Analysis of regulatory miRNA on nuclear encoded mitochondrial proteins

A MAJOR PROJECT REPORT in partial fulfillment of the requirement for the degree of

M. Tech (BIMTS)

Submitted by:

Prakriti Khurana

(2K14/BIO/17)

Delhi Technological University, Delhi, India

Under the supervision of:

Dr. Asmita Das

Asst. Professor

Deptt of Biotechnology

Delhi Technological University, Delhi, India

CERTIFICATE

This is to certify that the report entitled **"Computational Analysis of regulatory miRNA on nuclear encoded mitochondrial proteins"** in the fulfillment of the requirements for the reward of the degree of Master of Technology, Delhi Technological University, is an authentic record of the candidate's own work carried out by her under my guidance. The information and data enclosed in this report is original and has not been submitted elsewhere for honoring of any other degree.

Dr. Asmita Das
Asst. Professor
(Project Mentor)
Department of Bio-Technology
Delhi Technological University,
Delhi-110042, India

Dr. D. Kumar
Professor
(Head of Department)
Department of Bio-Technology
Delhi Technological University,
Delhi-110042, India

DECLARATION

I, Prakriti Khurana hereby declare that this thesis entitled **"Computational Analysis of regulatory miRNA on nuclear encoded mitochondrial proteins"** submitted to Delhi Technological University is the original and independent work carried out by me under the guidance of **Dr. Asmita Das**, Department of Biotechnology, Delhi Technological University and **Prof. Pawan Dhar**, Synthetic Biology group, School of Biotechnology, JNU, Delhi, in fulfillment of the requirements for the award of the degree of Master in Technology in Bioinformatics and this thesis or part thereof has not been submitted elsewhere for any other degree or diploma.

(Prakriti Khurana)

ACKNOWLEDGEMENT

Thanks to lord almighty for this valuable thesis as it would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of the study.

I would express my heartfelt thanks to ***Dr. D. Kumar****, HOD, Department of Biotechnology, Delhi Technological University, Delhi for providing me an opportunity to work in this way.*

I would like to render my genuine, heartfelt gratitude to my guide, ***Dr. Asmita Das****, Asst. Professor, Department of Biotechnology, Delhi Technological University, Delhi for her valuable support and encouragement. I would also like to thank her for the opportunity she provided me to work under her guidance.*

I am grateful to thank my co-guide ***Prof. Pawan Dhar****, Professor, Synthetic Biology group, School of Biotechnology, JNU, New Delhi for accepting me as a trainee in his lab. His constant support and guidance has been so fruitful for my work.*

My deep sense of gratitude lies for ***Dr. Kalaiarasan P****, Synthetic Biology group, School of Biotechnology, JNU, New Delhi who spared most of his valuable time in guiding me to this successful work. He taught me from the very basics, every step towards success.*

I also thank the whole faculty of Department of Biotechnology, Delhi Technological University who in one way or another contributed to my growth.

I thank my lab-mates and colleagues for being patient and supportive.

Last but the worthy, I thank my family and friends for always being there for me.

Prakriti Khurana

July 25, 2016

Delhi

CONTENTS

Topic **Page No**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1. ABSTRACT

MicroRNAs (miRNAs) play a vital role in regulating gene expression by inhibiting protein coding genes by repressing mRNA translation or promoting their degradation. Since miRNAs are known as major players in several biological processes, they are being considered as key to better understand, explain, and probably cure not only the pathogenesis of multifactorial diseases but also mitochondrial dysfunction and associated diseases. Several essential cellular functions of mitochondria depend on a high degree of functional interaction between nuclear and mitochondrial genomes. Mitochondria, is known to be the master of an array of cellular processes, may be effected by the intervention of miRNAs. Thus, it becomes important to find out the suspected effects of miRNAs on mitochondrial function as well. The goal of this study is to find out if the nuclear encoded mitochondrial proteins are being targeted by miRNA, and thus affecting their functions. Therefore the list of mitochondrial proteins was retrieved from HMPDb and was manually curated. Also proteins which did not have crystal structures available were modeled using SWISS-MODEL and were filtered out on the basis of high core coverage percentage and sequence identity followed by function prediction from EBI PRoFunc server, which identified sequence motifs from various databases. The modeled structures were then energy minimized on GROMACS, were submitted for plotting Ramachandran plots and subsequently miRNA targets were predicted using mirDB and miRanda to identify the miRNAs that are involved highly in regulation of genes.

CHAPTER 2

2. INTRODUCTION

2.1 The Mitochondrion and its genome

The mitochondrian has all the elements necessary for protein synthesis: it harbors its own genome-in the guise of circular 16.5 kb chromosome and transcription and translation apparatus. However mitochondrial DNA has just 37 genes, encoding 22 mitochondrial tRNAs, 2 mitochondrial rRNAs, and only 13 protein subunits belonging to respiratory complexes I, III, IV and V ; all the subunits of complex II. More than 1000 other proteins needed within the mitochondrian for its proper functioning are transcribed from nuclear genes, synthesized in the cytosol, and then transported into the organelle, where many are then posttranslationally processed before being located into position. The limited size of mtDNA also implies that most of the machinery needed to regulate mitochondrial gene expression is derived from nuclear genes.

Several essential cellular functions of mitochondria depend on a high degree of functional interaction between nuclear and mitochondrial genomes. Thus, communication from the nuclear genome to the mitochondrion involves proteins that are translated in the cytosol and imported into the mitochondrion.
To date, the mechanisms underlying the coordinated expression of mitochondrial-encoded and nuclear gene–encoded transcripts remain to be fully elucidated.

2.2 MicroRNA and gene expression

MicroRNAs (miRNAs) constitute a class of non-coding RNAs that play key roles in the regulation of gene expression. These fascinating molecules act at the post transcriptional level, and may fine –tune the expression of as much as 30% of all mammalian protein encoding genes.
Mature microRNAs are short, single stranded RNA molecules approximately 22 nucleotides in length. MicroRNAs are sometimes encoded by multiple loci, some of which are organized in tandemly co-transcribed clusters.

RNA polymerase II transcribe microRNA genes as large primary transcripts (pri-microRNA) that are processed by a protein complex which contains the RNase III enzyme Drosha, to form

an approximately 70 nucleotide precursor microRNA (pre-microRNA). This precursor is subsequently transported to the cytoplasm where it is processed by a second RNase III enzyme, DICER, to form a mature microRNA of approximately 22 nucleotides long. The mature microRNA is then incorporated into a ribonuclear particle to form the RNA-induced silencing complex, RISC, which mediates gene silencing.

MicroRNAs usually induce gene silencing by binding to target sites found within the 3'UTR of the targeted mRNA. This interaction prevents protein production by suppressing protein synthesis and/or by initiating mRNA degradation. Since most target sites on the mRNA have only partial base complementarity with their corresponding microRNA, individual microRNAs may target as many as 100 different mRNAs. Moreover, individual mRNAs may contain multiple binding sites for different microRNAs, resulting in a complex regulatory network.

2.3 The function of microRNAs

MicroRNAs have been shown to be involved in a wide range of biological processes such as cell cycle control, apoptosis and several developmental and physiological processes including stem cell differentiation, hematopoiesis, hypoxia, cardiac and skeletal muscle development, neurogenesis, insulin secretion, cholesterol metabolism, aging, immune responses and viral replication. In addition, highly tissue-specific expression and distinct temporal expression patterns during embryogenesis suggest that microRNAs play a key role in the differentiation and maintenance of tissue identity.

Dysfunction of mitochondria is related to a variety of pathological processes and diseases. Changes in mitochondrial function and ultrastructure have been observed in many disorders; and mitochondrial DNA (mtDNA) mutations and deletions arise as a result of dysfunction of the mitochondrial respiratory chain and contribute to the aging process. The decreased expression of mitochondrial genes and disrupted mitochondrial structure has been associated with abnormal human development.

CHAPTER 3

3. REVIEW OF LITERATURE

3.1 Brief history of MiRNA

The discovery of miRNAs and the mRNAs targeted by them has unveiled novel mechanisms that regulate gene expression beyond the central dogma. MiRNAs are basically noncoding small RNAs which are not translated into proteins. Instead the genes encoding for miRNAs are transcribed from DNA to produce a primary transcript (pri-miRNA) that is processed into a mature, single stranded miRNA that is 18-24 nucleotides long. A mature miRNA binds to its mRNA target at their complementary sequences to downregulate gene expression by inhibiting their RNA translation to proteins or by inducing mRNA degradation.

3.2 Crosstalk between Nuclear and Mitochondrial Genome

The inventory of nuclear and mitochondrially coded proteins required to assemble a functional mitochondrion shows clearly that nuclear and mitochondrial genomes interact in two ways. First, both nuclear and mitochondrial genes contribute to mitochondrial protein function. Second, both nuclear and mitochondrial genomes interact to affect the synthesis and assembly of mitochondrial proteins. Thus, Communication from the nuclear genome to the mitochondrion involves proteins that are translated in the cytosol and imported into the mitochondrion.

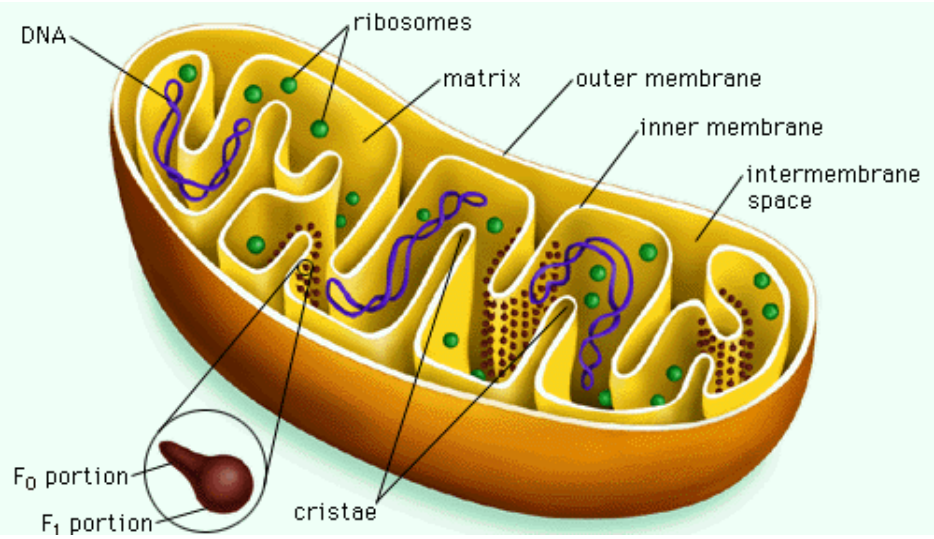


Fig 3.2.1: Mitochondria (Source: http://www.mthira.vic.edu.au/

3.3 MiRNA association with mitochondria

It is well established that miRNAs are transcribed in nucleus and transported into cytoplasm, which is a primary site of action. However, there is growing evidence that miRNAs also are present in or associated with other organelles or unstructured cytoplasmic foci, such as mitochondria, endoplasmic reticulum (ER), processing bodies (P-bodies), stress granules, multivesicular bodies, and exosomes (6). It can be suggested from these recent observations that miRNA-mediated gene regulation may be controlled within different cellular compartments, and that this miRNA-organelle crosstalk allows for more selective responses to specific cellular demands. Several studies have demonstrated the presence of miRNAs either within (11) or associated with (7) mitochondria isolated from various cell types and tissues, including the CNS (8). In addition, the miRNA machinery proteins, Argonaute (AGO) and Dicer have been detected in mitochondria (9), indicating the presence of an active miRNA ribonucleoprotein complex (miRNP). The majority of these mitochondria associated miRNAs are known to be nuclear-encoded, while a few are predicted to originate from the mitochondrial genome (10). Given the small genome of mitochondria and the presence of minimal non-coding DNA, the characterization and function of this group of intra-mitochondrial miRNA requires further investigation.

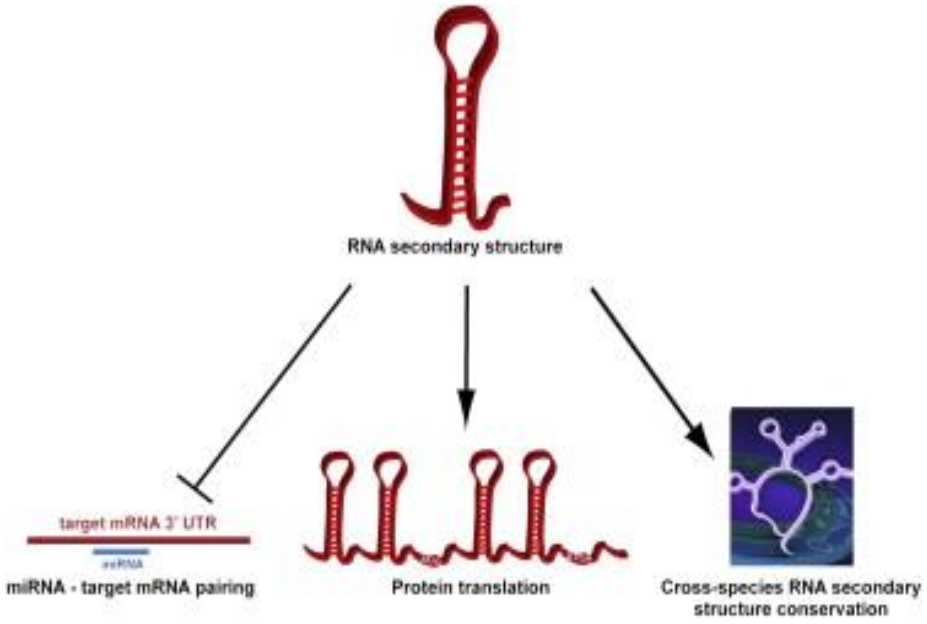


Fig 3.3.1: MicroRNA Structure (6)

One of the presumed functions of mitochondrial miRNA is the regulation of mitochondrial gene expression. In support of this, studies performed using muscle cells and heart tissues have identified nuclear-encoded miRNAs in mitochondria that directly regulate mitochondrial proteins. For example, miR-181c is enriched in mitochondria and it targets cytochrome c oxidase subunit 1 (COX1) in rat cardiomyotubes (12). Another nuclear-generated mitochondrial miRNA, miR-1 also targets COX1 and was found to increase in mitochondria during muscle cell differentiation (13). Surprisingly, miR-1 enhanced, rather than inhibited, COX1 protein translation. These same authors determined that this unconventional action of miR-1 requires AGO2 but not GW182 (glycine-tryptophan protein of 182 kDa), which is essential for cytoplasmic miRNA gene repression but is absent in mitochondria. In a separate study, computational analysis of mitochondria enriched miRNAs isolated from human skeletal muscle cells predicted 80 putative target sites in the mitochondrial genome (14). There is additional evidence supporting the role of miRNA in regulating mitochondrial gene expression and readers are referred to several excellent reviews (15). The remainder of this Perspective focuses on the role of mitochondria in regulating cellular miRNA activities.

