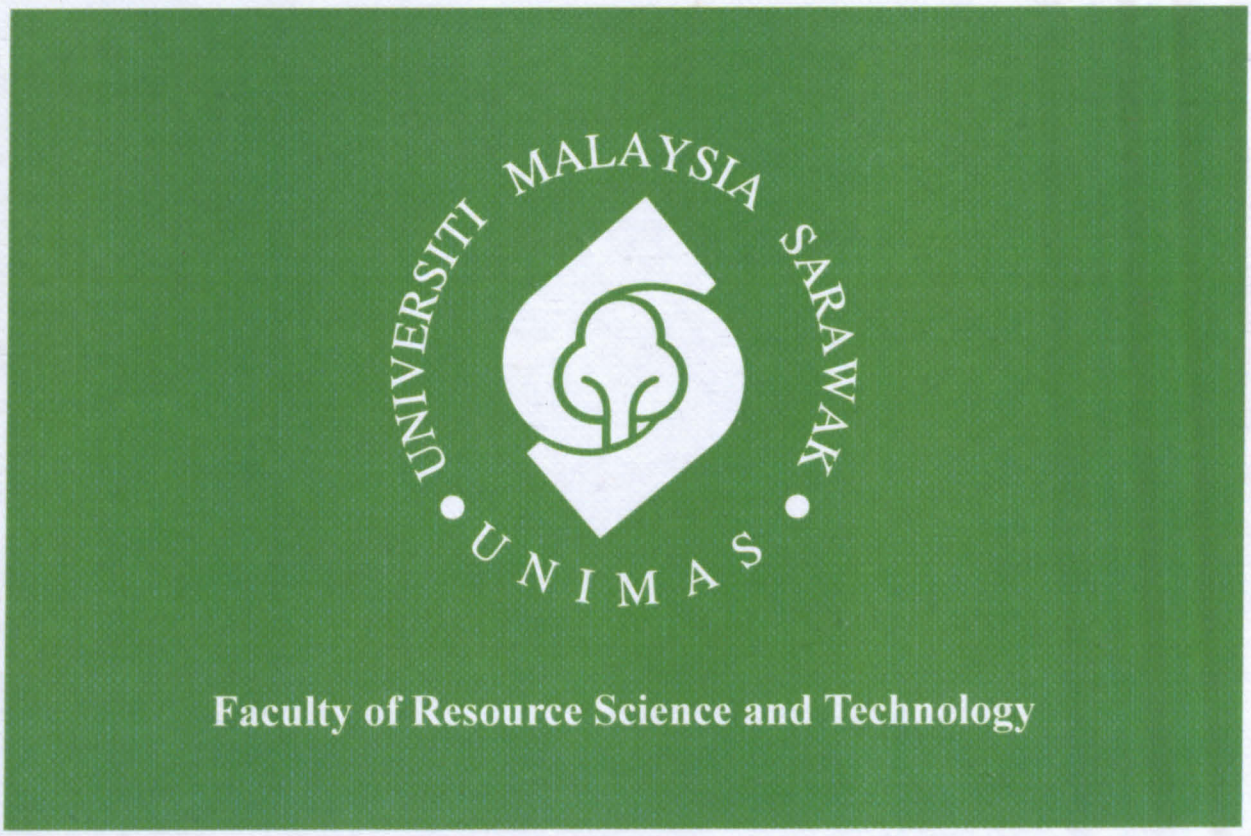

Identification of Protein Partners of the Human Ribosomal Protein, uS19 Using Bioinformatics Analysis

Daphne Olivia Jawai Anak Sadai

(50455)

Bachelor of Science with Honours
(Resource Biotechnology)
2018

Identification of Protein Partners of the Human Ribosomal Protein, uS19 Using Bioinformatics Analysis

Daphne Olivia Jawai Anak Sadai (50455)

This subject report is submitted

in fulfilment of the Final Year Project (STF 3013)

Resource Biotechnology

Supervisor: Dr. Edmund Sim Ui Hang

Resource Biotechnology

Faculty of Resource Science and Technology
University Malaysia Sarawak
2018

DECLARATION

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report [✓]

Masters [ ]

PhD [ ]

DECLARATION OF ORIGINAL WORK

This declaration is made on the 11th day of JUNE year 2018

Student's Declaration:

I DAPHNE OLIVIA JAWAI ANAK JADAI, 50455, FACULTY OF RESOURCE SCIENCE AND TECHNOLOGY

(PLEASE INDICATE NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, IDENTIFICATION OF PROTEIN PARTNERS OF THE HUMAN RIBOSOMAL PROTEIN, USI9 USING BIOINFORMATICS ANALYSIS is my original work. I have not copied from any other students' work or from any other sources with the exception where due reference or acknowledgement is made explicitly in the text, nor has any part of the work been written for me by another person.

