Project Report on

Insilico Structural and Functional Annotation of Hypothetical Protein from *Candida auris*

[To complete the degree of Bachelor of Pharmacy, a project report is submitted to the department of pharmacy, Daffodil International University]

Submitted To:

Department of Pharmacy

Faculty of Allied Health Sciences

Daffodil International University

Submitted By:

ID: 191-29-1436

Batch: 21

Department of Pharmacy

Daffodil International University

Date of Submission: April, 2023.

Approval

This project, Insilico structural and functional annotation of hypothetical protein from ***Candida auris***, submitted to the department of Pharmacy, Faculty of Allied Health Sciences, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Pharmacy and approved as to its style and contents.

Board of Examiners:

……..

Professor Dr. Muniruddin Ahmed

Head of the department of Pharmacy

Faculty of Allied Health Science

Daffodil International University

……. **Internal Examiner-I**

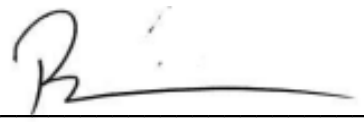
……. **Internal Examiner-II**

……. **Internal Examiner-III**

DECLARATION

I hereby declare that, this project, Insilico structural and functional annotation of hypothetical protein from Candida auris, is done by me under the supervision of **Mohammad Touhidul Islam**, Lecturer (Senior Scale), Department of Pharmacy, Daffodil International University, to complete the requirement for the degree of B.Pharm . I am declaring that this Project is my original work. I also declare that neither this project nor any part thereof has been submitted elsewhere for the award of Bachelor or any degree.

Supervise by



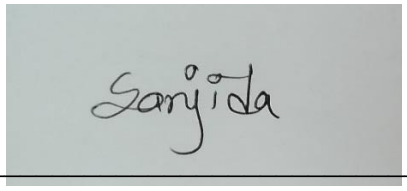
Mohammad Touhidul Islam

Lecturer (Senior Scale)

Department of Pharmacy

Daffodil International University

Submitted by



Sanjida Afrin

ID: 191-29-1436

Batch: 21

ACKNOWLEDGEMENT

I would like to express my profound gratitude to Almighty Allah for giving me enough courage and energy to carry out and complete this project successfully.

Then I would like to express my gratitude to my supervisor **Mr. Mohammad Touhidul Islam**, Lecturer (Senior Scale), Department of Pharmacy, Daffodil International University, for his continuous support and advise that help me to complete this project.

Finally, I would like to thank my parents, my sister and friends for all their encouragement and support during my project study which helped me in completion of this project.

I would also like to show my gratitude to my fellow classmates for all of their support.

DEDICATION

I would like to dedicate my work to the

Almighty Allah,

My Parents, My Sister

and

My supervisor

Abstract:

Candida auris is fungal pathogen that grow as yeast in the family of Saccharomycetaceae. It is worldwide outbound as a multidrug resistant fungal pathogen. This fungi can cause any type of diseases and can also spread easily among people and environment. The infection of this fungi can effect brain, blood, heart and can cause bloodstream infection and even death. So the main objective of this study is to find a structure and function of a hypothetical protein (**GBL47790)** that is important for Candida auris by using in silico method. Many computational tools has been used to identify domain family and function, secondary structure, 3D model and the overall quality of this protein is also checked by computational tools. Functional annotation reveal that this cytoplasmic protein is required for proper rRNA processing and maturation of 28s and 5.8s rRNA and catalyze the formation of peptide bond. This function seems quite important for the cell of *Candida auris*.

So further study about this protein can lead to design an antifungal medication that will treat diseases caused by *Candida auris*.

Keyword: Multidrug resistant, hypothetical protein, computational tools, functional annotation, antifungal medication.

Table of content

Table of Content

Chapter One: Introduction

Chapter Two: Materials and Methods

Chapter Three: Results

Chapter Four: Discussion

Chapter Five: Conclusion

Chapter Six: Reference

List Of Figure

List Of Table -

Chapter One
Introduction

Introduction:

There are some proteins in an organisms whose functions are predicted but a lack of experimental evidence of their existence, is known as hypothetical protein[1]. The sequence of this proteins are known but no experimental studies has been done to express their function. Though their functions are not characterized but these proteins play important role in biological and physiological pathway, to find new structure and functions, biomarkers and physiological targets, early detection for proteomic and genomic study and pharmacological target for drug design. The experimental study of hypothetical proteins showed effective functions in microbes, particularly in pathogens that are associated with human disease[2].

Hypothetical proteins are predicted by nucleic acid sequence and characterized by low identity to known, annotated proteins. Recently several bioinformatic database and tools such as BLASTp, UNIPORT, ProtParam, CELLO, Clustal Omega, PSIPRED, SOMPA, SWISS-Model, PROCHECK, ProSa-web are used to predict effective functions of hypothetical protein in microorganisms[3].

Fungi are eukaryotic organism that include yeast, molds and mushrooms. In current study estimate that, only 200 of the 150,000 fungul species are infectious to human. Most harmful fungi such as Cryptococcus, Aspergillus, Candida auris, Candida albicans etc cause lung infection, skin infection, allergic bronchopulmonary mycoses, Ringworm, Athlete's foot etc. But antifungal drugs are also discover to treat this disease such as clotrimazole, terbinafine, butenafine, miconazole, voriconazole etc.

Candidemia, a fungal infection is caused by some types of *Candida auris*. *Candida auris* was first isolated in Japan, 2009. It is responsible for 30 to 60% bloodstream infection. A recent study reported that more than 30 countries are infected by *Candida auris*.

Candida auris can easily cause nosocomial outbreak in five continent. This fungi has an ability to survive on humans and inert objects[4]. This fungi cause infection in respiratory and urine

specimens but it’s lung and bladder infection is still unclear. This fungi cause many disease such as bloodstream infection, wound infection, ear infection etc.

It is a great concern for all of us because this fungi is multidrug resistant, mode and pace of transmission, hard to identify in laboratory. Some strain of this fungi can resistant all three class of antifungal drugs[5]. Many clinical and public health labs' testing tools use reference datasets that don't fully include C. auris, leading to misdiagnosis[6].

Candida auris has cause major healthcare epidermis because it can transmitted directly from person to person[7]. Contaminated surfaces, medical gadgets, and tools can spread this fungi quickly. In 2009, Candida auris, a fluconazole-resistant strain, was discovered in East Asia[8].

Chapter Two
Material and Methods

Material and methods:

2.1 Sequence retrieval and similarity identification:

Recently 136 genomes assembly and annotation reports of Candida auris are available in NCBI (https://www.ncbi.nlm.nih.gov/) database. All of the hypothetical proteins are obtained from various functional annotation sources for the biological functional classification[9].

In this study, the sequence of hypothetical protein of *Candida auris* (accession no: **GBL47790.1**) contacting 136 amino acid residues. For further research, the protein's primary sequence was obtained in FASTA format[10].

