CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Structural and Functional Analysis of Translocation in DISC1 gene and Impact on Schizophrenia

by

Jaweria Zia

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2021

I dedicate this thesis to my parents and my teachers.

CERTIFICATE OF APPROVAL

Structural and Functional Analysis of Translocation in DISC1 gene and Impact on Schizophrenia

by

Jaweria Zia

(MBS191024)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Mazhar Qayyam	PMAS-UAAR, Rawalpindi
(b)	Internal Examiner	Dr. Erum Dilshad	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
April, 2021

Dr. Sahar Fazal
Head
Dept. of Biosciences & Bioinformatics
April, 2021

Dr. Muhammad Abdul Qadir
Dean
Faculty of Health & Life Sciences
April, 2021

Author's Declaration

I, **Jaweria Zia** hereby state that my MS thesis titled "**Structural and Functional Analysis of Translocation in DISC1 gene and Impact on Schizophrenia**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Jaweria Zia)

Registration No: MBS191024

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Structural and Functional Analysis of Translocation in DISC1 gene and Impact on Schizophrenia**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Jaweria Zia)

Registration No: MBS191024

Acknowledgement

In the name of Allah, the Most Gracious and the Most Merciful Alhamdulillah, all praises to Allah for giving me strength and for His blessings in completing my MS thesis. First, I would like to express my sincere gratitude to Capital University of Science and Technology (CUST) Islamabad for providing me an opportunity to do MS Biosciences and achieving my goal to pursue higher studies. I would like to start with a special appreciation that goes to my Supervisor, Dr. Syeda Marriam Bakhtiar, for her constant support, encouragement and guidance throughout this thesis. Her door was always open whenever I needed help, she always guided me as a mentor. I am unable to find words for expressing my heart feelings towards her for sincere encouragement, guidance, useful suggestions and trust in me throughout my graduation degree and also in this thesis research. Her observations and comments helped me to establish the overall direction of the research and to move forward with investigations in depth. I would like to thanks to my teachers Dr. Shaukat Iqbal, Dr. Erum Dilshad, Dr. Arshia Amin and Dr. Sahar Fazal. Special thanks to my friends and colleagues for supporting me throughout this time. Finally, I express my gratitude to my parents and siblings for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

(Jaweria Zia)

Abstract

Disrupted in schizophrenia 1 is a multifunctional scaffolding protein, which performs various activities by interacting with other molecular partners. The key functions of DISC1 involve neurogenesis, proliferation differentiation, migration and cell adhesion. It is ubiquitously expressed in the different body organs during early development and in later stages of life. The DISC1 protein is an 854 amino acids protein encoded by chromosome 1q42.2 and comprises of N-terminal head and C-terminal coiled coil tail. DISC1 is a candidate gene implicated in schizophrenia pathophysiology. Schizophrenia is a neuropsychiatric disease and hallmark symptoms include hallucinations, delusions and disorganized speech. In this study DISC1 translocation t (1;11) and related structural variants were explored to investigate the role of DISC1 in schizophrenia pathophysiology. These variants were analyzed to investigate disordered regions and effect of mutations. The DISC1 variant protein models were generated and analyzed for pathway analysis and enrichment. Then scrutiny of the DISC1 translocation and related variants provided understanding about the role of DISC1 in schizophrenia pathophysiology and how it can be targeted for future therapeutics. It was concluded that due to translocation, following 4 sequences (NP001158009.1, NP001158012.1, NP001012975.1, NP061132.2) out of 23 sequences were deleted. This deletion was the reason that resulted in damage to amino acid sequences hence the role of DISC I was disrupted.

Keywords: Translocation, Neurogenesis, proliferation, schizophrenia, hallucinations, pathway, therapeutics.

Contents

List of Figures

List of Tables

Abbreviations

AKT	Rac-alpha serine/threonine-protein kinase
ATF4/ATF5	Activating transcription factor 4/5
DBZ	DISC1-binding zinc-finger
DISC1	Disrupted-In-Schizophrenia 1
ERP	Event related potential
FEZ1	Fasciculation and elongation protein zeta-1
GSK3	Glycogen synthase kinase 3
LEF	Lymphoid enhancer factor
LOD	Logarithm of the odds ratio
MAP1A	Microtubule-associated protein 1A
MAPK	Mitogen activated protein kinase
MRS	Magnetic resonance spectroscopy
NDEL1	Nuclear distribution protein E homolog like-1
NDMA	N-methyl-d-aspartate
PCM1	Pericentriolar material 1
PDE4	Phosphodiesterase 4
PDs	Personality disorders
PKA	Protein kinase A
SiRNA	Small interference RNA
SNPs	Single nucleotide polymorphism
TCF	T cell factor

Chapter 1

Introduction

Disrupted in schizophrenia 1 (DISC1) is an intricate, large protein of 854 amino acids that has a 93,611 kDa molecular weight which occurs in humans and is coded by the DISC1 gene 1q42.2. . It performs various functions including cell proliferation, regulation, differentiation, migration, and cell to cell adhesion. There are different mutations reported in DISC1 gene but balanced t(1;11) DISC1 translocation leads to multiple neurological diseases and psychiatric conditions including schizophrenia, bipolar disorder, autism, Asperger's syndrome, and clinical depression. Various polymorphisms of DISC1 have indicated positive association with schizophrenia [1].DISC1 is located at the junction of many neu- rodevelopmental pathways and acts as a scaffold and binds to multiple proteins from which several of them are shown to be independent risk factors for major mental illness including schizophrenia. Therefore, , Disrupted in schizophrenia 1 and its associated protein interacting network is an achievable target for future therapeutic intervention [2, 3].

Several research projects on schizophrenia showed a linkage between different disorders specially schizophrenia and the q arm of the locus present on the chromosome 1 which is a centromere region to 1q42.1 chromosome. Numerous studies provides the evidence of the linkages between the schizophrenia and a balanced translocation involving chromosomes 1 and 11,and their q arms so that can be expressed as t (1;11) (q42.1;q14.3) [4]. This balanced translocation involves breakpoint between

genes DISC1 and DISC1FP1. It is associated with psychiatric disorders and was first discovered.

In a Scottish family about 20 years ago. Now, this practice is been replicated in different populations including American, Japanese, French, and Taiwan [2]. Deletion analyses suggest that DISC1 is a multifunctional and several protein which behaves as bridge between proteins at molecular level which are located and are involved in the signaling outside the cells, neurite outgrowth, and as well as the migration among the neurons. According to this model, DISC1 short arm disturb interaction with NUDEL and also in the cell culture, outgrowth is interrupted [5, 6]. The DISC1 protein is exclusively produced and released in the brain neurons, prominently in the regions specific for learning and processing memory mainly in hippocampus and cerebral cortex but these regions may become affected with schizophrenia disease. Variations in DISC1 gene can cause cognitive deficiencies, also affiliated with causing disturbances and influence functions of memory in schizophrenic patient [7].

DISC1 protein sequence consists of two regions: head region also known as N-terminal on which amino acid residues span and lack secondary structure elements, so it is predicted that this region is cause of disorder as it contains disordered stretches and the other region is called C terminal which contains alpha helix indicates conservation between adjoining orthologs as compared to another terminal. Researchers also found that DISC1 form dimers and oligomers [8].

The N terminal of Disrupted in schizophrenia 1 coordinates with the Microtubule-associated protein 1A, and the three central coiled domains of Disrupted in schizophrenia 1 binds with Mitogen activated protein kinase.These proteins bind with microtubules and play role in DISC1 microtubule association and so if function of microtubule is dis- rupted, it leads to adverse effects like abnormal neuronal architecture and receptor localization, which proceed to a schizophrenic brain [9]. DISC1 stimulates many neuronal signaling pathways by protein-protein interactions; but the mechanism's occurrence is still unclear, probably knowledge of the DISC1 structure is lacking [10]. Many studies provide evidence from different

sources suggests variation in the disrupted in schizophrenia 1 in the pathophysiology related to the mental illnesses including schizophrenia [11].

Rate of P300 is also reported to be reduced in the patients and they are expected to be the carriers of translocation and their activation leads to initiate memory in the patients [12].

DISC1 is a multifunctional protein which is examined by interactions of proteins, mapping of domain immunocytochemistry and subcellular localization, which interacts via distinct domains with different components of the intracellular machinery. How DISC1dysfunction relates to Schizophrenia needs to be understand by relying on above mentioned data and techniques and working model of DISC1can be generated [13]. The DISC1 protein is expressed among different species during brain development and in lifetime which include abnormal regions of the brain during schizophrenia like the prefrontal cortex, thalamus, and hippocampus. Neuronal migration, neurite outgrowth, and neurite extension are basic functions of DISC1 in the brain during developmental stages.DISC1 has been discovered in many populations of neurons and the associated structures along with synaptic function in adult individuals [14].

The DISC1 binding regions, subcellular localization, and known sequence variants are implicated in psychiatric diseases. The following features tell us about the origin of the DISC1 sequence. Firstly, the sequence of amino acids has rapidly evolved at an unexpected rate. For example, the sequence similarity of mouse & human evolved unusually rapidly. Amino acid sequence identity of DISC 1 for human and mouse is 50% which is quite different than all other genes. Secondly, DISC1 5´exons end encode sequences with high levels of specific residues of amino acids including serine, alanine, and glycine. Thirdly, several predicted coiled regions were encoded by the remaining DISC1 sequence [15].

The DISC1 protein acts as a scaffold protein, so it interacts with a huge number of protein binding partners to stimulate a large number of signaling pathways, most of them are of significant importance during neurodevelopment. Progress in understanding the mechanisms is on its way of development as information of basic

DISC1 structure is limited. Pathways and potential chances of bipolar disorder, schizophrenia, depression and associated disorders is identified by multiple DISC1 interacting proteins. Without proper information and knowledge of DISC1 mutant and native gene, structure, biological and physical properties, understanding of pathways involved is quite difficult and only limited performance can be done regarding DISC1 biology and relation of structure and function. cAMP and Wnt signaling, as well as AKT signaling, play role in glutamatergic and dopaminergic signaling while DISC1 pathway is helpful for drug targeting and development [16].

Pathways involving DISC1 and all relevant molecules must be related to mental health condition. As variations in DISC1 results in neuronal impairment of mild level. Ser-704 allele variation has effect on structure of hippocampus and its relevant functions. Mutant DISC1 expression leads to mild enlargement of lateral ventricles and development of neurite outgrowth in the region of primary cortical. These changes are due to DISC1, LIS1 and SNAP25‘s decreased level of proteins [17].

The Pathway of DISC1 in causing mental illness is unclear as the structure of DISC1 is not understood very well because the protein causes coils which are alike the composition of the domain. A lack of homology with known proteins made it difficult and has hindered attempts to define the composition of the domain properly.

DISC1 has the potential to simultaneously affect disease susceptibility on wide and large scale. PDE4B, PDE4D, NDE1, NDEL1, LIS1, FEZ1, PCM1 and TNIK are gene coding partners of DISC1. The interplay between single nucleotide polymorphism varies within DISC1 also affects the risk of schizophrenia mostly [18].

As we knew that DISC1 is phosphoprotein so it is regulated by the help of partner‘s binding and releasing, which further combine with microtubules. Because of shorten end of DISC1 and ATF binding the process and play role in inducing neurochemical abnormalities which can be easily seen in schizophrenic patient. Because DISC1 is dynamic macromolecule complex so it includes centrosomal proteins, NUDEL, MIPT3, MAP1A and microtubules, so shows multiple interactions

with several proteins. MAPs association with microtubules is regulated by phosphorylation, and so are the effects of NUDEL phosphorylation on microtubule has been illustrated. Interactions between DISC1 and binding proteins effects neuronal functions which may be responsible for physiological defects possible in schizophrenia [19].

DISC1 has vital and prominent role in neurosignaling and neurodevelopment, proved by performing experiments on laboratory animal models specially. DISC1 protein binds to diverse range of molecules which perform principal role in development of cerebral cortex signaling, like NUDE, zeta-1 protein for elongation and as well as Citron and phosphodiesterase-4B (PDE4B).

Mutant DISC1 affects neurons and behavior of an individual, which initiate negative mechanism leading to few features which are similar to schizophrenia. Male mice show hyperactivity and alterations in their social interactions while female mice are deficient in spatial memory. These neurophysiological and behavioral defects present in the hippocampus of rodents and primates are similar to human patients of this particular mental illness. According to human schizophrenics, these abnormalities of developing stage in the hippocampus are the major cause of schizophrenia [20]. The DISC1 protein is involved in neurite outgrowth, by its interaction with FEZ1. The findings of the author interpret that neurite outgrowth and synapse is not normal in schizophrenia. On the cellular level, DISC1 is associated with the cytoskeletal and centrosome components which are important in migration of neurons and as well as for outgrowth of neurite process [21].

Various parts of brain for example the hippocampus, prefrontal cortex, amygdala, and thalamus, schizophrenia attribute to the lack of the function. DISC1 is strongly expressed from the reticular nucleus of the thalamus which expresses DISC1 is a central integrating center for signals present between the multiple cortical and thalamic regions, and it plays analytical role in the control of information in the cortex region.

In schizophrenic brain, abnormalities in thalamic filtering of sensory input to the cortex are reported. N-terminal end of amino acids has nuclear localization signal

which is highly conserved among species. Whereas in case of C-terminal which is more conserved then above mentioned terminal of amino acid, is necessary for binding of many proteins including NUDEL with microtubules [22].

Reduced expression of DISC1 alters the microtubule structure which further exert negative effect on different parts of the neurons including extension of the axon, migration of neurons.

All this attribute to cause schizophrenia and also effect development of brain in early stages of an individual. Region of the brain known as orbitofrontal in which expression of DISC1 is altered and hence it is linked to other parts and promotes verbal memory working [22].

1.1 Aims of the Study

Schizophrenia is a polygenic multifactorial disease where multiple genes and pathways are reported to be involved in it. One of the key gene is DISC1 which is part of various pathways hence it plays a key role in schizophrenia pathogenesis.

The translocation in DISC1 is of great significance as this translocation can change the protein structure, therefore changing the major interactors of DISC I resulting in abnormal protein activity. This study is designed to evaluate the impact of translocation in DISC1 on various pathways involved in schizophrenia.

1.2 Objectives of the Study

The Study is designed to achieve given objectives.

1. To investigate deleterious and damaging regions in DISC1 protein variants.
2. To identify structural variations in DISC1 and associated pathways involved in schizophrenia.

3. To identify significant interactors of DISC1 protein and to perform pathway analysis to elucidate DISC1 and its variant in the pathophysiology of schizophrenia.

Chapter 2

Literature Review

2.1 Schizophrenia

It is mind damaging disease which is 1% present around the globe and it can be recognized by different signs and symptoms involving socially unfit and unable to socialize with people in proper manner, emotional mood swings and behaviors, and also deficient cognitively.

Several studies are conducted and it is illustrated that the role of genes and family history is very important in terms of causing schizophrenia and other related disorders. It's firstly identified gene has approximately 400 kb which was present at chromosome 1 and chromosome 11 at the q arm of both chromosomes when balanced translocation is identified and it was segregated with many mind health damaging disorders as observed in pedigree of Scottish family. Figure 2.1 shows symptoms of Schizophrenia [1, 23].

Schizophrenia's causing symptoms are very complicated and diversed so to understand them in a better way, they are divided into many categories and now it is accepted that the symptoms are negative or positive indicating the diversity of the schizophrenia. So,due to the phenomenon of release, positive symptoms arise and when function is lost specialy due to the loss of neuronal signals, negative symptoms occurs [24]. (Table 2.1)

FIGURE 2.1: shows some symptoms of Schizophrenia

TABLE 2.1: Showing Symptoms of Schizophrenia

Types of Symptoms	Characteristics	References
Positive Symptoms	Delusions, hallucinations, disorganized thinking	[25]
Negative Symptoms	Loss of spontaneity, impaired motivation, social withdrawal, diminished capacity of pleasure, cognitive dysfunctions, mood disturbances	[26, 27]

Positive symptoms can be treated by the use of drugs; specially antipsychotic drugs as the positive symptoms are also known as psychotic symptoms where there is loss of contact with reality. To reduce the risk and chance of suicidal attempts in schizophrenic patients on large scale, most commonly used drug antipsychotic drug is clozapine.

Basic or common reason of deficiency in the function of an individual is due to the impairment of the cognitives. Some other factors also play role in function deficiency like effects other members of the family also as they feel burden because of the schizophrenic patient.By looking at the genetics of the family it is concluded that role of defective genes is vital in passing of disorder from one to other generation so these studies can help to understand the pathology of the killing syndrome and also identify the targets which may help in the treatment of the disorder [28, 29].

2.2 Association Between Schizophrenia and Bipolar Disorder

Since past it is concluded and analyzed that schizophrenia and bipolar disorder and so depression are very different type of disorder and they are also not linked with each other clinically but when more studies are performed and conducted, it is indicated that they both belong to the common spectrum of disorder and linked to some extent. But such evidences which indicates the relation of bipolar disorder with family are not really exist still [30, 31]. Incase of schizophrenia, just 2% individuals are first degree relatives and if we look at individuals with schizoaffective disorder only present in above then 1% individuals as compared to bipolar disorder. Other conducted study also reveals that only less than 1% individuals show risk of schizophrenia along with bipolar disorder probands.One more study on the families indicates that approximately 1.5-2.5% bipolar individuals are also at a risk of affecting by schizophrenia and schizoaffective disorder. Collected evidences and data by different studies conducted by molecular genetics and other

studies including pharmacological studies performed in different schizophrenic individuals give us the idea of bipolar disorder and schizophrenia genetic etiology and similarity [32, 33].