12/06/18

Date submitted

Name of the student (Matric No.)
DAPHNE OLIVIA JAWAI
ANAK JADAI (50455)

Supervisor's Declaration:

I, ASSOCIATE PROFESSOR DR EDMUND SIM UI HANG (SUPERVISOR'S NAME), hereby certify that the work entitled, IDENTIFICATION OF PROTEIN PARTNERS OF THE HUMAN RIBOSOMAL PROTEIN, USI9 USING BIOINFORMATICS ANALYSIS (TITLE) was prepared by the aforementioned or above mentioned student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of BACHELOR OF SCIENCE WITH HONOURS (PLEASE INDICATE THE DEGREE TITLE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: _____ Date: 12-6-18

(Name of the supervisor)

Prof. Madya Dr. Edmund Sim Ui Hang
Jabatan Biologi Molekul
Fakulti Sains dan Teknologi Sumber
UNIVERSITI MALAYSIA SARAWAK

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*

☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*

☑ **OPEN ACCESS**

I declare this Project/Thesis is to be submitted to the Centre for Academic Information Services and uploaded into UNIMAS Institutional Repository (UNIMAS IR) (Please tick (√)):

☑ **YES**

☐ **NO**

Validation of Project/Thesis

I hereby duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic and research purposes only and not for other purposes.
- The Centre for Academic Information Services has the lawful right to digitize the content to be uploaded into Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis if required for use by other parties for academic purposes or by other Higher Learning Institutes.
- No dispute or any claim shall arise from the student himself / herself neither a third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student himself/herself without first obtaining approval from UNIMAS.

Student's signature: _____ (Date) 12/06/18

Supervisor's signature: _____ (Date) 12/06/18

Prof. Madya Dr. Edmund Sim Ui Hang
Jabatan Biologi Molekul
Fakulti Sains dan Teknologi Sumber
UNIVERSITI MALAYSIA SARAWAK

Current Address:

LOT 325 TINGKAT 2 NO 43 JALAN ANG CHENG HO
93100 KUCHING SARAWAK

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the date of restriction indicated, and the reasons for the confidentiality and restriction.

[The instrument was prepared by The Centre for Academic Information Services]

ACKNOWLEDGMENTS

I would like to thank my supervisor, Associate Professor Dr. Edmund Sim Ui Hang for the assistance, guidance and advice given throughout the conduction of my final year project and thesis writing. The completion of my project was due to his encouragement.

Next, I would like thank the seniors, the postgraduate and doctorate students especially Sim Chor Chien and Yew Keh Li for the assistance, guidance, advice and enduring my working attitude. I am glad to be given the opportunity to meet and work with the seniors which have assist me throughout this project.

Special thanks to my fellow friends and lab mates Fadzirah Izrahni, Leong Kar Xin, Anis Sabrina, Pithrianey Dyllarra and Rocky Vester for the guidance and endless physical and emotional supports throughout the ups and downs during the course of completing my project which have soothed me.

Furthermore, I would like to thank Faculty of Resource Science and Technology of UNIMAS for the facility such as the Immunology Laboratory which enable me to complete my project.

Lastly, I would like to thank both of my beloved parents for the support that have been given to me in emotional and monetary terms. The continuous console, word of encouragements and hope have been fueling me to complete my final year project.

Identification of Protein Partners of the Human Ribosomal Protein, uS19 Using Bioinformatics Analysis

Daphne Olivia Jawai Anak Sadai

Resource Biotechnology Programme
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

The role of human ribosomal protein, uS19 is not only bound in forming intersubunit between small and large subunits of a ribosome but also has extra-ribosomal roles which are causing a few type of diseases and cancers. Although many studies have been carried out to identify the functions of this protein, the information on the functions are still insufficient and limited. The approach to identify the functions of this protein is by identification of the predicted protein's partners of the human ribosomal protein, uS19 and docking stimulation of the protein with protein partners through bioinformatics analysis. This analysis is based on sequences, structure and associated proteins which is known as sequence-to-structure-to-function paradigm. Besides, this analysis involves the construction of predicted three-dimensional (3D) model of uS19 ribosomal protein and the identification of structural neighbour and protein partners. The identified protein partners are used in molecular docking stimulation to the uS19 ribosomal protein. After the analysis of docked uS19 ribosomal protein with the identified protein partners, three logical docked structures were revealed. This is the docking of uS19 ribosomal protein with two different domains of the Fragile X Mental Retardation protein (FMRP) and one domain of Transducer of ERBB2.1 (TOB1) protein. Hence, the study enable the recognition predicted extra-ribosomal functions of ribosomal protein, uS19.