3.4 Mitochondria are prime candidates for regulating miRNA function

Regulation of gene expression is a tightly controlled process that involves many factors both at the transcription and post-transcriptional levels. MiRNAs broadly participate in post-transcriptional gene regulation in almost all cellular events. MiRNAs regulate gene expression by complementary binding to their target transcript, with the critical region for miRNA binding being nucleotides 2–8 from the 5′ end of the miRNA, termed the ‘seed’ sequence (16). In addition to sequence matching, the thermodynamics of miRNA:mRNA interactions and the accessibility of the target mRNA are thought to be important factors (17). Nevertheless, the mechanisms by which a given miRNA determines its target and the events controlling miRNA function in response to cellular demands are not well understood.

Mitochondria have been shown to interact with several cellular compartments, organelles, and cytoplasmic foci including the cytoskeleton, ER, and P-bodies (18). These interactions create a dynamic network important for cellular energy distribution, signaling, and homeostasis. Given the rapid response capabilities of mitochondria in many cellular functions, we suggest that mitochondria are prime candidates for regulating miRNA activity and function. This is an

attractive hypothesis as it allows for control of translational specificity in response to unique cellular requirements across cellular domains. In this context, mitochondria can be viewed as local rheostats that respond quickly to changes in cellular demands, both physiological and pathophysiological. As such, this hypothesis postulates that appropriate miRNA responses are determined, in part, by rapid changes in mitochondrial function due to cellular demands, stresses or perturbations. In support of this, compromising mitochondrial function with an uncoupling agent has been shown to result in delocalization of AGO proteins from P-bodies leading to a subsequent decrease in miRNA mediated RNAi efficiency (19).

3.5 Potential roles of mitochondria in regulating miRNA functions

The association of miRNAs with mitochondria raises the possibility that mitochondria regulate miRNA activities in a manner that is specific to unique cellular demands. As such, we hypothesize that mitochondria can act in several ways including but not limited to: 1) Mitochondria, the warehouse; 2) Mitochondria, the vehicle; and 3) Mitochondria, the network.

Table 3.5.1 MicroRNA target prediction tools (http://www.exiqon.com)

Method	Type of Method	Method Availability	Resource
Stark et. al	Complementary	Online search	http://www.russell.embl.de/miRNAs
miRanda	Complementary	Download	http://www.microrna.org
miRanda MiRBase	Complementary	Online search	http://microrna.sanger.ac.uk
miRWalk	-	Online search	www.umm.uni-heideberg.de/apps/zmf/mirwalk/index.html
Target Scan	Seed Complementary	Online search	http://www.targetscan.org
DIANA mT	Thermodynamics	Download	http://diana.cslab.ece.ntua.gr/
PicTar	Thermodynamics	N/A	http://pictar.mdc-berlin.de/
RNAHybrid	Thermodynamics & Statistical model	Download	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid

miRGen++	Baynesian Inference	Mathlab Code	http://www.psi.toronto.edu/genmir
MiTarget	Support Vector Machine	Online search	http://cbit.snu.ac.kr/~miTarget
MiRtaget2	Support Vector Machine	Online search	http://mirdb.org
TarBase	Experimentally Validated Targets	N/A	http://diana.cslab.ece.ntua.gr/tarbase/

CHAPTER 4

4. MATERIALS AND METHODS

The materials and methods opted for accomplishing the desired results are:

4.1 Data retrieval and curation

Data for human mitochondrial proteins was retrieved from **Human Mitochondrial Protein Database (HMPDb)**, and the proteins were mapped to the **Uniprot Knowledgebase (UniprotKB)**. The resultant list of proteins was then manually curated to get the significant data. It was also scrutinized on the basis of availability of structures. MirBase was used to get the MiRNA data.

4.1.1 HUMAN MITOCHONDRIAL PROTEIN DATABASE (HMPDb)

The Human Mitochondrial Protein Database (HMPDb) from NIST provides comprehensive data on mitochondrial and human nuclear encoded proteins involved in mitochondrial biogenesis and function. This database consolidates information from SwissProt, LocusLink, Protein Databank (PDB), GenBank, Genome Database (GDB), Online Mendelian Inheritance in Man (OMIM), Human Mitochondrial Genome Database (mtDB), MITOMAP, Neuromuscular Disease Center and Human 2-D PAGE Databases. The mitochondrian plays a central role in cellular metabolism, and evidence of mitochondrial involvement in a number of different human diseases in increasing. This database is intended as a tool not only to aid in studying the mitochondrion but in studying the associated diseases.

4.2.2 UniProt Knowledgebase (UniprotKB)

It is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information is possible is added. UniProtKB consists of two sections: **Reviewed** (manually annotated) and **Unreviewed** (computationally analysed). Only reviewed entries were considered.

4.2 Protein Modeling

After scrutinizing the proteins for the structures available, proteins with no PDB structures were homology modeled using SWISS-MODEL and other necessary information was also retrieved for each modeled protein from the output HTML files.

4.2.1 SWISS-MODEL

It is a structural bioinformatics web-server which does homology modeling of protein 3D structures. Homology modeling is currently the most accurate method to generate reliable three-dimensional protein structure models and is routinely used in many practical applications. Homology (or comparative) modeling methods make use of experimental protein structures ("templates") to build models for evolutionary related proteins ("targets").

4.3 Function Prediction

Modeled structures were subjected to function prediction by Profunc server of EBI-EMBL. Results of **InterPro** scan for sequence motifs were analysed.

4.3.1 PROFUNC

ProFunc is a web server for predicting the likely function of proteins whose 3D structure is known but whose function is not. Users submit the coordinates of their structure to the server in PDB format. ProFunc makes use of both existing and novel methods to analyse the protein's sequence and structure identifying functional motifs or close relationships to functionally characterized proteins. A summary of the analyses provides an at-a-glance view of what each of the different methods has found.

4.3.2 INTERPRO SCAN

The InterPro scan search identifies sequence motifs from several databases, including PROSITE, GENE 3D, PRINTS, PFam-A, TIGERFAM, PROFILES and PRODOM. It shows the results diagrammatically and in the tabular form.

4.4 Energy Minimization using GROMACS

The modeled proteins were again filtered on the basis of sequence identity and core coverage. Only the proteins with a considerable high score were taken further i.e. the proteins with more than 90 percent sequence identity and more than 80 percent core coverage.

They were then energy minimized using GROMACS (GROningen MAchine for Chemical Simulations). It is a user friendly versatile package to perform the energy minimization by applying the Newton's equations of motion to calculate trajectories of particles, starting from a defined configuration. For each particle in the system, the total force acting on it is calculated from the interactions with other particles, described by the force field. The force divided by the mass of the particle gives the acceleration, which, together with the prior position and velocity, determines what the new position will be after a small time step.

4.4.1 SEQUENTIAL STEPS IN GROMACS ENERGY MINIMIZATION

Step1

The pdb file was first converted to the GROMACS readable file with the extension .gro where the topology is written for that file. Pdb2gmx command is used for this purpose.

Step2: editconf

In the next step the dimensions of the box where the initial simulation has to be done is set with the simple cubic periodic box type which is set according to the protein structure. As the protein is dynamic in structure, the distance from the box to the protein is set to 1nm (i.e. 10 angstroms from the molecule periphery) and the molecule is centered in the box by using –c flag. The GROMACS force field G43a1 was used for the minimization.

Step3: genbox

By using the genbox command, the box is solvated by using the Simple Point Charge (spc216) water model according to the dimensions set.

Step4: grompp

Grompp is a preprocessor program to set up the run for the input file to mdrun. The .mdp file for the energy minimization was set with the parameters of 50000 steps by steepest descent minimization method in the topology file giving the output file containing the ions to be neutralized in the following step.

Step5: genion

The water molecules were replaced by using counter ions to balance the charge by setting the system to net neutral where the tpr file is used to add ions to neutralize.

Step6: grompp

Energy minimization was performed for the resulted file in the previous step. The –f flag inputs the parameter file (.mdp). The –c flag is used to input the coordinate file (the pdb file), -p inputs the topology and –o outputs the input file (.tpr) needed for mdrun.

Step 7: mdrun

This is the first dynamics step for the .tpr file to be run with specified parameters in preceding steps.

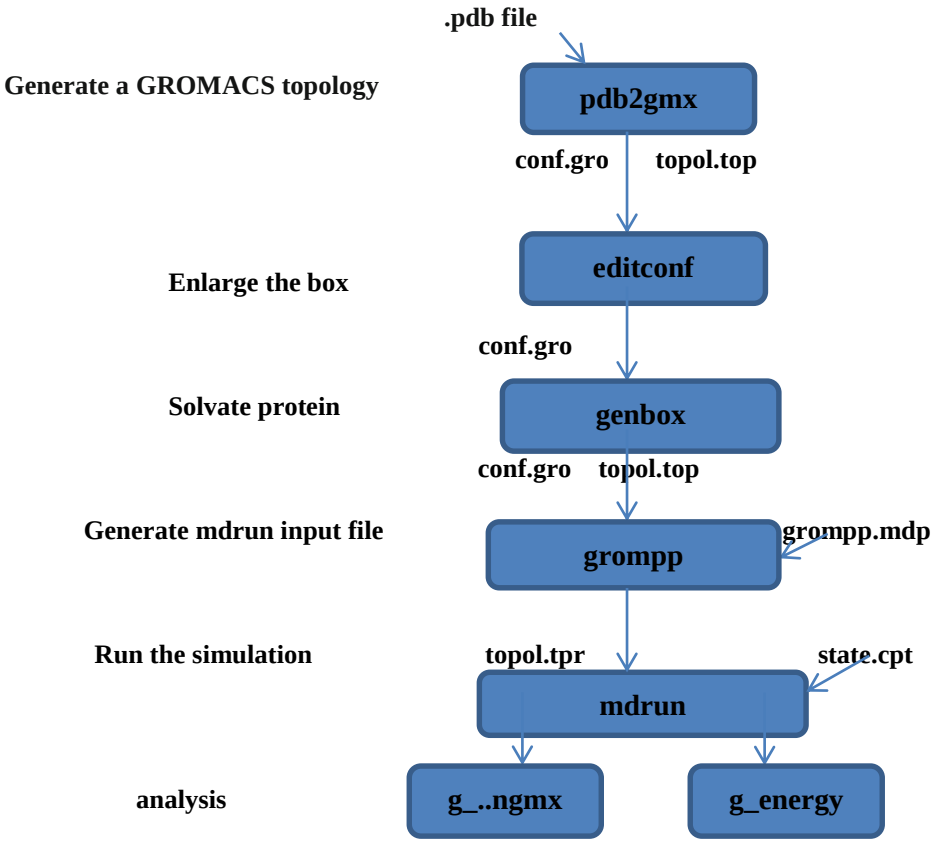


Fig 4.4.1.1: Sequential steps in minimization using GROMACS

4.5 Ramachandran plots

PDB files of the protein structures were then submitted to **PDBsum generate** of EBI to generate a full set of PDBsum structural analyses for it. Ramachandran plots were then downloaded for each of the protein and analysed.

4.6 MirNA Target Prediction

Identification of binding sites of microRNA was done using number of microRNA prediction tools listed below:

4.6.1 miRanda: miRanda is standalone tool which run on Linux Operating System. It is an algorithm written in C for finding genomic targets for microRNAs. It uses miRNA and 3'-UTR sequences for finding the targets.

4.6.2 mirDB: It is an online microRNA prediction tool. For prediction it takes miRNA name or Gene name as input. Pre predicted results can also be downloaded from mirDB.

CHAPTER *5*

5. RESULTS AND DISCUSSION

5.1 Data retrieval and curation

After manually curating the entries from HMPDb database, and mapping them to respective PDB structure IDs in UniProtKnowledgebase, it was found that among a total of 516 UniProt IDs, 233 had crystal structures, whereas for 283, structures were NOT available.

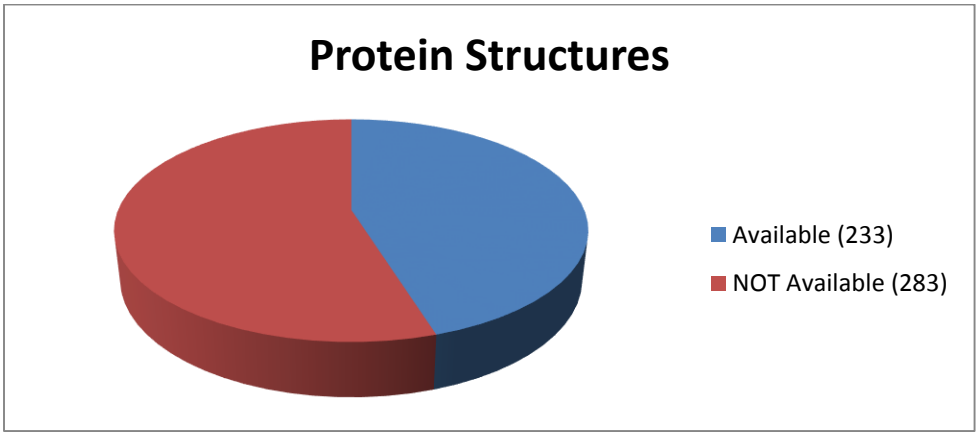


Fig 5.1.1: Pie chart depicting proteins structures' availability

280 proteins out of 283 were successfully modeled using SWISS-MODEL.