Copy number variants between CHD and SAD are identified by the cytogenetic linkages and relative associated studies so that, loss or gain of genes in phynotypic expression results in the variation of the individual and cause illness. Many anti autism treatments are founded but rate of cure is at very less level. [34, 35].

2.3 Relationship of Schizophrenia and Family History

The first study was conducted in 1916 by Rudin so that famililal nature of schizophrenia can be understood in a better way. First degree schizopherenic relatives of the individuals have a higher morbidity risk for schizophrenia of about 10% compared to the controlled risk in the relatives [36].

In 2002, Chang and his fellows reported a report after studying a family living in Taiwan and concluded that by using Weinberg method, risk ratio is about 10 and by Kaplan Meier method, it reaches at 15.They also studied the risk in second degree relatives which was nearly 4% and in third degree relatives it is just 2% [37]. Family studies have been utilized to examine hypotheses involving a difference in rates of schizophrenia by sex or age of onset. In female relatives, the percentage of schizophrenia occurance is higher as compared to males according to this study. In addition, the adoptive relatives had a risk for developing schizophrenia that was similar to the overall risk in the general population [38, 39].

Twin, and adoption studies suggest a familial association between schizophrenia and PDs including schizotypal, schizoid, paranoid, anti-social and avoidant PDs. However, findings on schizophrenia's familial relationship with PDs vary by study [40]. Descriptions of a schizophrenia spectrum vary by study, as no concrete classification for this term currently exists. [41].

2.4 Discovery of DISC1 Translocation

St Clair et al. 1 first reported that a Scottish family is more suspectible with risk of severe health related issues in which translocation of chromosome 1 and 11 occurs. After following the family history for long time, some translocations in the family were noticed and different psycatric issues can be arised at different levels basically schizophrenia [42]. DISC1 was firstly discovered in the translocation of the chromosome 12 as indicated by previously cited literature.This was balanced translocation present at the breakpoint.This translocation causes depression,bipolar disorder and mainly schizophrenia. [43]. The importance of the translocation is very significant as it shows wide range spectrum of psycatric and mental health issues which run in the families with different LOD Scores. Blackwood et al. further reported that these mental health and psychiatric problems are specific and no other distinguishing clinical features are present.

2.5 The t(1;11) (q42.1,q14.3) Translocation

When a survey is conducted relevant to the rearrangements of the chromosomes,an interesting result is seen that translocation of chromosome 1 and 11 in an individual can be analysed by the cytogenetic analysis at clinical level. After the conducted study it is assumed that disrupted in schizophrenia 1 and 2 are the breakpoints on the chromosome 1 where diseased carrying individual can develop schizophrenia,uni and bipolar disorders. [45].

2.6 Effects of Translocation

Disruptor In Schizophrenia 1 is a scaffolding protein, which interact and combine over more than 200 interacting proteins which were originated from yeast model in the beginning but now it is proved by the use of mammal cells and tissues which are used to predict the DISC1 in mammals and humans also.Interestingly,DISC1

interacting interactors are also closely linked and related to each other showing that at some point they have mutual interactions and pathways. Now a days, some scientists discover that the DISC1 gene pathways are also relevant to many other paths of different diseases including Huntington's disease and epilepsy. This all research shows and suggest that the biology of psycatric, cognitive behavior and mood swings are the cause and reason of wide range of disorders and hence increase the importance of the pathways in neurology and pathophysiology. [46, 47]. In protein protein interactions, when, some regions of DISC1 are deleted,mutation arises and region of the interaction is not avalible for the interactions anymore.So, it is not important that binding site of protein is not sured in such protein protein interactions.Some important features like localization signals at subcellular levels and sites of oligomerization are such important features which are represented by Disrupted in schizophrenia 1 gene [48, 49]

2.7 DISC1 (Location, Weight, SNPs Amount)

Schizohrenia,bipolar disorder and clinical depression can be caused by the translocation or mutation of the chromosome 1 and 11. When studies are conducted on large genomics scale,variations and mutations in mental health issues were also studied [50, 51]. Several conducted studies confirmed that the weight of Disrupted in Schizopherenia 1 is near 101kDa which was translated from the 14 exons of the long arm of the gene and about80kDa belongs to the one part of the protein species. DISC1 is abundant in the nucleus and mitochon- dria. [52, 53].

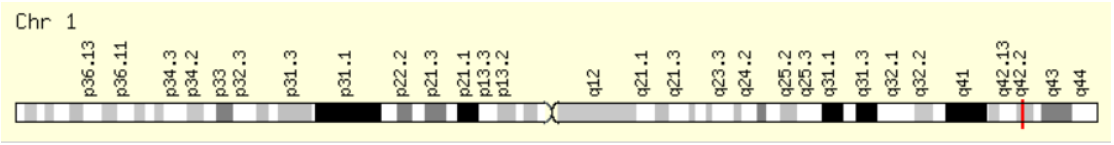


FIGURE 2.2: showing location of DISC1 gene location on Chromosome 1

Several conducted studies confirmed that the weight of Disrupted in Schizopherenia 1 is near 101kDa which was translated from the 14 exons of the long arm of the gene and about80kDa belongs to the one part of the protein species.

DISC1 is abundant in the nucleus and mitochondria. Moreover, a number of DISC1 interactors have been identified using yeast two-hybrid assays and confirmed in follow-up cell-based studies, including FEZ1, platelet-activating factor acetylhydrolase, isoform Ib, PAFAH1B1 or lissencephaly 1 protein and nuclear distribution element-like [52, 53].

2.8 DISC1 Gene Structure

The human DISC1 gene spans approximately 415 kb of genomic DNA and consists of 13 exons producing a full-length transcript of approximately 7.5 kb. The genomic structure of DISC1 is well conserved amongst most species identified to date [26].

In humans, this intron encompasses approximately one third of the whole gene at around 140 kb in length. In humans, this may come in the form of DISC2, the antisense RNA gene that overlaps exon 9 of DISC1, and with its 50 located within DISC1 intron 9, but as yet not formally defined [54, 55].

2.9 DISC1 Protein Structure

The Disrupted in schizophrenia1 sequence has 854 amino acids and has following two regions (I) An N-terminal is a head region which has amino acids from 1-325 aminoacids residues and lack secondary structure containing elements.

(II) An alpha-helix which contains C-terminal :containing coils and a lot of conservation than other end of the terminal [56].

Figure 2.3 shows overall structure of DISC1 protein [57]. Narayanan et al. predicted that the disrupted in schizophrenia S704 variants are in oligomeric state [58].

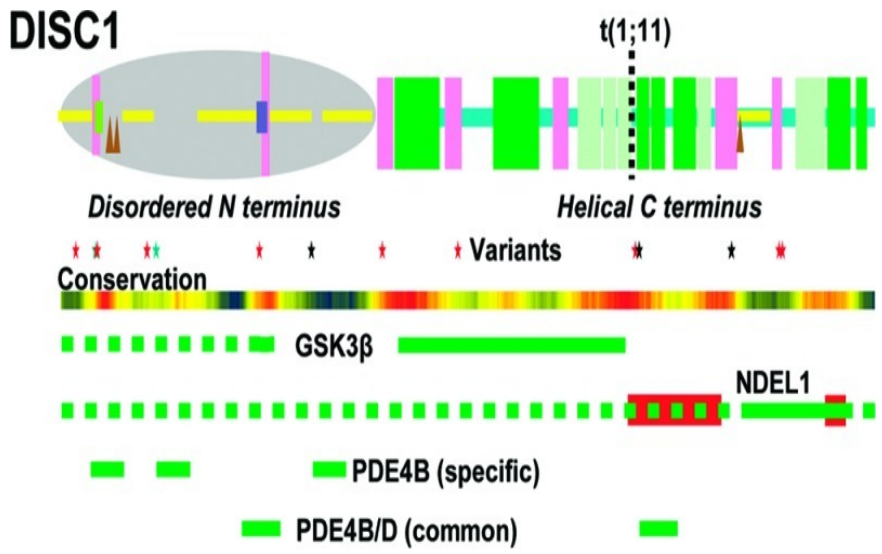


FIGURE 2.3: Structural overview of Human DISC1 protein

2.10 Role of DISC1 Protein

Many efforts in human genetics have been made to test the general validity of the role of Disrupted in schizophrenia 1 in mental health related different issues and psychiatric disorders. Moreover the role for Disrupted in schizophrenia 1 is still under observation. By comparing with other animal models,by different groups it is confirmed that the role of Disrupted in schizophrenia 1 is involved in many important processes involving and including neuronal development and synaptic functions [43]. Disrupted in schizophrenia 1 actively perform multiple steps during the neural development,which may include neuronal migration, neuronal architecture and many other important processes including synaptic plasticity, intracellular transport, and neural signaling. [59, 60].

2.11 DISC1 Protein Sequence Variation: Structural and Functional Impact

When talking about mental illness and disorders, noncoding regions of disrupted in schizophrenia 1 are blamed or responsible for this variation. So the writer assessed and analyzed the protein coding regions which are related with mental health

related issues and location is mapped and predicted sequences in the structure and function are assessed so that secondary structure location,sequence conservation and their overlap on the binding sites and other motifs are analysed easily and more [15, 51, 61, 63].

2.12 DISC1 Variations in Humans.

Uptill now, R264Q,S704C and L607F are identified variations of DISC1 gene relevant to the different mental health related issues.R264Q has vital role and effect on the cortical thickness of the occipital gyrus wheras S704C and L607F shows biological evidence for causing growth of the growing body [66]. Some other variations are also identified but their role is still unknown and they are not important in causing bipolar disorder and schizophrenia[65]. While F607 gene allele carriers shows reduction in the different parts of brain i.e anterior cingulate cortex, left supramarginal gyrus and superior frontal gyrus which are involved in causing disorders like schizophrenia. Whereas in case ofL607F patients show more severe positive symptoms and then in memory it shows increase activity of dorsolateral prefrontal cortex activity [67, 68].

Underlying biological mechanisms behind the observations need to be understood in the better way as variations in F607 is associated with the reduced release of adrenaline which decrease the level of disrupted in schizophrenia's interacting protein at the centrosome and the defects of the mitochondria [69, 70]. Allele of F607 shows low level of the transcript of the Disrupted in schizophrenia 1 and so its level in decreased and altered in the patients of Schizophrenia [71].

In the structure of the brain, the association of the S704 with the C704 allele has increased the volume of the regions present in the brain [72, 73]. Variations in amino acid 704 effect the volume of the gray matter which may include formation of the Para hippocampal and also effect the integrity of the white matter[74-76]. Studies of the functional imaging has predict that homozygotes of S704 has decreased the activity of the hippocampal during different memory tests, greater

hippocampal and paracampal activation during encoding of memory and activation of prefrontal cortex as compared to C704 carriers [77].

The C704 allelle is linked with different test scores and cognitive abilities in different age groups especially in old aged men and mostly in the patients with the positive symptoms[64, 78]. At the molecular level, variations of aminoacid which are changed and also cause change in the expressions of DISC1 transcripts. Moreover, the association of C704 allele is with the reduced activity of the different kinases;serine-protein kinase and alter the activity of the disrupted in schizophrenia for Nuclear distribution protein E homolog like-1 and Nuclear distribution protein E1 and also causes changes in the oligomeric status of the disrupted in schizophrenia 1 [79].

2.13 Ultrarare DISC1 Mutations in Humans

Song et al. sequenced different regions of DISC1 in different patients in which some are patients and other are controlled and include specially the regions like coding exons and splice junctions.

In some patients ultrarare missense are found and in other ultrarare missense mutations are not found so overall it is indicated that risk of schizophrenia is near 2%.So, that Q264R and S704F increase the risk of the schizopheria. G14A, R37W, S90L, R418H, and T603I are cohort specific nonsynonymous variations which are found in schizopherenia [29, 69].

2.14 DISC1 splicing in Humans

The four identified splice forms of human DISC1 are mostly nominated in the following pattern as L means long, Lv means long variant, S means short, and Es means extremely short. The Long form of the variations is coded by the other forms of the exons; the long variant form is different from the long form by the

use of a splice donor site in the 11th exon, which leads to the exclusion of many distal exons alongwith the nucleotides. The Short form of the variant splices from 9th exon to an alternate terminal exon and 3rd untranslated region which is located in the 9th intron. In the mouse, two splice forms have been identified and also verified by different experiments [66, 80, 81]. 2.17. Signaling pathways of DISC1 in the neurogenesis, synaptic plasticity and neural development, many receptors including N-methyl-D-aspartate type glutamate receptor is involved in many different processes. It also play role in cell survival and proliferation and decrease such cells which synthesize DNA and increase number of these cells [82].

2.15 DISC1 and AKT Signaling

AKT is a serine/threonine-specific protein kinase, whose activation in neurons can phosphorylate different substrates and thereafter regulate multiple cellular processes or neuronal development, such as glucose metabolism, apoptosis, cell proliferation, transcription, cell migration, morphogenesis, dendritic development, synapse formation and synaptic plasticity.

Figure 2.4 shows different signaling pathways which are influenced by DISC1. In primary neuronal culture, small interference RNA (siRNA) knockdown of endogenous DISC1 leads to the suppression of phosphorylation of Ras-extracellular signal-regulated kinase (pERK) and AKT (pAKT), suggesting that DISC1 is involved in ERK and AKT activation.

In particular, two DISC1 genetic variants, Ser704Cys and Cys704, appear to participate in ERK and AKT pathways, because over-expression of these two variants by viral transduction in cortical culture results in an increase of pERK and pAKT [71, 87].

Green arrows depict activation enzymes, or otherwise enhancement of the target functions. Red arrows depict inhibition. Black arrows depict effects which do not fall easily into one of the above categories or that are not yet fully understood.

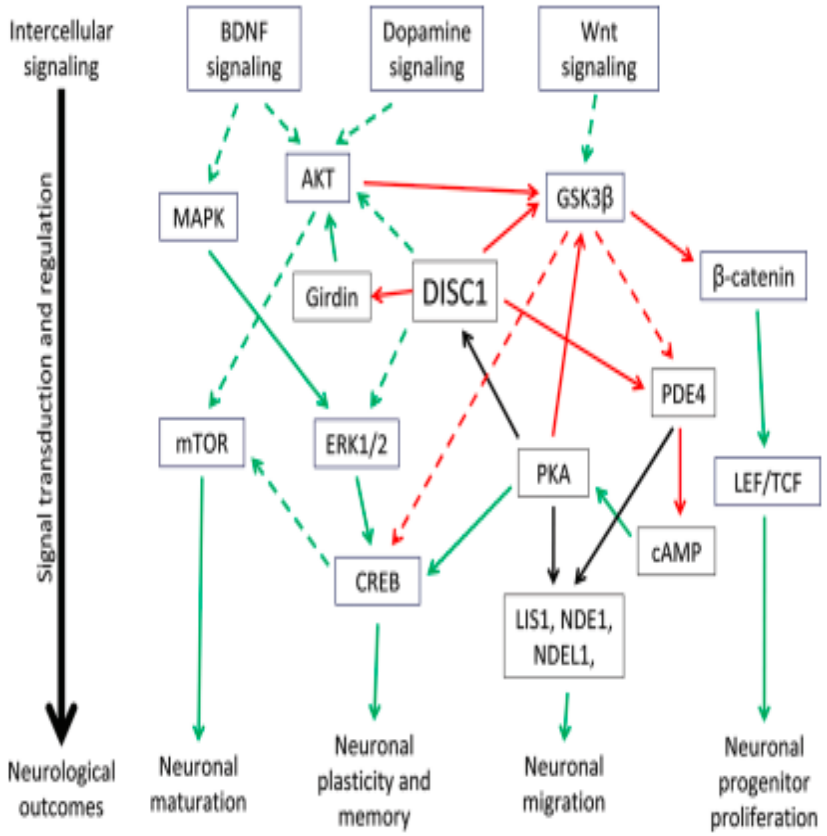


FIGURE 2.4: Flowchart showing cellular signaling pathways influenced by DISC1.

2.16 DISC1 and GSK3β/β-Signaling Pathway

Key regulators in metabolism of glucose are proteins which are multifunctional kinases serine and threonine.In the eukaryotes,GSK3 β is expressed and play role in many important functions like cell adhesion, cell division and differentiation and proliferation [88]. Also, it is expressed widely in the central nervous system so that it is expressed in the brain during the developmental stages and remains till adulthood. In the neurons of hippocampal, GSK3β is expressed I the whole cell [89].

Chapter 3

Materials and Methods

3.1 Detection of Deleterious and Damaging Regions in DISC1 Protein Variants

Following methodology is followed in order to achieve our aims and objectives.

3.1.1 To Detect the Variant Sequences of DISC1 Gene

Genetic variations including indels, transversions or translocations can result in variations in protein sequencing hence the 3D structure of proteins are also changed. DISC I is one of the central gene reported to be associated with schizophrenia. To detect the variant sequences of DISC1 gene were retrieved from NCBI (https://www. ncbi. nlm. nih. gov/protein/) and 23 such sequences were obtained from NCBI which were saved for further process. NCBI includes more than 3700 organisms of all types and more than 2879800 proteins. Resulting sequences retrieve coding regions, conserved domains, names and much more relevant information obtained by scientific community, propagated from GenBank and NCBI [91].

3.1.2 Detecting Disordered Protein Region in DISC1 Protein

In the next step, disordered Protein region was detected by the use of DisEMBL(Intrinsic Protein Disorder Prediction 1. 5(http://dis. embl. de) tool. This method rely on artificial neural networks used to predict the disorder in the protein sequence segment.

