Theoretical ribosomal protein mass fingerprint database

Wenfa Ng

Unaffiliated researcher, Singapore, Email: ngwenfa771@hotmail.com

Abstract

Ribosomes are highly conserved macromolecular machines whose critical function is protein synthesis. However, existence of unique molecular mass of the same type of ribosomal protein for individual species in the same domain of life raises the interesting question concerning the interaction between natural selection forces and the conservation of structure and function of ribosomal proteins. Thus, given differentiated molecular mass and sequence of ribosomal proteins across species, the structures of ribosomes are correspondingly differentiated even though the general structure and function of the macromolecular machine is conserved across species in the same domain of life. The collection of molecular mass of all ribosomal proteins in the large and small ribosome subunits can be understood as the ribosomal protein mass fingerprint of the species useful for gaining fundamental knowledge of ribosomal proteins, as well as serving as tools for species identification through comparison of ribosomal protein mass spectra. This preprint introduces the Theoretical Ribosomal Protein Mass Fingerprint database that comprises the theoretical molecular mass of all ribosomal proteins of a species calculated based on available amino acid sequence information of the ribosomal proteins. Using amino acid sequence information from the Ribosomal Protein Gene Database, the Theoretical Ribosomal Protein Mass Fingerprint database (https://ngwenfa.wordpress.com/database/) spans species from cyanobacteria, fungus, bacteria, archaea, nematodes, diatoms, micro-algae, and various model organisms. The database should be useful as a resource for gaining fundamental understanding of the mass distribution of ribosomal proteins of a species, or serving as a limited reference database for identifying species based on comparing experimental ribosomal protein mass fingerprint of unknown species against theoretically calculated ones of known species. Future expansion of the database will aim to catalogue the theoretical ribosomal protein mass fingerprint of more microbial species using amino acid sequence information from UniProt.

Keywords: ribosomal proteins, ribosome, molecular mass, biological conservation, natural selection, structure/function, ribosomal protein mass fingerprint, mass spectrometry, mass spectra, microbial identification,

Subject areas: biochemistry, biotechnology, ecology, biodiversity, bioinformatics,

Introduction

Ribosomes are essential to protein synthesis; thus, from the evolutionary perspective, the structure and function of ribosomes should be highly conserved across species in the three domains of life. This is essentially true except for observations of domain-specific structure of ribosomes of species in bacterial, archaeal and eukarya domain. However, structure and function of ribosomes

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26878v1 | CC BY 4.0 Open Access | rec: 22 Apr 2018, publ: 23 Apr 2018

of species in the same domain are highly conserved. Given that the ribosome comprises a large compendium of ribosomal proteins, such ribosomal proteins should also be highly conserved in amino acid sequence and structure.

However, recent theoretical calculations of molecular mass of ribosomal protein of species from all three domains of life revealed the existence of unique molecular mass of the same type of ribosomal protein from individual species in the same domain of life.[1] Since unique molecular mass suggests distinctive amino acid sequence and structure, the same type of ribosomal proteins of different species should thus be distinct from each other. Aggregating over the ensemble of ribosomal proteins that constitute the ribosome, there should be small differences to the structure and function of ribosomes even though the general structure and function of the critical macromolecular complex should be highly conserved.

Since ribosomes from different species are differentiated in structure and function due to the presence of a unique set of ribosomal proteins of distinctive mass distribution,[1] opportunities arise for using the set of ribosomal proteins as species identifiers during species identification.[2] Specifically, presence of unique sets of ribosomal proteins in each species meant that a distinctive ribosomal protein mass fingerprint exists for each species, which could inform species identification efforts given that tools for profiling such a mass fingerprint exists.[2] For example, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) could be utilized in gel-free mass spectrometric profiling of solubilized ribosomal proteins.[2 3 4 5] Such preparation of ribosomal proteins could be achieved via a two-step approach of density gradient centrifugation for isolating ribosomes, followed by the addition of chemical agents for solubilizing the ribosomal proteins.[2] Obtained mass spectrum of ribosomal proteins of a species could be compared with those of known species catalogued in a reference database for identification purposes.[2]

To support such an effort at species identification coupled with the goal of providing fundamental knowledge of the ribosomal protein mass distribution of individual species, this study sought to generate a theoretical ribosomal protein mass fingerprint database whose data are available free online. Specifically, the set of ribosomal protein and their amino acid sequence were obtained from the Ribosomal Protein Gene Database (http://ribosome.med.miyazaki-u.ac.jp/).[6] Molecular weight of the ribosomal proteins was calculated using the Compute pI/Mw tool at (https://web.expasy.org/compute_pi/). Relevant information was encapsulated in Excel files ready for download at the database site for the user (https://ngwenfa.wordpress.com/database/).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26878v1 | CC BY 4.0 Open Access | rec: 22 Apr 2018, publ: 23 Apr 2018

Methods

The Theoretical Ribosomal Protein Mass Fingerprint database was based on the Ribosomal Protein Gene Database (http://ribosome.med.miyazaki-u.ac.jp/). Specifically, amino acid sequence of individual ribosomal proteins were obtained for specific organism, and serve as input to an online molecular weight calculator (Compute pI/Mw Tool, https://web.expasy.org/compute_pi/). Molecular weight of the ribosomal proteins was calculated and collated into a theoretical ribosomal protein mass fingerprint of a species.