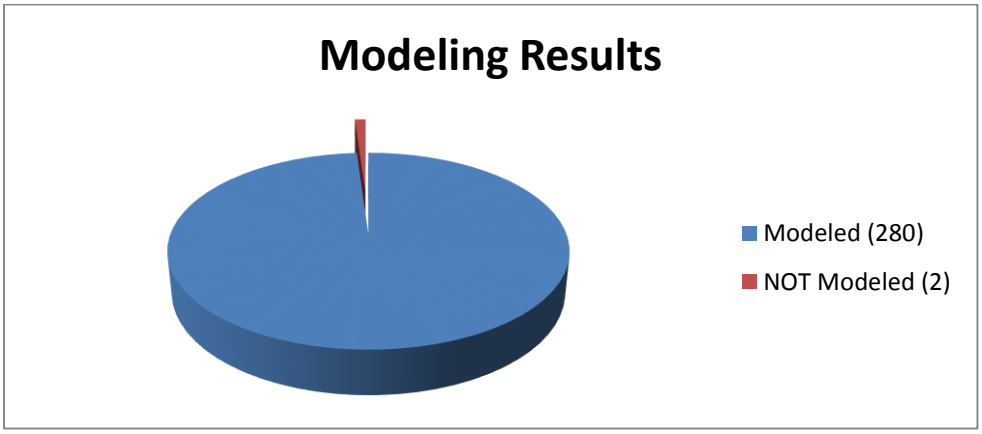


Fig 5.1.2: Pie chart depicting modeling statistics

135 of them showed less than 35% **sequence identity** whereas 145 showed more than or equal to 35% identity.

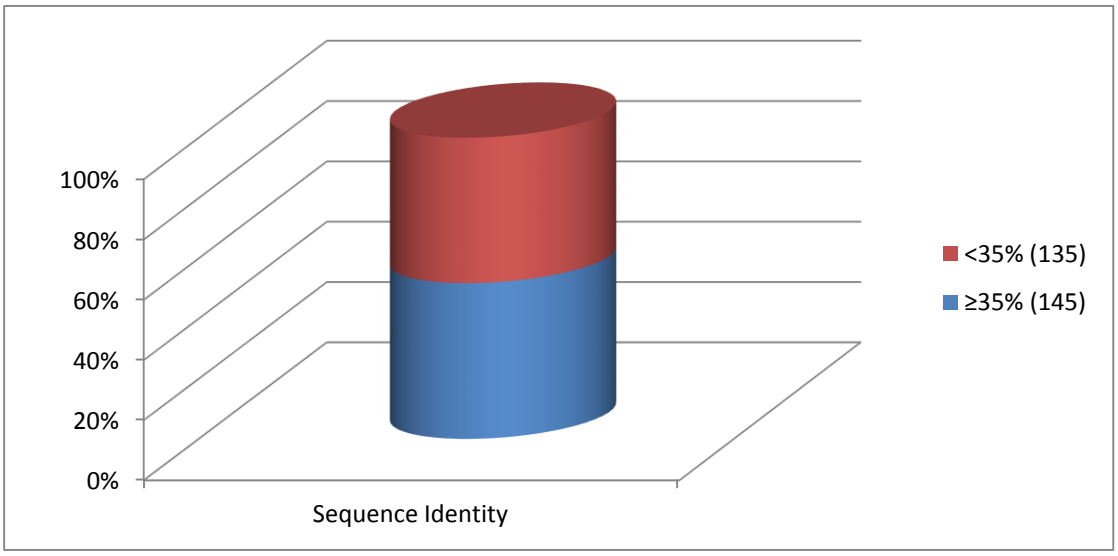


Fig 5.1.3: Graph depicting Sequence Identity Statistics

The start and end codon length was known only for 187 such proteins, out of them 26 showed less than 40% core coverage, whereas 161 showed more than or equal to 40% core coverage.

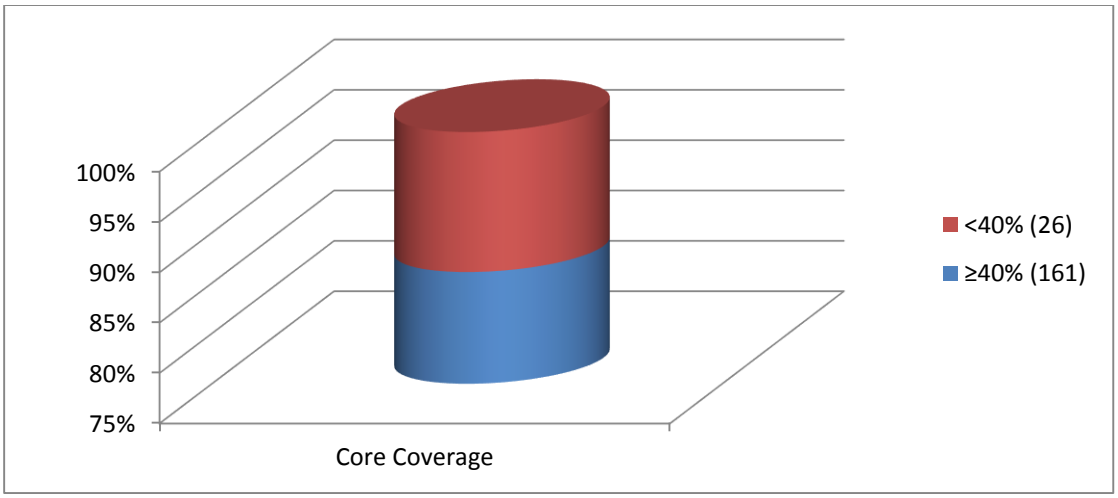


Fig 5.1.4: Graph depicting Core Coverage Statistics

Also 129 of them showed good core coverage (more than or equal to 70%), and 58 showed less than 70% core coverage.

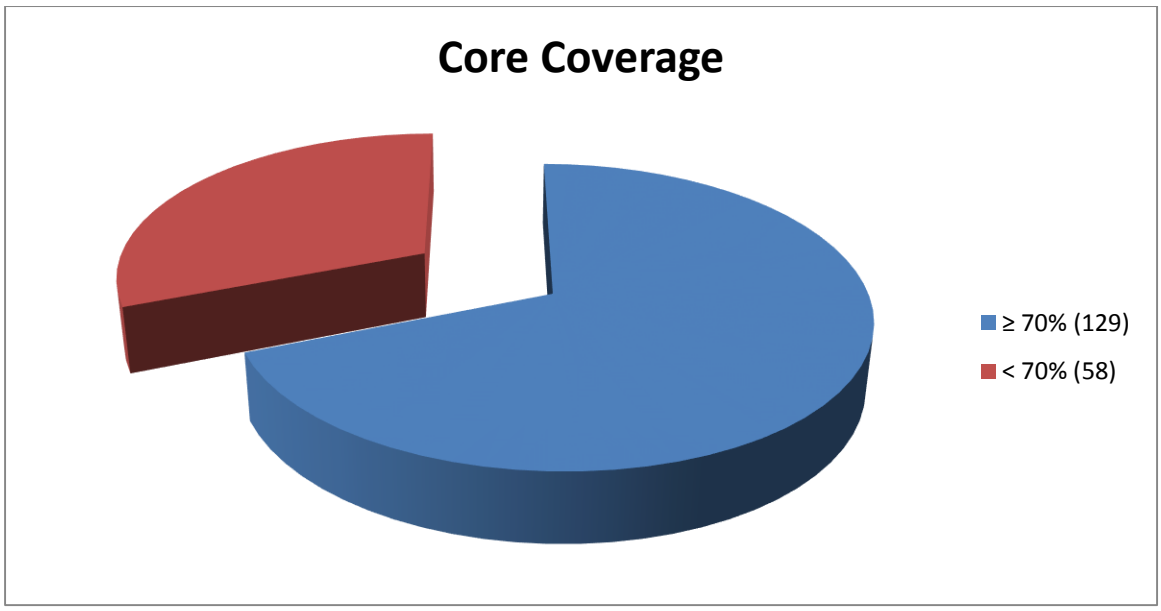


Fig 5.1.5: Pie chart depicting Core Coverage Statistics

Modeled proteins were then filtered based upon their sequence identity (those showing identity more than 90.00 were taken) and their core coverage percentage (having more than 80% core coverage were taken).

19 such proteins were further analyzed with UniProt IDs as follows:

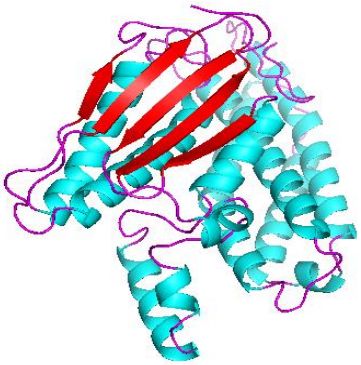
O14874, O75390, P00395, P00505, P12235, P25705, P31040, P31930, P49448, P53597, P55851, P56277, P80404, Q9BYT8, Q9P0J1, Q9UI17, Q9Y2Q3, Q16134, Q99798.

5.2 Protein Modeling

The structures for the proteins along with their names, template IDs and sequence identity are as follows:

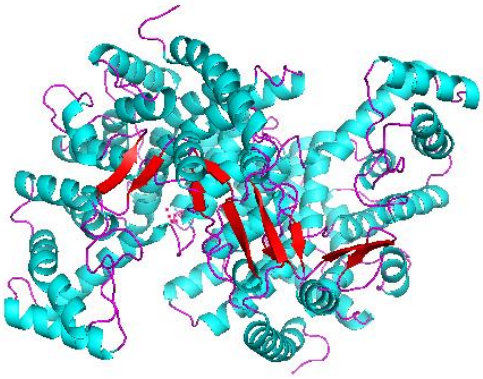
Protein Name **Template ID** **Sequence Identity**

A.



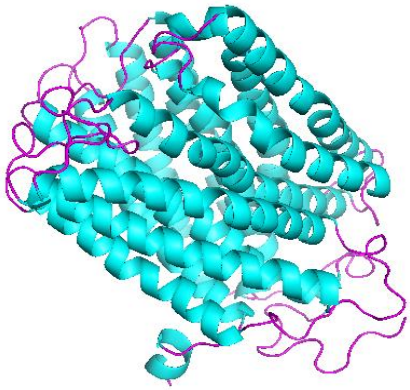
BCKDK
1.gkz.1.A
98.16

B.



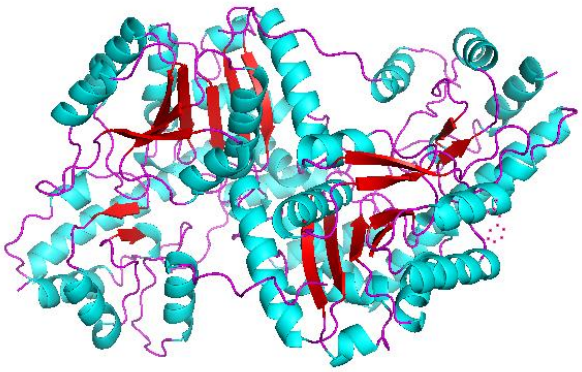
CS
3enj.1.A
96.34

C.

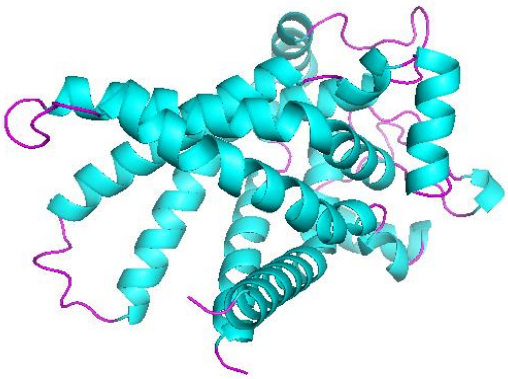


MT-CO1
3abm.1.A
91.59

D.

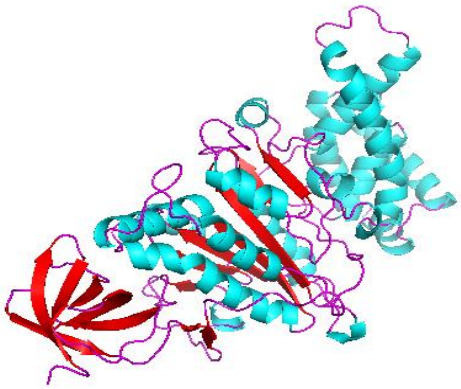


GOT2
5ax8.1.A
99.75



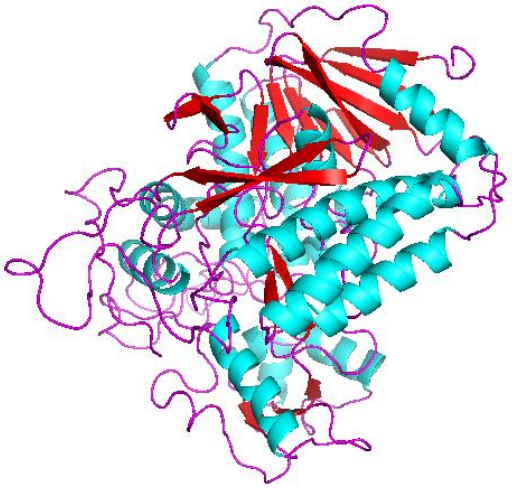
E.

SLC2544
1.okc.1.A
96.28



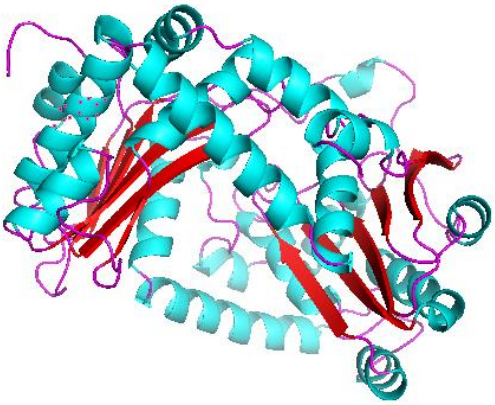
F.