This tool is used for target selection mostly used in structural genomics. Here such proteins are considered disordered which allow more sites for and interacting partners. Certain parameters were set to get accurate result. These parameters include firstly Loops/coils threshold whose value is 1. 20, Hot- loops threshold is 1. 40 and Remark-465 threshold is 1. 20. Temperature is set to 298. 15 kelvin, ionic strength is 0. 02M and Ph is 7. 40. Values of all these parameters are set by default so it is fine to move ahead with them [92].

3.1.3 DSSP Loop in DISC1 Protein (Secondary Structure)

According to DSSP loops or coils are defined as the residues which belong to secondary structure. Residues include H, G and E as ordered ones. Coils are denoted as T, S, B and I. In the loops, protein disorder is found. [93]. Hot loops include above mentioned loops with mobility of high degree along with B factor. This B factor is used to determine, define and predict protein disorder [94]. Remark defined missing 465 entries in coordinates in the structure of X-ray. It reflect intrinsic disorder which is also used to predict disorder [95].

3.1.4 Predicting Disordered Protein Region in DISC1 Protein

Results of DisEMBL tool are further verified by PrDOS server (Protein Dis Order Prediction System) (http://prdos. hgc. jp). PrDOS also predicts disordered

region of protein from the given sequence of the sequence. Protein sequence is pasted in the form of FASTA format. Certain parameter like false prediction rate is set at 5% which is already set by the server as it gives perfect results [96].

3.1.5 Predicting Functional Variation in Amino Acids of DISC1 Protein

Funtrp(Function Neutral/ Toggle/ Rheostat Predictor)(https://services. brombergl ab. org/funtrp/) is used to create functional maps of sequences of proteins and to predict the functional variations in the amino acid sequence by giving input in fasta sequence and email id is provided in order to get the results. Neutral has mostly no effects; rheostat has different functional variations while toggle has mostly strong effects. This tool also predicts the possible mutations in the amino acids position [97]. The highlighted amino acid mutations containing sequences were used along with fasta sequence of proteins to add input for provean tool. According to human gene

mutation database(HGMD), at the gene level when human disease variation is sequenced more than half variations are single nucleotide substitutions and 25% of mutations are linked with small indels [98].

3.1.6 Investigating the Substitutions in the DISC1 Protein

So to predict the variations in amino acid sequences, we used PROVEAN algorithm (Protein Variation Effect Analyzer) (http://provean. jcvi. org/)which predict insertions, deletions and multiple substitutions in the structure of protein [99].

It is software tool which is used to detect damage protein in the protein structure of any organism. Input is variant protein sequence of schizophrenic humans obtained from NCBI and amino acid variants are also added as input. Query sequence is in the form of fasta sequence. Result may take few minutes [100]. Results of PROVEAN are further analyzed by Polyphen-2 tool (polymorphism phenotyping

v2) (http://genetics. bwh. harvard. edu/pph2/) which predicts about impact of substitution of amino acid on the function, stability and structure of human proteins. It is unique from other tools in case of predictive features, alignment and classification methods [101].

This software is available both as web server and as software. It finds the annotation of single nucleotide polymorphism, protein structural annotations and many other such annotations related to proteins. After the annotation detection, missense mutations probability is estimated. Polyphen score, sensitivity and specificity is calculated and hence status of protein is calculated. Firstly fasta sequence of protein is pasted then the position of amino acid substitution is added. Then amino acids substitutions are selected then query is submitted. Result page is refreshed after few seconds. By clicking on the results, result can be displayed [102].

3.2 Identification of Structural Variations in DISC1 and Associated Pathways Involved in Schizophrenia

To identify the DISC1 structural variations and associated pathways involved in schizophrenia following tools and softwares were used.

3.2.1 Prediction of Thermal Stability of the Protein

I-Stable 2. 0 (http://ncblab. nchu. edu. tw/iStable2) is basically used to predict the thermal stability of the protein sequences by using many characteristic modules. Different machine learning methods include support vector machine (SVM),neural network (NN)which are used in i-Stable 2. 0 tool. I-Stable can be operated by two input types: on the base of structure and sequence. Sequence based methods include I-Mutant 2. 0 and MU pro as an example.

I proceed on the protein sequence as my protein structure is not predicted yet [104]. In the first step protein sequence is added in the fasta format, in the next step mutation site is selected and residue position is also located than in third step, wild type amino acid is added along with position of amino acid. Certain parameter like temperature is set at 25℃ and pH is set at 7. Then it is submitted, after few minutes result appeared.

3.2.2 Investigating the Structural and Functional Properties of DISC1 Protein

MutPred 2(http://mutpred. mutdb. org/)is effectively used to detect functional and structural properties of amino acids. It is web based tool which identify amino acid substitutions whether they are benign or damaging. It predict 50 different properties which may include altered disordered interface, altered DNA binding, gain of helix, loss of strand and loss of phosphorylation site etc. so that molecular mechanisms of pathogenicity can be interpreted. Protein sequence in fasta format and substitutions in amino acids are submitted. p value is fixed at 0. 05 by default [103].

3.2.3 Predicting Functional Regions in DISC1 Protein

Consurf server (http://consurf. tau. ac. il/) is used to predict functional regions present in the protein sequence. It also checks and find evolutionary conservative positions of amino acid in the protein sequence. The evolutionary conservation of amino acid depends upon the structural and functional importance. Analysis is done in order to build phylogenetic tree by using homologous sequences of proteins. This server can be used to find sequence homologs, align sequences and then select best evolutionary model [105]. It can also calculate conservation score and project these scores on the molecule. It can search for 3D structure for protein sequence. Firstly, home page of the server asks about what we want to analyze amino acid or nucleotides than is protein structure is already known or

not? is multiple sequence alignment is available or not? Than protein sequence in the fasta format is submitted. Following parameters were set to search homolog algorithms: Multiple sequence alignment was done by using MAFFT method. Homologues sequences were collected from UNIREF90 and search algorithm used for this purpose was BLAST. E-value set for PSI-BLAST is 0. 0001 and number of these interactions were 3. Maximal ID% between sequences is 95 and minimal ID% for homologs is set at 50. According to our query, 10 closest sequences were required and set as a limit [106].

3.2.4 Constructing Phylogenetic Tree

Phylogenetic tree is constructed by using Neighbor joining with maximum likelihood distance method. Method of calculation was Bayesian and Best fit was model which was substitution for proteins. At the end, mail id and job title (sequence detail) was given in order to get the result [107].

Swiss Model (https://swissmodel. expasy. org/) is used to predict the protein models which are approachable to the researchers all over the world. This is web based server on which 3D protein homology modeling is done [108].

SWISS MODEL Repository is a database which contains about 4000 protein models which are generated automatically. SWISS MODEL follows the following methodology. Input data is provided in the form of amino acid sequence in the form of fasta format or as simple text then template of the provided data is searched from the template library of SWISS MODEL which is abbreviated as SMTL [109].

This uses BLAST; fast and accurate method for closely related templates and other search method is HHblits which is best for remote protein structures. After template searching, they are ranked according to desired quality and user has option to view all available templates and can construct model of choice. Than for every selected template 3D protein model is generated automatically which can be viewed [110]. To know whether known structure of my sequences are present

or not i started the process in Swiss model expasy server. In the first step, on the home page different options are present. I opted for start modeling then in the next step; i pasted my sequence and also added my email id so that results can be viewed later on also. Then finally i click find template option. No specific parameters were present at this point [111].

3.2.5 Phyre2, RaptorX Servers Used to Construct DISC1 Structure

Phyre2(ProteinHomology/analogYRecognitionEngineV2. 0.)(http://www. sbg. bio. ic. ac. uk/phyre2)is used to detect the 3D protein models of mutant sequences. In the initial step sequence of desired protein is pasted in fasta format, job description and to provide email is mandatory. Modeling mode is set to normal. After submitting the information, the result is available after 30 minutes to 3 hours depending upon the query sequence and also on the number of queries submitted by other users [112]. By following the given steps result of given sequence compiles: finding homologues with PSI-Blast, building hidden Markov model of given sequence then 3D model of the desired protein is constructed on the base of alignments between hidden markov model and the desired sequence. Then finally top matched model is submitted which can be viewed easily. When the submitted job is completed, then email is received which contains summary, job identifier and link is provided which open ups the main result page and an attachment file containing top scoring models in PDB format [113].

Raptor X (http://raptorx. uchicago. edu/) is used for the prediction of secondary structure and tertiary structure modeling of the templates. I basically used this World Wide Web based server to validate the structures which were earlier detected by Phyre2 server. This server also predicts binding sites, disordered regions, solvent accessibility along with secondary and tertiary structure prediction [114]. On the home page of Raptor X there are many options like structure prediction, structure alignment and property prediction. I opt for structure prediction, then by clicking on the submit option, next page appears; here i submit my protein

sequence in fasta format along with job name and email id. Here i noticed interesting parameter that they are currently accepting only that proteins which have amino acids less than 1000 and predicting their 3D models. After submitting the sequence, a page appears which tells about status of my job and also amount of jobs which are currently running on the server around the globe. After the completion of the job result was emailed to me [115].

3.2.6 Saves 5.0 Server and TM-align Used to Evaluate the Predicted Models

Saves 5. 0 (https://saves. mbi. ucla. edu/) is use to confirm the protein structure stability and by providing graph in the result, compares expected and observed value. Verify determines the compatibility of 3D model with the input amino acid and after comparing with good structures, also assign scores to them [116]. ERRAT is used to analyses the statistics of different type of atoms that are non-bonded interactions and also plot the values of the errors and determine the quality factors. PROVE calculates the atoms volume by calculating and running algorithms so that it calculates the deviation of the Z-score from highly refined and resolved PDB structures and also calculate the error present. WHATCHECK is derived from subset of protein verification tools from the WHATIF program. It checks sterochemical parameters of the residues present in the model. When proceeding to the methodology, this server accepts pdb file only as a input so on the home page I entered the pdb file of the sequence then click on run saves. After few minutes I got the results [117].

TM-align(https://zhanglab. ccmb. med. umich. edu/TM-align/)is specific algorithm used to sequence independent protein structure and compare the disordered and wild protein sequence. It first aligns the structure on the base of similarity in the structure. TM score has two values which are 1 and 0. 1 shows the perfect structure match [118]. On the other hand, score below 0. 2 indicate unrelated proteins which were randomly choosed and if score is above 0. 5 then indicates presence in same fold. TM-algorithm computes TM-score(template modeling

score),RMSD(root mean square deviation, aligned or identical score and length of aligned sequence. Higher value of RMSD is result of greater or more variation between the disordered or mutant and wild protein structures [119].

On the home page of TM Align server option to add input sequence is present which accept only PDB file. So, in the first input box, I added the pdb file of reference or wild type protein sequence predicted in previous step. Now, I moved to other input accepting box and here I added the pdb file of mutant or disordered protein sequence. By proceeding further I added my email id and click on Run TM-align. Result compilation took few minutes [120].

3.3 Identification of Significant Interactors of DISC1 Protein

To identify the interactors involved in the pathways and interactions of dislocation of DISC1 protein in the schizophrenia, I used and apply given tools and soft wares:

3.3.1 Protein Protein Network Development

PICKLE 2. 0 (Protein InteraCtion KnowLedgebasE) (http://www. pickle. gr/) is meta-database which is used to detect direct protein-protein interaction network in humans.

It uses the reviewed human complete proteome of Uniprot as a standard. To find the protein-protein interactions in the DISC1 and associated genes, I opted for Pickle 2.0 database as it perform the action and find protein-protein interactions of mouse and humans directly so as I am working on human gene so I go to home page of Pickle 2. 0. Here general description of database, contributing databases, general statists along with the option to search for PPI is present. On clicking the given option, next page appears. Here search bar for the specie and gene name is present so I entered them and proceed further by clicking on the search for the

identifiers in the PICKLE RHCP/RMCP ontology network [121]. GENEMANIA (https://genemania. org /) is used to predict the function of genes and set of genes. It is user friendly and flexible web interface which can generate functions of genes and select genes for functional assay. It can operate for single gene queries, multiple gene queries and for searching network. Genemania has high accuracy rate algorithm, large database so it is very useful for analyzing function genes and vice versa [122].

To know about the function and interactions of gene I entered the link of Genemania, then a page opens on which on the top right side of home page, option to add name of gene is present so I added my gene of interest and click on GO tab. Result processing take some time [123].

3.3.2 Analysis of Functional Module within Network

Gephi 0. 9. 2 is software which is used to visualize and analyze the graphs and networks freely. It helps the user to explore and manipulate the interacting networks. It can deal with 20,000 nodes at a time. To operate the Gephi, it is compulsory to save edges and nodes of DISC1 in which gene1 acts as source and gene2 is target one [124]. Gephi is downloaded from http://gephi. org and then after downloading, when I open the software, home page appears showing different options, I select start new project, click on the file option then I selected nodes file from my previously saved data. After few moments screen showing nodes appear, than I click on next and then finish option. Then after few seconds, new screen appears with many options, I selected append to existing and then click on ok option. Then a screen appears showing nodes of DISC1 network. In the similar manner I added data of edges into the gephi and suddenly network of nodes and edges appears which is showing interaction of DISC1 with other genes. From choose layout option, option of Fruchterman Rheingold is selected and then it is run. Option present at lower side is used in order to colour the nodes and also label the interacting nodes. Network overview is present at left corner so I run the values and saved the data in MS Word file. When I clicked on colour, then

on nodes after then on degree, appears showing the percentage of the nodes. The graph is coloured now according to the modules. [125].

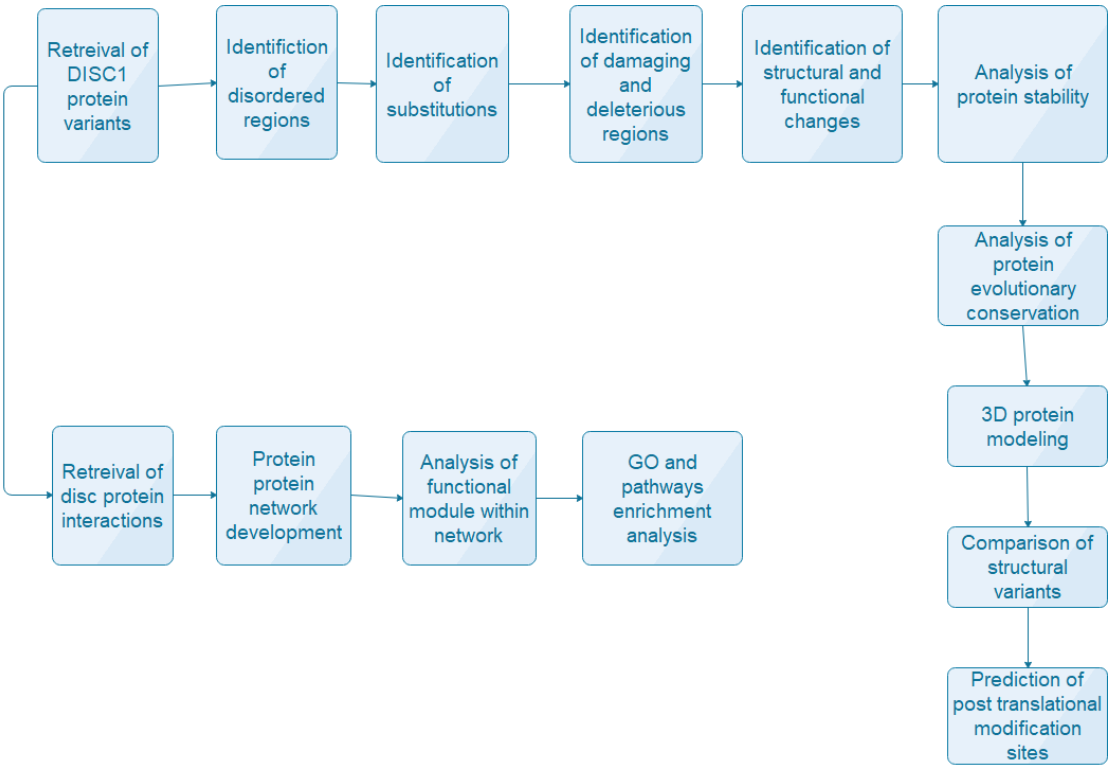


FIGURE 3.1: Methodological Steps to Evaluate the Impact of Translocation in DISC1 on Schizophrenia

3.3.3 GO and Pathway Enrichment Analysis

EnrichNet(http://www. enrichnet. org/) is used for enrichment analysis of the network. It is web based tool which evaluate function, components, processes and pathways among the proteins and genes. So, to evaluate the pathway I moved towards the procedure and In the first step, I clicked on the given above link then home page of EnrichNet appears on which option of entering input gene/protein is given. Molecular network is set to STRING as default and Identifier format is set to ENSEMBL ID. I entered list of nodes as input which was obtained earlier. Then I clicked on go to next step option. Then next page appears, here choose an annotation database option appears on which different pathways and processes options are present. So, firstly I opt for KEGG then on start analysis. After

few minutes, results appears [126]. After then I choose for GO gene (Biological process) and repeat the above methodology. Then proceeding forward, to know about molecular function I select GO gene ontology(Molecular functions) and wait for the result for few minutes and after getting that I move ahead to the last step in order to know about subcellular localization, I choose GO gene ontology(cellular components)option and after 2-3 minutes I got my final result [127].

Chapter 4

Results and Analysis

Genetic variations including indels, transversions or translocations can result in variations in protein sequencing hence the 3D structure of proteins are also changed. DISC I is one of the central gene reported to be associated with schizophrenia.