Key words: uS19, bioinformatics, three-dimensional (3D) model, protein interaction, candidate partners, molecular docking, functional prediction.

ABSTRAK

Fungsi protein ribosom manusia, uS19 bukan hanya tertakluk dalam pembentukan antara subunit diantara subunit-subunit kecil dan besar ribosom tetapi juga mempunyai fungsi-fungsi lebihan iaitu menyebabkan beberapa jenis penyakit dan kanser. Walaupun banyak kajian telah dijalankan untuk mengenalpasti fungsi-fungsi protein tersebut, informasi diatas fungsi-fungsi protein tersebut masih tidak mencukupi dan terhad. Pendekatan untuk mengenalpasti fungsi-fungsi protein tersebut ialah dengan mengenalpasti rakan-rakan kongsi ramalan protein ribosom, uS19 dan stimulasi cantuman protein tersebut dengan rakan-rakan kongsi ramalan melalui analisis bioinformatik. Analisis tersebut ialah berasakan urutan, struktur dan protein yang terlibat yang dikenali sebagai paradigma urutan ke struktur ke fungsi. Selain itu, analisis tersebut melibatkan konstruksi ramalan model tiga dimensi (3D) protein ribosom uS19 dan pengenalpastian struktur jiran dan rakan-rakan kongsi. Rakan-rakan kongsi yang dikenalpasti digunakan dalam stimulasi cantuman molekular ke protein ribosom uS19 tersebut. Selepas penganalisisan cantuman protein ribosom uS19 dengan rakan-rakan kongsi yang telah dikenalpasti, tiga cantuman struktur yang logik telah dapat didedahkan. Cantuman tersebut adalah cantuman protein ribosom uS19 dengan dua domain Fragile X Mental Retardation protein (FMRP) yang berlainan dan satu domain protein Transducer of ERBB2.1 (TOB1). Oleh itu, kajian ini membolehkan pengenalan fungsi-fungsi lebihan ramalan protein ribosom, uS19.

***Kata kunci**: uS19, bioinformatik, model tiga dimensi (3D), protein interaksi, calon-calon rakan kongsi, cantuman molekular, ramalan fungsi*

TABLE OF CONTENTS

Declaration	I
Acknowledgment	III
Abstract	IV
Abstrak	IV
Table of Content	V
List of Tables	VIII
List of Figures	IX
List of Abbreviation	XI
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	3
2.1 Ribosome	3
2.1.1 Eukaryote Ribosome	4
2.1.1.1 Small and Large Subunit	5
2.1.2 Ribosomal RNA (rRNA)	5
2.1.3 Ribosomal Protein (r-protein)	6
2.1.3.1 Ribosomal Protein (r-protein), uS19	7
2.2 Bioinformatics	8
2.3 The National Center for Biotechnology Information	9
2.4 Database	9
2.4.1 Ribosomal Protein Gene (RPG) Database	10
2.4.2 Pfam Database	11

