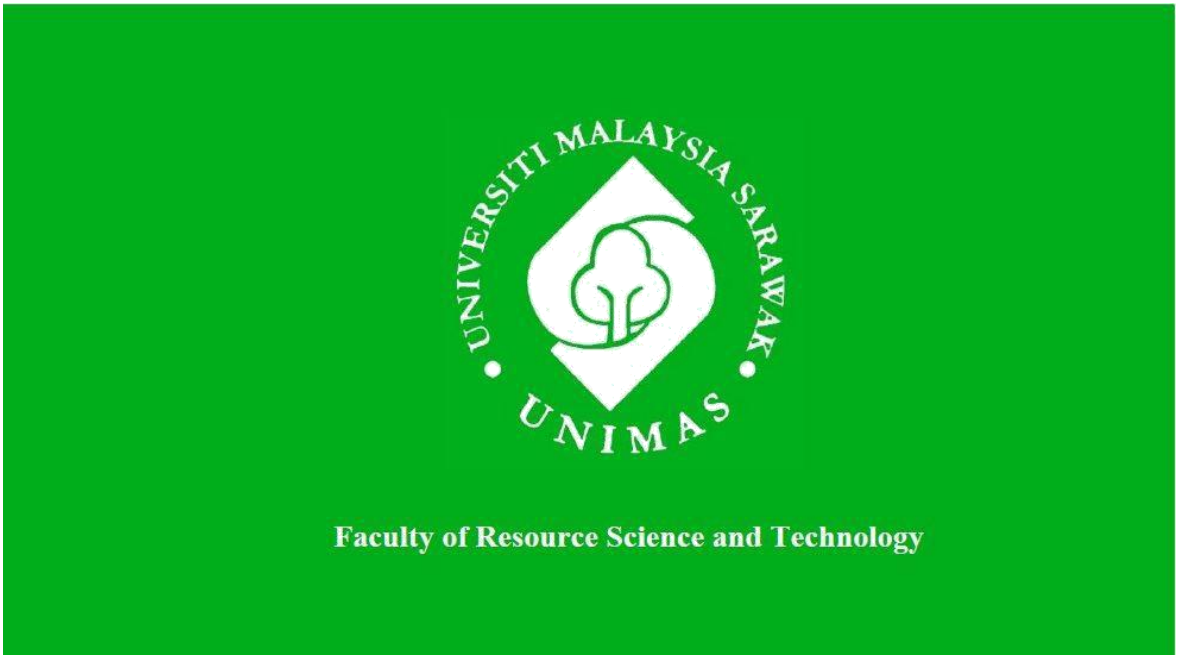

Identification of logical extra-ribosomal functions mediated by the ribosomal protein genes, RPS3 and RPS12 via *In Silico* approach.

Nazatul Syahira Bt. Roslan (37447)

Bachelor of Science with Honours
(Resource Biotechnology)
2015

Identification of Logical Extra Ribosomal Functions mediated by the Ribosomal Protein Genes, RPS3 and RPS12 via *In Silico* Approach

Nazatul Syahira Bt Roslan

A thesis submitted in partial fulfilment of the Final Year Project 2 (STF 3015) course

Supervisor: Associate Professor Dr. Edmund Sim Ui Hang

Programme of Biotechnology Resource

Department of Molecular Biology

Faculty of Resource Science and Technology

University Malaysia Sarawak

ACKNOWLEDGEMENT

First, I would like to thank to the God for giving me strength and the ability to complete this project. I would also like to express my deepest gratitude to my supervisor, Associate Professor Dr. Edmund Sim Ui Hang who guiding and encouraging me throughout the completion of my final year project successfully.

Besides, I am thankful to the postgraduate students, Ms. Kherlee Ng, Ms. Stella Chan Li Li and Ms. Shruti Talwar, who always sharing their knowledge and advise me while conducting this final year project.

Moreover, I am grateful to my lab mates and my fellow friends who always willing to help and give their best suggestions. Finally, I would like to thank to my parents, Mr. Roslan bin Abrahim and Ms. Hanimah bt. Mohamed for their endless support and prayed throughout the duration of my final year project. Their support and their trust in me have made me not to give up easily in completing this project to a successful.