Translocations in DISC I gene not only change the protein sequence but also can impact 3D structure of protein. When the 3D structure is changed, the interactors are also changed and mutated protein results in activation of abnormal pathways. This thesis was designed with an aim to explore these impacts of mutations of DISC I protein structure and function. Following are results which were obtained:

4.1 Detection of Deleterious and Damaging Regions in DISC1 Protein Variants

Normal DISC1 Protein is involved in performing many important functions including neurogenesis regulation, development of brain during early embryonic stage. It has also vital role in different signaling pathways.

DISC1 Sequence has two regions which contain 854 amino acids. Deletions and damaging in DISC1 gene causes many psychiatric conditions like bipolar disorder, severe depression and schizophrenia.

4.1.1 Detection of Variant Sequences in DISC1 Gene

23 disordered sequences of DISC1 gene present in the FASTA format were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/protein/)and saved in MS word file as shown in the table below.

TABLE 4.1: showing protein sequences of DISC1

S no	Protein sequence	Fasta Format from NCBI
1	NP_001158021.1 disrupted in schizophrenia 1 protein isoform l [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRRYC
2	NP_001158020.1 disrupted in schizophrenia 1 protein isoform k [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRRKPFLDG
3	NP_001158019.1 disrupted in schizophrenia 1 protein isoform i [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRRNKCEGKYYEVHGNT
4	NP_001158018.1 disrupted in schizophrenia 1 4protein isoform i [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRRNKCEGKYYEVHGNT

Table 4.1 continued from previous page

S no	Protein sequence	**Fasta Format from NCBI**
5	NP_001158017.1 disrupted in schizophrenia 1 protein isoform h [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKETISGRLKTSPRRLDH
6	NP_001158016.1 disrupted in schizophrenia 1 protein isoform g [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYGR
7	NP_001158014.1 disrupted in schizophrenia 1 protein isoform f [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYETSVKENTMKYMETLKNKLCR
8	NP_001158013.1 disrupted in schizophrenia 1 protein isoform e [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPG(VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI
9	NP_001158012.1 disrupted in schizophrenia 1 protein isoform d [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQHLQERIKSLNLSLKEITTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIR SLTSEREGLEGLLSKLLVLSSRNVKKLGSVKEDYNRLRREVEHQETAYETSVKENTMKYMETLKNKLCSC KCPLLGKVWEADLEACRLLIQSLQLQEARGSLSVEDERQMDDLEGAAPPIPPRLHSEDKRKTPLKVLEEW KTHLIPSLHCAGGEQKEESYILSAELGEKCEDIGKKLLYLEDQLHTAIHSHDEDLIQSLRRELQMVKETL QAMILQLQPAKEAGEREAAASCMTAGVHEAQA

Table 4.1 continued from previous page

S no	Protein sequence	**Fasta Format from NCBI**
10	NP_001158011.1 disrupted in schizophrenia 1 protein isoform c [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYDGVSLCRPVWSAVVRSCSLQPLPPEFKQFSCLSLRSSWDY RCPPPCLANFVFLVEMGFYHVDQTGLKLLTSSDPPSSASQSAGITDMSHCAWPLQ
11	NP_001158010.1 disrupted in schizophrenia 1 protein isoform b [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYETSVKENTMKYMETLKNKLCSCKCPLLGKVWEADLEACRL LIQSLQLQEARGSLSVEDERQMDDLEGAAPPIPPR
12	NP_001158009.1 disrupted in schizophrenia 1 protein isoform a [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKVETGFHYVGQAGLELLTSSNPPASASQSAGITAETLQQRLEDLEQEKI SLHFQLPSRQPALSSFLGHLAAQVQAALRRGATQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWL LQEKQQLQKEIEALQARMFVLEAKDQQLRREIEEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLA SAGQIPFHAEPPETIRSLQERIKSLNLSLKEITTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISG NHFWTAKDLTEEIRSLTSEREGLEGLLSKLLVLSSRNVKKLGSVKEDYNRLRREVEHQETAYETSVKENT MKYMETLKNKLCSCKCPLLGKVWEADLEACRLLIQSLQLQEARGSLSVEDERQMDDLEGAAPPIPPRLHS EDKRKTPLKVLEEWKTHLIPSLHCAGGEQKEESYILSAELGEKCEDIGKKLLYLEDQLHTAIHSHDEDLI QSLRRELQMVKETLQAMILQLQPAKEAGEREAAASCMTAGVHEAQA
13	NP_001158028.1 disrupted in schizophrenia 1 protein isoform t [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRRKPFLDG
14	NP_001158027.1 disrupted in schizophrenia 1 protein isoform r [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEL RRYNKD

Table 4.1 continued from previous page

S no	Protein sequence	**Fasta Format from NCBI**
15	NP_001158026.1 disrupted in schizophrenia 1 protein isoform q [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKGEF
16	NP_001158025.1 disrupted in schizophrenia 1 protein isoform p [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAGTNCFGSTMEASTS
17	NP_001158024.1 disrupted in schizophrenia 1 protein isoform o [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKGLLEEVATSHLTLHT
18	NP_001158023.1 disrupted in schizophrenia 1 protein isoform n [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQ
19	NP_001158022.1 disrupted in schizophrenia 1 protein isoform m [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKGSHWKGYIFIWGEQQLWIRIMKIDNGKWACHSGTFPSFFPEPAGTNCF

Table 4.1 continued from previous page

S no	Protein sequence	**Fasta Format from NCBI**
20	NP_001012977.1 disrupted in schizophrenia 1 protein isoform S [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYGYKYCDAESWTQRSQQLA
21	NP_001012976.1 disrupted in schizophrenia 1 protein isoform Es [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEL EPIALDPPWKPRHPEPNSY
22	NP_001012975.1 disrupted in schizophrenia 1 protein isoform Lv [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYETSVKENTMKYMETLKNKLCSCKCPLLGKVWEADLEACRL LIQSLQLQEARGSLSVEDERQMDDLEGAAPPIPPRLHSEDKRKTPLKESYILSAELGEKCEDIGKKLLYL EDQLHTAIHSHDEDLIQSLRRELQMVKETLQAMILQLQPAKEAGEREAAASCMTAGVHEAQA
23	NP_061132.2 disrupted in schizophrenia 1 protein isoform L [Homo sapiens]	>MPGGGPQGAPAAAGGGGVSHRAGSRDCLPPAACFRRRRLARRPGYMRSSTGPGIGFLSPAVGTLFRFPGG VSGEESHHSESRARQCGLDSRGLLVRSPVSKSAAAPTVTSVRGTSAHFGIQLRGGTRLPDRLSWPCGPGS AGWQQEFAAMDSSETLDASWEAACSDGARRVRAAGSLPSAELSSNSCSPGCGPEVPPTPPGSHSAFTSSF SFIRLSLGSAGERGEAEGCPPSREAESHCQSPQEMGAKAASLDGPHEDPRCLSRPFSLLATRVSADLAQA ARNSSRPERDMHSLPDMDPGSSSSLDPSLAGCGGDGSSGSGDAHSWDTLLRKWEPVLRDCLLRNRRQMEV ISLRLKLQKLQEDAVENDDYDKAETLQQRLEDLEQEKISLHFQLPSRQPALSSFLGHLAAQVQAALRRGA TQQASGDDTHTPLRMEPRLLEPTAQDSLHVSITRRDWLLQEKQQLQKEIEALQARMFVLEAKDQQLRREI EEQEQQLQWQGCDLTPLVGQLSLGQLQEVSKALQDTLASAGQIPFHAEPPETIRSLQERIKSLNLSLKEI TTKVCMSEKFCSTLRKKVNDIETQLPALLEAKMHAISGNHFWTAKDLTEEIRSLTSEREGLEGLLSKLLV LSSRNVKKLGSVKEDYNRLRREVEHQETAYETSVKENTMKYMETLKNKLCSCKCPLLGKVWEADLEACRL LIQSLQLQEARGSLSVEDERQMDDLEGAAPPIPPRLHSEDKRKTPLKVLEEWKTHLIPSLHCAGGEQKEE SYILSAELGEKCEDIGKKLLYLEDQLHTAIHSHDEDLIQSLRRELQMVKETLQAMILQLQPAKEAGEREA AASCMTAGVHEAQA

The obtained results from disEMBL shows some amino acids in small and other in Capital letters letters, one which are in small letters with black color are disordered proteins as compared to normal proteins which are coloured and in capital letters. The given results of loops/coil disorder also contain none loops which indicates ordered amino acids [128].

4.1.2 Detection of Disordered Protein Regions

fuNTRp is important tool as it can predict and uncover the association of disease and also evolution mechanism and the output of the fuNTRp server is a graph showing neutral, toggle and rheostat values along with prediction scores. And a table is appeared which can be downloaded in the Microsoft excel file form and it give us the scores prediction, neutral value, toggle score and rheostat score. This table will also show amino acids in which mutations are occurred. So, we highlighted the amino acids with the mutations and then proceed further to next Step [129].

The given results are present in annexure 1 which contain tables and talks about the position of amino acid on which mutations occurs, along with different score values. In my obtained result Neutral score indicate mostly no or weak effects, toggle score predicts variant range of function tuning positions whereas rheostat score shows strong effects on the positions of amino acids in the protein sequence.

4.1.3 Deleterious Regions Detection and Substitutions Prediction in DISC1 Protein

Output of provean is the PROVEAN score and prediction. Default threshold of prediction is -2.5 so that variants with -2.5 score or below are deleterious whereas variants with score of -2.5 or above are neutral and are deleted from protein sequence. Result page of Polyphen 2 tool give details about the query sequence. It describes length of sequence, amino acid substitutions position. Prediction score are from 0-1.

Sensitivity and specificity depends upon Polyphen score. The value more nearer to zero is considered to be benign so that this protein is not damaged or mutated. If Polyphen score is 1 or nearer to 1 than chance of damage is more in this protein sequence [129]. Tables are present in annexure II which shows results of all 23 sequences and also tells about damaging amino acids.

4.2 Identification of Structural Variations in DISC1 and Associated Pathways Involved in Schizophrenia

(A) Functional and structural properties analysis of DISC1 Protein Output of Mutpred 2 gives p-value where values less than 0.05 are considered as hypothetically correct and confident [130].

The result of I-Stable server will show confidence score of the predicting protein sequence. I-Mutant 2.0, MUproSVM, MUproNN are element predictors which show increase or decrease along with different positive and negative values in front of mutations.

Values which show decrease are noted in the table as they show decrease in the stability. Some graphs and structural information is also present in the result but they are not noted in our methodology [131]. Tables in Annexure II show the mutant amino acid and its value by different predictors.

4.2.1 Functional Regions Prediction in DISC1 Protein

(B) The given result of consurf server includes multiple sequence alignment, phylogenetic tree. UniRef90 consists of 11 or more residues which shows 90% or more sequence identity with other organisms.

So, I saved the desired data gathered by UniRef90 which include multiple sequence alignment which was differentiated by different colour coded by conservation.

Obtained result data can be divided into three categories on the base of conservation: variable (blue colour), average (white colour) and conserved (pink colour). The obtained result is the comparison of our input sequence with related protein sequences of other organisms as shown in the table 4.2 [132].

TABLE 4.2: showing protein sequences of DISC1

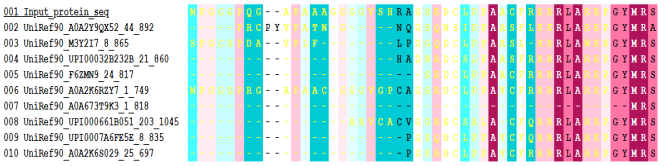
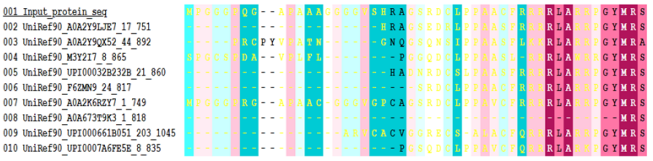
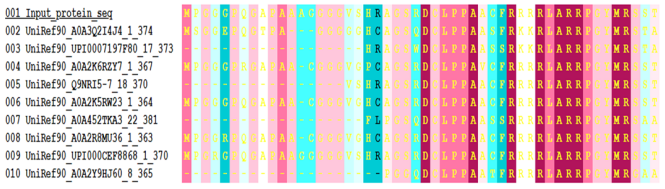
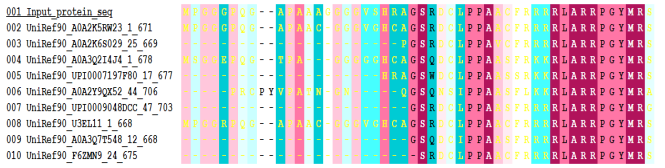
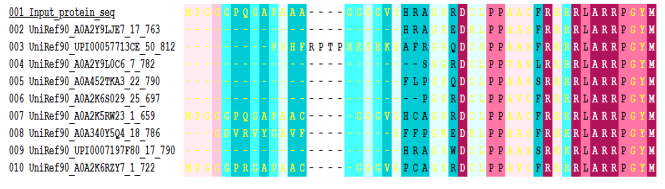
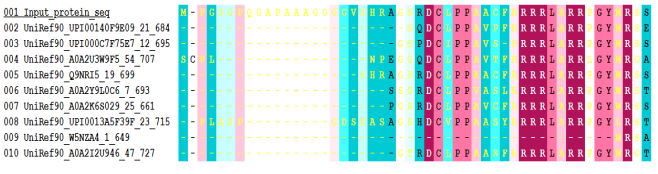
S no	Protein Sequences	Multiple Sequence Alignment(MSA)
1	NP061132.2	
2	NP001012975.1	
3	NP001012976.1	
4	NP001012977.1	
5	NP001158009.1	
6	NP001158010.1	
7	NP001158011.1	

Table 4.2 continued from previous page

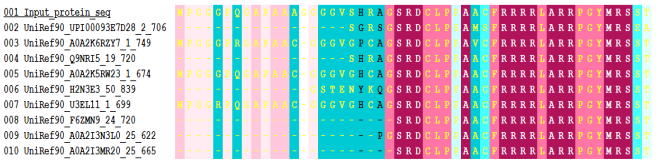
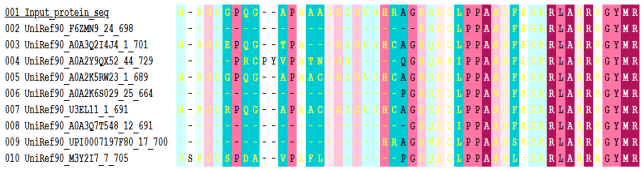
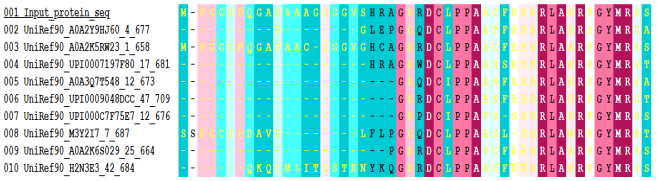
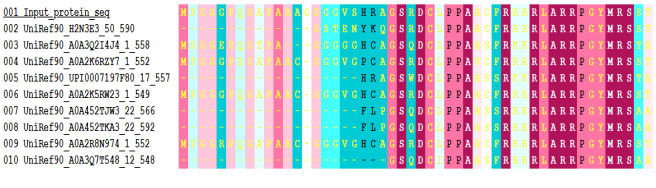
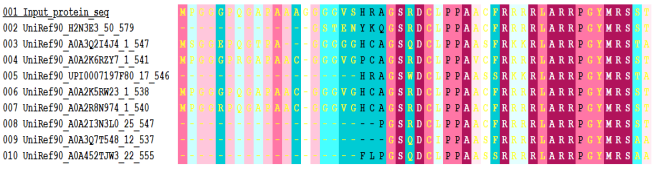
S no	Protein Sequences	Multiple Sequence Alignment(MSA)
8	NP001158012.1	
9	NP001158013.1	
10	NP001158014.1	
11	NP001158016.1	
12	NP001158017.1	
13	NP001158018.1	
14	NP001158019.1	

Table 4.2 continued from previous page

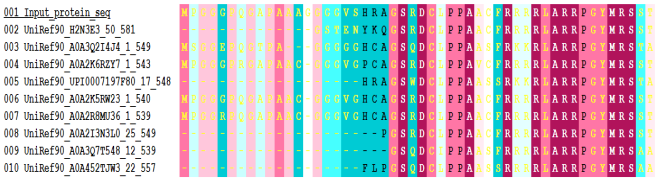
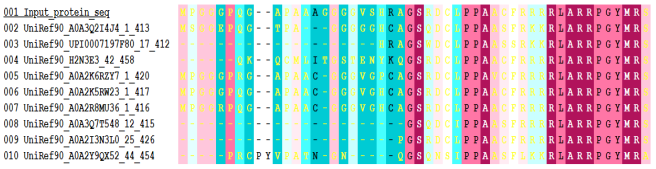
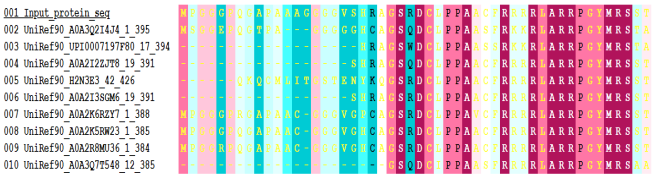
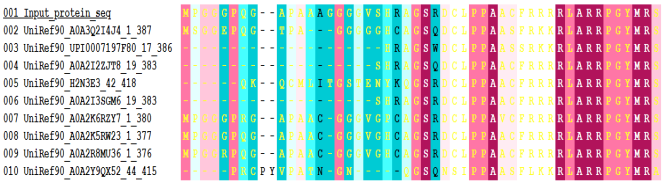
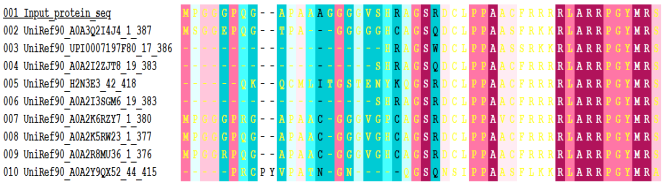
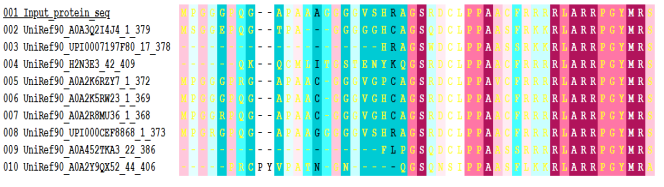
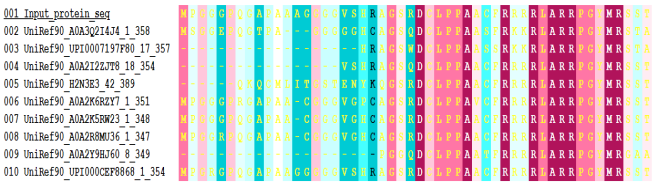
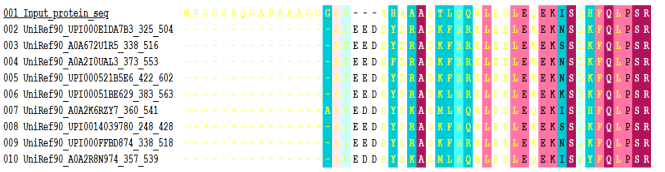
S no	Protein Sequences	Multiple Sequence Alignment(MSA)
15	NP001158020.1	
16	NP001158021.1	
17	NP001158022.1	
18	NP001158023.1	
19	NP001158024.1	
20	NP001158025.1	
21	NP001158026.1	

Table 4.2 continued from previous page

S no	Protein Sequences	Multiple Sequence Alignment(MSA)
22	NP001158027.1	
23	NP001158028.1	

Multiple sequence alignment (MSA) is the alignment of three or more biological sequences of same length arranged in a column. The obtained result of multiple sequence alignment, predicts homology of the sequences and also the evolutionary, structural and functional relationship.