2.4.3 InAct 11

2.4.4 PROSITE 12

2.4.5 Protein Data Bank (PDB) Database 13

2.5 Sequence Alignment 13

2.5.1 Basic Local Alignment Sequence Tool (BLAST) 14

2.5.2 Clustal Program 15

2.6 Three-Dimensional (3D) Structure Modeling 16

2.6.1 SWISS-Model Workspace 17

2.6.1.1 Qualitative Model Energy ANalysis (QMEAN) 18

2.6.1.2 PROCHECK by Ramachandran Plot Analysis 19

2.7 Three-Dimensional (3D) Visualization 19

2.7.1 Swiss-Pdb Viewer v4.1 20

2.7.2 Cn3d ("See in 3D") 20

2.8 Protein Docking 20

2.8.1 ClusPro 2.0 21

CHAPTER 3: MATERIALS AND METHODS 23

3.1 Structural Template Identification 23

3.2 Three-Dimensional (3D) Protein Modeling 24

3.3 Structural Neighbours and Candidate Partners Identification 24

3.4 Protein-Protein Docking 25

CHAPTER 4: RESULTS 26

4.1 Three Dimensional (3D) Modeling of uS19 Structure 26

4.2 Structural Neighbours and Candidate Partners Identification 29

4.3 Docking of 3D Model of uS19 With Candidate Partners 33

CHAPTER 5: DISCUSSION 38

CHAPTER 6: CONCLUSION 44

REFERENCES 46

APPENDICES 53

List of Tables

Table 1	:	BLAST programmes that are generally used depending on the query and target sequences type.	15
Table 2	:	The raw score and Z-score of the constructed 3D model of uS19 protein from QMEAN analysis.	29
Table 3	:	Structural neighbours of constructed uS19 3D model against VAST.	30
Table 4a	:	Seven proteins interact with 2QND protein which involved in binary protein-protein interactions.	31
Table 4b	:	Eight proteins interact with 2F7S protein which involved in binary protein-protein interactions.	31
Table 4c	:	Seventeen proteins interact with 4CQO protein which involved in binary protein-protein interactions.	31
Table 5a	:	Eight candidate protein partners interacting with 2QND protein.	32
Table 5b	:	Eight candidate protein partners interacting with 4CQO protein.	32
Table 6a	:	Number of clusters of docking of uS19 protein with each of the eight candidate partners of the 2QND protein.	33
Table 6b	:	Number of clusters of docking of uS19 protein with each of the eight candidate partners of the 4CQO protein.	33
Table 7	:	ClusPro scores for docked uS19-candidate partners.	34

List of Figures

Figure 1	:	The result of alignment of uS19 target protein sequence with template sequences.	28
Figure 2	:	The constructed 3D model of uS19 target protein.	29
Figure 3	:	Result of constructed 3D model of uS19 analysed by Ramachandran Plot and Procheck analysis.	30
Figure 4a	:	uS19_2BKD docked model.	34
Figure 4b	:	Logical interaction between the ribosomal protein uS19 signature and Agenet-like motif 1 of FMRP depicted in blue ribbon.	34
Figure 4c	:	Logical interaction between the ribosomal protein uS19 signature and Agenet-like motif 1 of FMRP depicted in wireframe.	34
Figure 4d	:	uS19_2QND docked model.	34
Figure 4e	:	Logical interaction between the ribosomal protein uS19 signature and Type-1 KH motif 1 of FMRP depicted in blue ribbon.	34
Figure 4f	:	Logical interaction between the ribosomal protein uS19 signature and Type-1 KH motif 1 of FMRP depicted in wireframe.	34
Figure 4g	:	uS19_2Z15 docked model.	34
Figure 4h	:	Logical interaction between the ribosomal protein uS19 signature and BTG family signature 1 and 2 of Tob1 protein depicted in blue ribbon.	34
Figure 4i	:	Logical interaction between the ribosomal protein uS19 signature and BTG family signature 1 and 2 of Tob1 protein depicted in wireframe.	34
Figure 5a	:	The template sequences searched from Position-Specific Iterated BLAST (PSI-BLAST).	50
Figure 5b	:	The template proteins (2ZKQ_s) sequence and structure linked from PSI-BLAST.	50
Figure 5c	:	The template proteins (4KZX_p) sequence and structure linked from PSI-BLAST.	50

Figure 6	:	Multiple alignment between the target sequence (uS19) and the template sequence (2ZKQ_s and 4KZX_p) by using ClustalX.	51
Figure 7a	:	The secondary structure of 3D modeling of the predicted structure of the uS19 by using the template structure, 2ZKQ_s generated by using SWISS-MODEL workspace and viewed by using Swiss-pdb Viewer.	52
Figure 7b	:	The secondary structure of 3D modeling of the predicted structure of the uS19 by using the template structure, 4KZX_p generated by using SWISS-MODEL workspace and viewed by using Swiss-pdb Viewer.	52
Figure 8a	:	Secondary structure of docked uS19_2BKD.	53
Figure 8b	:	Secondary structure of docked uS19_2FMR.	53
Figure 8c	:	Secondary structure of docked uS19_2QND.	53
Figure 8d	:	Secondary structure of docked uS19_4DKK.	53
Figure 8e	:	Secondary structure of docked uS19_1G47.	53
Figure 8f	:	Secondary structure of docked uS19_1NYP.	53
Figure 8g	:	Secondary structure of docked uS19_2COR.	53
Figure 8h	:	Secondary structure of docked uS19_2D8X.	53
Figure 8i	:	Secondary structure of docked uS19_2DKL.	53
Figure 8j	:	Secondary structure of docked uS19_2Z15.	53
Figure 8k	:	Secondary structure of docked uS19_3E9V.	53
Figure 8l	:	Secondary structure of docked uS19_4OLA.	53
Figure 8m	:	Secondary structure of docked uS19_3LUK.	53
Figure 8n	:	Secondary structure of docked uS19_1E4U.	53
Figure 8o	:	Secondary structure of docked uS19_2RXR.	53
Figure 8p	:	Secondary structure of docked uS19_1LBD.	53