DECLARATION

I declare that this thesis entitled "*Identification of logical extra-ribosomal functions mediated by the ribosomal protein genes, RPS3 and RPS12 via in Silico approach*" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

Signature:

Name : Nazatul Syahira Bt. Roslan

Date : JUNE 2015

TABLE OF CONTENT

Content	Page
ACKNOWLEDGEMENT	**I**
DECLARATION	**II**
TABLE OF CONTENT	**III**
LIST OF ABBREVIATIONS	**V**
LIST OF TABLES AND FIGURES	**VI**
ABSTRACT	**VIII**
1.0 INTRODUCTION	**1**
2.0 LITERATURE REVIEW	**2**
2.1 Ribosomal Protein	**2**
2.2 Ribosomal Protein S3	**2**
2.3 Ribosomal Protein S12	**3**
2.4 Bioinformatics Approach	**3**
2.4.1 Ribosomal Protein Gene Database	**3**
2.4.2 PSI-BLAST	**4**
2.4.3 CLUSTALX	**4**
2.4.4 SWISS-MODEL	**5**
2.4.5 VAST	**5**
2.4.6 IntAct Database	**6**
2.4.7 ClusPro 2.0	**6**
3.0 METHODS AND MATERIALS	**7**
3.1 Materials	**7**
3.2 Methods	**8**
3.2.1 Multiple Sequence Alignment	**8**

3.2.2 Protein Modeling **8**

3.2.3 Protein-protein Interaction **9**

4.0 RESULTS AND DISCUSSION **10**

5.0 CONCLUSION **33**

6.0 REFERENCES **34**

LIST OF ABBREVIATIONS

Ribosomal Protein	RPs
Ribosomal Protein S3	RPS3
Ribosomal Protein S12	RPS12
Ribosomal Protein Gene database	RPG
Position-Specific Iterative Basic Local Alignment Search Tool	PSI-BLAST
Expected value	E-value
Vector Alignment Search Tool	VAST
Protein Data Bank	PDB
Quality Model Energy Analysis	QMEAN

LIST OF TABLES

Content		Page
TABLE 1	Sequences producing significant alignment with E-value better than threshold for RPS3. The identity of sequences was selected above 60%.	11
TABLE 2	Sequences producing significant alignment with E-value better than threshold for RPS12. The identity of sequences was selected above 60%.	12
TABLE 3	Possible domain and motifs on RPS3	16
TABLE 4	Possible domain and motifs on RPS12	16
TABLE 5	Selected structure neighbors and candidate partners of RPS3 and RPS12	23
TABLE 6	ClusPro scores of RPS-candidate partner dock model	24
TABLE 7	RMSD for RPS3 and EGFR	25
TABLE 8	RMSD for RPS3 and YWHAE	27
TABLE 9	RMSD for RPS12 and ELOB	29
TABLE 10	RMSD for RPS12 and SKIL	31

LIST OF FIGURES

Content		Page
Figure 1	Multiple sequence alignment of template sequences to RPS3 protein sequence. The figure has been viewed using Jalview program	14
Figure 2	Multiple sequence alignment of template sequences to RPS12 protein sequence. The figure has been viewed using Jalview program	15
Figure 3	The Qmean4 and Z-score of RPS3	18
Figure 4	The Qmea 4 and Z-score of RPS12	19
Figure 5	Ramachandran plot analysis on RPS3	20
Figure 6	Ramachandran plot analysis on RPS12	21
Figure 7	RPS3 docked with EGFR	26
Figure 8	RPS3 docked with YWHAE	28
Figure 9	RPS12 docked with ELOB	30
Figure 10	RPS12 docked with SKIL	32

ABSTRACT

Ribosomal proteins (RPs) are the main component of ribosomes and play important roles in protein biosynthesis. Besides, several of them are shown to have extraribosomal functions. In this research, we used bioinformatics approach in finding the interaction between the ribosomal protein S3 (RPS3) and ribosomal protein S12 (RPS12) with their protein partners. Our results discovered that RPS3 interacts with Epidermal Growth Factor Receptor (EGFR) and 14-3-3 protein epsilon (YWHAE). RPS3 is predicted to inhibit the function of EGFR and YWHAE. RPS12 interacts with Transcription Elongation Factor B Polypeptide (ELOB) and Ski-like protein (SKIL). Hence, it is predicted to interrupt the function of ELOB and SKIL.

Keywords: Ribosomal protein, Ribosomal protein S3, Ribosomal protein S12, Extraribosomal functions.