ATP5A1
2w6e.1.A
97.74



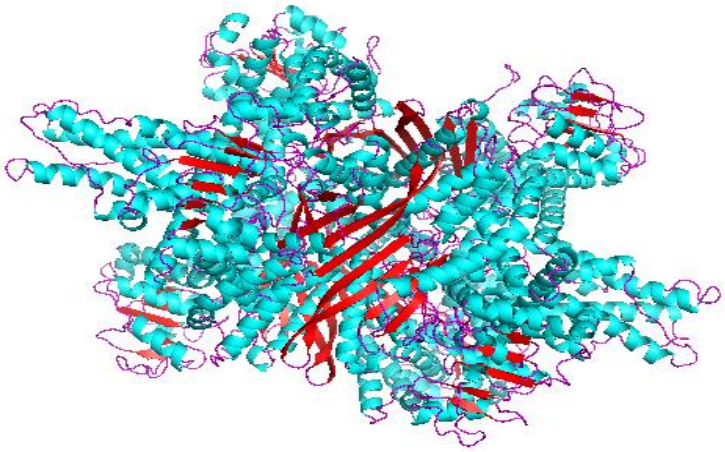
G.

SDHA
1z0y.1A
96.30



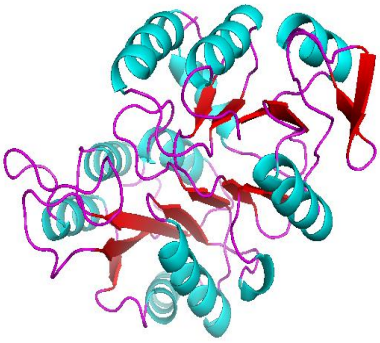
H.

UQCRC1
1be3.1.A
90.36



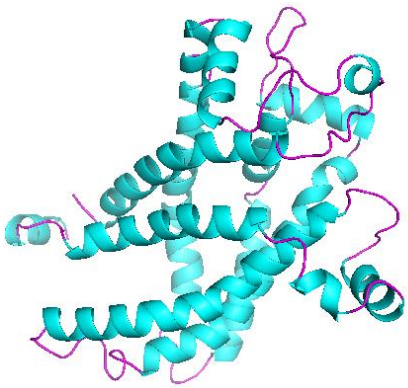
I.

GLUD2
3etg.1.D
96.21



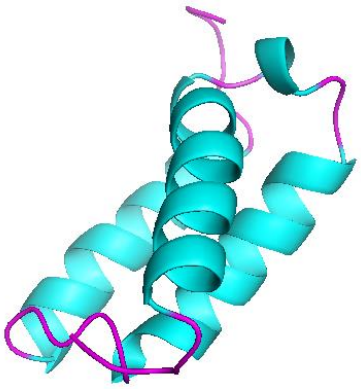
J.

SUCLG1
1eud.1.A
96.12



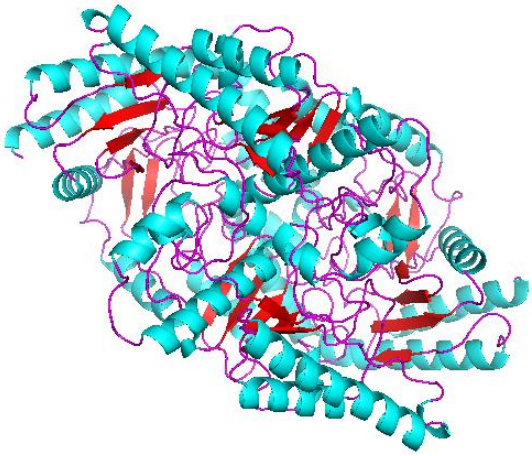
K.

UCP2
21ck.1.A
95.95



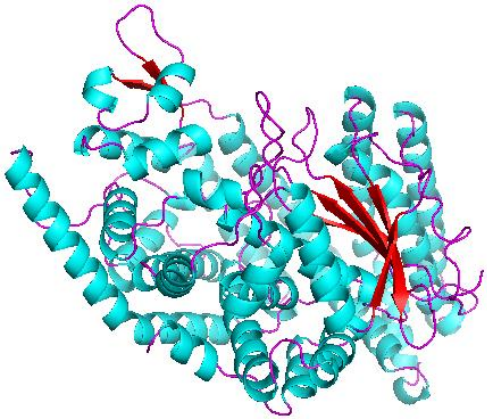
L.

CMC4
2hp8.1.A
100.00



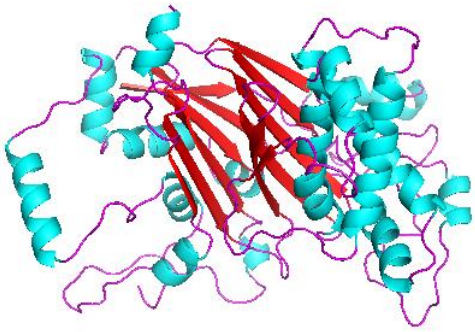
M.

ABAT
1ohw.1.A
95.97



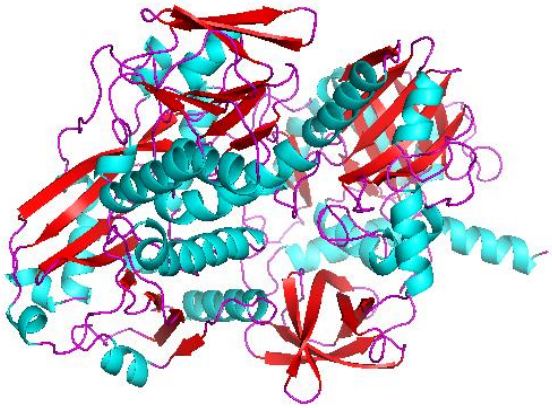
N.

NLN
1i1i.1.A
91.02



O.

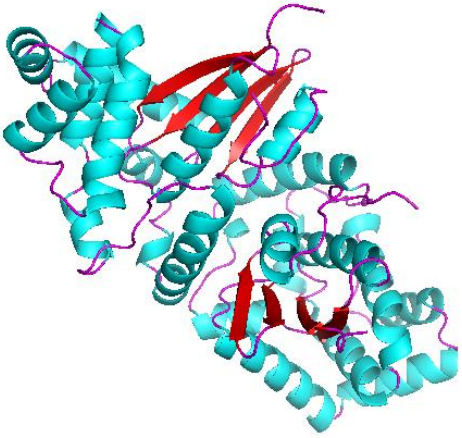
PDP1
3n3c.1.A
98.71



P.

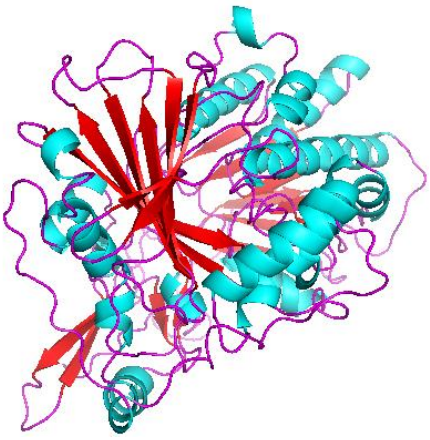
DMGDH
4pab.1.A
92.02

Q.



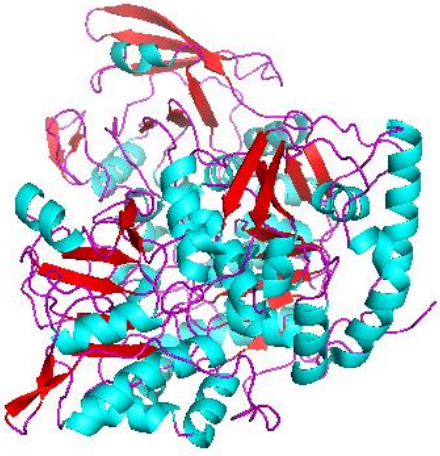
GSTK1
3rpn.1.A
100.00

R.



ETFDH
2gmh.1.A
95.38

S.



ACO2
1b0j.1.A
96.68

5.2 Function Prediction

The ProFunc's InterPro results showed the sequence motifs from various databases listed below. Different colors are depicting results from different databases as follows:

Red – SUPERFAMILY
Orange - Prosite
Aqua – SMART
Purple – Pfam
Green – Prosite profiles
Dark green – Gene3D
Pink – PANTHER