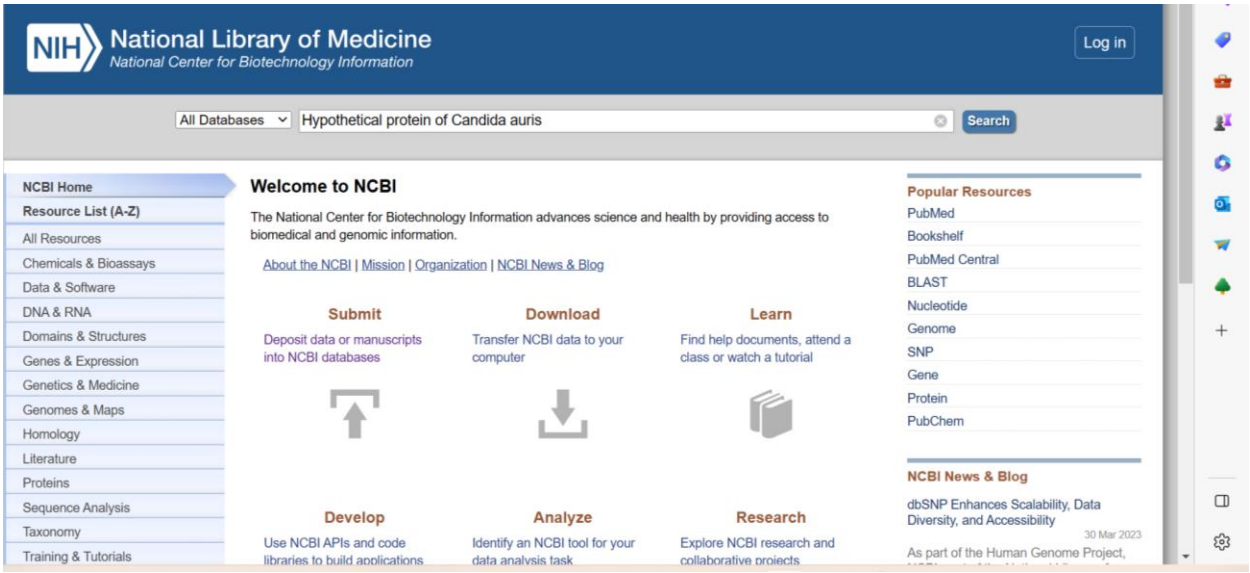


Figure 2.1.1: National Center for Biotechnology Information is used for sequence retrieval

2.2 Analysis of physiochemical properties:

Physiochemical properties of this protein sequence such as number of amino acid, molecular weight, theoretical pH, amino acid composition, atomic composition, estimated half life (should be more than 10 hours, in vivo), instability index (protein is stable if this score is 40 or more than 40), aliphatic index, GRAVY, isoelectric point were determined by using ExPASy ProtParam tool (https://web.expasy.org/protparam/)[11].

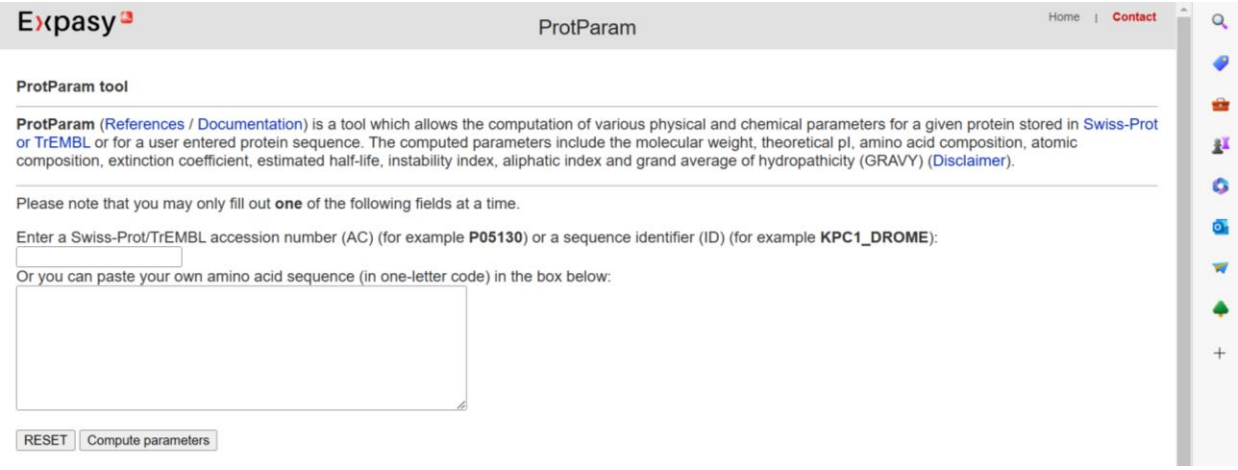


Figure 2.2.1: ExPASy ProtParam tool

2.3 Subcellular localization:

For subcellular localization, PSLpred tool (https://webs.iiitd.edu.in/raghava/pslpred/submit.html) was used to predict the exact position of (**GBL47790.1**) in a cell. PSLpred can accurately predict the location of protein in membrane, extracellular, cytoplasm, mitochondria[12].

CELLO, PSORT II, SOSUI tools were also used to check the result accurately and to calculate the solubility of the protein.

Figure 2.3.1: PSLpred tool for subcellular localization prediction

2.4. Function prediction by conserved domain analysis:

NCBI conserved domain search service (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) was used for function prediction. This confirm the presence of conserve domain in the protein sequence. Pfam (https://pfam.xfam.org/), InterProscan (http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=iprscan5) was also used for domain analysis and function prediction.

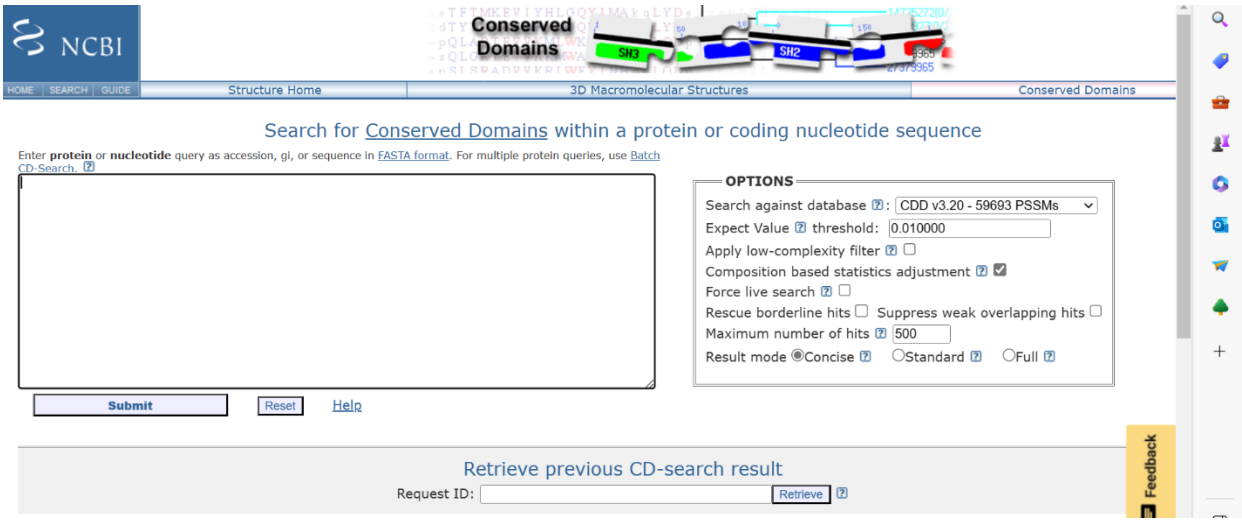


Figure 2.4.1: National Center for Biotechnology Information for conserve domain analysis

2.5 Homology Search:

Homology search was done to find homologues sequence of the protein. For this, BLASTp search tool of NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used against nonredundant database to discover homologues sequence.