Colour coding bar indicate the conservation of sequences. Turquoise colour shows variable conservation,White colour indicate average conservation, maroon colour predicts highly conservation while Yellow colour shows insufficient data and low confidence.

4.2.2 Construction of Phylogenetic Tree

Whereas phylogenetic tree is also predicted by using WASABI. It is such browser-based application which is used and is helpful for analysis and visualization of multiple alignment sequence data at molecular level. So, I saved the desired phylogenetic tree which show the conservation of our sequence with other sequences [133]. Table 4.3 given below shows phylogenetic trees of 23 sequences of DISC1 Protein and shows their relation with closest protein sequences of other species. Table 4.3 shows phylogenetic tree of protein sequences of DISC1

TABLE 4.3: shows phylogenetic tree of protein sequences of DISC1

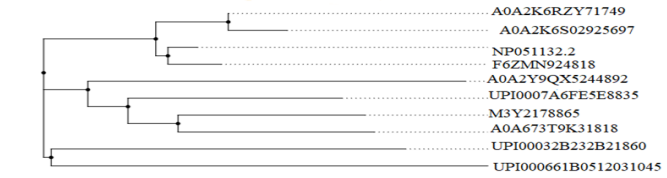
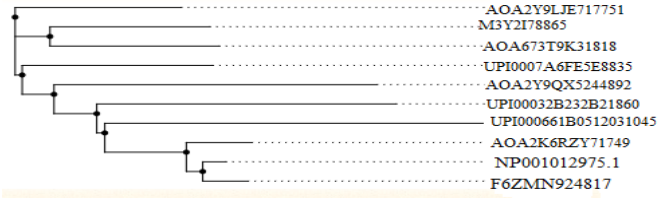
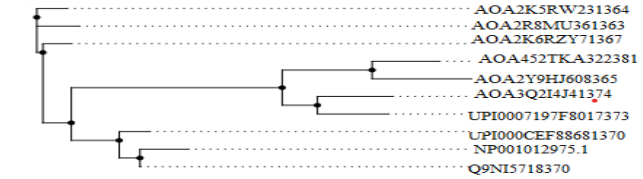
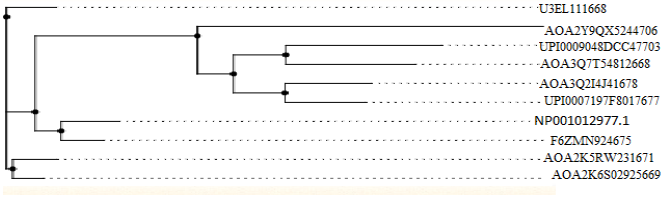
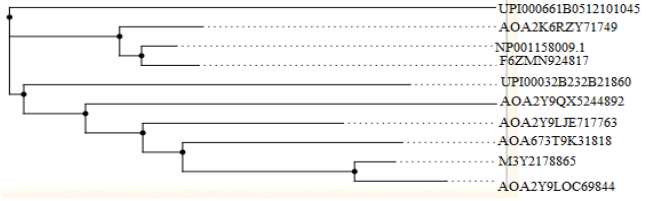
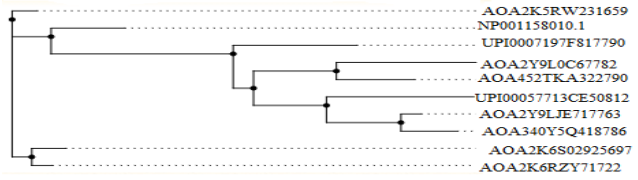
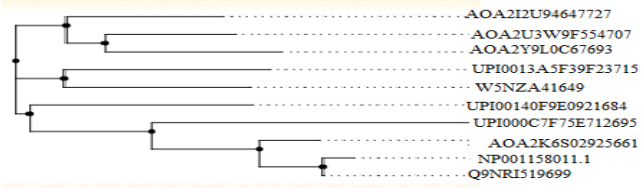
S no	Protein Sequences	phylogenetic tree of protein sequences
1	NP061132.2	
2	NP001012975.1	
3	NP001012976.1	
4	NP001012977.1	
5	NP001158009.1	
6	NP001158010.1	
7	NP001158011.1	

Table 4.3 continued from previous page

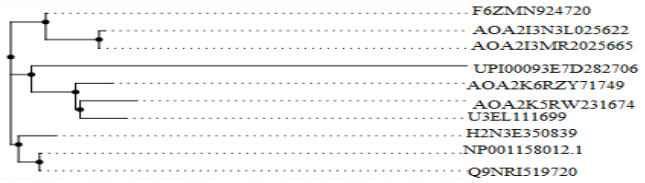
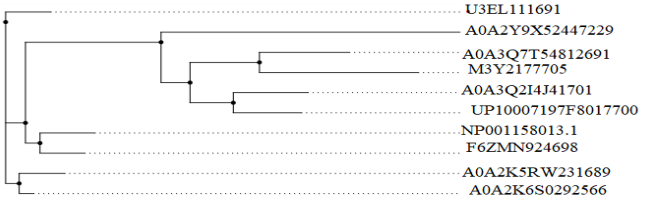
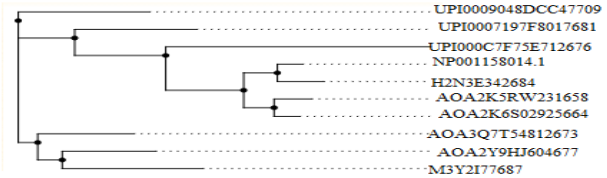
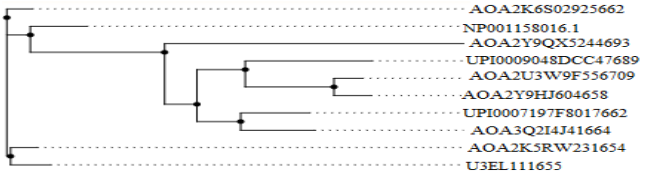
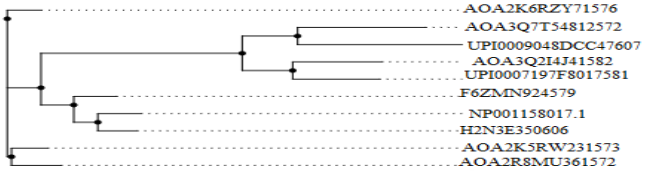
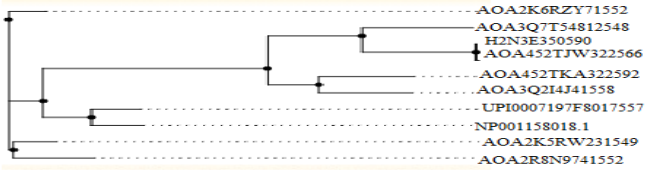
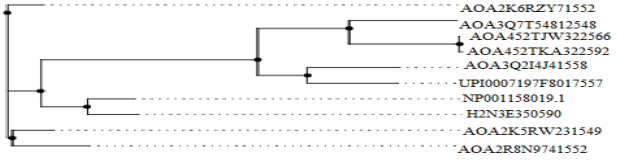
S no	Protein Sequences	phylogenetic tree of protein sequences
8	NP001158012.1	
9	NP001158013.1	
10	NP001158014.1	
11	NP001158016.1	
12	NP001158017.1	
13	NP001158018.1	
14	NP001158019.1	

Table 4.3 continued from previous page

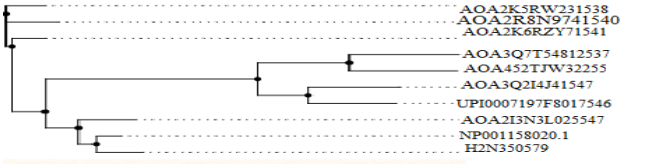
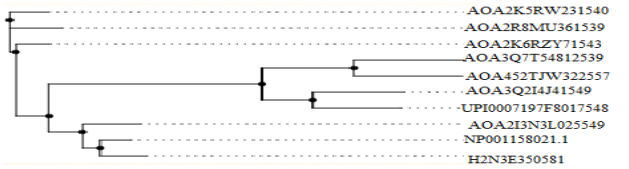
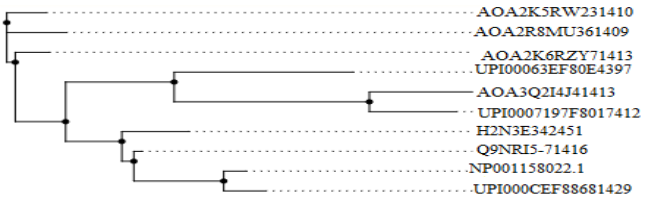
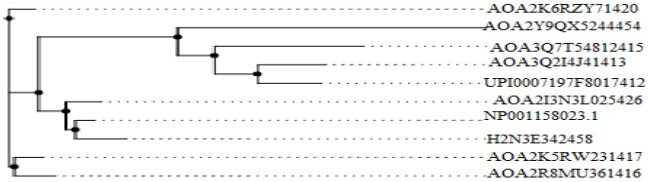
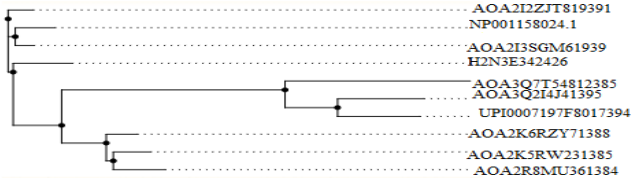
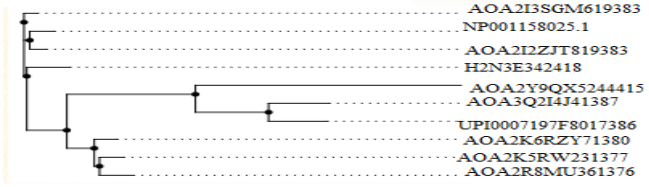
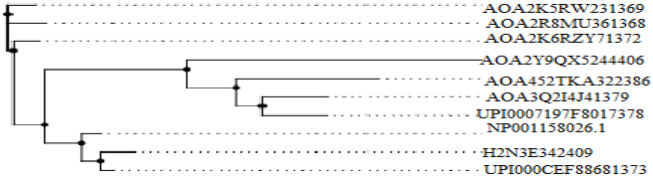
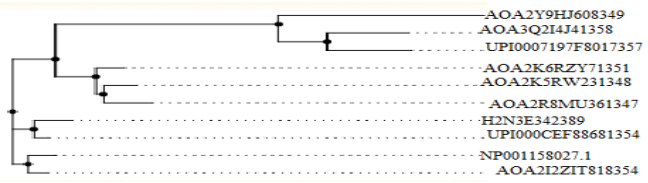
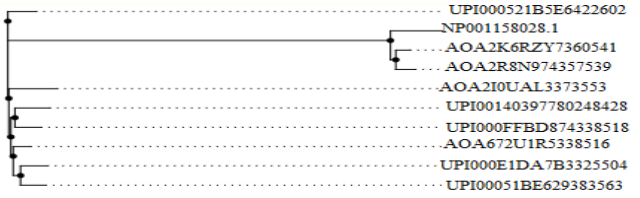
S no	Protein Sequences	phylogenetic tree of protein sequences
15	NP001158020.1	
16	NP001158021.1	
17	NP001158022.1	
18	NP001158023.1	
19	NP001158024.1	
20	NP001158025.1	
21	NP001158026.1	

Table 4.3 continued from previous page

S no	Protein Sequences	phylogenetic tree of protein sequences
22	NP001158027.1	
23	NP001158028.1	

Phylogenetic tree is used to represent evolutionary relationship between different organisms. The branching pattern of the tree predicts that how the species are evolved from the ancestor. They can be constructed by use of different bioinformatics tools, for example Consurf tool uses neighbor joining method along with maximum likelihood to construct phylogenetic tree. Neighbor joining is quick and fast method while maximum likelihood decrease the distance so as a whole, this tool construct best phylogenetic trees.

4.2.3 3D Protein Model Prediction

After few minutes results of SWISS MODEL EXPASY appeared which include models and templates detail. Then i clicked on template to view the detail. Huge detail in the form of table appears. Now, i scroll down to look onto values of similarity identity and then noted the highest value in the table which i separately created in MS word includes the sequence, name and identity value. Meanwhile data of all 23 sequences is noted down as shown in the given table 4.4.

Then the result is analyzed and some values which are above the value 75 are excluded from the result as for them abinitio modeling will be performed in next upcoming steps [134].

TABLE 4.4: shows highest value obtained by Swiss model expasy

#	Sequence	Name	Identity (highest value)	Conclusion
1	NP_001158021.1		21.62	Damaged/Mutated seq
2	NP_001158020.1		21.62	Damaged/Mutated seq
3	NP_001158019.1		21.92	Damaged/Mutated seq
4	NP_001158018.1		21.92	Damaged/Mutated seq
5	NP_001158017.1		25.00	Damaged/Mutated seq
6	NP_001158016.1		23.47	Damaged/Mutated seq
7	NP_001158014.1		23.47	Damaged/Mutated seq
8	NP_001158013.1		23.23	Damaged/Mutated seq
9	NP_001158012.1	6irr.1.A	72.29	Non mutated seq
10	NP_001158011.1		23.47	Damaged/Mutated seq
11	NP_001158010.1		23.47	Damaged/Mutated seq
12	NP_001158009.1	6irr.1.A	72.29	Non mutated seq
13	NP_001158028.1		24.14	Damaged/Mutated seq
14	NP_001158027.1		17.86	Damaged/Mutated seq
15	NP_001158026.1		24.44	Damaged/Mutated seq
16	NP_001158025.1		23.40	Damaged/Mutated seq
17	NP_001158024.1		17.86	Damaged/Mutated seq
18	NP_001158023.1		20.97	Damaged/Mutated seq
19	NP_001158022.1		24.44	Damaged/Mutated seq
20	NP_001012977.1		23.23	Damaged/Mutated seq
21	NP_001012976.1		32.00	Damaged/Mutated seq
22	NP_001012975.1	6irr.1.A	72.15	Non mutated seq
23	NP_061132.2	6irr.1.A	72.29	Non mutated seq

The obtained result of RaptorX contains the submitted sequence, references and two downloadable attachments which are actually the protein models. One predicted model is in the form of pdb file and other one is also the model of full

sequence. Full results can also be viewed by clicking on the provided link; this result also contain the information about sequence and the predicted model details like the number of domains, best template information, disordered positions, secondary structure detail, access to the solvent and amount of modeled residues [135].

4.2.4 Checking Stability of Protein Models

Result of Saves server comprises of different factors which determine the protein stability like VERIFY, ERRECT and WHATCHECK. They all show different values and parameters indicating and confirming the amount of error and disorder in the protein sequence. The obtained graph also shows the difference between the expected and observed amino acids in the sequence of interest. In the similar way result of all 19 sequences is recorded and saved in MS word for further analysis [136]. Figure 4.1-4.19 shows values obtained by Saves server, graphs are also present which shows the comparison of expected and obtained data by means of obtained peaks.