Fig 5.2.1: InterPro Scan of BCKDK

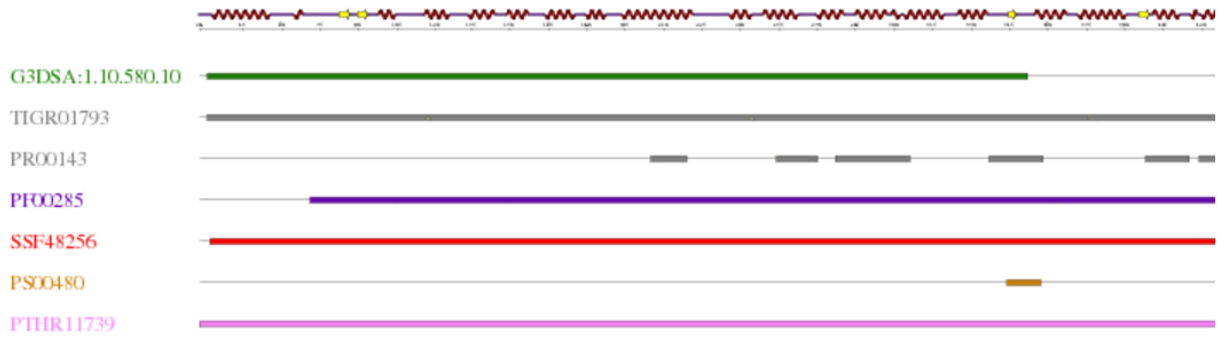


Fig 5.2.2: InterPro Scan of CS

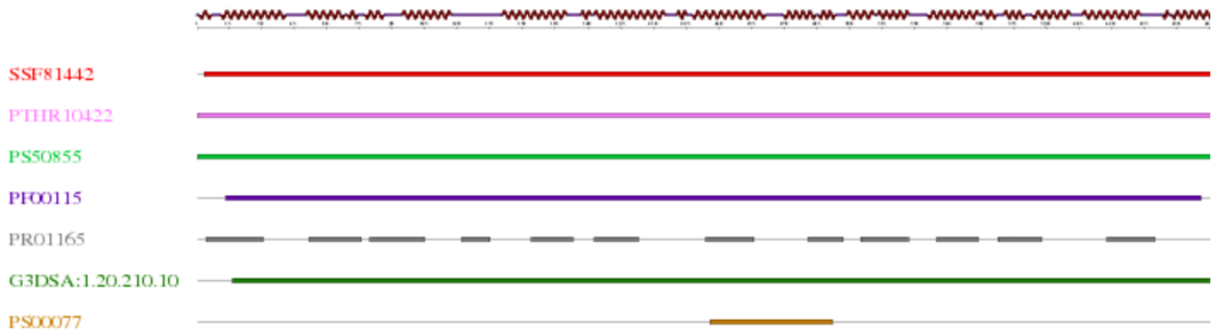


Fig 5.2.3: InterPro Scan of MT-CO1

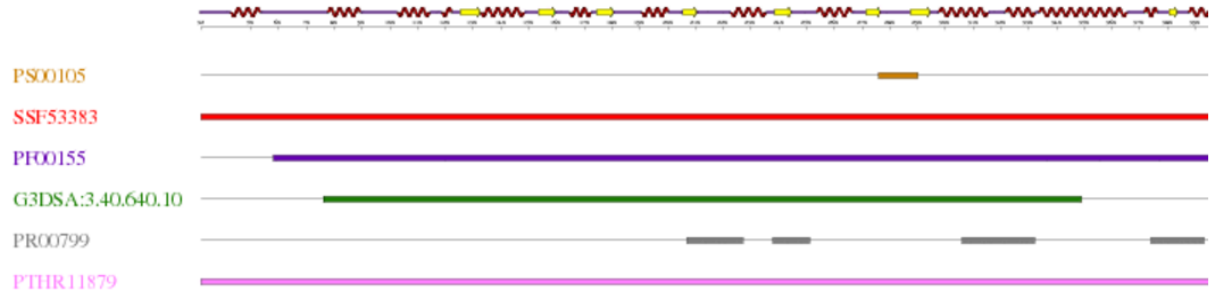


Fig 5.2.4: InterPro Scan of GOT2

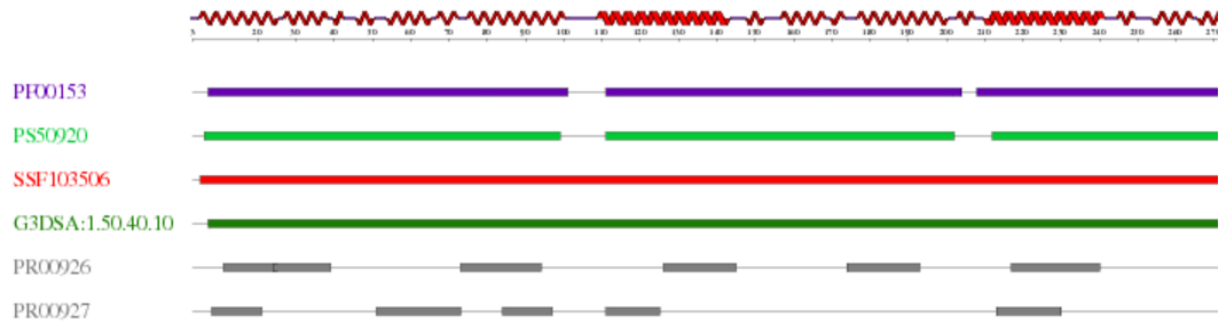


Fig 5.2.5: InterPro Scan of SLC25A4

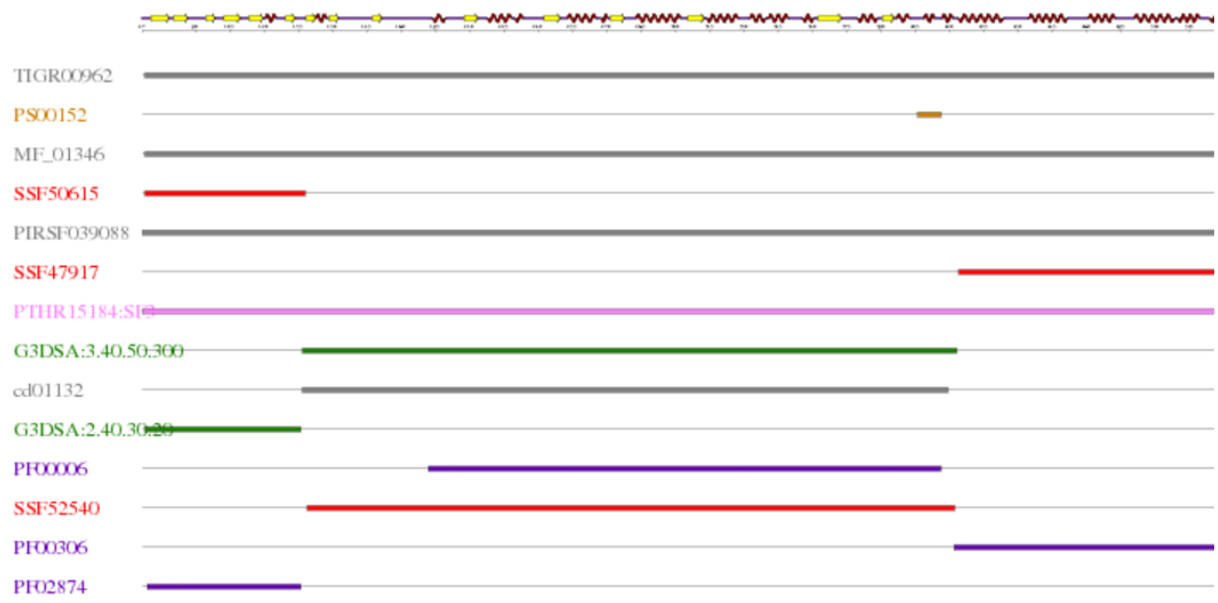


Fig 5.2.6: InterPro Scan of ATP5A1

Fig 5.2.7: InterPro Scan of SDHA

Fig 5.2.8: InterPro Scan of UQCRC1

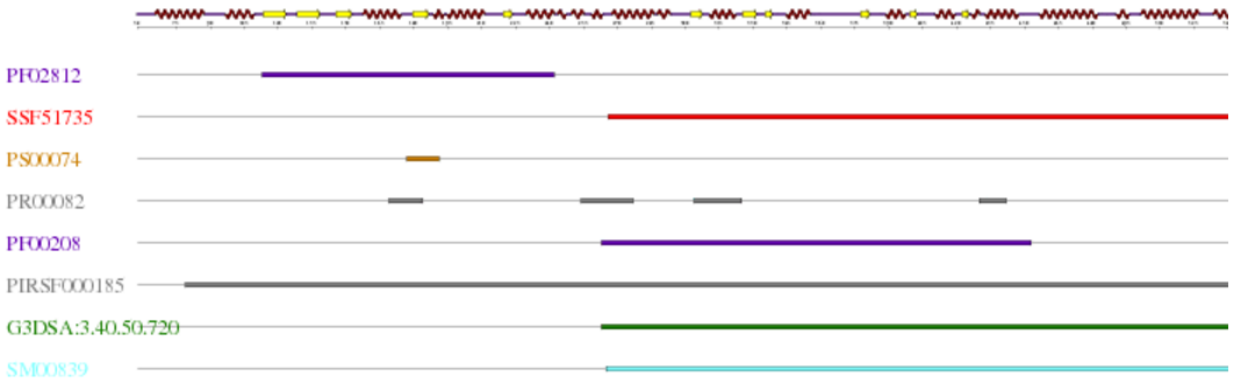


Fig 5.2.9: InterPro Scan of GLUD2

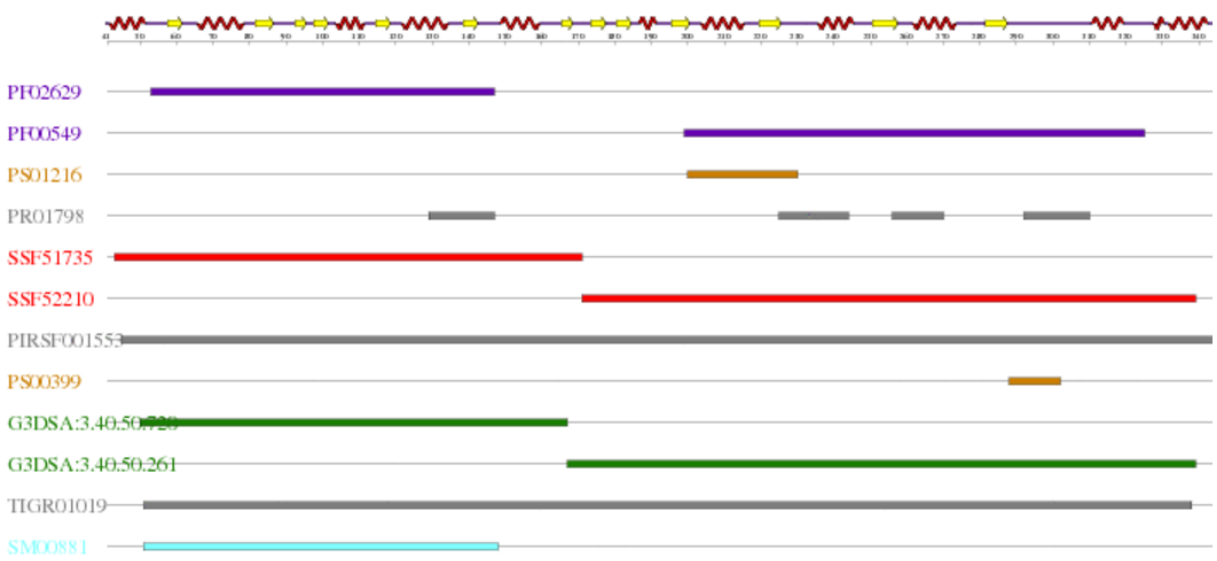


Fig 5.2.10: InterPro Scan of SUCLG1

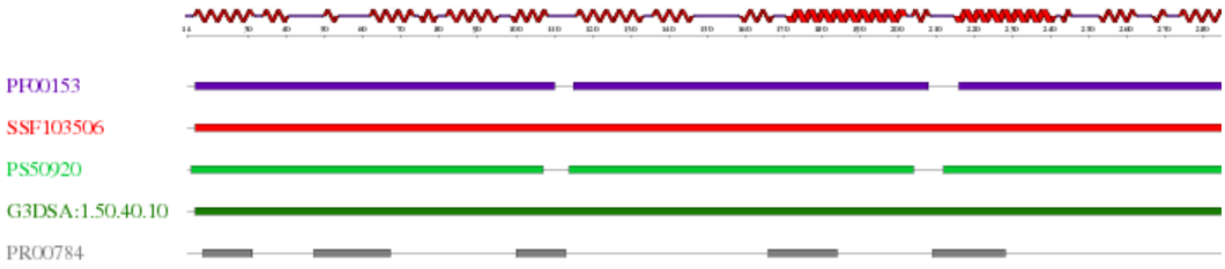


Fig 5.2.11: InterPro Scan of UCP2

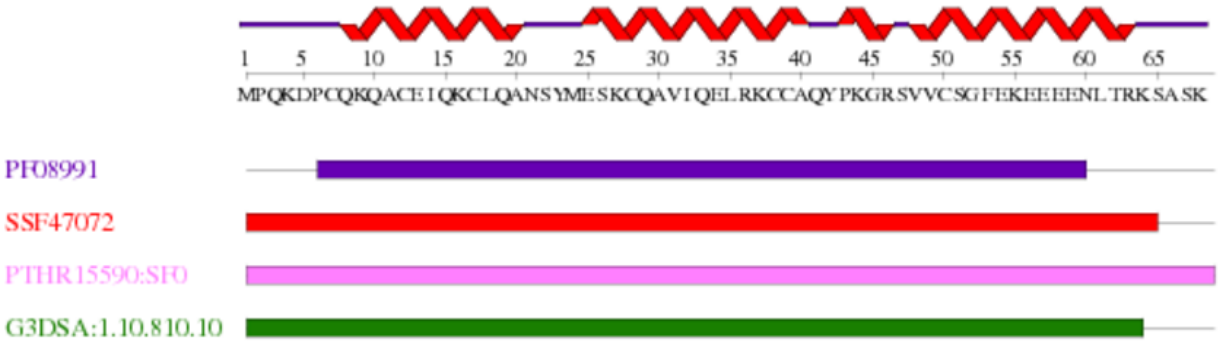


Fig 5.2.12: InterPro Scan of CMC4

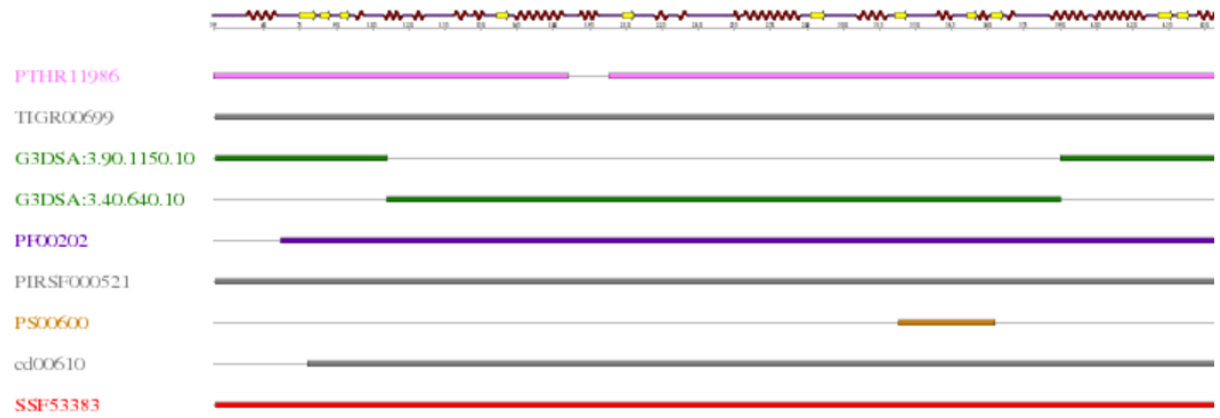


Fig 5.2.13: InterPro Scan of ABAT

Fig 5.2.14: InterPro Scan of NLN

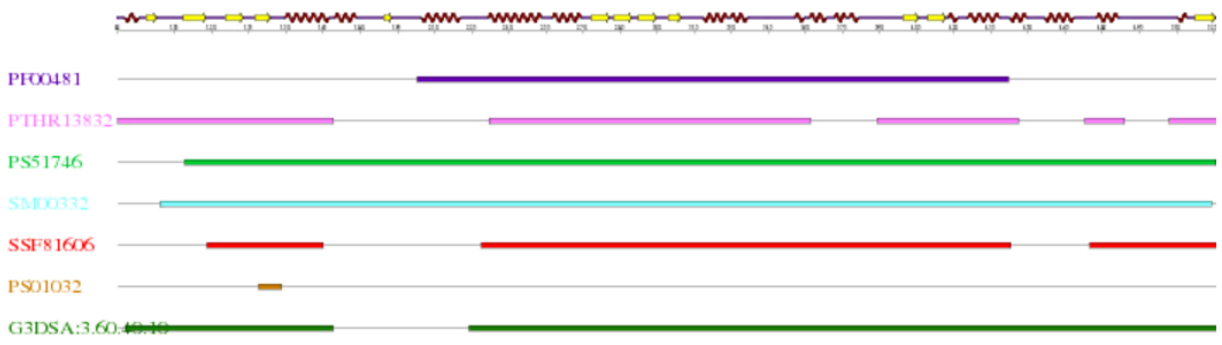


Fig 5.2.15: InterPro Scan of PDP1

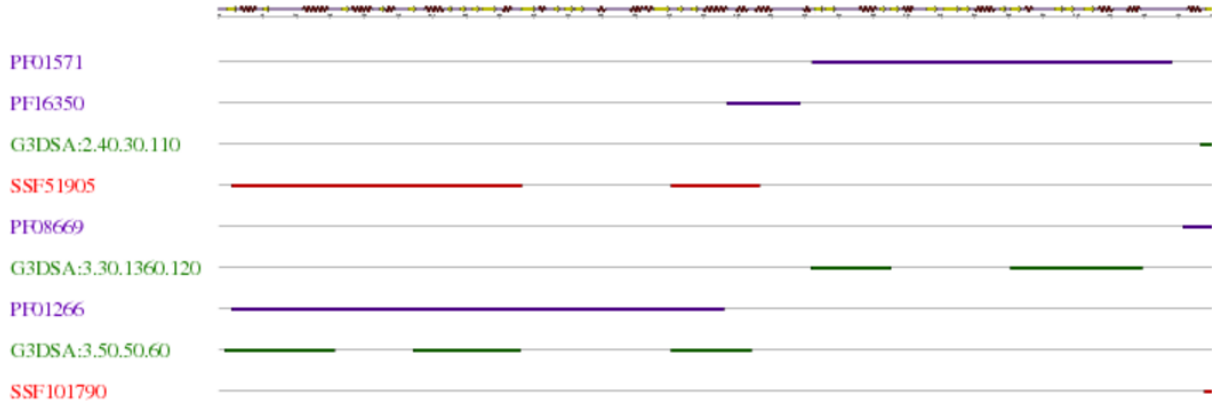


Fig 5.2.16: InterPro Scan of DMGDH

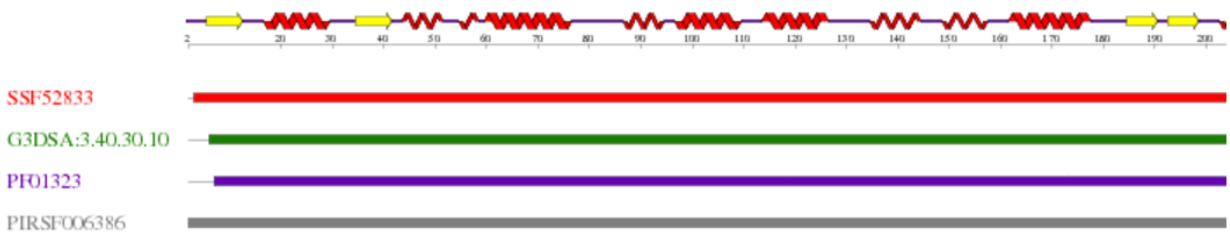


Fig 5.2.17: InterPro Scan of GSTK1

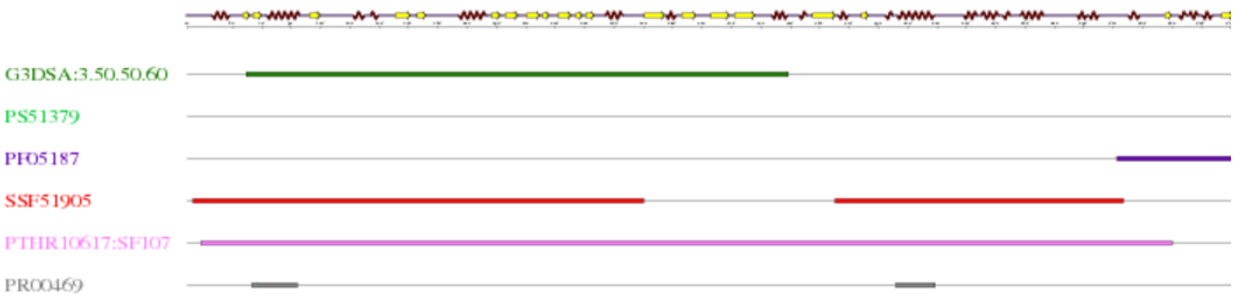
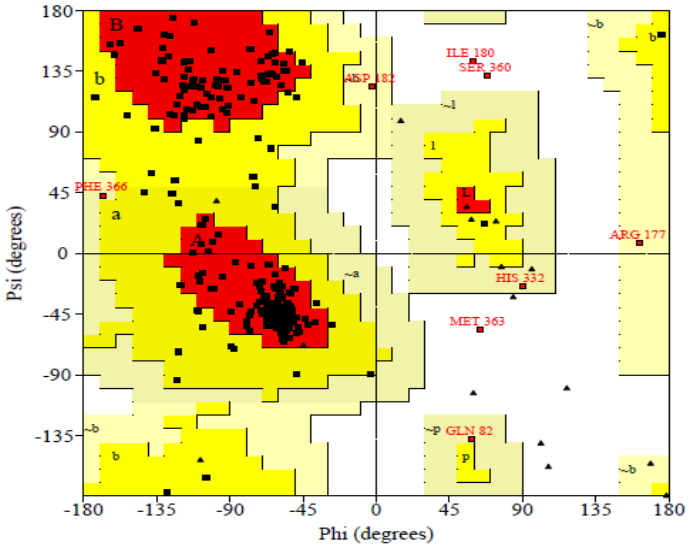


Fig 5.2.18: InterPro Scan of ETFDH

Fig 5.2.19: InterPro Scan of PDP1

5.3 Ramachandran plots

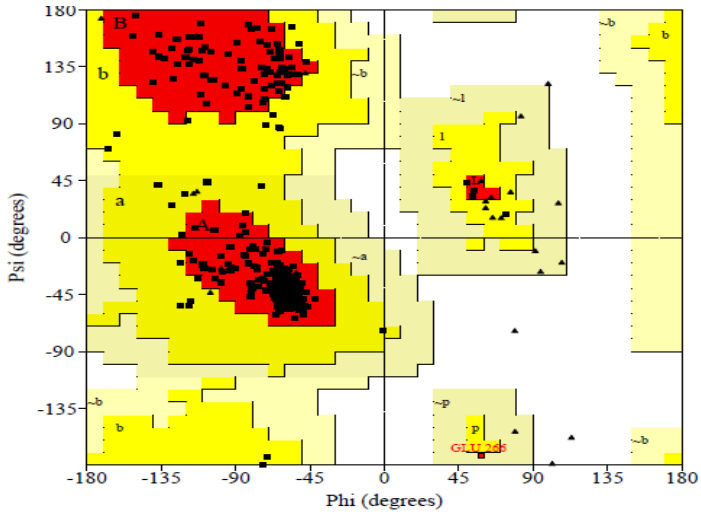
The Ramachandran plots for the proteins are depicted below:



Plot statistics

Residues in most favoured regions [A,B,L]	258	86.9%
Residues in additional allowed regions [a,b,l,p]	31	10.4%
Residues in generously allowed regions [~a,~b,~l,~p]	5	1.7%
Residues in disallowed regions	3	1.0%
Number of non-glycine and non-proline residues	297	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	23	
Number of proline residues	19	
Total number of residues	341	