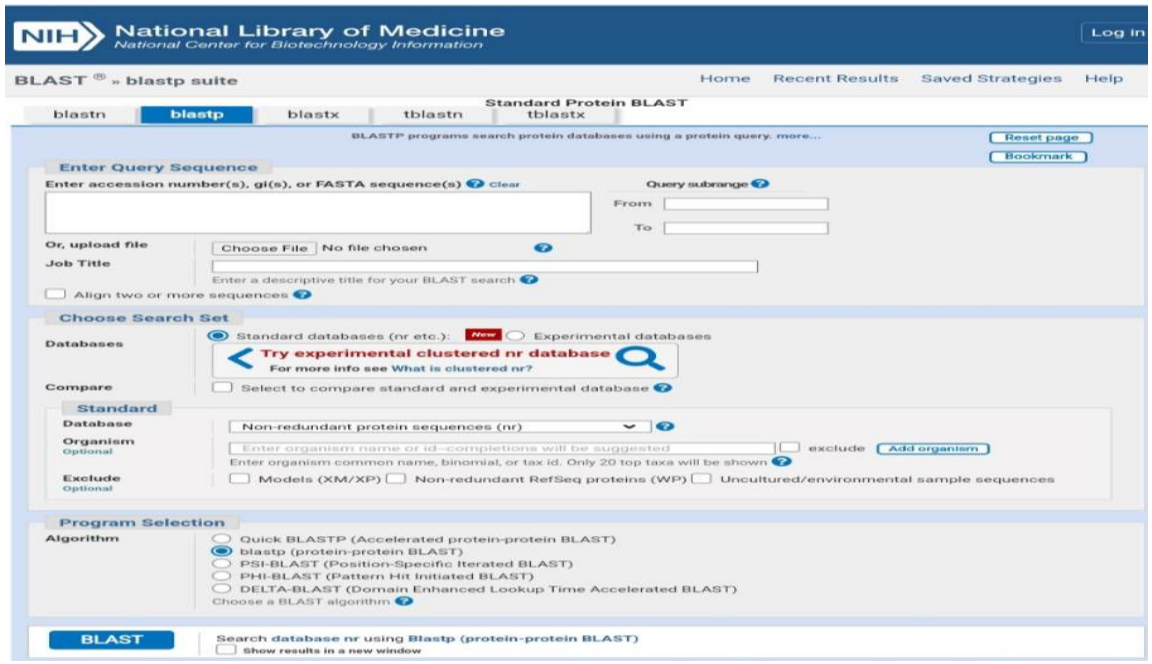


Figure 2.5.1: NCBI BLASTp tool for homology search

2.6. Multiple sequence alignment and phylogenetic tree prediction:

For multiple sequence alignment and phylogenetic tree, clustal omega https://www.ebi.ac.uk/Tools/msa/clustalo/) was used. Clustal Omega is a multiple sequence alignment software that can accurately and efficiently match numerous sequences together using a computer's processing power. It deal with large number of protein sequence for calculating phylogenetic tree[13].

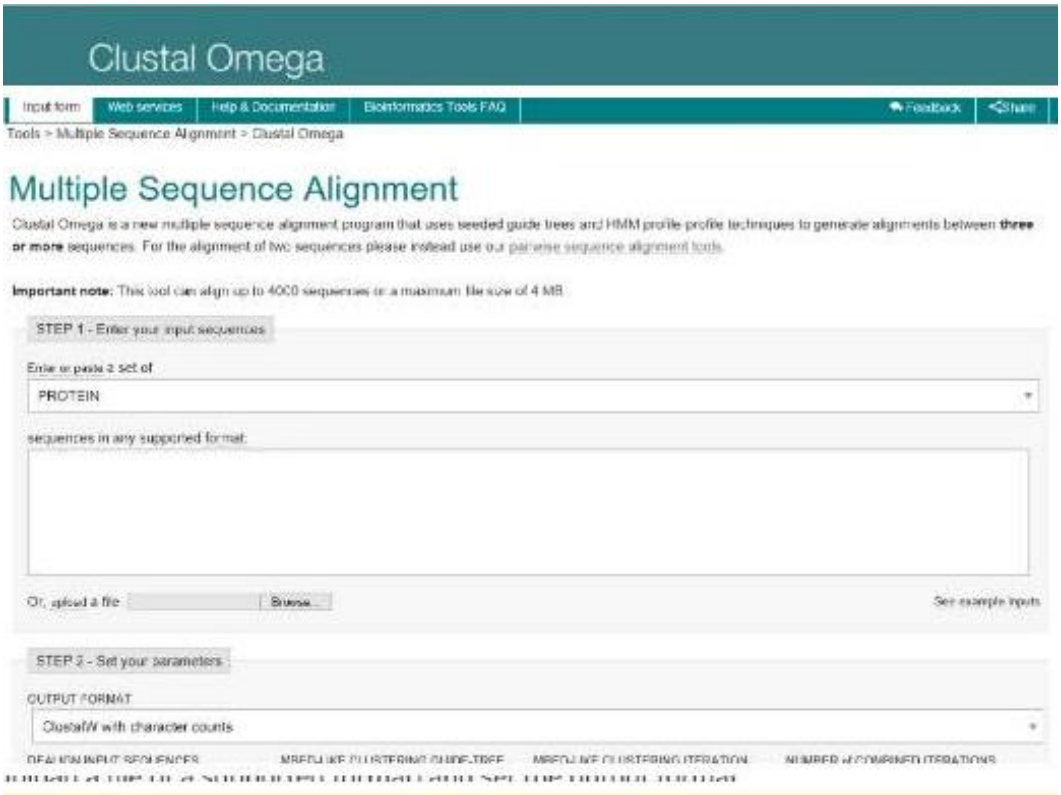


Figure 2.6.1: Clustal omega tool for Multiple sequence alignment and phytogenic tree

2.7 Secondary Structure Determination:

Secondary structure was predicted by using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) (predicted from amino acid sequence) and SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=%2FNPSA%2Fnpsa_sopma.html) . Hypothetical protein **GBL47790.1** was input on FASTA format in PSIPRED and SOPMA.

Figure 2.7.1: PSIPRED tool for predicting secondary structure

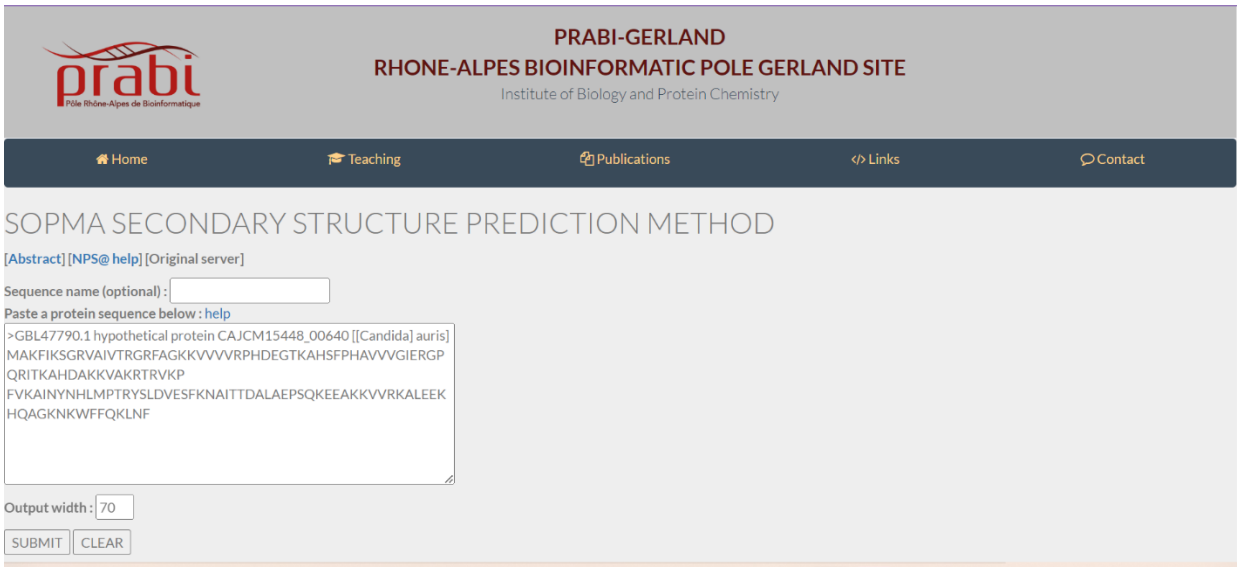


Figure 2.7.2: SOPMA tool for predicting secondary structure

2.8 Tertiary Structure Prediction:

The 3D structure of hypothetical protein (**GBL47790.1)** was predicted by using SWISS-MODEL (https://swissmodel.expasy.org/interactive) server that depend on the similarity between the target protein and available template structure options.