List of Abbreviations

3D	:	Three-Dimensional
82-FIP	:	82 KDa FMRP interacting Protein
Å	:	Angstrom
AFCS	:	Alliance for Cellular Signaling
APRO	:	Anti-proliferative
BLAST	:	Basic Local Alignment Sequence Tool
BIND	:	Biomolecular Interaction Network Database
BTG	:	TOB/B cell translocation gene
CATH	:	Class, analogy, topology, homologous superfamily
CLL	:	Chronic Lymphocytic Leukemia
Cn3D	:	See in 3D
DBA	:	Diamond-Blackfan Anemia
DDBJ	:	DNA Data Bank of Japan
DNA	:	Deoxyribonucleic Acid
DIP	:	Database of Interacting Protein
EMBL	:	Ensembl
ER	:	Endoplasmic Reticulum
FASTA	:	Fast Alignment
FFT	:	Fast Fourier Transform
FMRP	:	Fragile X Mental Retardation Protein
GO	:	Gene Ontology
HETATM	:	Heterogens atoms
HMM	:	Hidden Markov model
hnRNP K	:	Human heterogeneous nuclear ribonucleoprotein K
HPRD	:	Human Protein Reference Database
KH	:	hnRNP K homology
LSU	:	Large Subunit
MMDB	:	Molecular Modeling Database
mRNA	:	Messenger Ribonucleic Acid
mRNP	:	mRNA ribonucleoprotein
NCBI	:	National Center for Biotechnology Information
NDF	:	N-terminal domain of FMRP
NLS	:	Nuclear localization sequence
NMR	:	Nuclear Magnetic Resource
NUFIP	:	Nuclear FMRP interacting Protein
PDB	:	Protein Data Bank
PIR	:	Protein Information Resource
PSI-BLAST	:	Position-Specific Iterated BLAST
PTC	:	Peptidyltransferase Center
QMEAN	:	Qualitative Model Energy Analysis
RCSB	:	Research Collaboratory for Structural Bioinformatics
RMSD	:	Root-mean-square-deviation
RNA	:	Ribonucleic Acid
RNP	:	Ribonucleoprotein

RP	:	Ribosomal Protein
RPG	:	Ribosomal Protein Gene
R-protein	:	Ribosomal Protein
rRNA	:	Ribosomal RNA
SCOP	:	Structural Classification of Proteins
SEQRES	:	Reference Sequence
SSU	:	Small Subunit
TOB1	:	Transducer of ERBB2.1
Tm	:	Melting Temperature
tRNA	:	Transfer Ribonucleic Acid
UniProtKB	:	UniProt Knowledgebase
VAST	:	Vector Alignment Sequence Tool

CHAPTER 1: INTRODUCTION

Ribosome is a large ribonucleoprotein (RNP) that functions in synthesizing protein in cells from genetic material encoded in messenger ribonucleic acids (mRNA). All ribosomes are basically made up of two unequal size subunits which are small subunit and large subunit. These subunits are made up of ribosomal RNA (rRNA) and ribosomal proteins (r-protein). Both prokaryotes and eukaryotes ribosomes differ in structure due to rRNA and r- proteins types. Prokaryote ribosome, known as 70S, is made up of 30S small subunit (SSU) and 50S large subunit (LSU) (Bhavsar, Makley & Tsonis, 2010). The subunits are made up of three rRNAs (16S, 5S and 23S) with 54 r-proteins (Wilson & Cate, 2012).

In contrary, eukaryote ribosome, known as 80S, is made up of 40S small subunit (SSU) and 60S subunit (SSU) (Bhavsar *et al.*, 2010). In yeast, the subunits are made up of four rRNAs (5S, 5.8S, 18S and 25S) with 79 r-protein (Ben-Shem, Jenner, Yusupova & Yusupov, 2010). In human, the subunits are made up of four rRNAs (5S, 5.8S, 18S and 28S) with 80 r-proteins (Khatter, Myasnikov, Natchiar & Klaholz, 2015).