ABSTRAK

Protein ribosom (RP) adalah komponen utama ribosom dan memainkan peranan penting dalam biosintesis protein. Selain itu, beberapa daripada protein ini yang terbukti mempunyai fungsi extraribosomal. Dalam kajian ini, kami menggunakan pendekatan bioinformatik dalam mencari interaksi antara protein ribosom S3 (RPS3) dan protein ribosom S12 (RPS12) dengan pasangan protein mereka. Keputusan RPS3 kami mendapati bahawa berinteraksi dengan Epidermal pembesaran Faktor Reseptor (EGFR) dan 14-3-3 epsilon protein (YWHAE). RPS3 diramalkan untuk menghalang fungsi EGFR dan YWHAE. RPS12 adalah berinteraksi dengan Transkripsi Pemanjangan Factor B polipeptida ELOB) dan protein Ski-suka (SKIL). Oleh itu, adalah diramalkan untuk mengganggu fungsi ELOB dan SKIL.

Kata kunci: Protein ribosom, Protein ribosom S3, Protein ribosom S12, Fungsi extraribosomal.

1.0 INTRODUCTION

Ribosomal proteins (RPs) play important roles in protein biosynthesis. All eukaryotic 80S ribosome have a large 60S subunit and a small 40S subunit. The 60S subunit consists of three rRNAs (5S, 28S and 5.8S) and about 50 proteins while the 40S subunit consists of one 18S molecule of rRNA and about 33 proteins (Bhavsar *et al*., 2010). RPs can also function independently during various cellular processes of protein biosynthesis. RPs is also called as extraribosomal functions whereby it can function during various cellular processes such as replication, gene transcription, apoptosis, DNA repair, and even inflammation (Kasai *et al*., 2003).

Ribosomal protein S3 function as a part of the nuclear factor-kB complex that interacts with specific sites in the genome, on the tumour necrosis factor stimulation while the ribosomal protein S12 function as RNA splicing and modification. Over-expressed of these proteins which are RPS3 and RPS12 may lead to colon cancer (Bhavsar *et al*., 2010).

However, there is limited research on the protein-protein interaction that made extraribosomal function difficult to understand. This may be due to difficulties in obtaining ribosome structure through experiment. Therefore, to overcome this problem, we are using bioinformatics approach to predict the protein structure and protein-protein interaction of ribosomal protein S3 and ribosomal protein S12 between their putative protein partners.

The main objective includes:

1) To generate the protein structure model of ribosomal protein S3 and ribosomal protein S12.
2) To predict the interaction of ribosomal protein S3 and ribosomal protein S12 between their putative protein partners.

2.0 LITERATURE REVIEW

2.1 Ribosomal Protein

The eukaryotic 80S ribosome consists of 60S large subunit and 40S small subunit. This eukaryotic ribosome is more complex and consists more components than prokaryotic ribosome and significantly larger in size (Vallabhaneni, 2009). The 60S subunit consists of three rRNAs (5S, 28S and 5.8S) and about 50 proteins while the 40S subunit consists of one 18S molecule of rRNA and about 33 proteins (Bhavsar *et al*., 2010).

Ribosomal proteins have the second function from both the ribosome and protein synthesis. Ribosomal proteins can also function in other cellular processes such as transcription, replication, DNA repair, apoptosis, RNA processing and inflammation (Kasai *et al*., 2003).

2.2 Ribosomal Protein S3

Ribosomal protein S3 (RPS3) belongs to the S3P family of ribosomal proteins. Previous studies on mouse and rat proteins have showed that it have an extraribosomal role as an endonuclease involved in the repair of UV-induced DNA damage. The protein located in both cytoplasm and nucleus but not in nucleolus. Through several observations, expression of this gene is high in colon adenocarcinomas and adenomatous polyps compared to adjacent normal colonic mucosa (NCBI, 2014). In addition, RPS3 involved in the nuclear factor-kappaB (NF-kB)-mediated gene regulation (Ahn *et al.,* 2011) and also respond to DNA damage signals by activating cycle checkpoints which arrest the cell cycle and activating DNA repair systems or inducing apoptosis (Jang *et al.,* 2004).