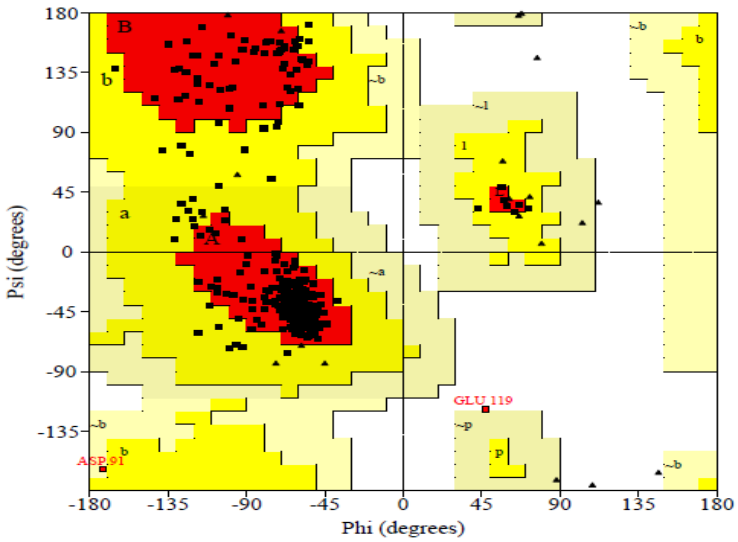
Fig 5.3.1:BCKDK Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	353	93.9%
Residues in additional allowed regions [a,b,l,p]	22	5.9%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.3%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	376	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	22	
Total number of residues	436	

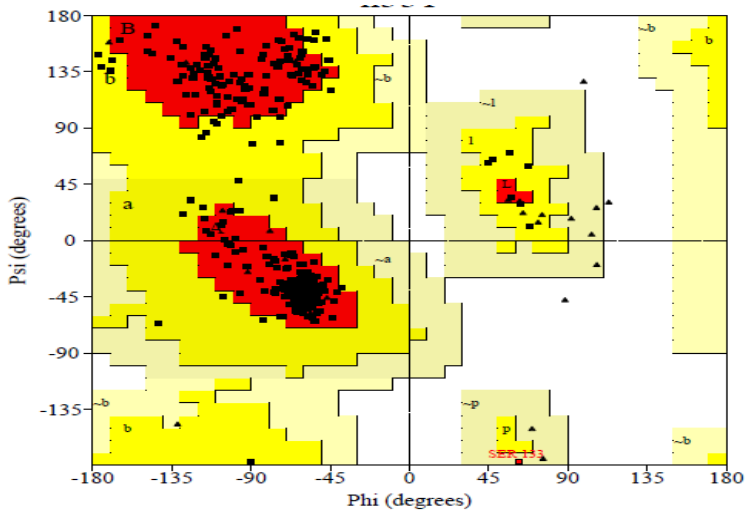
Fig 5.3.2: CS Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	399	91.9%
Residues in additional allowed regions [a,b,l,p]	33	7.6%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.2%
Residues in disallowed regions	1	0.2%
Number of non-glycine and non-proline residues	434	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	46	
Number of proline residues	29	
Total number of residues	511	

Fig 5.3.3: MT-CO1 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	308	89.3%
Residues in additional allowed regions [a,b,l,p]	36	10.4%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.3%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	345	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	33	
Number of proline residues	19	
Total number of residues	399	

Fig 5.3.4: GOT2 Ramachandran Plot

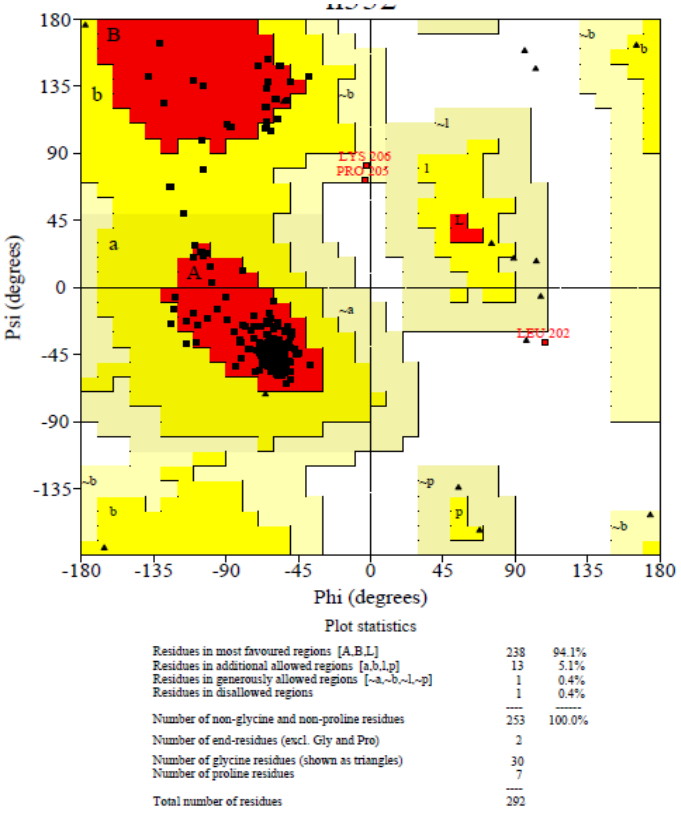


Fig 5.3.5: SLC25A4 Ramachandran Plot

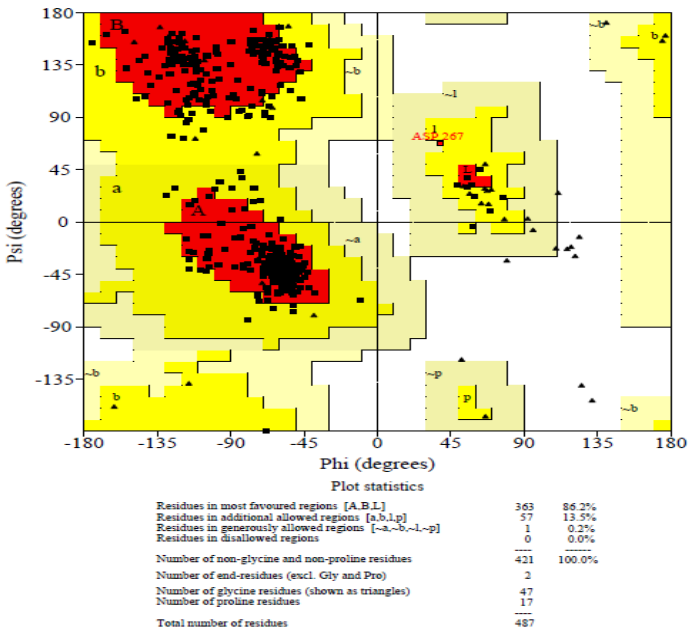
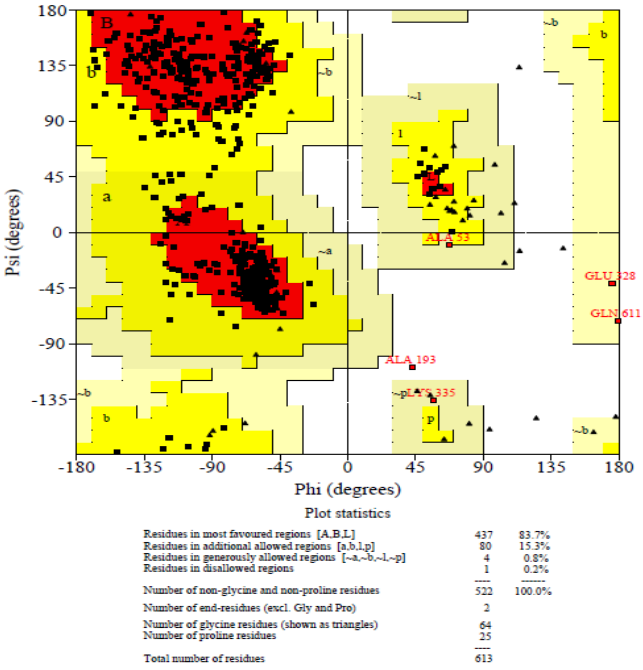


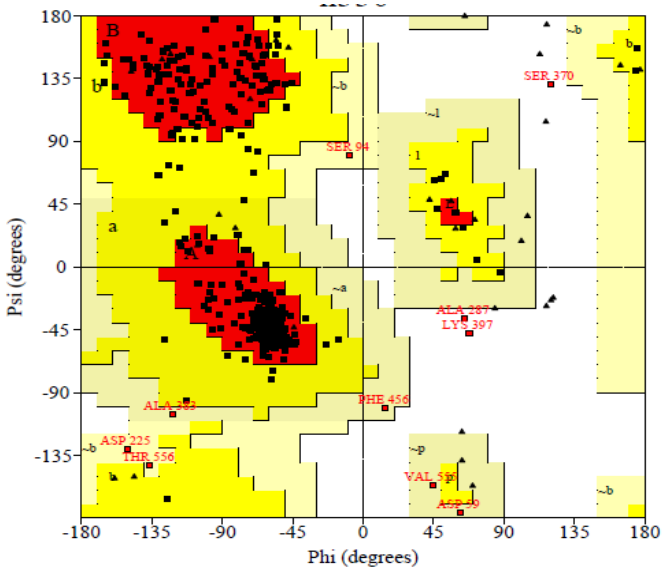
Fig 5.3.6: ATP5A1 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	437	83.7%
Residues in additional allowed regions [a,b,l,p]	80	15.3%
Residues in generously allowed regions [~a,~b,~l,~p]	4	0.8%
Residues in disallowed regions	1	0.2%
Number of non-glycine and non-proline residues	522	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	64	
Number of proline residues	25	
Total number of residues	613	