Human ribosomal protein, uS19 have multiple functions. Recent studies show r-proteins are not only involved in protein translation but also involved in extra-ribosomal functions such as DNA repair, apoptosis and cellular homeostasis (Shenoy *et al.*, 2012). Moreover, the overexpression of specific r-proteins leads to liver, colon, prostate and other tumors. For example, overexpression of RPL19 gene has been found in prostate cancer cell line (Shenoy *et al.*, 2012).

The knowledge on the extra-ribosomal functions of the uS19 ribosomal protein is inadequate and limited thus leads to lack of understanding of this protein. Hence, the study enable the identification of the functions of this r-protein through the identification of predicted protein partners. The prediction can be done by using bioinformatics analysis by analysing the functions of the protein partners based on the sequences, structure and associated proteins of the r-protein. This analysis is based on the sequence-to-structure-to-function paradigm.

The identification of structural proteins neighbours and candidate protein partners of human ribosomal protein, uS19 by using bioinformatics will enable the identification of predicted functions of human ribosomal protein, uS19 through docking stimulation of uS19 with candidate protein partners. Therefore, the objectives of this study are:

(i) To identify protein partners (candidate partners) of human ribosomal protein, uS19 by using structural neighbours identification and docking stimulation and

(ii) To predict the function(s) of human ribosomal protein, uS19 through neighbour protein(s) and candidate partner(s) identification.

CHAPTER 2: LITERATURE REVIEW

2.1 Ribosome

Ribosome is a small particle which are found in all living cells in large numbers. This particle occurs as free particles in both prokaryotic and eukaryotic cells whereas attached to the membranes of the endoplasmic reticulum (ER) in eukaryotic cells. The main function of ribosome is to serve as protein synthesis site, translating nucleotide sequence of messenger RNA (mRNA) to the amino acid sequence of a protein. Besides, protein synthesis also involves transfer RNA (tRNA) and other translation factors. The discovery of ribosome was done by a cell biologist named George E. Palade, where he described ribosomes as a small particles of the cytoplasm and associated with the membrane of the endoplasmic reticulum (Jamieson, 2008).

Generally, ribosome is made up of ribosomal RNA (rRNA) and ribosomal proteins (r-proteins). Furthermore, ribosome also consists of two subunits which are small subunit and large subunit. The small subunit serves as recognition of accurate codon-anticodon between mRNA and tRNA molecules whereas the large subunit serves as site for peptide bond formation of amino acid.

Prokaryote and eukaryote ribosomes differ in structure where eukaryote ribosome is larger and complex. Prokaryote ribosome, known as 70S, consists of a 30S small subunit (SSU) and a 50S large subunit (LSU). In contrary, eukaryote ribosome, known as 80S, consists of a 40S SSU and a 60S LSU. The S designated in both ribosomal subunits and rRNA molecules stands for Svedbergs, a unit, which is the measure of the sedimentation rate of suspended particles

under standard conditions (Lodish *et al.*, 2000). In contrary, the numerical value stands for region of the sedimentation gradient where the cell components settle in the tube. The sedimentation affected by the size, shape and density of the components.

2.1.1 Eukaryote Ribosome

Eukaryote is organisms where the cells consist of a nucleus, enclosed by membranes. The example of eukaryotic organisms are human and yeast. The ribosome of eukaryote, known as 80S, is made up of 40S small subunit (SSU) and 60S large subunit (LSU) (Bhavsar, Makley & Tsonis, 2010).

In yeast, both of the SSU and LSU subunits are made up of four ribosomal RNAs (5S, 5.8S, 18S and 25S) and 79 ribosomal-proteins (r-protein) (Ben-Shem, Jenner, Yusupova & Yusupov, 2010). The SSU is made up of 18S rRNA and 33 r-proteins whereas the LSU is made up of 5S, 5.8S and 25S rRNAs and 46 r-proteins (Ben-Shem *et al.*, 2010).

For human, the subunits are made up of four rRNAs (5S, 5.8S, 18S and 28S) with 80 r-proteins (Khatter, Myasnikov, Natchiar & Klaholz, 2015). The SSU is made up of 18S rRNA and 33 r-proteins whereas the LSU is made up of 28S, 5S and 5.8S rRNAs with 47 proteins (Khatter *et al.*, 2015). Compared to yeast, LSU of human ribosome's consists of 47 r-proteins due to an extra r-protein which is the eL28 (Khatter *et al.*, 2015).