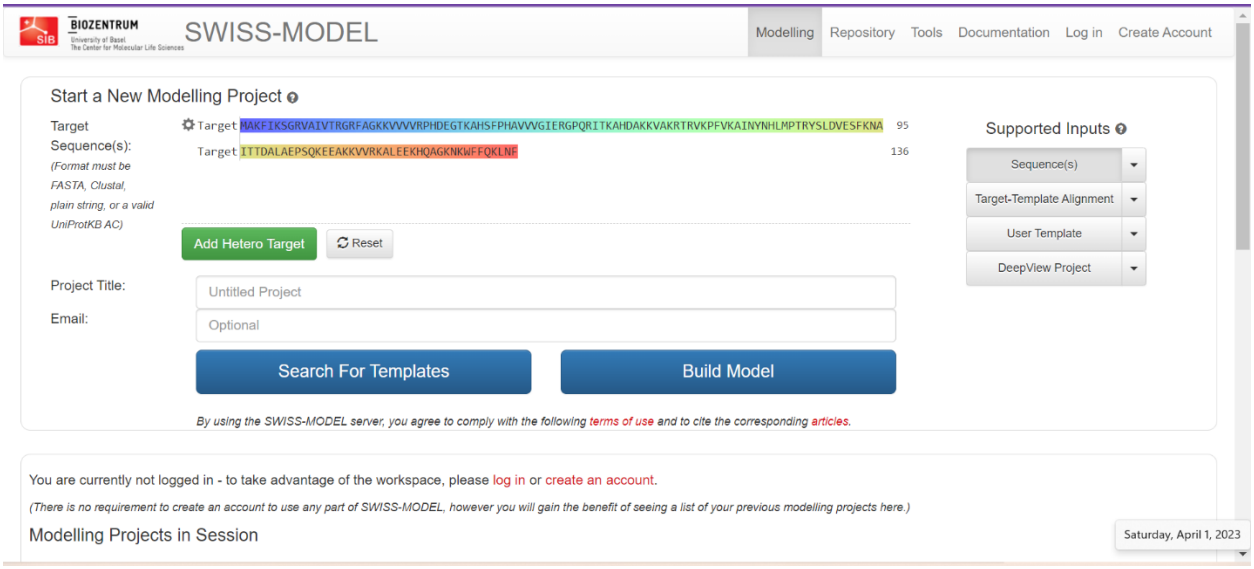


Figure 2.8.1: SWISS_MODEL for predicting 3D structure

2.9 Tertiary Structure Validation:

Tertiary structure validation was done by analyzing Ramachandran plot, Z score (should be more than -5), ERRAT and Verify3D which are evaluated by using PROCHECK (https://saves.mbi.ucla.edu/) and ProSa-web (https://prosa.services.came.sbg.ac.at/prosa.php)

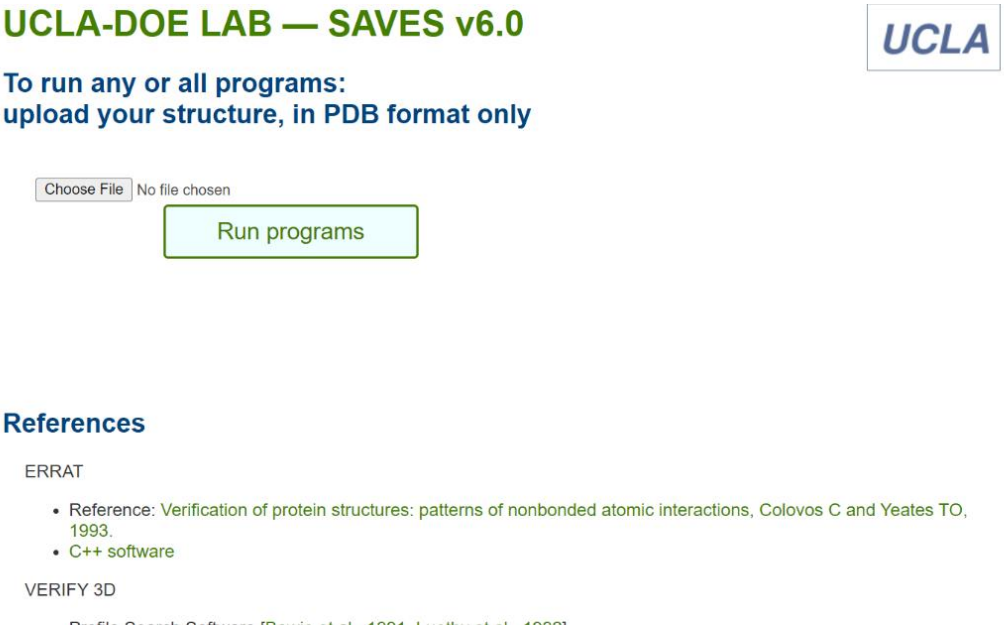


Figure 2.9.1: PROCHECK server check stereochemical quality of protein structure

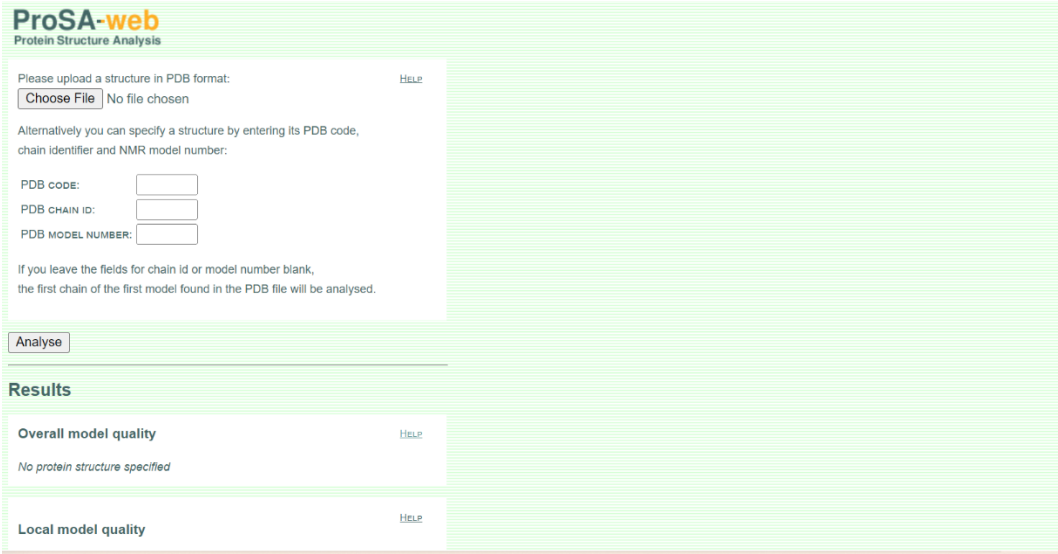


Figure 2.9.2: ProSA-web used for recognizing error in tertiary structure of a protein

Chapter Three
Result

Result:

Sequence retrieval:

The hypothetical protein (**GBL47790.1**)of *Candida auris* has been obtain from National Center for Biotechnology Information (NCBI) and get a protein sequence in FASTA format which is given below:

>GBL47790.1 hypothetical protein CAJCM15448_00640 [[Candida] auris]
MAKFIKSGRVAIVTRGRFAGKKVVVVRPHDEGTKAHSFPHAVVVGIERGPQRITKAHD
AKKVAKRTRVKP
FVKAINYNHLMPTRYSLDVESFKNAITTDALAEPSQKEEAKKVVRKALEEKHQAGKNK
WFFQKLNF

This protein does not have any 3D structure but by using many bioinformatic database and tools this protein sequence was selected which has effective functional properties[14].

