Tu	4.81	0.9861	5.77	0.7839
EF-G	4.78	0.9573	5.54	0.7272
EF-4	4.99	0.5337	6.36	0.0909
EF-P	4.77	0.9978	5.75	0.5473
RF1	4.81	0.9746	6.06	0.8042
RF2	4.62	0.1005	5.47	0.5
RF3	5.03	0.6555	6.14	0.8347
Translational Factors (RRF)	5% sample quantiles	p-values ($H_0: q_{05} \geq 5.1$)	95% sample quantiles	p-values ($H_0: q_{95} \leq 9.25$)
RRF	5.063	0.1999	9.028	0.9905

Amino acid frequency distributions of elongation and release factors

Interestingly, when we randomly chose 60 amino acid sequences (represents 60 bacterial species) of each of the elongation and release factors and calculated their amino acid frequencies, we found the occurrence of a high frequency of glutamic acid in all of those factors, (Figure 2A-G). In 2001, Schwartz et al. [7] also observed that the cytosolic acidic proteins also found to have a high frequency of glutamic acid

Relation of pI values of IF2 and RRF proteins with phylogeny

Since IF2 had a wide range of pI value distribution from acidic to basic, we performed phylogenetic analysis (Figure 3A) of the IF2 proteins (Additional file 1) to investigate the relation of its pI value distribution with the phylogeny. In the case of the Phylum Proteobacteria, we found that the class of Gammaproteobacteria (blue) and Betaproteobacteria (verdigris) were acidic (with only a few exceptions). Whereas the class Alphaproteobacteria (brown) had few genera as acidic (i.e., Ehrlichia) and some genus as basic (i.e., Brucella and Bartonella) and others had both acidic and basic (i.e., Rickettsia) pI values. In the case of other Phyla, Chlorobi (cyan), Cyanobacteria (red), Thermotogae (yellow), and Deinococcus-Thermus (light grey), they had mostly acidic pI values, whereas the Chlamydiae (Safron) and Spirochaetes (light green) had basic pI values. The pI values of the IF2 protein in Phylum Firmicutes (pink) and Actinobacteria (light blue) and Tenericutes (purple) had both the acidic and basic pI values.

The phylogenetic analysis of RRF (which had a wide range of pI value distribution also) showed that the pI value distribution of RRF (Figure 3B) (Additional file 1) like IF2 (Figure 3A) also linked to the phylogeny. We found that different classes of Proteobacteria had different pI value distributions. The Gammaproteobacteria (blue), and Alphaproteobacteria (brown) (with a few exceptions e.g. Genus; Salmonella of Gammaproteobacteria and Genus; Rickettsia and Ehrlichia of Alphaproteobacteria) had acidic pI values. However, Betaproteobacteria (verdigris) (i.e., Bordetella - acidic, Burkholderia - basic) and Deltaproteobacteria (apple green) (i.e., Desulfococcus – acidic, Geobacter – basic,) had acidic and basic pI values as well. In the case of other Phyla, Chlamydiae (Safron), Chlorobi (cyan), and Spirochaetes (light green), they had basic pI values. In contrast, Phylum, Actinobacteria (light blue), and the Phylum Firmicutes (pink) had both the acidic and basic pI values.

Conclusion

Along with the translation factors, we did additional studies on the pI value distributions of the protein factors of two other universal processes of central dogma i.e., replication and transcription processes in domain bacteria (Additional files 2 and 3, respectively). For that we studied 529 bacterial sp., and 1707 amino acid sequences for the replication factors, and 488 bacterial sp., and 1998 amino acid sequences for transcription factors, respectively. Some of the proteins involved in replication and transcription showed a narrow range of pI value distributions, and some showed a wide range (Additional file 2 and 3 respectively). Unlike translation factors, we found no specific pattern of pI value distributions of the proteins involved in the individual steps of the initiation, elongation, and release in those two processes. So, in conclusion, the observation of our study of the precise pI value distribution of the translation factors throughout the domain bacteria indicates that the overall acidity or basicity of translation factors is an essential feature in the process of translation. Our study found a unique pattern of pI value distributions of translation protein factors in domain bacteria. The proteins involved in the initiation event i.e., initiation factors, were basic, whereas elongation and release factors were strictly acidic due to the high frequency of negatively charged amino acids, like glutamic acids (Fig. 2A-G). If we focus on the mode of interaction of these factors with the ribosome, we can categorize the facts i.e., initiation factors, IF1, IF2, and IF3 are involved in the formation of the 30S initiation complex, which is an open complex. On the other hand, during elongation, the elongation factors interact at the inter subunit space of the