2.1.1.1 Small and Large Subunits

The small subunit (SSU) of eukaryote ribosome functions in decoding the genetic code by bringing in messenger RNA (mRNAs) and aminoacyl-transfer RNAs (tRNAs) (de la Cruz *et al.*, 2015). During translation, coordinated conformational changes occurs within SSU which enables (tRNA) and (mRNA) translocation through the ribosome.

In contrary, the large subunit (LSU) contains the peptidyltransferase center (PTC) which functions in catalyzing peptide bond formation. The protein synthesis occurs when both SSU and LSU are connected by interacting regions known as intersubunit bridges.

The intersubunit bridges consists of a series of 17 intersubunit bridges which involve either interactions between RNA with RNA and RNA with protein or protein with protein (Rhodin & Dinman, 2011). In addition, these bridges are ratchet-like which enables ribosome to have the ability to reorganize for mRNA and tRNA translocation during protein synthesis accompanied with the intersubunit movement.

2.1.2 Ribosomal RNA (rRNA)

Ribosomal RNA (rRNA) is known to serve the role in stabilizing ribosome structure and the function of ribosome. The secondary and tertiary structure of the rRNAs enable the construction of the ribosome subunits by cross linking of nucleotide sequences between the base-paired stem loops of the rRNAs. Besides, the secondary structure of rRNAs function in recognizing conserved portions of mRNAs and tRNAs (Encyclopaedia Britannica, 2016).

Moreover, the stabilization of ribosome is deduced by the presence of three pesudouridine residues in the H69 region of the LSU (28S) of human shown by the UV melting (*Tm*) experiment where the melting point of H69 with the pseudouridine is higher compared to the one modified with three uridines (Sumita *at el.*, 2005).

The rRNAs role in the function of ribosome is directly associated with the formation of subunits by the rRNAs, such as the catalyzing the peptidyl transferase reaction during protein synthesis (Encyclopaedia Britannica, 2016).

2.1.3 Ribosomal Protein (r-protein)

Ribosomal proteins (r-protein) are known to function in acting as the central role in rRNA processing, protein assembly, RNA folding, transporting ribosomal precursors, stabilization of the subunit structure and interacting with other factors required for either ribosome biogenesis or translation. The r-proteins play role as pivot by attaching rRNA domains together by threading across the surface of the subunits and in contact with multiple domains of rRNA for ribosome stabilisation (Ferreira-Cerca, Pöll, Gleizes, Tschochner & Milkereit, 2005).

Moreover, ribosomal proteins also have extra-ribosomal functions such as replication, transcription, splicing or even ageing depending on the r-proteins (Bhavsar *et al.*, 2010). For example, S12 of human r-protein function in splicing and modifiying RNA (Bhavsar *et al.*, 2010).

In contrary, the mutation of r-proteins lead to diseases and cancer. A study was conducted on two different families with autism leads to the identification of two missense mutations in the RPL10 ribosomal protein gene which is located in Xq28 (Klauck *et al.*, 2006).

2.1.3.1 Ribosomal Protein (r-protein), uS19

Human ribosomal protein, uS19 is previously known as S15 in eukaryotes and S19 in prokaryotes due to the change in naming system to avoid confusion. The 'S' in the naming system meaning the protein made up the small subunit whereas the 'u' stands for universal meaning the protein is found in all three domains of life which are Archaea, Bacteria and Eukarya (Ban *et al.*, 2014).

The ribosomal protein, uS19 is the neighbouring protein of uS13 in eukaryotes where uS13 participates in the B1b/c bridge formation (Bowen, 2015). Besides, uS19 involves in the formation of B1a bridge with H38 (Bowen *et al.*, 2015). The gene encoded the uS19 protein is the RPS15 gene where the protein made up the 40S ribosome subunit besides being the member of the S19P ribosome protein family (Sim, Ma, Chan, Lee & Narayanan, 2016).

Mutation of RPS15 (uS19) have been found in Chronic Lymphocytic Leukemia (CLL) and more frequently in relapsed CLL with up to 19.5% of cases (Goudarzi & Lindström, 2016). Moreover, the mutation of this gene along with other r-proteins (RPS7, RPS10, RPS17, RPS24, RPS26, RPS27A, RPS28, RPS29, RPL5, RPL11, RPL15, RPL35A and RPL36) also caused Diamond-Blackfan Anemia (DBA) as it was later identified in DBA patients (Zhou, Liao, Liao, Liao & Lu, 2015).