2.3 Ribosomal Protein S12

Ribosomal protein S12 (RPS12) belongs to the S12E family of ribosomal proteins and is located in the cytoplasm. Several observation shows that expression of this gene increased in colorectal cancer compare to normal colonic mucosa (NCBI, 2014). Furthermore, RPS12 is most effective to stimulate splicing of phage T4 introns (Coetzee *et al*., 1994). Ribosomal protein mutation in *Escherichia coli* have been found to confer resistance to streptomycin and these mutation frequently exist in the RPS12, encoded by *rpsL* and result in streptomycin resistance or streptomycin dependence. Mutations that correspond to RPS12 have been characterized in *E.coli*. Thus, it can cause preservation of the translation activity and enhance the expression of enzymes that involved in antibiotic production in the late stationary phase (Chumpolkulwong *et al.*,2004).

2.4 Bioinformatics Approach

2.4.1 Ribosomal Protein Gene database (RPG)

Ribosomal Protein gene (RPG) is a new database that provides a details information about ribosomal proteins genes. The database contains data from humans and also other organism such as *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Escherichia coli and Caenorhabditis elegans*. The user can search the database by gene name and specific organism. Each record includes sequences, intron/exon structures, chromosome locations and information about the orthologs. Besides, the user can use map viewer to view and compare the gene structures of the above organisms and make multiple amino acid sequence alignments. In addition, RPG also provides information on small nucleolar RNAs (snoRNAs) that are encoded in the introns of the ribosomal protein genes (Nakao *et al*., 2004).

2.4.2 Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST)

PSI-BLAST is an iterative program to search a database for proteins with distantly similarity to a query sequence. It is widely used programs for detecting sequence similarities including subtle ones, in searches of protein sequence database. This program is to construct a multiple alignment from the output of a BLAST protein database similarity search, abstract a position-specific score matrix from this multiple alignment and search the database using the score matrix as query (Schäffer *et al.,* 2001). PSI-BLAST program run approximately the same speed per iteration as gapped BLAST but it is much more sensitive weak but biologically relevant sequence similarities (Altschul *et. al.,* 1997).

2.4.3 CLUSTAL X

Clustal X is a multiple sequence alignment programs that have been completely rewritten in C++ with a simple object model in order to make it easier to maintain the code and more importantly to make it easier to modify or even replace some of the alignment algorithm (Larkin *et al.,* 2007). Clustal X displays the sequence alignment in a window on the screen. A versatile sequence colouring scheme will allowed the user to highlight conserved features in the alignment. The quality of alignment analysis can be performed and low-scoring segments or exceptional residues can be highlighted (Thompson *et al*., 1997). In Clustal, the old version using Neighbor-Joining (NJ) method to calculate the guide trees but now the earliest version of the program UPGMA was used. UPGMA is faster than NJ but prone to cluster long branches together when the evolutionary rates are very unequal in different lineages (Larkin *et al.,* 2007).

2.4.4 SWISS MODEL

SWISS-MODEL is a server for automated comparative modelling of three dimensional (3D) protein structures. It is the most widely-used free web-based automated modelling. Template selection, alignment and model building are completely done by the server. For the complex modelling tasks it can be done using DeepView (Swiss-PDBViewer), an integrated sequence-to-structure workbench (Schwede *et al*., 2003). The Swiss-PDB viewer is not only acts as a client for SWISS-MODEL but also provides a large selection of structure analysis and display tools. In addition, SWISS-MODEL Repository is a database containing more than 3500 automatically generated protein models which can be readily downloaded.

2.4.5 Vector Alignment Search Tool (VAST)

Vector Alignment Search Tool is a fast and efficient way to perform structural similarity comparisons. These protein structure comparison methods can provide structural and functional similarities from the remote homologs which cannot be detected by sequence information. The VAST algorithm uses vectors as secondary structure elements (SSEs) in the protein structures and the comparisons that align only one or two vectors between two proteins are never significant (Tsang, 2007). VAST has been used to compare all known Protein Data Bank (PDB) domains to each other. The alignment results are presented in the NCBI's Molecular Modeling Database (Hung & Lin, 2013).

2.4.6 IntAct Database

IntAct provides an open source database and toolkit for the storage, presentation and analysis of protein interactions. A web service allows the direct computational access to retrieve interaction networks in XML format (Hermjakob *et al.,* 2004). IntAct contains over 200 000 curated binary interaction evidences. Currently, IntAct provides a two-tiered view of the interaction data due to growing data volume and user requests. The search interface will allowed the user to iteratively develop complex queries, exploiting detailed annotation with the hierarchical controlled vocabularies. The results that are provided at any stage in a simplified and tabular view (Aranda *et al.,* 2010).