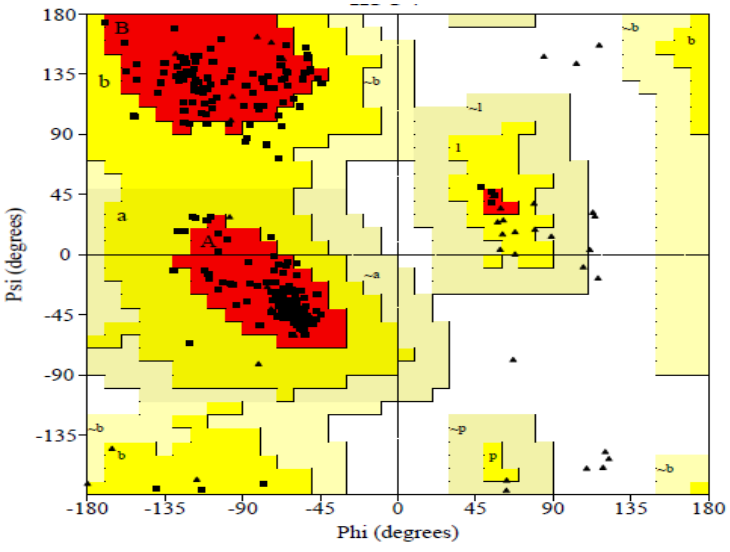
Fig 5.3.7: SDHA Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	379	87.7%
Residues in additional allowed regions [a,b,l,p]	43	10.0%
Residues in generously allowed regions [~a,~b,~l,~p]	7	1.6%
Residues in disallowed regions	3	0.7%
Number of non-glycine and non-proline residues	432	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	45	
Number of proline residues	22	
Total number of residues	501	

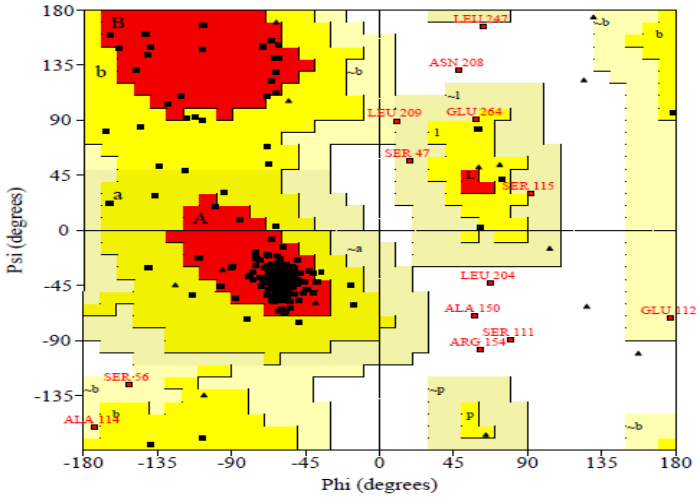
Fig 5.3.8: UQCRC1 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	221	89.8%
Residues in additional allowed regions [a,b,l,p]	25	10.2%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	246	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	41	
Number of proline residues	17	
Total number of residues	306	

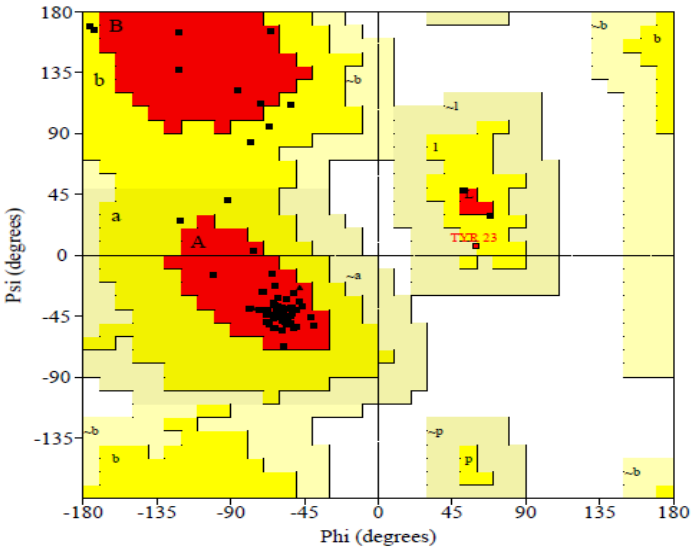
Fig 5.3.9: GLUD2 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	217	84.4%
Residues in additional allowed regions [a,b,l,p]	27	10.5%
Residues in generously allowed regions [~a,~b,~l,~p]	7	2.7%
Residues in disallowed regions	6	2.3%
Number of non-glycine and non-proline residues	257	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	10	
Total number of residues	295	

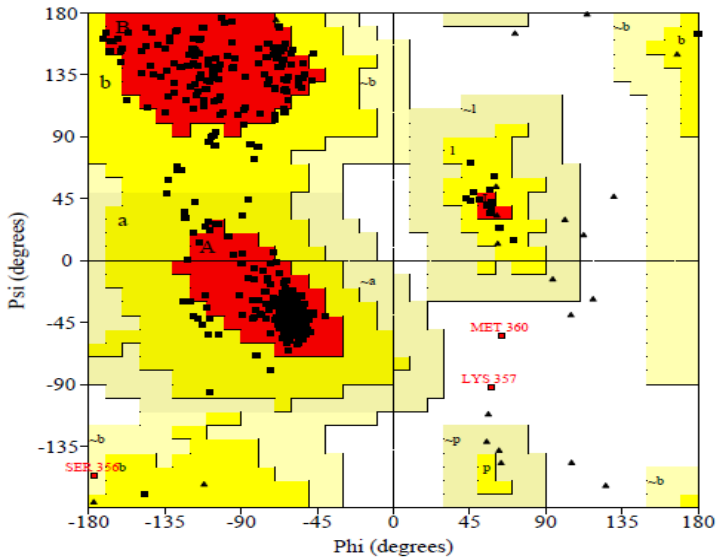
Fig 5.3.10: SUCLG1 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	53	86.9%
Residues in additional allowed regions [a,b,l,p]	7	11.5%
Residues in generously allowed regions [~a,~b,~l,~p]	1	1.6%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	61	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	2	
Number of proline residues	3	
Total number of residues	68	

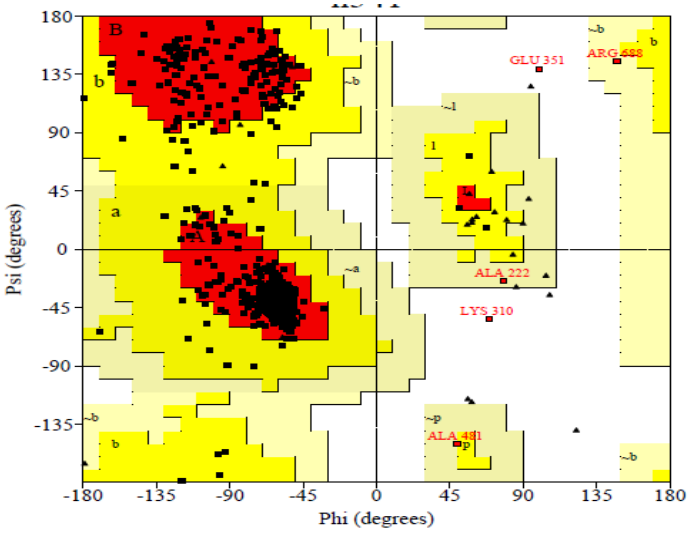
Fig 5.3.11: UCP2 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	347	86.8%
Residues in additional allowed regions [a,b,l,p]	50	12.5%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.2%
Residues in disallowed regions	2	0.5%
Number of non-glycine and non-proline residues	400	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	25	
Total number of residues	461	

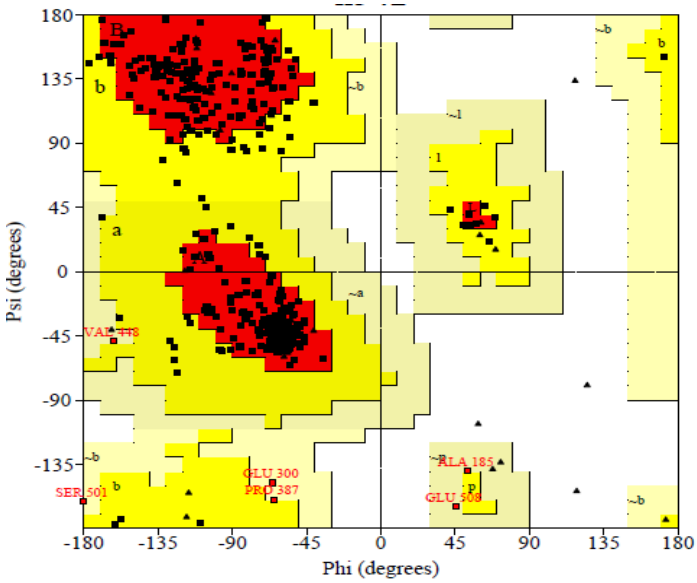
Fig 5.3.12: CMC4 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	548	91.0%
Residues in additional allowed regions [a,b,l,p]	49	8.1%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.5%
Residues in disallowed regions	2	0.3%
Number of non-glycine and non-proline residues	602	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	38	
Number of proline residues	23	
Total number of residues	665	

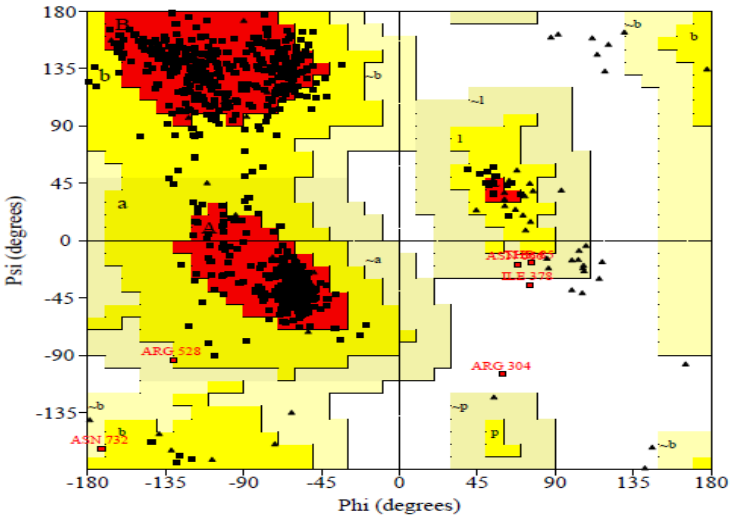
Fig 5.3.13: ABAT Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	350	86.4%
Residues in additional allowed regions [a,b,l,p]	50	12.3%
Residues in generously allowed regions [~a,~b,~l,~p]	5	1.2%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	405	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	20	
Total number of residues	453	

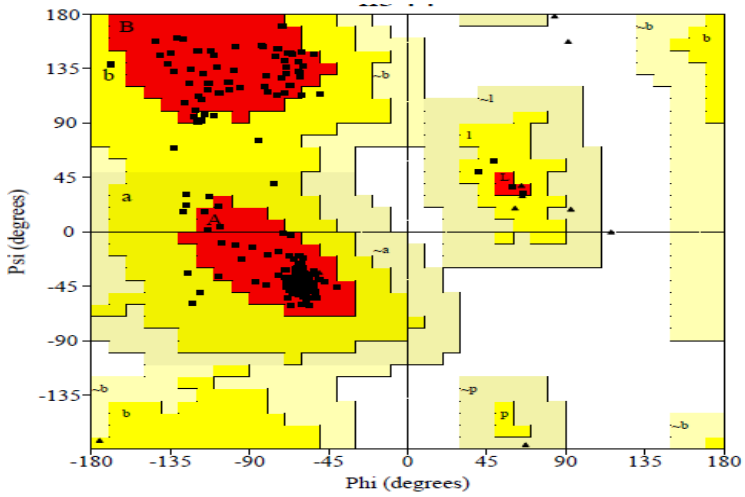
Fig 5.3.14: NLN Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	613	87.6%
Residues in additional allowed regions [a,b,l,p]	81	11.6%
Residues in generously allowed regions [~a,~b,~l,~p]	4	0.6%
Residues in disallowed regions	2	0.3%
Number of non-glycine and non-proline residues	700	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	76	
Number of proline residues	43	
Total number of residues	821	

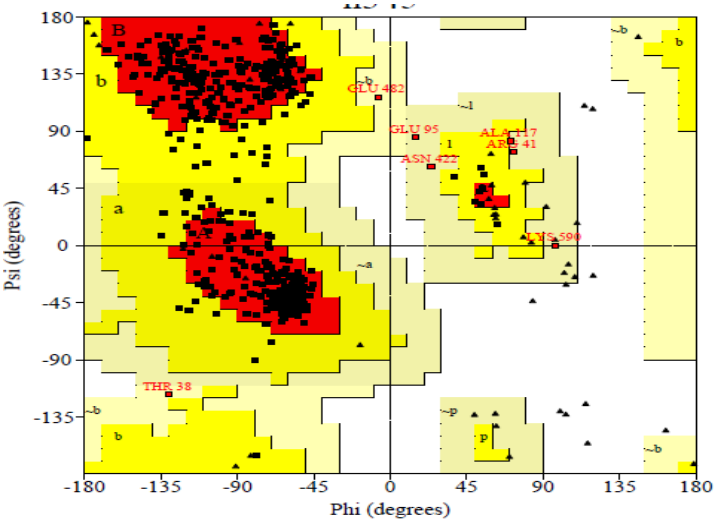
Fig 5.3.15: PDP1 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	172	91.5%
Residues in additional allowed regions [a,b,l,p]	16	8.5%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	188	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	15	
Number of proline residues	16	
Total number of residues	220	