relatively closed translating 70S complex that is composed of 30S and 50S ribosomal subunits plus mRNA and tRNAs. Both the elongation and release factors, irrespective of these protein's different shapes, sizes, and functions (Fig. 4) interact with the A site of the semi-enclosed inter-subunit space of the translating 70S ribosome. If we try to find out the reasons for our observation, we can focus on the important fact i.e., the process of initiation of translation takes some seconds [8–10] to assemble ribosome on the mRNA with the accordance of initiation factors but the elongation process happens at a faster rate than initiation. Several amino acids are incorporated within a second [8–10] and it continues until the whole mRNA gets read and stop codon appears. The criteria of the interaction of these translation factors into the inter-subunit space (A-site) of the ribosomal subunits maybe like that the proteins (elongation and release factors) must be acidic to make the process happen most accurately along with the faster rate. The observation of the strictly acidic pI value distributions of the two other auxiliary elongation factors, EF-P and EF-4 (which also interacts with the inter-subunit space of the ribosomal subunits during the translation elongation cycles at certain conditions) also support our conclusion (Fig. 1A). The wide range of pI value distributions of RRF reveals that the conservedness of the acidic and basic properties of this translation factor is not as important as the other translation factors in bacteria.

In this study we took into account of a wide range of bacterial species that belong to the entire domain of bacteria on earth. For the sake of survival, bacteria evolve numerous mechanisms to adapt to that environment. The habitat of these bacteria vary widely, viz: soil, water, food, industrial waste, deep ocean, acidic hot springs, in symbiotic and parasitic relationships with animals and plants, and radioactive waste also [11]. The nature of these bacteria are also different (i.e., acidophiles, alkaliphiles, aerobic, anaerobic, phototrophs, chemotrophs, nitrogen-fixing bacteria, nitrifying and denitrifying bacteria, bioluminescent bacteria, free-living bacteria, enteric bacteria, and obligate intracellular parasites) [11]. Irrespective of the wide range of phylogeny, habitat, and nature of these bacteria, our statistical test showed that except IF2 and RRF, all the initiation, elongation, and release factors are conserved in terms of pI values all over the domain bacteria.

In the case of IF1 and IF3, basic pI value distributions all over the domain bacteria revealed that the pI value is not dependant on the phylogeny, nature, or habitat of the bacterial species. Similar to these two initiation factors, acidic pI value distributions of the elongation and release factors throughout the domain bacteria indicated that the pI values of those translation factors are also not affected by phylogeny, nature, or habitat. The wide range of pI value distributions of IF2 and RRF (Fig. 3A and 3B respectively) unveiled that different phylum of bacteria had different traits of pI value distributions.

We conclude our study with a pictorial description of our findings in Fig. 4, where we depict the mean pI value distributions of the translation factors in bacteria that show distinct translation event specificity.

Figure 4: Graphical representation of translating intact 70S ribosome and initiation, elongation, release, and recycling factors of the studied bacterial species along with the mean pI value and standard deviation value. Red and blue color referred to the acidic and basic mean pI value, respectively. The dark

grey color of IF2 and RRF represented the fact that these factors had mean pI values close to neutral and had a higher standard deviation value than others.

Methods

Data collection

We studied the following translation factors viz., IF1, IF2, IF3, EF-Tu, EF-G, EF-4, EF-P, RF1, RF2, RF3, and RRF from bacteria that directly interact with ribosome. Between the reviewed and unreviewed categories of the protein sequences of the UniProt [12] database, we collected the reviewed only for the accuracy of sample data. We removed all the incomplete fragments of amino acid sequences to circumvent erroneous assumptions.

We calculated pI and MW values from 4262 reviewed amino acid sequences (Additional file 1) of the bacterial translation factors, and those values and the corresponding accession numbers of those proteins have been provided with the Additional file 1.

Method of pI value and MW value calculation

We used the "Compute pI/MW tool" (http://web.expasy.org/compute_pi/) of the ExPaSy-Bioinformatic resource portal to calculate the pI value and MW value. We chose this "Compute pI/MW tool" webserver because it shows reasonable good agreements of the calculated pI values with the experimentally determined pI values [13-15].

Statistical tests

We performed the asymptotic test [6] for the translation proteins for 5% quantile and 95% quantile. We calculated the p values corresponding to the null hypotheses for 5% and 95% quantiles for the translation proteins in MATLAB (R2019b) software. We generated all the graphs of this study in OriginPro 8.5 software (Origin (Pro), "Version 2019b") [16].

Phylogenetic analysis

We used MEGA7 software [17-18] to investigate the distribution of the pI values of IF2 and RRF protein in the bacterial taxonomy. We used all the available primary amino acid sequences to construct the phylogenetic trees in both the cases of IF2 and RRF protein. We used 500 bootstrap replicates to analyze the phylogenetic tree, and we presented here the tree having the highest log-likelihood.

Declarations

Notes: The authors declare no competing financial interest.

Author contributions: Concept, planning, design, analysis and manuscript writing by PPD. Inventory, data mining, execution, analysis, figures and manuscript writing by Soma Jana.