2.4.7 ClusPro 2.0

ClusPro is first fully automated web-based program for the computational docking of protein structures. The users can upload the coordinate files of two proteins structures through ClusPro's web interface or enter the PDB codes of respective structures which Cluspro will download from PDB server. The docking algorithms evaluated billions of putative complexes, retaining a number with favourable surface complementarities. A filtering method is used to this set of structure and selecting those with good electrostatic and desolvation free energies for further clustering. The program output is a short listed of putative complexes rank according to their clustering properties (Comeau *et al*., 2004).

3.0 MATERIALS AND METHODS

3.1 MATERIALS

Tools	Host/Developer (s)
Ribosomal Protein Gene (RPG) (http://ribosome.med.miyazaki-u.ac.jp/)	Nakao *et al*. (2004)
Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi)	Altschul *et al.* (1990)
ClustalX (http://www.clustal.org/clustal2/)	Jeanmougin *et al.* (1998)
SWISS-MODEL (http://swissmodel.expasy.org/)	Arnold *et al.* (2006)
Vector Alignment Search Tool (VAST) (http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml)	National Center for Biotechnology Information (NCBI)
IntAct database (http://www.ebi.ac.uk/intact/)	European Bioinformatics Institute
ClusPro 2.0 (http://cluspro.bu.edu/login.php)	Structural Bioinformatics Lab
PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/)	European Bioinformatics Institute
Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do)	Hamilton, W. (1971)
PROSITE (http://prosite.expasy.org/)	Bairoch, A. (1988)

3.2 METHODS

3.2.1 Multiple Sequence Alignment

The protein sequences of ribosomal protein S3 (RPS3) and ribosomal protein S12 (RPS12) were retrieved from Ribosomal Protein Gene (RPG) database (Nakao *et al*., 2004). The protein identifier for RPS3 was NP_001243731 and for RPS12 was NP_001007. After that, the query protein sequences against other sequences from Protein Data Bank (PDB) were performed using Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) with parameter to search for the templates. The template structures were analysed with the expected value (E-value) that above a certain threshold for high sequence identity and structure resolution. Next, the template were selected for multiple sequence alignment and aligned with their respective target sequences using ClustalX program.

3.2.2 Protein Modeling

The RPS template alignment was constructed into models by using SWISS-MODEL software. By using alignment mode in SWISS-MODEL, RPS3-template alignments were submitted to SWISS-MODEL with PDB ID (4BSZ_A) while RPS12-template alignments was modeled with PDB ID (4KZX_M) using the same alignment mode. The quality of the generated models was evaluated by Qualitative Model Energy Analysis (QMEAN) (Benkert *et al*., 2009) and PROCHECK that corresponds to Ramachandran plot from SWISS-MODEL (Hooft *et al*., 1997). All models that show an average QMEAN score in range 0-1 were selected for further analysis.

3.2.3 Protein-protein Interaction

By using Vector Alignment Search Tool (VAST), the protein models were searched against medium redundancy subset of PDB structures for structural neighbors. All the structural neighbors were filtered for human proteins only. The putative protein partners to structural neighbors were retrieved from IntAct database. Then, the numbers of putative partner for docking were listed down based on similarity of the sequence patterns using PROSITE search. ClusPro 2.0 server was used to dock RPS model to putative partners. RPS model was input as receptor and putative partners as ligand. The returned dock complexes were evaluated based on the energy score. Molecular surface of dock complex was scanned for potential interaction sites.

4.0 RESULT AND DISCUSSION

4.1 Analysis of the Multiple Sequence Alignment

Protein sequences of ribosomal protein S3 (RPS3) and ribosomal protein S12 (RPS12) were retrieved in FASTA format from Ribosomal Protein Gene (RPG) database. The protein identifiers used are NP_001243731 for RPS3 and NP_001007 for RPS12. After that, by using Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) to search each query for RPS sequences against Protein Data Bank (PDB) was performed with parameter to search for templates. The template structures were analyzed with expected value (E-value) above certain threshold which is 0 for high sequence identity. The default threshold for E-value is 10.

For chosen the templates, the sequence identity should more than 80%, query cover 99% and all remaining templates were rejected due to poor score (Rounak *et al*., 2014). The identity of sequence was selected above 60% and query coverage above 90%. This is because some sequence identity is low, but the query coverage is high and this might give a good template. Table 1 shows the sequences producing significant alignment with E-value better than threshold for RPS3 while table 2 shows the sequences producing significant alignment with E-value better than threshold for RPS12.