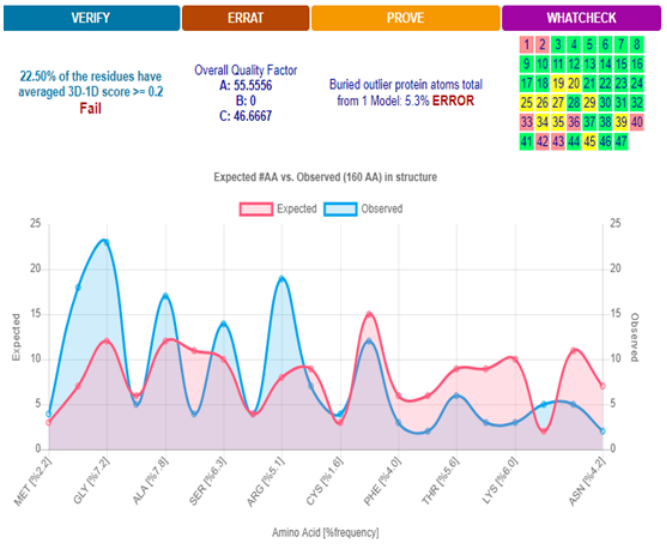


FIGURE 4.1: shows results for the sequence NP001012976.1

In the obtained result [Figure 4.1] for sequence NP001012976.1,VERIFY result shows 22.50% residues score below 0.2,ERRAT shows quality factor, PROVE also

shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

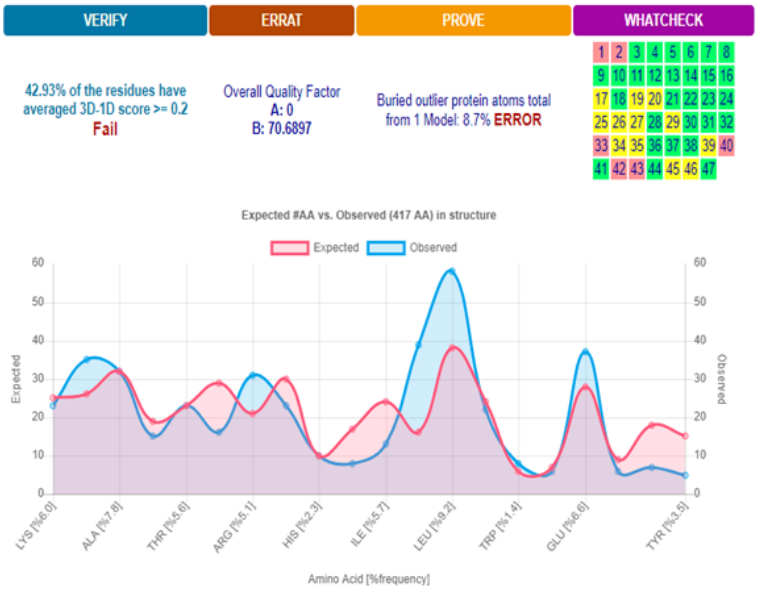


FIGURE 4.2: shows results for the sequence NP001012977.1

In the obtained result [Figure 4.2] for sequence NP001012977.1,VERIFY result shows 42.39% residues score is below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.3] for sequence NP001158010.1,VERIFY result shows 28.51% of residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.4] for sequence NP001158011.1,VERIFY result shows 27.62% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

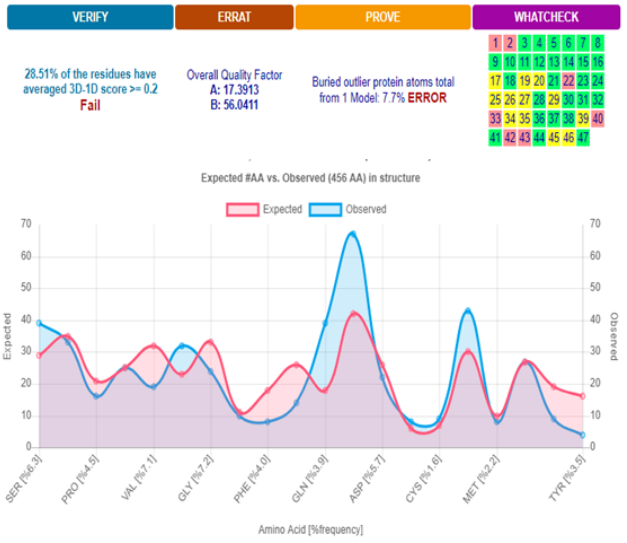


FIGURE 4.3: shows results for the sequence NP001158010.1

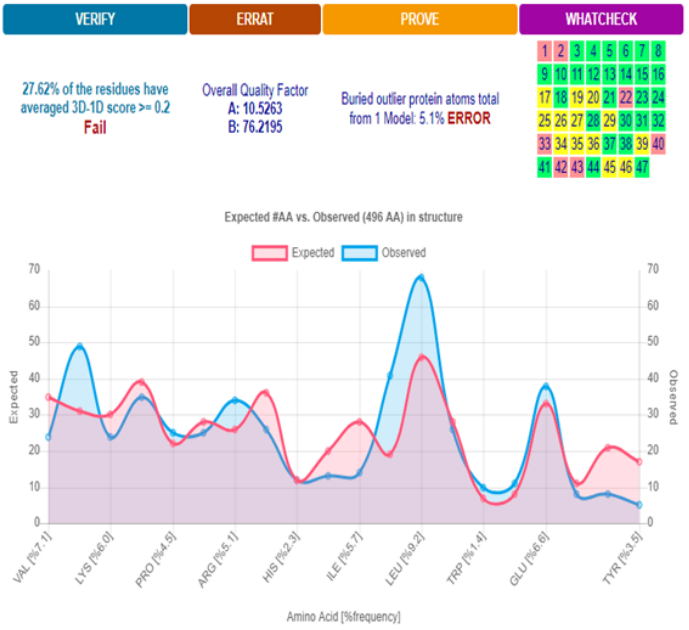


FIGURE 4.4: shows results for the sequence NP001158011.1

In the obtained result [Figure 4.5] for sequence NP001158012.1,VERIFY result shows 52.73% aminoacids residues score is below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.6] for sequence NP001012976.1,VERIFY result shows 25.46% residues score isbelow 0.2,ERRAT shows quality factor, PROVE

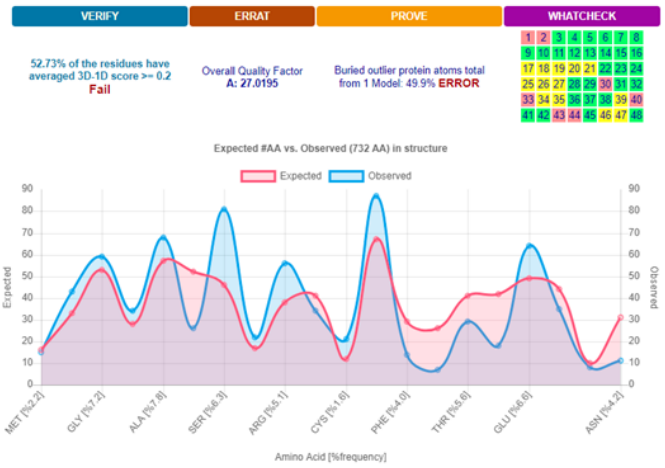


FIGURE 4.5: shows results for the sequence NP001158012.1

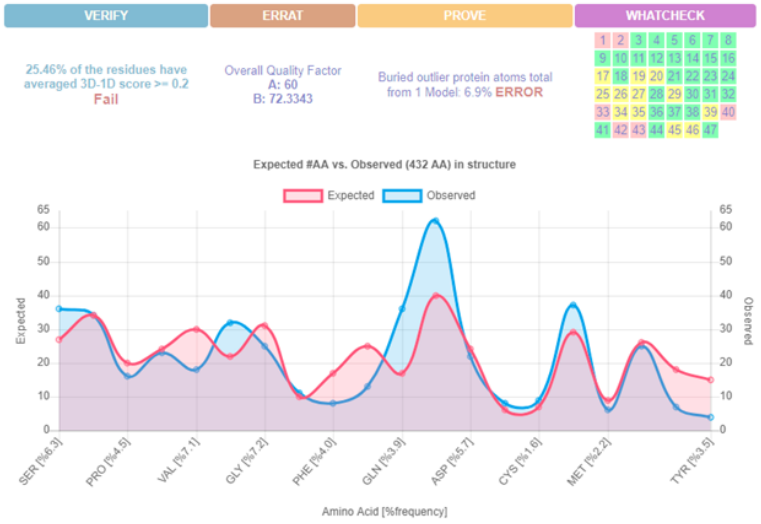


FIGURE 4.6: shows results for the sequence NP001158013.1

also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.7] for sequence NP001012976.1,VERIFY result shows 29.90% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.8] for sequence NP001012976.1,VERIFY result shows 33.17 residues score below 0.2,ERRAT shows quality factor, PROVE also

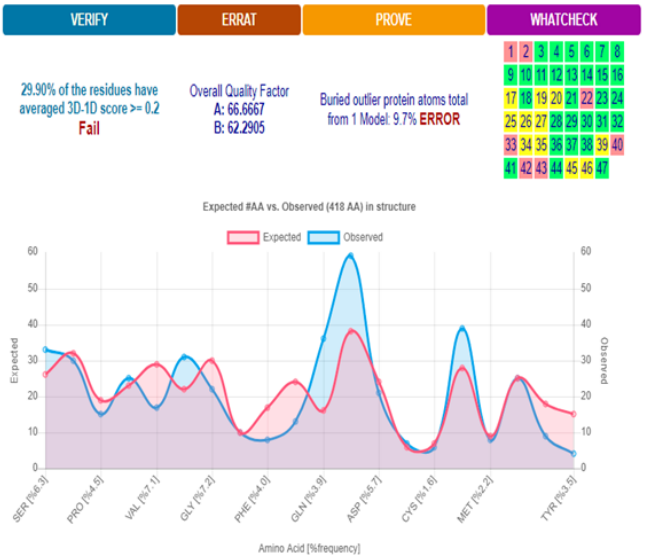


FIGURE 4.7: shows results for the sequence NP001158014.1

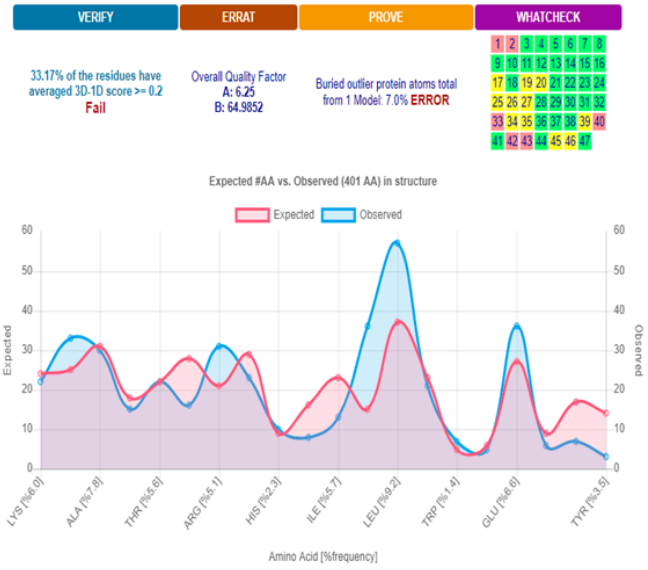


FIGURE 4.8: shows results for the sequence NP001158016.1

shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.9] for sequence NP001158017.1,VERIFY result shows 19.06% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

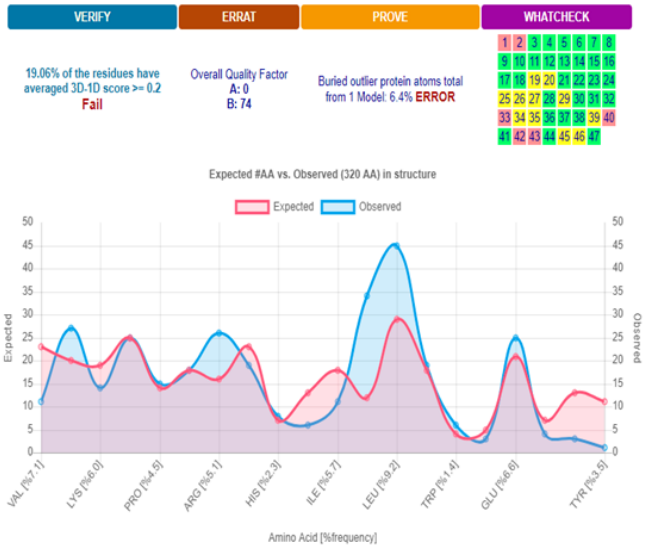


FIGURE 4.9: shows results for the sequence NP001158017.1

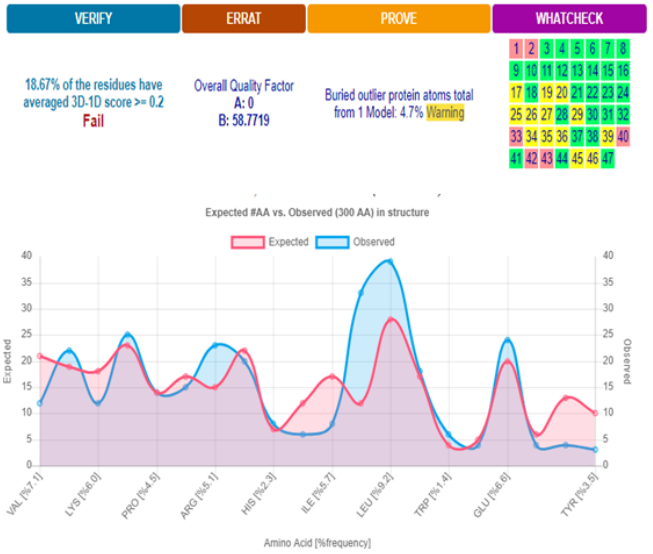


FIGURE 4.10: shows results for the sequence NP001158010.1

Figure 4.10 shows results [Figure 4.10] for the sequence NP001158018.1 In the obtained result for sequence NP001158018.1,VERIFY result shows 18.67% residues score below 0.2,ERRAT shows quality factor.

PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.11] for sequence NP001012976.1,VERIFY result shows 18.67% residues score below 0.2,ERRAT shows quality factor, PROVE also

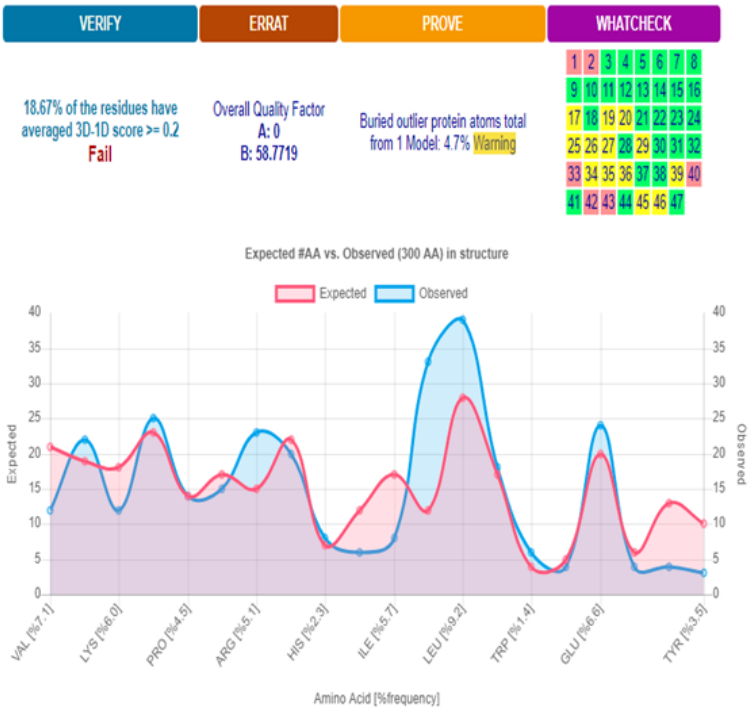


FIGURE 4.11: shows results for the sequence NP001158019.1

shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

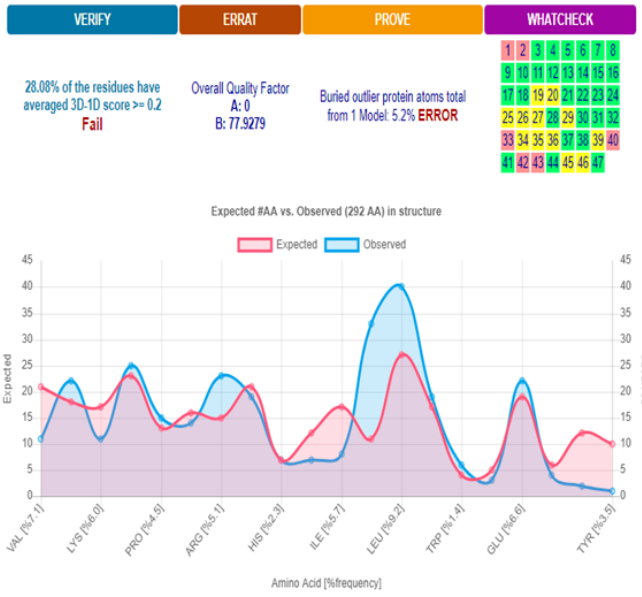


FIGURE 4.12: shows results for the sequence NP001158020.1

In the obtained result [Figure 4.12] for sequence NP001012976.1,VERIFY result shows 28.08% residues score below 0.2,ERRAT shows quality factor, PROVE also

shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

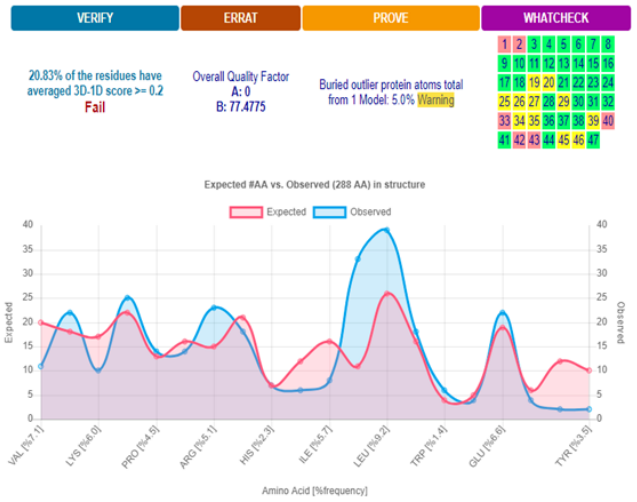


FIGURE 4.13: shows results for the sequence NP001158021.1

In the obtained result [Figure 4.13] for sequence NP001012976.1,VERIFY result shows 20.83% residues score below 0.2,ERRAT shows quality factor.

PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

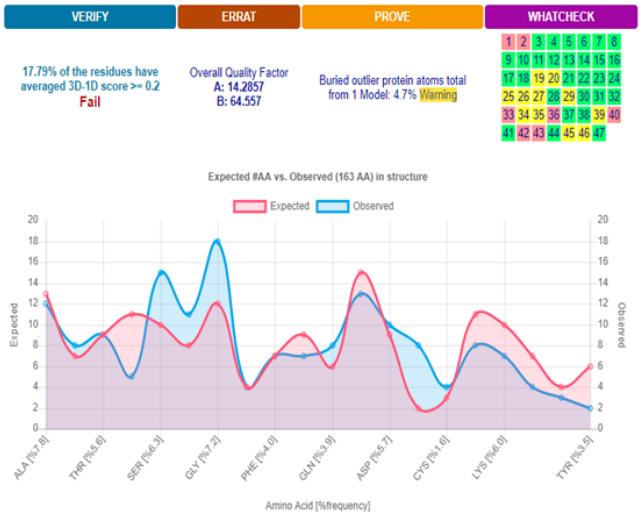


FIGURE 4.14: shows results for the sequence NP001158022.1

In the obtained result [Figure 4.14] for sequence NP001158022.1,VERIFY result shows17.79% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

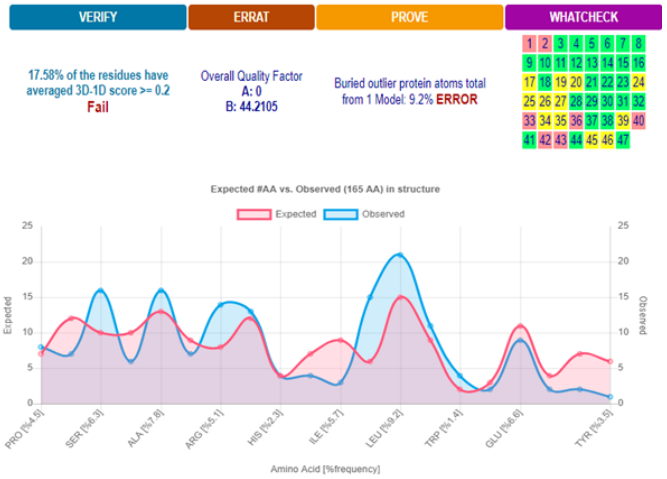


FIGURE 4.15: shows results for the sequence NP001158023.1

In the obtained result [Figure 4.15] for sequence NP001012976.1,VERIFY result shows 17.58% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.16] for sequence NP001012976.1,VERIFY result shows 10.74% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.17] for sequence NP001012976.1,VERIFY result shows score 32.23% residues below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

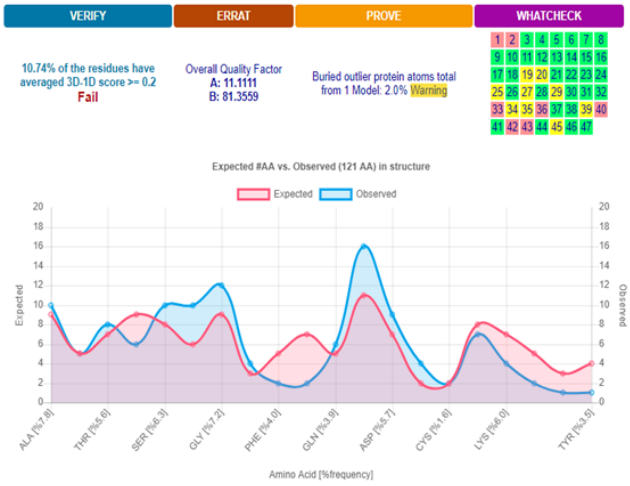


FIGURE 4.16: shows results for the sequence NP001158024.1

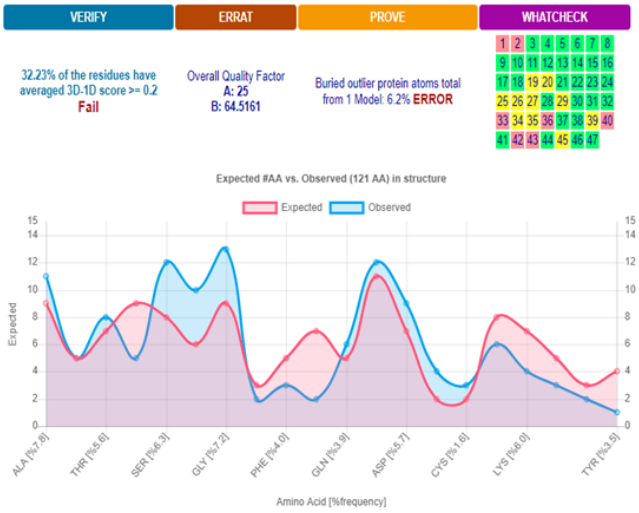


FIGURE 4.17: shows results for the sequence NP001158025.1

In the obtained result [Figure 4.18] for sequence NP001012976.1,VERIFY result shows score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

In the obtained result [Figure 4.19] for sequence NP001012976.1,VERIFY result shows 27.03% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference b/w the expect and observe values indicate that obtain sequence is damaged.

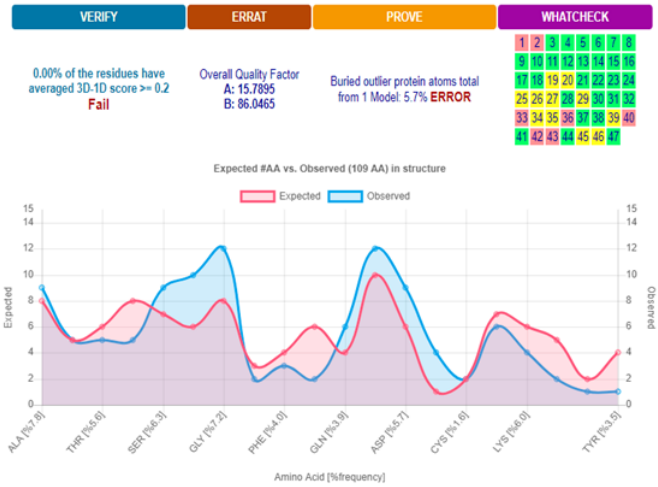


FIGURE 4.18: shows results for the sequence NP001158026.1

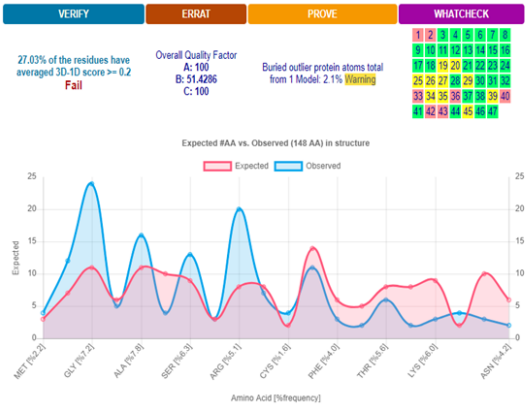


FIGURE 4.19: shows results for the sequence NP001158027.1

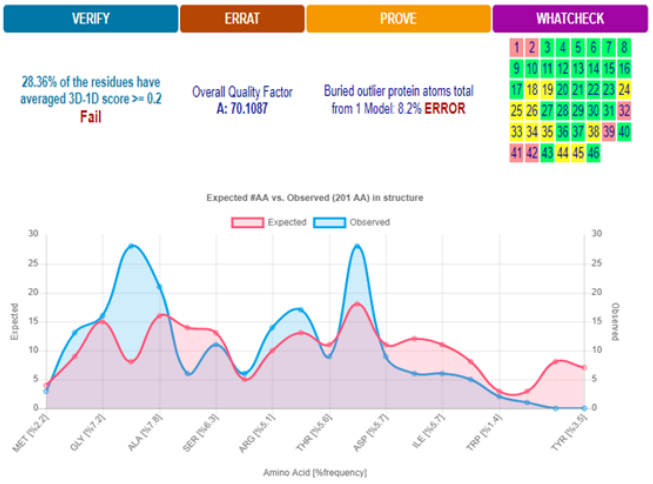


FIGURE 4.20: shows results for the sequence NP001158028.1

In the obtained result [Figure 4.20] for sequence NP001012976.1,VERIFY result shows 28.36% residues score below 0.2,ERRAT shows quality factor, PROVE also shows error and WHATCHECK also indicate error and the graph shows the difference between the expected and observed values which indicates that the obtained sequence is damaged.

4.2.5 Comparison of Protein Models

Result of TM-align server deeply analyze and compare both given sequences and then give many values including length of the sequence, RMSD value, Identical or aligned value and TM-score. All of these values for each of the sequence are noted and then compared in order to get the more refined result. Moreover in the results, comparison of wild and mutated protein model is also present in which red coloured protein is wild type and blue coloured protein is mutated one. These proteins can be viewed in 3D in RasMol etc. I saved these structures and related values in the MS-Word file for later analysis and result compilation. Table 4.5 showing comparison of different mutated sequences while mutated structures are presented in blue colour and wild sequence is in red colour [137]. Table4.5 showing TM-score, identical score, aligned length and RMSD value of sequences

TABLE 4.5: showing TM-score, identical score, aligned length and RMSD value of sequences

#	Sequence	Aligned length	RMSD	Identical/Aligned	TM-score
1	NP_001158021.1	126	3.97	0.07	0.33459
2	NP_001158020.1	141	5.66	0.043	0.31842
3	NP_001158019.1	152	5.24	0.066	0.34999
4	NP_001158018.1	152	5.24	0.066	0.34999
5	NP_001158017.1	171	5.39	0.047	0.36355
6	NP_001158016.1	165	6.46	0.079	0.27329
7	NP_001158014.1	162	6.70	0.074	0.25162
8	NP_001158013.1	190	6.73	0.105	0.27625

Table 4.5 continued from previous page

#	Sequence	Aligned length	RMSD	Identical/Aligned	TM-score
9	NP_001158011.1	152	4.72	0.099	0.24818
10	NP_001158010.1	186	5.83	0.070	0.29224
11	NP_001158028.1	109	4.48	0.073	0.37535
12	NP_001158027.1	84	5.22	0.131	0.31866
13	NP_001158026.1	74	4.30	0.122	0.3589
14	NP_001158025.1	86	4.90	0.081	0.36439
15	NP_001158024.1	80	4.07	0.100	0.44187
16	NP_001158023.1	95	5.39	0.095	0.33844
17	NP_001158022.1	85	5.24	0.094	0.29733
18	NP_001012977.1	200	7.12	0.085	0.28427
19	NP_001012976.1	71	4.44	0.028	0.29532

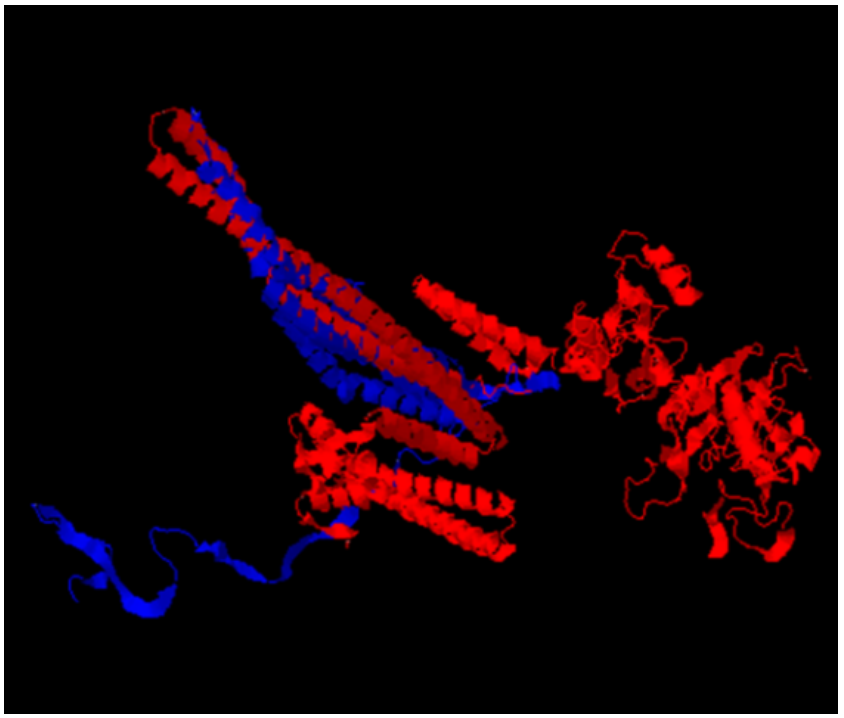


FIGURE 4.21: showing structure of NP001012976.1 shows aligned length of,71RMSD score of 4.44,aligned score of0.028,TM score of0.29532

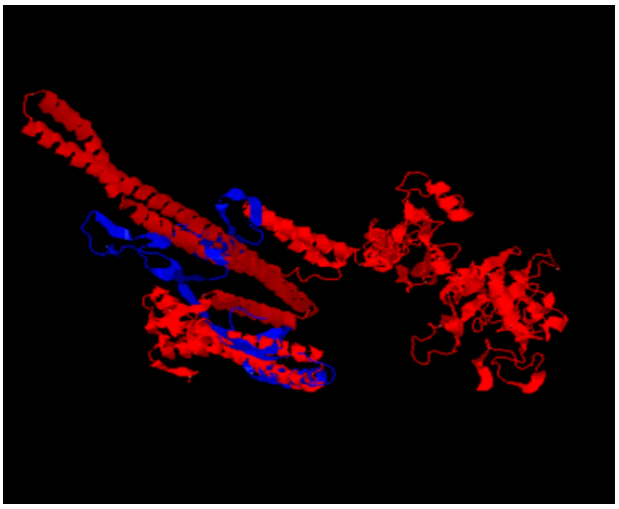


FIGURE 4.22: showing structure of NP001012977.1 shows aligned length of 200,RMSD score of 7.12,aligned score of 0.085,TM score of 0.29532

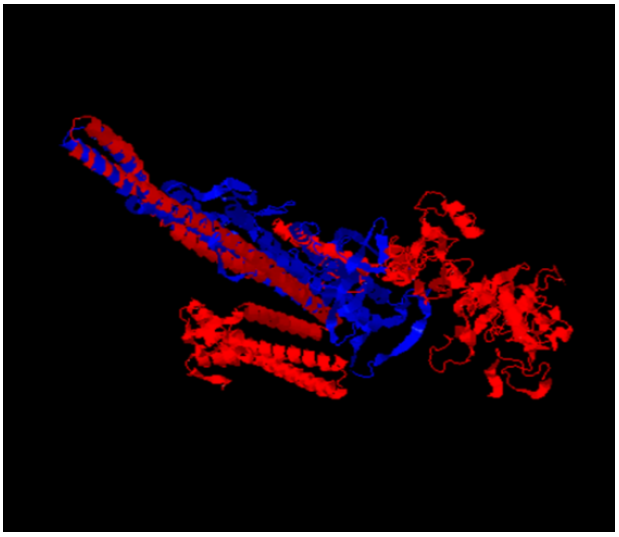


FIGURE 4.23: showing structure of NP001158010.1 shows sequenceof 186 aligned length,RMSD score of 5.83,aligned 0.070,TM score of 0.29224.