3.1 Analysis of physiochemical properties:

The hypothetical protein (**GBL47790.1)** of *Candida auris* has many physiochemical properties which were evaluated by ProtParam tool that are mentioned in the table below. The predicted value of the amino acid is 136, molecular weight of 15418.02, pH value 10.50, **Aliphatic index is** 74.56, grand average of hydropathicity (GRAVY) of -0.576 so the protein is not water soluble and The instability index (II) is 23.83 which classify the protein as stable.

No. of amino Acid no	Molecular weight	Estimated half life	Theoretical pH	Asp + Glu	Arg + Lys	Aliphatic index	Instability index:	Grand average of hydropathicity (GRAVY)
136	15418.02	30 hours	10.50	12	30	74.56	23.83	-0.576

Table 3.1.1: Physiochemical properties of hypothetical protein (**GBL47790.1)**

3.2 Subcellular localization:

Subcellular localization is important for showing their function and used in drug design against target protein. Subcellular localization of this target protein was predicted as **"Cytoplasmic"** by PSLpred. The result of PSORT II and SOSUIGRAMN tool are also given below.

PSLpred result:

Score of Different Subcellular Location	
Localization	Score
Cytoplasm	-0.46327099
Extracellular	-0.62137649
Inner-membrane	-0.64448712
Outer-membrane	-0.69866937
Periplasmic	-0.48862994

Predicted Subcellular Localization

Cytoplasmic Protein

PSORT II result:

```
60.9 %: cytoplasmic
26.1 %: nuclear
 8.7 %: mitochondrial
 4.3 %: peroxisomal

>> prediction for QUERY is cyt (k=23)
```

SOSUIGRAMN result:

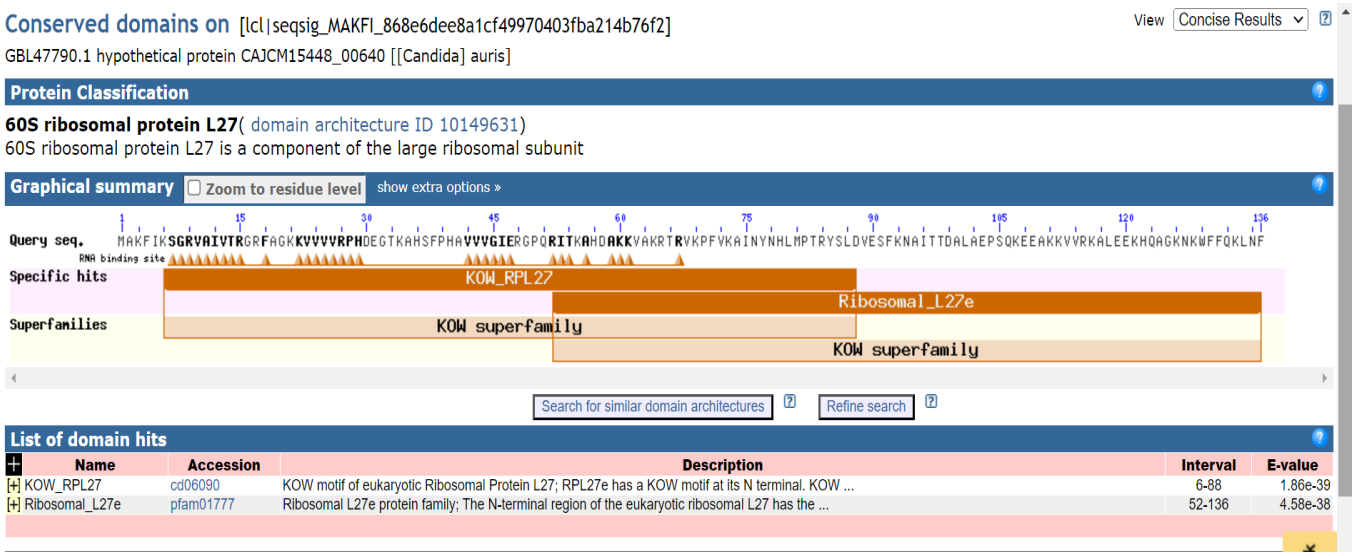
SOSUI$_{GramN}$ Result

No.	seg.Length	subcellular Localization site	ID
0001	136a.a.	C (cytoplasmic)	GBL47790.1 hypothetical protein CAJCM15448_00640 [[Candida] auris]

3.3 Function prediction :

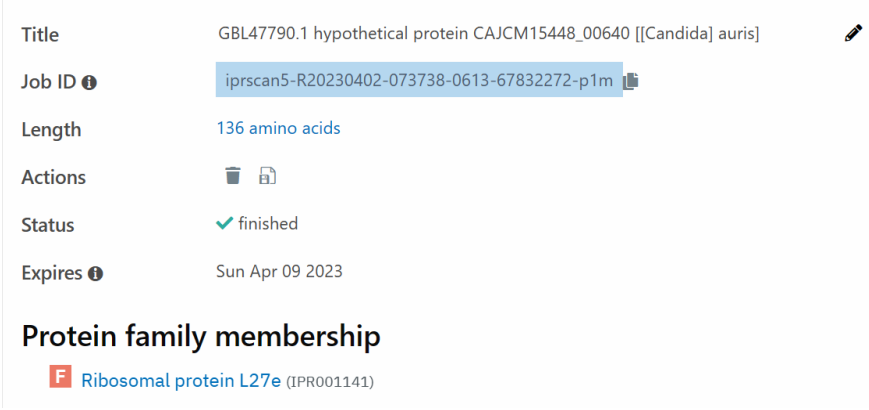
The conserve domain and effective function of this protein is predicted by NCBI-CD. The predicted function is 60S ribosomal protein L27 with an E value of 1.86e-39 and interval 6-88.

60S ribosomal protein L27 is a protein that are encoded by RPL27 gene. This protein are component of the large ribosomal subunit which is required for proper rRNA processing and maturation of 28s and 5.8s rRNA (by similarity). This protein contain ribosomal catalytic site termed the peptidyl transferase center, that catalyze the formation of peptide bond.



This experiment also done by pfam and InterPro server which show same result.

InterPro result:



3.4 Homology searching:

Homology searching was done to find homologues sequence of hypothetical protein by using BLASTp tool from NCBI. Top 10 results of homology searching is shown in the table given below.

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 10

select all *10 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
60S_ribosomal_protein_L27 [[Candida] auris]	[Candida] auris	276	276	100%	2e-93	100.00%	136	XP_028891315.1
60S ribosomal protein L27 [[Candida] pseudohaemulonii]	[Candida] pseudohaemulonii	264	264	100%	1e-88	92.65%	136	XP_024713202.1
60S ribosomal protein L27 [[Candida] auris]	[Candida] auris	257	257	100%	8e-86	94.12%	135	QEL58046.1
hypothetical protein FT662_01981 [[Candida] haemuloni var. vulneris]	[Candida] haemuloni var. vulneris	256	256	100%	2e-85	89.71%	136	KAF3990924.1
60S ribosomal protein L27 [Clavispora lusitaniae]	Clavispora lusitaniae	254	254	100%	7e-85	88.97%	136	KAF5209478.1
60S ribosomal protein L27 [Clavispora lusitaniae]	Clavispora lusitaniae	248	248	100%	5e-82	83.45%	145	KAF7581492.1
CIC11C00000004600 [[Candida] intermedia]	[Candida] intermedia	247	247	100%	6e-82	85.29%	136	SGZ50284.1
hypothetical protein JCM33374_g5772 [Metschnikowia sp. JCM 33374]	Metschnikowia sp. JCM 33374	240	240	100%	4e-79	82.35%	136	GEQ72086.1
hypothetical protein CA7LBN_000066 [[Candida] auris]	[Candida] auris	263	263	96%	1e-78	96.95%	1233	QWW21320.1
hypothetical protein HF325_001866 [Metschnikowia persimmonesis]	Metschnikowia persimmonesis	238	238	100%	3e-78	81.62%	136	KAF8004418.1

Figure 3.4.1: Top 10 result of homology search by using BLASTp tool

3.5 Multiple sequence alignment and Phylogenetic Tree Prediction:

For multiple sequence alignment and phylogenetic tree, clustal omega online tool was used. Top 10 multiple sequence alignment is shown in the figure 3.6.1.