Description	Max score	Total score	Query cover	E-value	Identity	Accession Number
Chain D, Rabbit 40s Ribosomal Subunit In Complex With Eif1	375	375	100%	2e-131	99%	4KZX_D
Chain c, Structure Of A Mammalian Ribosomal 40s Subunit Within An 80s Complex Obtained By Docking Homology Models Of The RNA And Proteins Into An 8.7 A Cryo-em Map	375	375	100%	2e-131	100%	2ZKQ_CC
Chain D, Structure Of The D. Melanogaster 40s Ribosomal Proteins	346	346	95%	5e-120	84%	3J38_D
Chain B, Localization Of The Small Subunit Ribosomal Proteins Into A 6.1 A Cryo-Em Map Of Saccharomyces Cerevisiae Translating 80s Ribosome	330	330	98%	1e-113	65%	3IZB_B
Chain A, Crystal Structure Of The Yeast Ribosomal Protein Rps3 In Complex With Its Chaperone Yar1	330	330	98%	1e-113	65%	4BSZ_A
Chain E, Crystal Structure Of Narciclasine Bound To The Yeast 80s Ribosome	327	327	98%	9e-113	65%	4UJF_E
Chain D, Cryo-em Structure Of 40s-eif1-eif1a Preinitiation Complex	317	317	96%	1e-108	64%	3J80_D
Chain D, Kluyveromyces Lactis 80s Ribosome In Complex With Crpv-ires	315	315	92%	5e-108	68%	4CUY_D

Table 1. Sequences producing significant alignment with E-value better than threshold for RPS3. The identity of sequences was selected above 60%.

Description	Max score	Total score	Query cover	E-value	Identity	Accession Number
Chain M, Rabbit 40s Ribosomal Subunit In Complex With Eif1.	181	181	100%	2e-58	100%	4KZX_M
Chain M, Structure Of The Human 40s Ribosomal Proteins	178	178	100%	1e-57	98%	3J3A_M
Chain M, Structure Of The 80s Mammalian Ribosome Bound To Eef2 (this Entry Contains The Small Ribosomal Subunit)	167	167	93%	2e-53	98%	4W23_M
Chain M, Localization Of The Small Subunit Ribosomal Proteins Into A 5.5 A Cryo-em Map Of Triticum Aestivum Translating 80s Ribosome	154	154	91%	2e-47	63%	3J60_M
Chain M, Structure Of The D. Melanogaster 40s Ribosomal Proteins	152	152	90%	3e-47	68%	3J38_M

Table 2. Sequences producing significant alignment with E-value better than threshold for RPS12. The identity of sequences was selected above 60%.

The Expected value (E-value) is the number of hits one can expect to see by chance when searching a database of a particular size (Entelechon, n.d.). E-value of 4e-35 is the standard way to write low numbers. For example 1×10^ (-4) = 1e-4 = 0.0001. Mostly, a lower e-value indicates a better quality in the search or alignment. Next, the templates were selected for multiple sequence alignment and aligned with their respective target sequences using ClustalX program. Figure 1 and 2 showed the complete alignment for each RPS3 and RPS12. In the figure, alignments were coloured using ClustalX scheme shows orange: glycine (G), gold: proline (P), blue: small and hydrophobic amino acids (A, L, V, I, M, W, F), green: hydroxyl and amine amino acids (S, N, T, Q), magenta: negative-charged amino acids (D,E), red: positive-charged amino acids (R, K), and dark blue: histidine (H) and tyrosine (Y).

After that, by using PROSITE database each RPs were search for protein domains and motifs. The motifs descriptors used in the PROSITE were either patterns or profiles which were derived from multiple sequence alignments of homologous sequences (Sigrist *et al.*, 2002). Table 3 and 4 showed the possible domain and motifs on the protein RPS3 and RPS12 with the amino acid name such as alanine (A), arginine (R), asparagine (N), aspartic acid (D), cysteine (C), glutamine (Q), glutamic acid (E), glycine (G), histidine (H), isoleucine (I), leucine (L), lysine (K), methionine (M), phenylalanine (F), proline (P), serine (S), threonine (T), tryptophan (W), tyrosine (Y) and valine (V). Based on the table, RPS3 and RPS12 showed the domain and motifs which function as the site of attachment that are responsible in the cell signal transduction pathway. For example, both RPs have casein kinase II phosphorylation site and aids in phosphorylates a large number of substrates containing acidic residues C-terminal to the phosphorylated serine or threonine. Besides, serine phosphorylation was related with regulation of p53 which is a tumor suppressor protein (Keller *et al.,* 2001).