Funding sources: None.

Declaration of competing interest: The authors declare no competing interest.

Acknowledgements: We thank Dr. Satyaki Mazumder, Tiasha Saha Roy and Avishek Chatterjee for helping us in data execution. We are also thankful to Ananya Chatterjee, Ajaynath E, Saurav Rohilla, and Sukanya Bhowmik for their help in data mining.

Data availability: All data supporting the findings in this study are provided in the main text and in “Additional data file”.

References

1. Wilson DN, Nierhaus KH: The weird and wonderful world of bacterial ribosome regulation. *Crit Rev Biochem Mol Biol* 2007, 42:187-219.
2. Ramakrishnan V: Ribosome structure and the mechanism of translation. *Cell* 2002, 108:557-572.
3. O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975, 250:4007-4021.
4. Prescott LM, Harley GP, Klein DE, Microbiology, W. Brown Publishers, Dubuque, Iowa, USA, 2nd edition, 1993.
5. Michael M. Brock Biology of Microorganisms. San Francisco: Pearson/Benjamin Cummings. ISBN 978-0-13-232460-1.
6. Serfling RJ: Approximation Theorems of Mathematical Statistics. 1980.
7. Schwartz R, Ting CS, King J: Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome. Res* 2001, 11:703–709.
8. Wintermeyer W, Peske F, Beringer M, Gromadski KB, Savelsbergh A, Rodnina MV: Mechanisms of elongation on the ribosome: Dynamics of a macromolecular machine. *Biochem Soc Trans* 2004, 32:733-737.
9. Lovmar M, Ehrenberg M: Rate, accuracy and cost of ribosomes in bacterial cells. *Biochemie* 2006, 88:951-961.
10. Rodnina MV, Beringer M, Wintermeyer W: How ribosomes make peptide bonds. *Trends Biochem. Sci* 2007, 32:20-26.
11. Gupta A, Gupta R, Singh RL: Microbes and environment. In *Principles and Applications of Environmental Biotechnology for a Sustainable Future* 2017, 43–84.
12. Pichler K, Warner K, Magrane M: UniProt Consortium, SPIN: Submitting Sequences Determined at Protein Level to UniProt *Curr. Protoc* 2018, Bioinformatics. 62:52.

13. Bjellqvist B, Hughe GJ, Pasquali Ch, Paquet N, Ravier F, Sanchez J-Ch, Frutiger S, Hochstrasser DF: The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 1993, 14:1023-1031.
14. Bjellqvist B, Basse B, Olsen EJ, Celis E: Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions *Electrophoresis* 1994, 15:529-539.
15. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A: Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press. 2005
16. Origin (Pro), "Version 2019b". OriginLab Corporation, Northampton, MA, USA.
17. Jones DT, Taylor WR, Thornton JM: The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 1992, 8:275-282.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 2013, 30:2725-2729.

Figures

A

B

Figure 1

Box plot diagram of pI value (A) and molecular weight value (B) of the protein factors in the process of translation in bacteria. In the box plot diagrams, the lower hinge showed the first quartile (25%), whereas the upper hinge represented the third quartile (75%). The sign (-) above and below the box diagrams represented the maximum and minimum values respectively. The upper and lower solid triangle

represented 99% and 1% values of the data set respectively. The vertical line and the box inside the box plot represented the median and mean values of samples respectively.

Figure 2

Individual amino acid frequencies of EF-Tu (A), EF-G (B), EF-4 (C), EF-P (D), RF1 (E), RF2 (F), and RF3 (G). The X-axis is referred to as the individual amino acid of translation elongation and release factors. Y-axis is referred to as the amino acid frequencies. We depicted here the single amino acid frequencies from the

randomly selected 60 amino acid sequences of the elongation and release factors. Here, each color represented each randomly selected bacterial species.

A

B

Figure 3

Phylogenetic analysis based on the primary amino acid sequences of IF2 (A) and RRF (B) proteins of reviewed bacterial species. We analyzed the evolutionary history using the Maximum Likelihood method based on the JTT matrix-based model. We took 500 bootstrap replicates to build the phylogenetic tree.

Numbers near the branches refer to the bootstrap percentages (greater than 50% bootstrap replicates only shown here). The tree had been drawn to scale after eliminating all positions containing gaps and missing data. The branch lengths were measured by the number of substitutions per site. Blue triangles and red circles refer to the basic and acidic pI values respectively.

Figure 4

Graphical representation of translating intact 70S ribosome and initiation, elongation, release, and recycling factors of the studied bacterial species along with the mean pI value and standard deviation value. Red and blue color referred to the acidic and basic mean pI value, respectively. The dark grey color of IF2 and RRF represented the fact that these factors had mean pI values close to neutral and had a higher standard deviation value than others.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [JanaDattaAdditionalfile.doc](#)