Results and Discussion

Collated theoretical ribosomal protein mass fingerprint of individual species should be useful as a reference for understanding the experimental ribosomal protein mass spectrum of a species. Thus, the database of theoretical ribosomal protein mass fingerprints of species in the three domains of life should find use in enabling the understanding of experimental mass spectra of ribosomal protein profiling exercise. Such experiments typically would generate many mass spectra from ribosomal proteins of a single or different species. In the case of single species, the theoretical ribosomal protein mass fingerprint could serve as a useful check for the reliability and performance of the MALDI-TOF mass spectrometer in enabling ribosomal protein profiling. The theoretical data obtained could also be useful in gaining fundamental knowledge of ribosomal proteins mass distribution of a species. Finally, the theoretical ribosomal protein mass fingerprint database could serve as a reference database against which obtained experimental ribosomal protein mass spectrum could be compared for identification of possible species. Though the types of species profiled in the theoretical ribosomal protein mass fingerprint database is currently limited, it nevertheless span important species in the three domains of life such as bacteria, archaea, cyanobacteria, fungus, diatoms, nematodes, eukaryotic microbial species and various model organisms. Thus, the database should be useful either for understanding the mass distribution of ribosomal proteins of a species, or as a reference database for identifying species based on a comparison of obtained experimental ribosomal protein mass fingerprint with those of theoretically calculated ones in the database.

Conclusions

Ribosomal proteins differ in molecular mass for the same protein across different species in the three domains of life. Thus, the collection of molecular mass of different ribosomal protein of a species could constitute a ribosomal protein mass fingerprint unique to individual species, which offers possibilities for its use in species identification. Specifically, experimentally derived ribosomal protein mass fingerprints from mass spectrometry studies could be compared against those profiled in a theoretical ribosomal protein mass fingerprint database. Using amino acid sequence information from the Ribosomal Protein Gene Database, and an online molecular weight calculator, the theoretical ribosomal protein mass fingerprints of many species from the three domains of life were calculated and profiled in an online open access database, the Theoretical

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26878v1 | CC BY 4.0 Open Access | rec: 22 Apr 2018, publ: 23 Apr 2018

Ribosomal Protein Mass Fingerprint database. The database should enable greater understanding of the molecular mass distribution of the ensemble of ribosomal proteins of individual species, as well as serve as a quality check for experimental ribosomal protein mass fingerprints obtained by MALDI-TOF MS profiling of solubilized ribosomal proteins. Finally, the database could serve as a reference for identifying the species provenance of experimentally derived ribosomal protein mass fingerprint of individual species. But the limited species coverage of the database may hamper species identification. Thus, the database will be expanded progressively in future to catalogue more bacteria and archaea species that should aid the development of microbial identification based on comparison of ribosomal protein mass spectrum fingerprint between unknown species and known ones with a theoretically calculated ribosomal protein mass fingerprint. Amino acid sequence information of species for the stated expansion will be obtained from UniProt.

References

1. Ng, W. Existence of theoretical ribosomal protein mass fingerprints in bacteria, archaea and eukaryotes. *PeerJ Prepr.* **6,** e26511v1 (2018).
2. Ng, W. Theoretical ribosomal protein mass distribution of *Pseudomonas aeruginosa* PAO1. *PeerJ Prepr.* **6,** e3500v1 (2018).
3. Nakamura, S. *et al.* Ribosomal subunit protein typing using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) for the identification and discrimination of *Aspergillus* species. *BMC Microbiol.* **17,** 100 (2017).
4. Hotta, Y., Sato, H., Hosoda, A. & Tamura, H. MALDI-TOF MS analysis of ribosomal proteins coded in *S10* and *spc* operons rapidly classified the *Sphingomonadaceae* as alkylphenol polyethoxylate-degrading bacteria from the environment. *FEMS Microbiol. Lett.* **330,** 23–29 (2012).
5. Nakamura, S., Sato, H., Tanaka, R. & Yaguchi, T. Verification of ribosomal proteins of *Aspergillus fumigatus* for use as biomarkers in MALDI-TOF MS identification. *Mass Spectrom.* **5,** A0049–A0049 (2016).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26878v1 | CC BY 4.0 Open Access | rec: 22 Apr 2018, publ: 23 Apr 2018

6. Nakao, A., Yoshihama, M. & Kenmochi, N. RPG: the Ribosomal Protein Gene database. *Nucleic Acids Res.* **32,** D168–D170 (2004).

Conflicts of interest

The author declares no conflicts of interest.

Author's contribution

The author conceived the idea, collected and analysed the data, and wrote the manuscript.

Funding

No funding was used in this work.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26878v1 | CC BY 4.0 Open Access | rec: 22 Apr 2018, publ: 23 Apr 2018