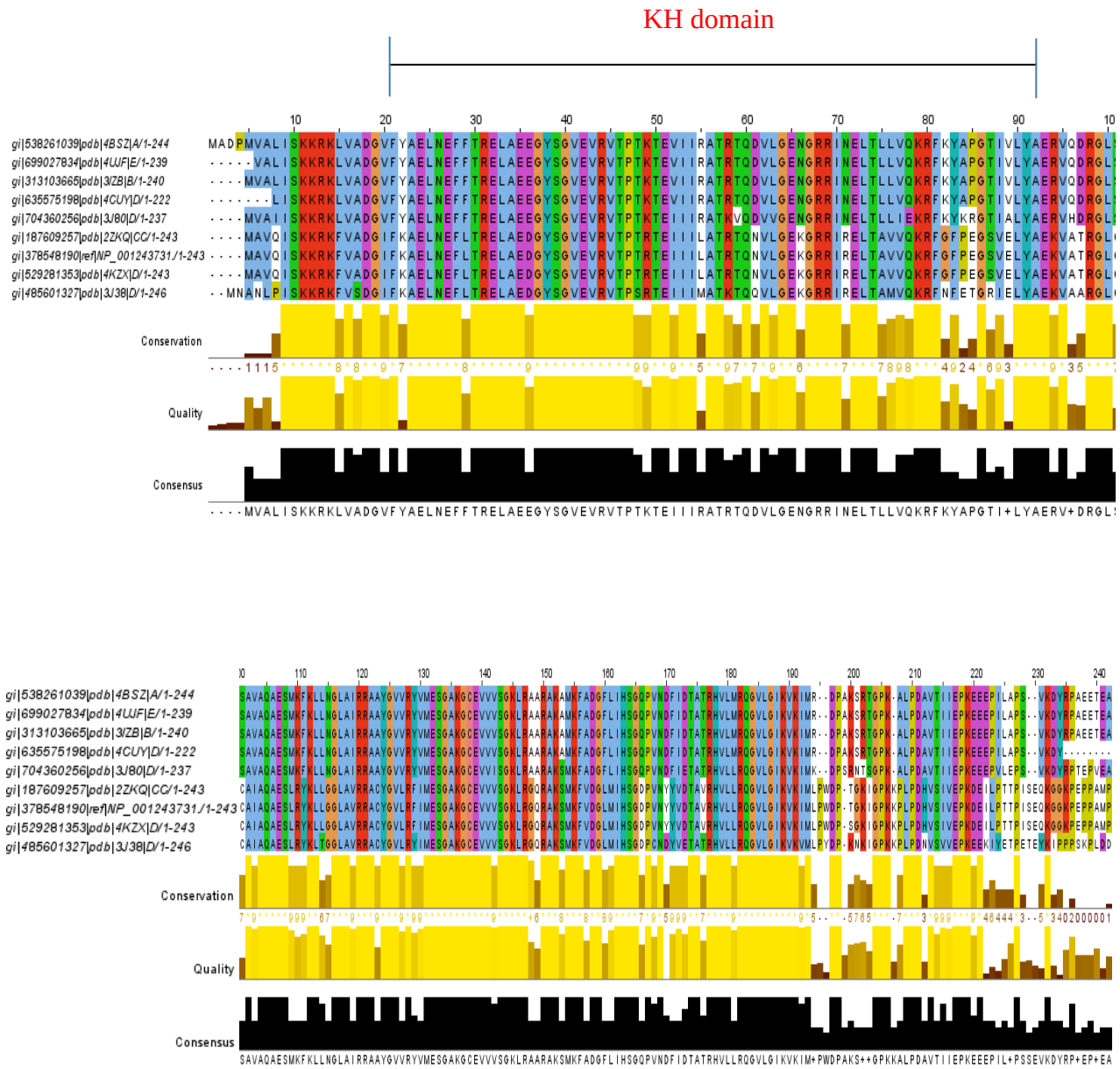

Figure 1. Multiple sequence alignment of template sequences to RPS3 protein sequence. The figure has been viewed using Jalview program (Waterhouse *et al.*, 2009).