2.2 Bioinformatics

Bioinformatics is define as biology in terms of molecules (bio) and informatics is the application of the techniques such as applied maths, computer science and statistics to understand and organize information on molecules on a large scale (Luscombe, Greenbaum, & Gerstein, 2001).

Besides, bioinformatics is the process where sequences of deoxyribonucleic acid (DNAs), ribonucleic acid (RNAs) and protein besides structural information are recorded, annotated, stored, analyzed and retrieved (O'Leary-Driscoll, 2015). In addition, bioinformatics involves the use of computer software tools to create database, manage, warehousing and mining data and for global communication networking. Moreover, bioinformatics are also used to detect molecular interaction and the function complexes by analyzing the protein's sequence, structure and the association of the protein with other proteins. This is known as sequence-to-structure-to-function paradigm (Sim *et al.*, 2016).

The functions of bioinformatics are divided into three-fold as described by Luscombe, Greenbaum & Gerstein (2001). The first fold is the organization of data which enables researchers to access information and to submit new information in databases. The example is the Protein Data Bank for the 3D structures of macromolecules. The second fold is the development of tools and resources which aid in data analysis which requires extensive knowledge of computational theory and thorough understanding of biology. The example is the PSI-BLAST and FASTA which are used to compare newly sequenced protein sequences to previously characterized sequences. The third fold is the using of developed tools to analyse the data and interpret the results.

The applications of bioinformatics are in identifying homologous of poorly characterized protein or genome sequences, rational drug design based on sequence-to-structure-to-function paradigm, large-scale censuses providing information on genomes, structures and expression databases and advancing applications in medical sciences by gene expression analysis.

2.3 The National Center for Biotechnology Information (NCBI)

The National Center for Biotechnology Information (NCBI) is a server established by the National Institute of Health which provides biomedical and research professionals the access to public databases and software tools to analyze biological data besides performing research in computational biology.

The search engine of this server is Entrez which enable the search and retrieval of the databases. The services provided by NCBI are bibliographic database PubMed, GenBank, and Basic Local Alignment Search Tool (BLAST) (McEntyre & Ostell, 2002). Besides, three-dimensional (3D) structure database such as Molecular Modeling Database (MMDB) is available in NCBI where it provides the structure of molecular 3D and linked articles of the searched molecule (Chen *et al.*, 2003).

2.4 Databases

The databases of generally used in bioinfromatics is known as biological database. The databases are further classified into primary, secondary and composite databases. The primary database contains raw experimentally data such as nucleotide sequences and three-dimensional structures where the repository is archived. The example of primary databases are GenBank, DNA Data Bank of Japan (DDBJ), Ensembl (EMBL), Swiss-Prot, Protein Information

Resource (PIR), TrEMBL and Protein Data Bank (PDB) (Babu, 1997). Moreover, protein-protein interaction primary databases are human protein reference database (HPRD), InAct database, Molecular Interaction (MINT) database, Database of Interacting Protein (DIP), Munich Information Center for Protein Sequences (MIPS) database, Alliance for Cellular Signaling (AFCS), Biomolecular Interaction Network database (BIND), Reactome, and PDZBase.

In contrary, the secondary database contain curated information of the primary database where information on conserved sequence, signature sequence and residues of active site of the protein families. The example of secondary databases are Pfam, PROSITE, BLOCKS, Structural Classification of Proteins (SCOP) and Class, architecure, topology, homologous superfamily (CATH) (Babu, 1997).

The composite database is a database which combined the variety of primary database and secondary database sources which enables the search of multiple information. The example of composite database is The National Center for Biotechnology Information (NCBI).

2.4.1 Ribosomal Protein Gene Database (RPG)

Ribosomal Protein Gene (RPG) database (http://ribosome.med.miyazaki-u.ac.jp) is an online database which is used to retrieve genomic, cDNA and amino acid sequences, intron/exon structures, genomic locations and orthologs information (Nakao, Yoshihama & Kenmochi, 2004). This database provides detailed information on the searched ribosomal protein genes. Moreover, RPG also enables the comparison of gene structure across different species and alignment of multiple amino acid sequences (Nakao, Yoshihama & Kenmochi, 2004). In