Through phylogenetic tree analysis, we get a protein **XP_028891315.1_1-136 0** which is more similar to our target protein **GBL47790.1**

```
GEQ72086.1:1-136          MVKFIKSGKV---------AIVTRGRFAGKKVVVVRPHDDGTKSHSFSHAIVVGVERGPL  51
KAF8004418.1:1-136        MAKFIKSGKV---------AIVTRGRFAGKKVVVVRPHDDGTKSHSFSHAIVVGVERAPL  51
SGZ50284.1:1-136          MAKFIKSGKV---------AIVTRGRFAGKKVVVVRPHDDGTKSHSFPHAVVVGIERGPL  51
KAF5209478.1:1-136        MAKFIKSGKV---------AIVTRGRFAGKKVVVVRPHDDGTKSHSFPHAVVVGIERGPL  51
KAF7581492.1:1-145        MAKFIKSGKVVLTFILTSLAIVTRGRFAGKKVVVVRPHDDGTKSHSFPHAVVVGIERGPL  60
XP_024713202.1:1-136      MAKFIKSGKV---------AIVTRGRFAGKKVVVVRPHDDGTKAHSFPHALVVGIERGPQ  51
KAF3990924.1:1-136        MAKFIKSGKV---------AIVTRGRFAGKKVVVVRPQDDGTKGHSFPHALVVGIERGPQ  51
QEL58046.1:1-135          MAKFIKSGRV---------AIVTRGRFAGKKVVVVRPHDEGTKAHSFPTSWLS-ALRGPQ  50
GBL47790.1                MAKFIKSGRV---------AIVTRGRFAGKKVVVVRPHDEGTKAHSFPHAVVVGIERGPQ  51
XP_028891315.1:1-136      MAKFIKSGRV---------AIVTRGRFAGKKVVVVRPHDEGTKAHSFPHAVVVGIERGPQ  51
QWW21320.1:1103-1233      -----QSSSN---------AIVTRGRFAGKKVVVVRPHDEGTKAHSFPHAVVVGIERGPQ  46
                               :*.           *****************:*:***.***  : :    *.*

GEQ72086.1:1-136          KVTKAHSAKQLAKRTKVKPFVKLVNYNHLMPTRYSLDVESFKNVVTPETLKEPTQKEEAK  111
KAF8004418.1:1-136        KVTKAHSAKQLAKRTKVKPFVKLVNYNHMMPTRYSLDVESFKNAVTPETLKEPTQKEEAK  111
SGZ50284.1:1-136          KVTRAHDEKKVAKRTKIKPFVKLVNYNHLMPTRYSLDVESFKNVVTPDSLNEPSQREEAK  111
KAF5209478.1:1-136        KVTKAHDAKKVAKRTRIKPFVKLVNYNHLMPTRYSLDVESFKNAVTPDSLNEPSQREEAK  111
KAF7581492.1:1-145        KVTKAHDAKKVAKRTRIKPFVKLVNYNHLMPTRYSLDVESFKNAVTPDSLNEPSQREEAK  120
XP_024713202.1:1-136      RITKAHDAKKIAQRTRVKPFVKVINYNHLMPTRYSLDVESFKNAVTSDALNEPSQREEAK  111
KAF3990924.1:1-136        RITKAHDAKKIAQRTKVKPFVKVINYNHLMPTRYSLDVESFKNAVTSDALNEPSQRVEAK  111
QEL58046.1:1-135          RITKAHDAKKVAKRTRVKPFVKAINYNHLMPTRYSLDVESFKNAITTDALAEPSQKEEAK  110
GBL47790.1                RITKAHDAKKVAKRTRVKPFVKAINYNHLMPTRYSLDVESFKNAITTDALAEPSQKEEAK  111
XP_028891315.1:1-136      RITKAHDAKKVAKRTRVKPFVKAINYNHLMPTRYSLDVESFKNAITTDALAEPSQKEEAK  111
QWW21320.1:1103-1233      RITKAHDAKKVAKRTRVKPFVKAINYNHLMPTRYSLDVESFKNAITTDALAEPSQKEEAK  106
                          ::*:**. *.:*:**.:***** :****:**************.:* ::* **:*: ***

GEQ72086.1:1-136          KVVRKALEEKHQAGKNKWFFSKLNF     136
KAF8004418.1:1-136        KVVRKALEEKHQAGKNKWFFSKLHF     136
SGZ50284.1:1-136          KVVKKALEEKHQAGKNKWFFTKLNF     136
KAF5209478.1:1-136        KVVRKALEEKHQAGKNKWFFTKLNF     136
KAF7581492.1:1-145        KVVRKALEEKHQAGKNKWFFTKLNF     145
XP_024713202.1:1-136      KVVRKALEEKHQAGKNKWFFQKLNF     136
KAF3990924.1:1-136        KVVRKALEEKHQAGKNKWFFQKLNF     136
QEL58046.1:1-135          KVVRKALEEKHQAGKNKWFFQKLNF     135
GBL47790.1                KVVRKALEEKHQAGKNKWFFQKLNF     136
XP_028891315.1:1-136      KVVRKALEEKHQAGKNKWFFQKLNF     136
QWW21320.1:1103-1233      KVVRKALEEKHQAGKNKWFFQKLNF     131
```

Figure 3.5.1: Top 10 multiple sequence alignment get by using Clustal omega

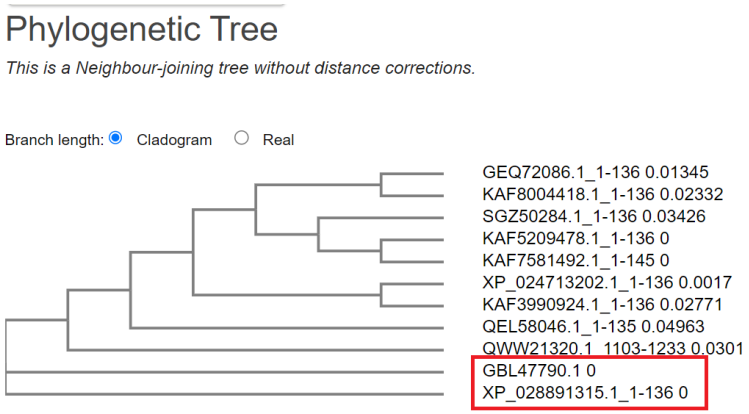


Figure 3.5.2: Phylogenetic tree prediction

3.6 Secondary Structure determination:

SOPMA analysis of the hypothetical protein (**GBL47790.1**) reveled the percentage of alpha helix (40.44), extended strand (22.79%), and beta turn (7.35%), and random coil (29.41%). Secondary structure of this protein is also predicted by PSIPRED, which show similar result.

Predicted secondary structure of this hypothetical protein is given below.