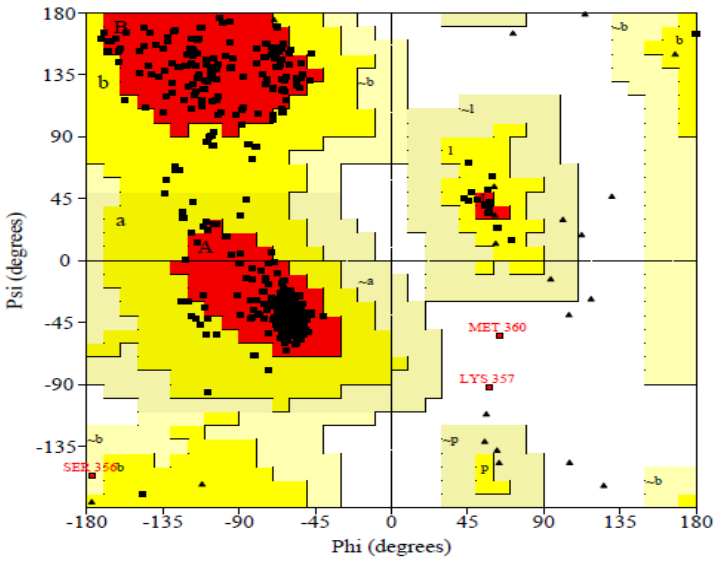
Fig 5.3.16: DMGDH Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	418	87.3%
Residues in additional allowed regions [a,b,l,p]	54	11.3%
Residues in generously allowed regions [~a,~b,~l,~p]	7	1.5%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	479	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	61	
Number of proline residues	39	
Total number of residues	581	

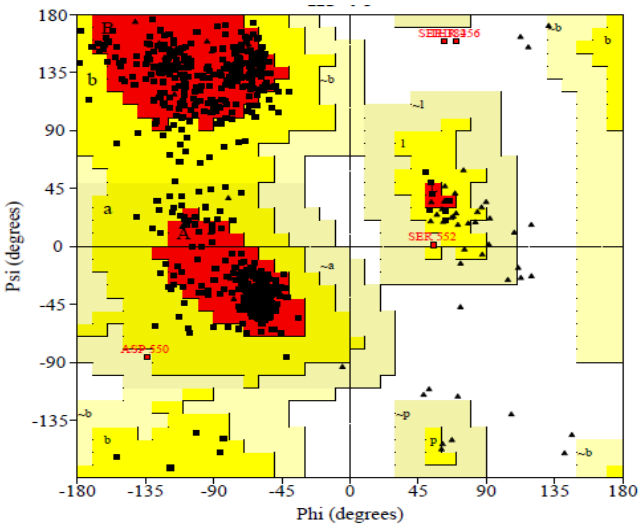
Fig 5.3.17: GSTK1 Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	347	86.8%
Residues in additional allowed regions [a,b,l,p]	50	12.5%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.2%
Residues in disallowed regions	2	0.5%
Number of non-glycine and non-proline residues	400	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	25	
Total number of residues	461	

Fig 5.3.18: ETFDH Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	546	86.1%
Residues in additional allowed regions [a,b,l,p]	84	13.2%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.3%
Residues in disallowed regions	2	0.3%
Number of non-glycine and non-proline residues	634	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	74	
Number of proline residues	42	
Total number of residues	752	

Fig 5.3.19: ACO2 Ramachandran Plot

5.4 MicroRNA binding with 3' UTR

With the help of various target prediction tools the binding sites of miRNA on mRNA were identified. After the analysis of all the predicted results, it was observed that some of miRNA potentially bind to mRNA to inhibit translation. Result of each prediction tool is described as follows:

5.4.1 mirDB

A file of microRNA target prediction result was downloaded. After its analysis it was noticed that a single microRNA has more than one target, even some have targets in the range of hundreds. On calculation for each miRNA, a range of targets was selected to present the results.
11 of miRNAs were there which had more than 200 targets, 18 miRNAs had targets between 150-200, 24 miRNAs showed binding to 150-100 targets, 402 miRNAs had more than 100 but less than 50 targets and 1252 miRNAs showed binding to less than 50 targets.
The graph depicts a clear picture of the results analysed through mirDB:

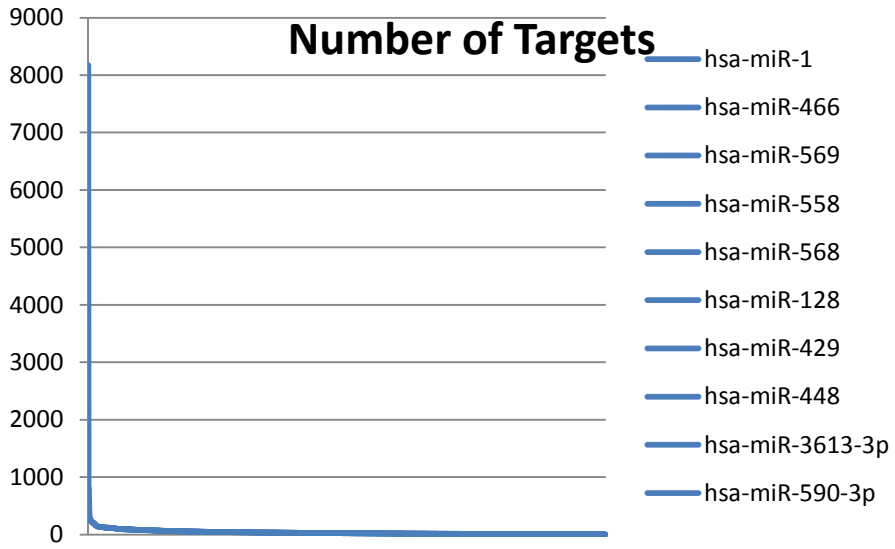


Fig5.4.1.1: Result of mirDB: Y-axis shows the number of genes targeted by the microRNAs on X-axis

5.4.2 miRanda

Through miRanda program, 7680 miRNAs were found targeting the nuclear coded mitochondrial genes, but the top 5 miRNAs were analyzed that were highly regulating the genes. Fig: 5.4.2.1 is showing the graphical representation if miRanda result.

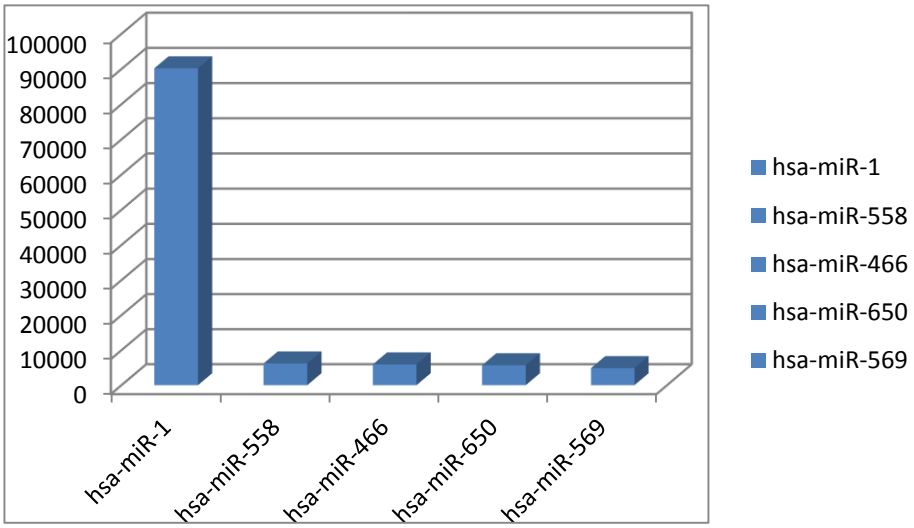


Fig5.4.2.1: miRanda Result: miRNAs that bind to maximum number of targets.

CHAPTER 6

6. SUMMARY AND CONCLUSION

In this dissertation, firstly all the mitochondrial localised proteins were identified with the information of their gene name, chromosome position, start and end location of chromosome, sequences, molecular function and biological processes and manually curated. Out of 516 proteins, 233 had crystal structures, and the remaining 283 proteins' structures were modeled using the SWISS-MODEL so that their function could be analysed with the help of predicted structures. EBI Profunc was used to predict functions of the modeled proteins. InterPro Scan identified sequence motifs from several databases.

The most promising protein structures showing high core coverage as well as high sequence identity were filtered out and energy minimised using Gromacs In linux OS.

Ramachandran plots were plotted for them to analyse the structures further.

In microRNA target prediction around 2000 mitochondrial microRNA were collected that were responsible for gene regulation of mitochondrial protein in Humans. Further, using a number of target prediction tools these microRNAs were analysed to find their statistics in controlling gene expression. It was observed that approximate 20 miRNAs regulate more than 200 genes. Also, hsa-miR-1, hsa-miR-558, hsa-miR-650 and hsa-miR-569 are the miRNAs that involved in highly regulation of gene.

Analysis of microRNA results presented within this dissertation and integration of mitochondrial localised protein information in database would prove useful for analysis of human microRNAs, which are responsible for controlling the gene expression of nuclear coded mitochondrial proteins.

CHAPTER 7

7. REFERENCES

1. Yamamoto, H., et al., *MicroRNA-494 regulates mitochondrial biogenesis in skeletal muscle through mitochondrial transcription factor A and Forkhead box j3.* Am J Physiol Endocrinol Metab, 2012. **303**(12): p. E1419-27.
2. Poyton, R.O. and J.E. McEwen, *Crosstalk between nuclear and mitochondrial genomes.* Annu Rev Biochem, 1996. **65**: p. 563-607.
3. Li, P., et al., *Control of mitochondrial activity by miRNAs.* J Cell Biochem, 2012. **113**(4): p. 1104-10.
4. Latronico, M.V. and G. Condorelli, *The might of microRNA in mitochondria.* Circ Res, 2012. **110**(12): p. 1540-2.
5. Ardekani, A.M. and M.M. Naeini, *The Role of MicroRNAs in Human Diseases.* Avicenna J Med Biotechnol, 2010. **2**(4): p. 161-79.
6. Spinazzola, A. and M. Zeviani, *Mitochondrial diseases: a cross-talk between mitochondrial and nuclear genomes.* Adv Exp Med Biol, 2009. **652**: p. 69-84.
7. Li, F., et al., *Global analysis of RNA secondary structure in two metazoans.* Cell Rep, 2012. **1**(1): p. 69-82.
8. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
9. Boldogh IR, Pon LA. Interactions of mitochondria with the actin cytoskeleton. Biochim Biophys Acta.2006;1763:450–462.
10. Wexler Y, Zilberstein C, Ziv-Ukelson M. A study of accessible motifs and RNA folding complexity. J Comput Biol. 2007;14:856–872.
11. Lung B, Zemann A, Madej MJ, Schuelke M, Techritz S, Ruf S, Bock R, Huttenhofer A. Identification of small non-coding RNAs from mitochondria and chloroplasts. Nucleic Acids Res. 2006;34:3842–3852.
12. Barrey E, Saint-Auret G, Bonnamy B, Damas D, Boyer O, Gidrol X. Pre-microRNA and mature microRNA in human mitochondria. PLoS One. 2011;6:e20220.
13. Lackner LL. Determining the shape and cellular distribution of mitochondria: the integration of multiple activities. Curr Opin Cell Biol. 2013;25:471–476.
14. Zhang X, Zuo X, Yang B, Li Z, Xue Y, Zhou Y, Huang J, Zhao X, Zhou J, Yan Y, Zhang H, Guo P, Sun H, Guo L, Zhang Y, Fu XD. MicroRNA directly enhances mitochondrial translation during muscle differentiation. Cell. 2014;158:607–619.
15. Hata A. Functions of microRNAs in cardiovascular biology and disease. Annu Rev Physiol.2013;75:69–93.
16. Shinde S, Bhadra U. A complex genome-microRNA interplay in human mitochondria. Biomed Res Int 2015. 2015 206382
17. Lackner LL. Determining the shape and cellular distribution of mitochondria: the integration of multiple activities. Curr Opin Cell Biol. 2013;25:471–476

18. Liu NK, Xu XM. MicroRNA in central nervous system trauma and degenerative disorders. Physiol Genomics. 2011;43:571–580
19. Sullivan PG, Rabchevsky AG, Waldmeier PC, Springer JE. Mitochondrial permeability transition in CNS trauma: cause or effect of neuronal cell death? J Neurosci Res. 2005;79:231–239
20. McEwen ML, Sullivan PG, Rabchevsky AG, Springer JE. Targeting mitochondrial function for the treatment of acute spinal cord injury. Neurotherapeutics. 2011;8:168–179.
21. Bienertova-Vasku J, Sana J, Slaby O. The role of microRNAs in mitochondria in cancer. Cancer Lett.2013;336:1–7
22. Kren BT, Wong PY, Sarver A, Zhang X, Zeng Y, Steer CJ. MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. RNA Biol. 2009;6:65–72
23. Wang WX, Visavadiya NP, Pandya JD, Nelson PT, Sullivan PG, Springer JE. Mitochondria-associated microRNAs in rat hippocampus following traumatic brain injury. Exp Neurol.2015;265:84–93.
24. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007;39:1278–1284
25. Bandiera S, Ruberg S, Girard M, Cagnard N, Hanein S, Chretien D, Munnich A, Lyonnet S, Henrion-Caude A. Nuclear outsourcing of RNA interference components to human mitochondria. PLoS One.2011;6:e20746.
26. Wexler Y, Zilberstein C, Ziv-Ukelson M. A study of accessible motifs and RNA folding complexity. J Comput Biol. 2007;14:856–872
27. Nelson PT, Wang WX, Rajeev BW. MicroRNAs (miRNAs) in neurodegenerative diseases. Brain Pathol. 2008;18:130–138.
28. Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood AM, Haugen E, Bracken CP, Rackham O, Stamatoyannopoulos JA, Filipovska A, Mattick JS. The human mitochondrial transcriptome. Cell. 2011;146:645–658
29. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer. 2006;6:857–866.
30. Das S, Ferlito M, Kent OA, Fox-Talbot K, Wang R, Liu D, Raghavachari N, Yang Y, Wheelan SJ, Murphy E, Steenbergen C. Nuclear miRNA regulates the mitochondrial genome in the heart. Circ Res.2012;110:1596–1603
31. Nunnari J, Suomalainen A. Mitochondria: in sickness and in health. Cell. 2012;148:1145–1159.
32. Sripada L, Tomar D, Prajapati P, Singh R, Singh AK, Singh R. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. PLoS One. 2012;7:e44873.