Figure 3.6.1 Secondary structure of hypothetical protein

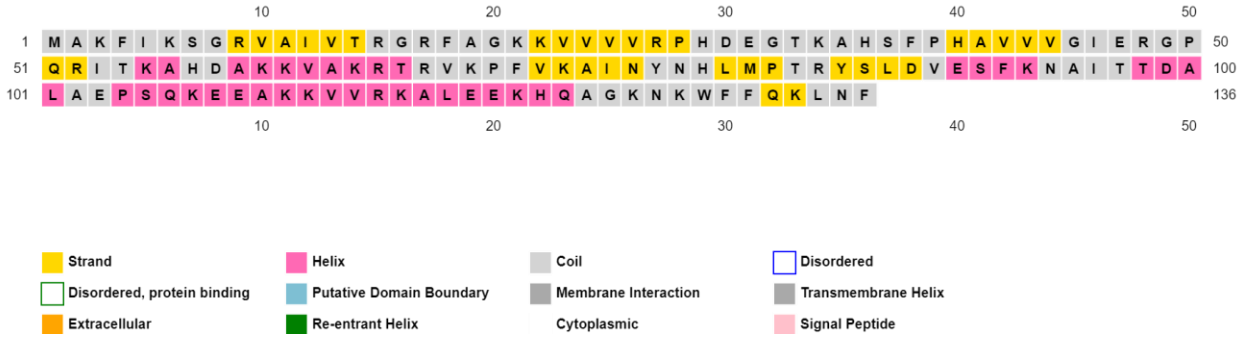


Figure 3.6.2: Individual parts of secondary structure is shown by using PSIPRED

3.7 Tertiary structure prediction:

The 3D structure of hypothetical protein (**GBL47790.1)** was predicted by using SWISS-MODEL, which show 79.41% sequence identity with the target protein. The predicted 3D structure is given below.

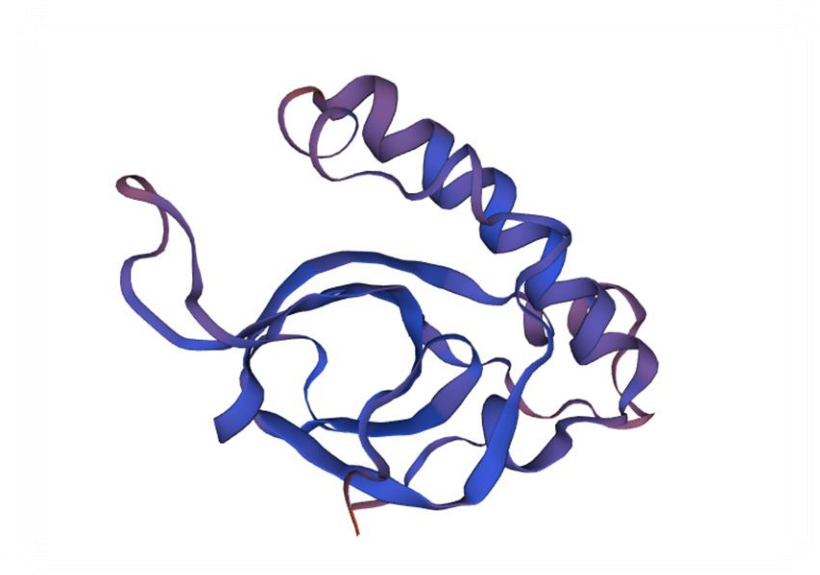


Figure 3.7.1: 3D structure of hypothetic protein

3.8 Tertiary Structure Validation:

PROCHECK, ERRAT, Verify3D evaluated the quality of 3D structure. According to PROCHECK analysis, in Ramachandran plot show 92.5% residues in the most favoured region. In verify 3D tool show 89.63% of the residues have averaged 3D-1Dscore >= 0.1 . ERRAT value of this protein was 98.374% that also predict the quality of 3D structure. Z score measure the total energy and overall model quality of the 3D structure. The predicted value of Z score is -7.24 by using ProSA-web.

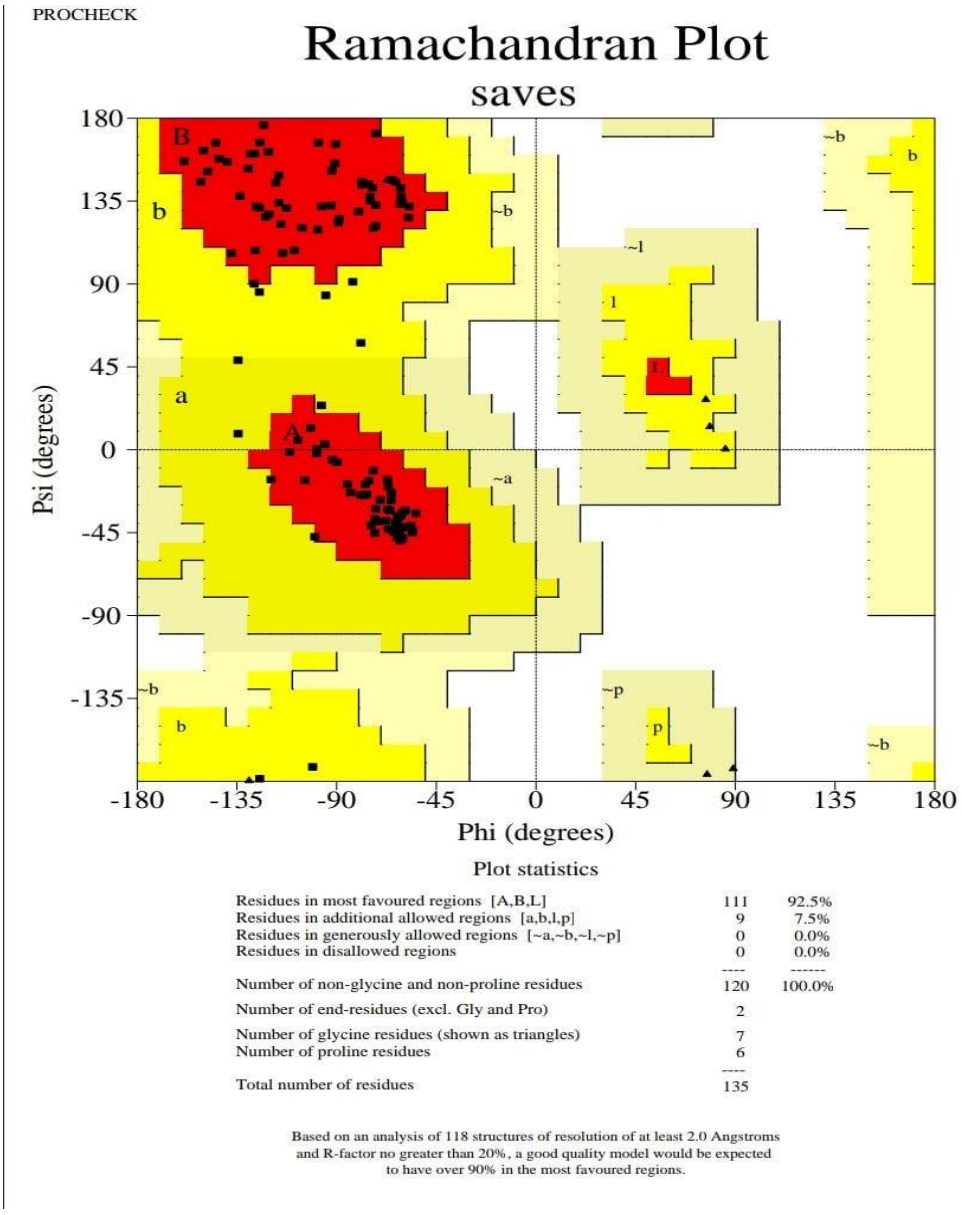


Figure 3.8.1 Ramachandran Plot Statistic

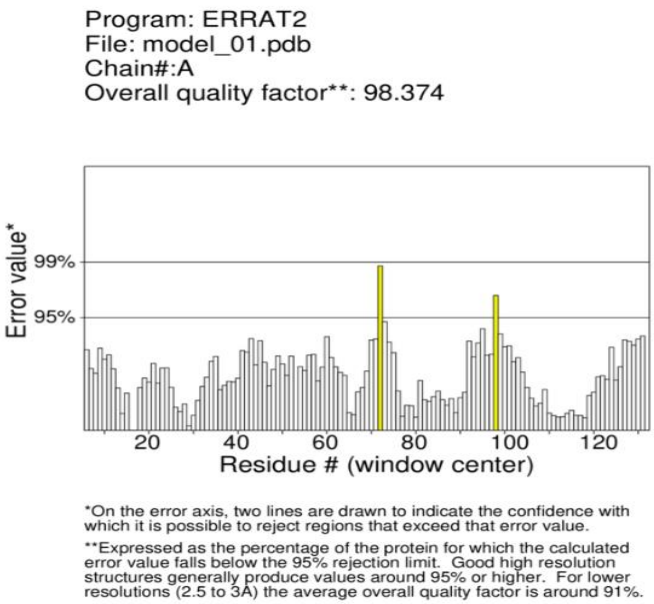


Figure 3.8.2: ERRAT valur is 98.37, yellow color indicate less problematic region, red color indicate problematic region and grey color indicate non problematic region.

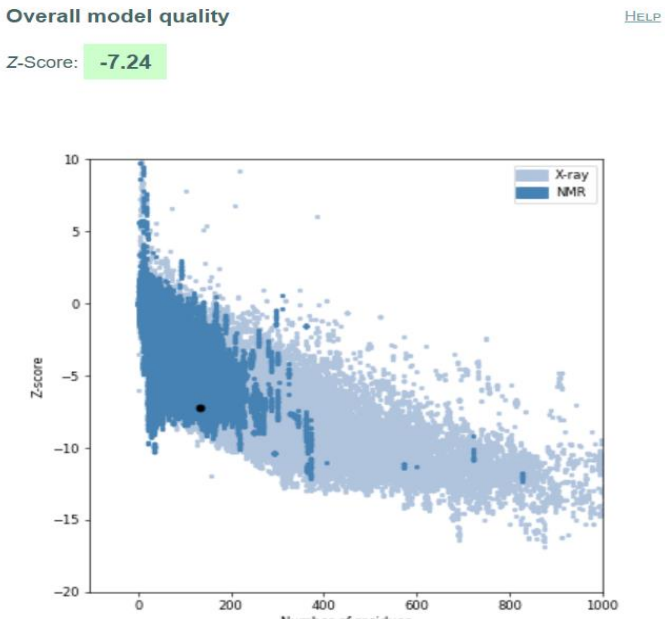


Figure 3.8.3: z-score prediction where HP shows in black dots

Chapter Four
Discussion

Discussion:

In hypothetical protein (**GBL47790)** of *Candida auris* analysis, one or more server was used for each prediction such as ExPASy's ProtParam for physiochemical properties that was estimate to contain 136 amino acid, pH value is 10.50, Aliphatic index is 74.56, grand average of hydropathicity (GRAVY) of -0.576 so the protein is not water soluble and the instability index (II) is 23.83 which classify the protein as stable. The hypothetical protein sequence was obtain from NCBI server. PSLpred serve predict this water insoluble protein that was found in Cytoplasm.

Functional study of this hypothetical protein include protein domain and function prediction. By using NCBI CD, predicted function is 60S ribosomal protein L27 which is required for proper rRNA processing and maturation of 28s and 5.8s rRNA (by similarity) and catalyze the formation of peptide bond with an E value of 1.86e-39 and interval 6-88.

Clustal omega online tool was used for multiple sequence alignment and phylogenetic tree and the target protein GBL47790 got a similar protein XP_028891315.1_1-136 0 so the function can be similar also.

SOPMA and PSIPRED was used to predict the secondary structure of the hypothetical protein. SWISS MODEL was first completely automated homological server that predict 3D structure[15]. SWISS MODEL predict 3d structure with 79.41% sequence identical with the target protein. The 3D model quality was validated by using PROCHECK and ProSA-web.

According to PROCHECK analysis, Ramachandran plot show 92.5% residues in the most favored region which rated as reliable and good. In verify 3D tool show 89.63% of the residues have averaged 3D-1Dscore >= 0.1 which indicate that the model was high quality[16.17]. ERRAT value was 98.374 which indicate that the model has good high resolution [18] . Z score measure the total energy and overall model quality of the 3D structure. The predicted value of Z score is -7.24 which means the score was found within the range.

So by analyzing the hypothetical protein (**GBL47790)** of *Candida auris*, we get to know that it has good quality and function.

Chapter Five
Conclution

Conclusion:

The purpose of the analysis of hypothetical protein (**GBL47790**) of *Candida auris*, prediction of similar function and determine a 3D structure. The prediction of physiochemical properties and subcellular localization helps to understand the character and location of this protein. The predicted function of this protein was 60s ribosomal proteinL27 that helps rRNS processing and maturation and also catalyze the formation of peptide bond which seems very important for cells. If we can conceal the activity of this function in the cell then the cell growth will be inhibited and the cell will die. Further study about this protein's function can help to invent a new anti-fungal drug.

Chapter Six
Reference

Reference:

1. Ijaq J, Malik G, Kumar A, Das PS, Meena N, Bethi N, Sundararajan VS, Suravajhala P. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. BMC bioinformatics. 2019 Dec;20:1-8.

2. Malik G, Agarwal T, Raj U, Sundararajan VS, Bandapalli OR, Suravajhala P. Hypothetical Proteins as Predecessors of Long Non-coding RNAs. Current Genomics. 2020 Nov 1;21(7):531-5.

3. Varma PB, Adimulam YB, Kodukula S. In silico functional annotation of a hypothetical protein from Staphylococcus aureus. Journal of infection and public health. 2015 Nov 1;8(6):526-32.

4. Rhodes J, Fisher MC. Global epidemiology of emerging Candida auris. Current opinion in microbiology. 2019 Dec 1;52:84-9.

5. Schelenz S, Hagen F, Rhodes JL, Abdolrasouli A, Chowdhary A, Hall A, Ryan L, Shackleton J, Trimlett R, Meis JF, Armstrong-James D. First hospital outbreak of the globally emerging Candida auris in a European hospital. Antimicrobial Resistance & Infection Control. 2016 Dec;5(1):1-7.

6. Many clinical and public health labs' testing tools use reference datasets that don't fully include C. auris, leading to misdiagnosis

7. Govender NP, Magobo RE, Mpembe R, Mhlanga M, Matlapeng P, Corcoran C, Govind C, Lowman W, Senekal M, Thomas J. Candida auris in South Africa, 2012–2016. Emerging Infectious Diseases. 2018 Nov;24(11):2036.

8. Spivak ES, Hanson KE. Candida auris: an emerging fungal pathogen. Journal of clinical microbiology. 2018 Feb;56(2):e01588-17.

9. Prabhu D, Rajamanikandan S, Anusha SB, Chowdary MS, Veerapandiyan M, Jeyakanthan J. In silico functional annotation and characterization of hypothetical proteins from Serratia marcescens FGI94. Biology Bulletin. 2020 Jul;47:319-31.

10. Rabbi MF, Akter SA, Hasan MJ, Amin A. In silico characterization of a hypothetical protein from shigella DYSENTERIAE ATCC 12039 reveals a pathogenesis-related protein of the type-VI secretion system. Bioinformatics and Biology Insights. 2021 Apr;15:11779322211011140.

11. Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. Humana press; 2005.

12.] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. Cenk Sahinalp, M. Ester, L.J. Foster, F.S.L. Brinkman, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics 26 (2010) 1608–1615

13. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. Multiple sequence alignment methods. 2014:105-16.

14. Mou M, Islam S, Mahfuj M. In Silico Functional Annotation of VP 128 Hypothetical Protein from Vibrio parahaemolyticus.

15. In silico structural modeling and quality assessment of Plasmodium knowlesi apical membrane antigen 1 using comparative protein models Haron, F.N.1, Azazi, A.1, Chua, K.H.2, Lim, Y.A.L.3, Lee, P.C.4,5, Chew, C.H.1*

16. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics. 2008 Apr;71(1):261-77.

17. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic acids research. 2009 Jul 1;37(suppl_2):W510-4.

18. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511–9.