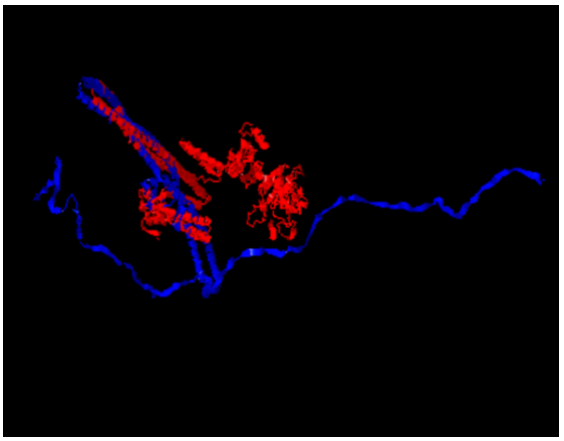


FIGURE 4.24: showing structure of NP001158011.1 shows sequence length of 152,RMSD score aligned of 4.72,aligned 0.099TM score of 0.24818

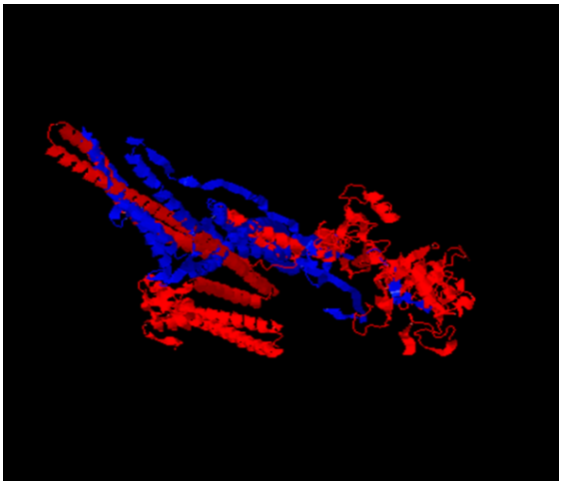


FIGURE 4.25: showing structure of NP001158013.1 shows aligned length of 190,RMSD score of 6.73,aligned score of0.105 ,TM score of 0.27625

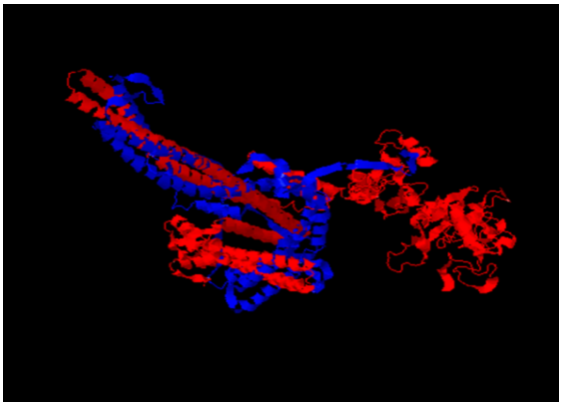


FIGURE 4.26: showing structure of NP001158014.1 shows aligned length of 162,RMSD score of 6.70,aligned score of 0.074,TM score of 0.25162

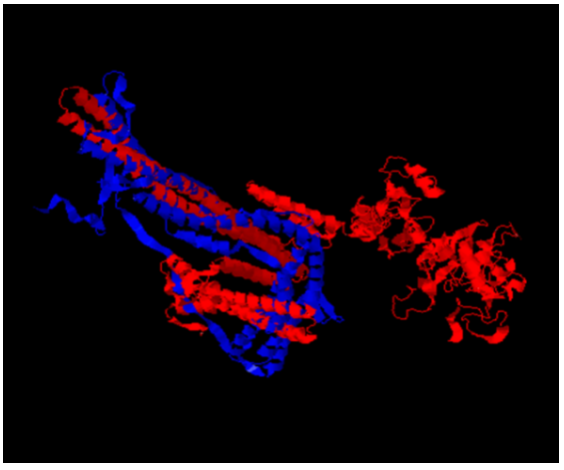


FIGURE 4.27: showing structure of NP001158016.1 shows aligned length of165,RMSD score of 6.46 ,aligned score of 0.079,TM score of0.27329

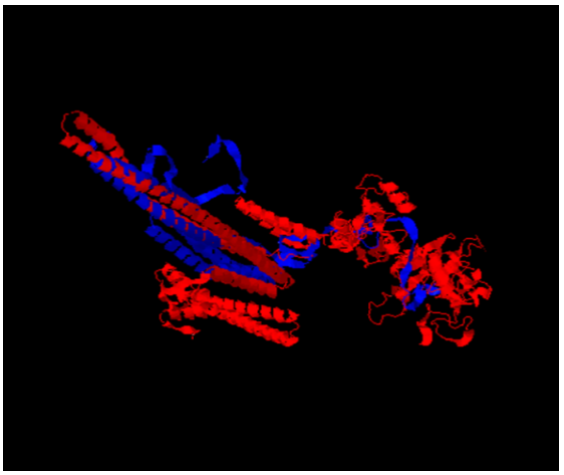


FIGURE 4.28: showing structure of NP-001158017.1 shows aligned length of171,RMSD score of 5.39, aligned score of 0.047,TM score of 0.36355

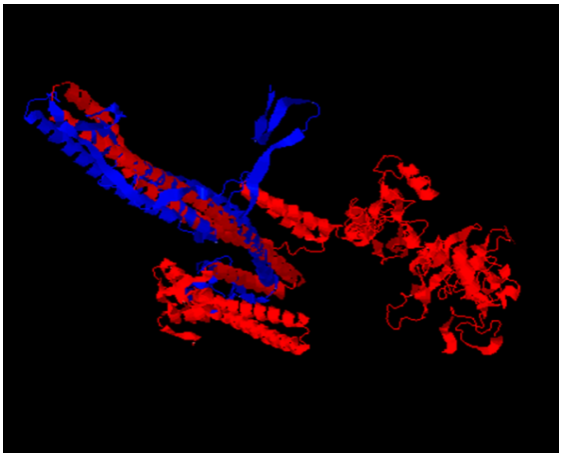


FIGURE 4.29: showing structure of NP-001158018.1 shows aligned length of 152,RMSD score of5.24,aligned score of 0.066,TM score of0.034999

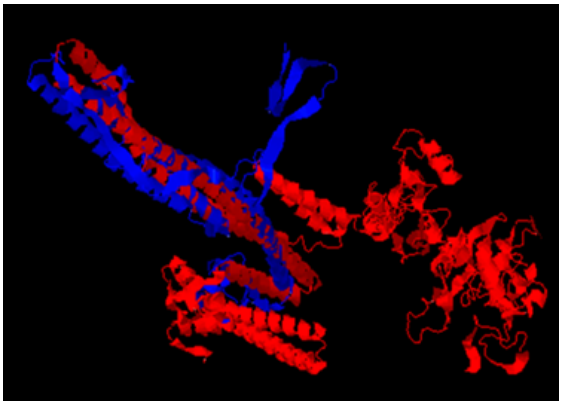


FIGURE 4.30: showing structure of NP-001158019.1 shows aligned length of152,RMSD score,5.24aligned score of 0.066,TM score of0.34999

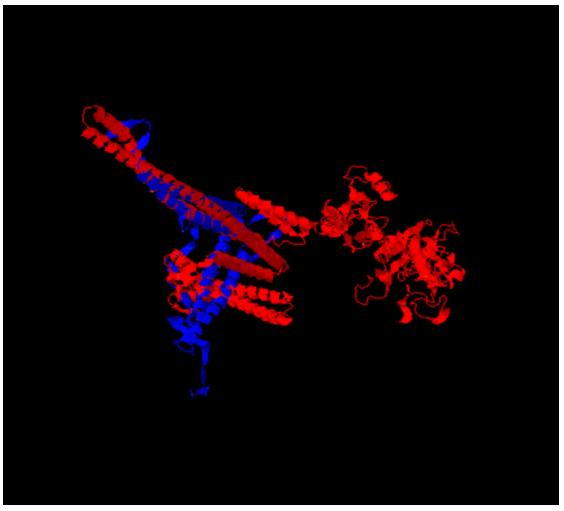


FIGURE 4.31: showing structure of NP-001158020.1 shows aligned length of141,RMSD score of 5.66,aligned score of 0.043,TM score of0.31842

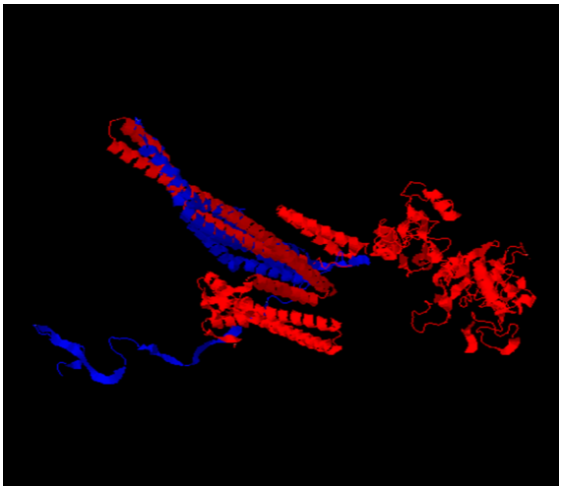


FIGURE 4.32: showing structure of NP-001158021.1 shows aligned length of 126,RMSD score of 3.97,aligned score of 0.07,TM score of0.33459

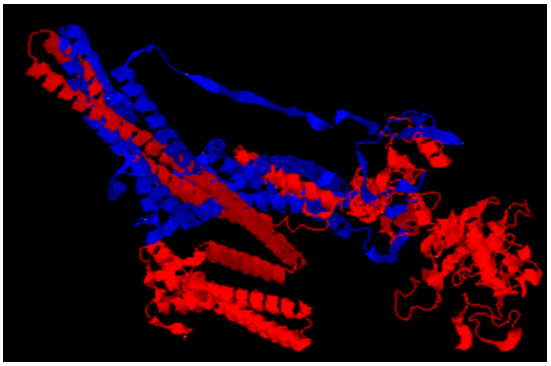


FIGURE 4.33: showing structure of NP-001158022.1 shows aligned length of 85,RMSD score of 5.24,aligned score of 0.094,TM score of 0.29733

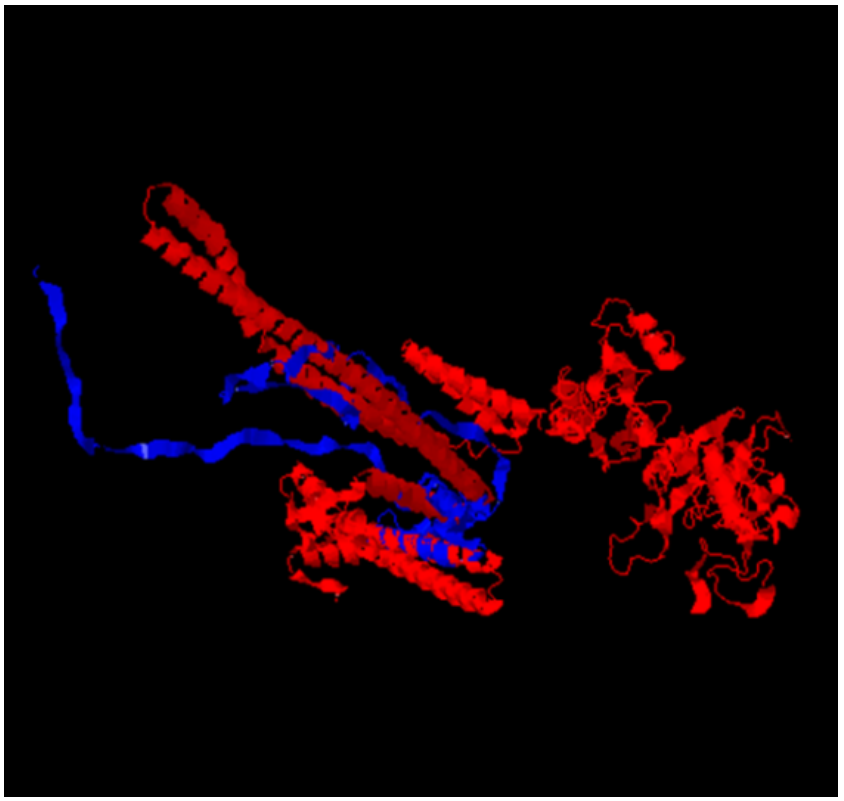


FIGURE 4.34: showing structure of NP-001158023.1 shows aligned length of 95,RMSD score of 5.39,aligned score of0.095,TM score of0.33844

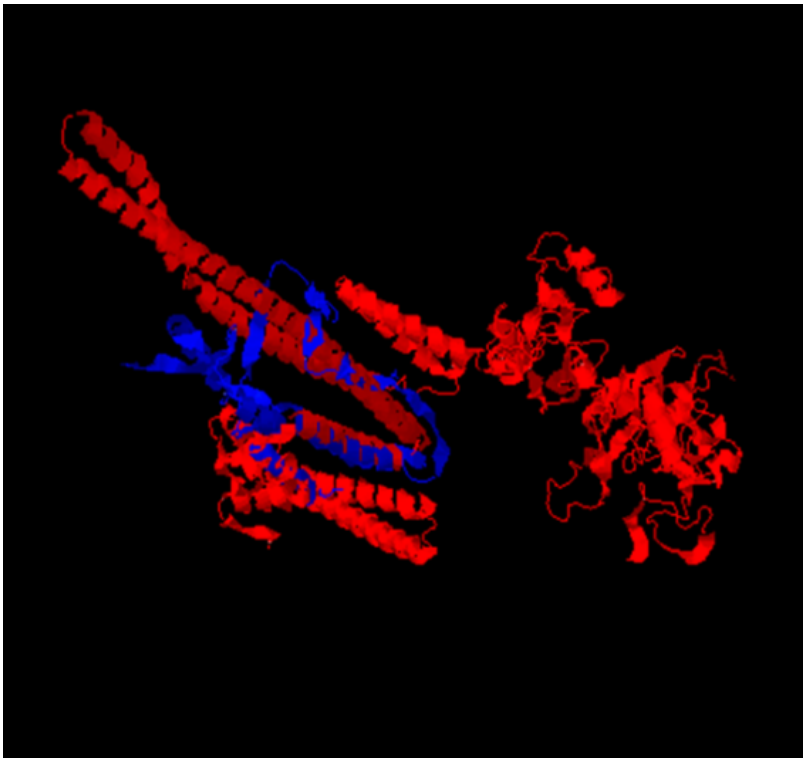


FIGURE 4.35: showing structure of NP001158024.1 shows aligned length of80,RMSD score of 4.07,aligned score of 0.100,TM score of0.44187

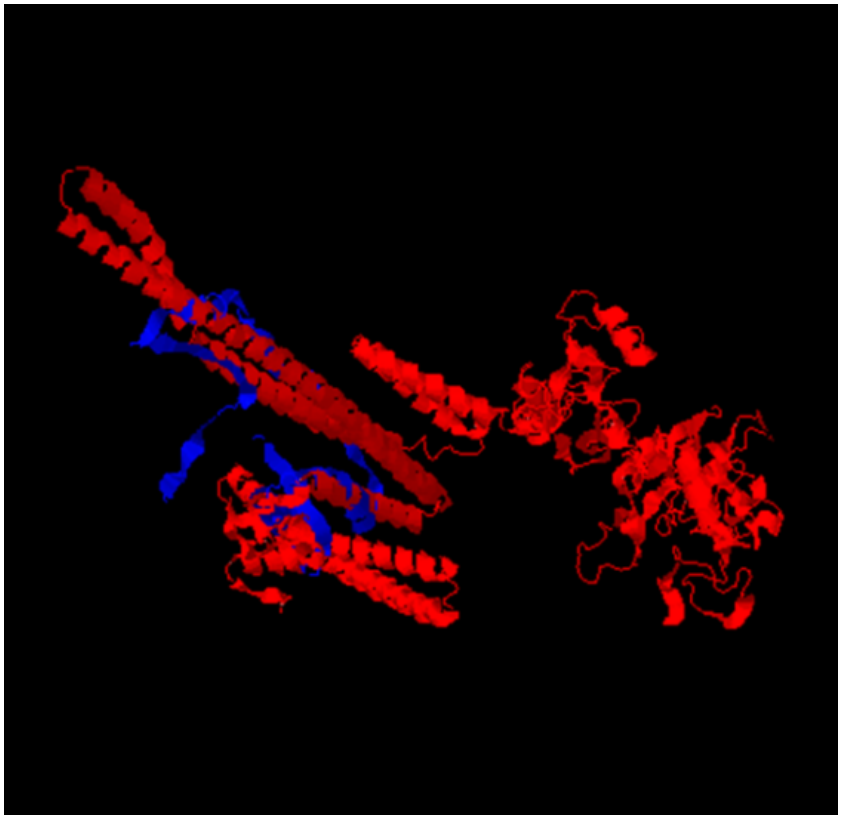


FIGURE 4.36: showing structure of NP001158025.1 shows aligned length of 86,RMSD score of 4.90,aligned score of 0.081,TM score of 0.364439

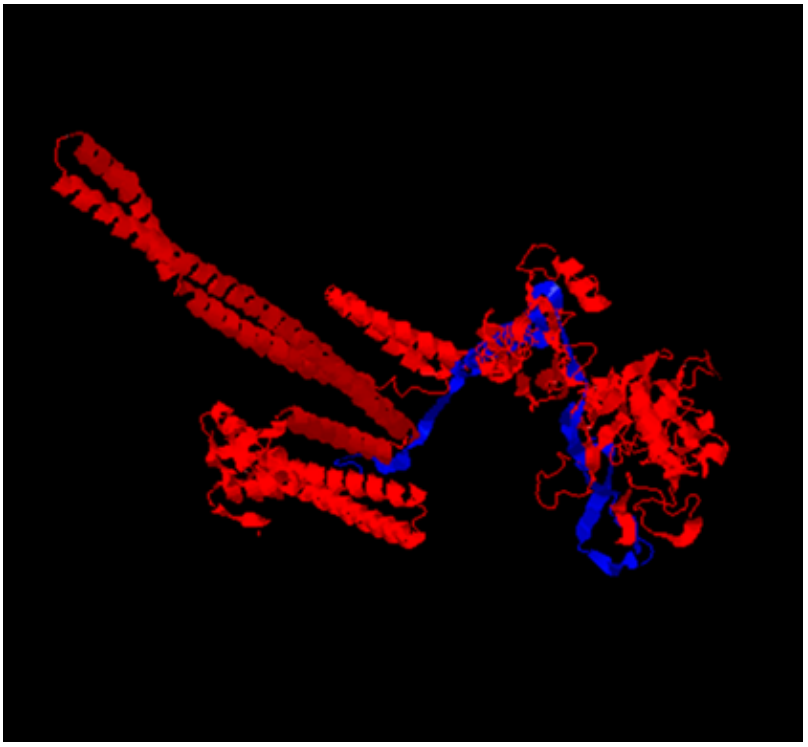


FIGURE 4.37: showing structure of NP001158026.1 shows aligned length of 74,RMSD score of 4.30,aligned score of 0.122,TM score of 0.3589

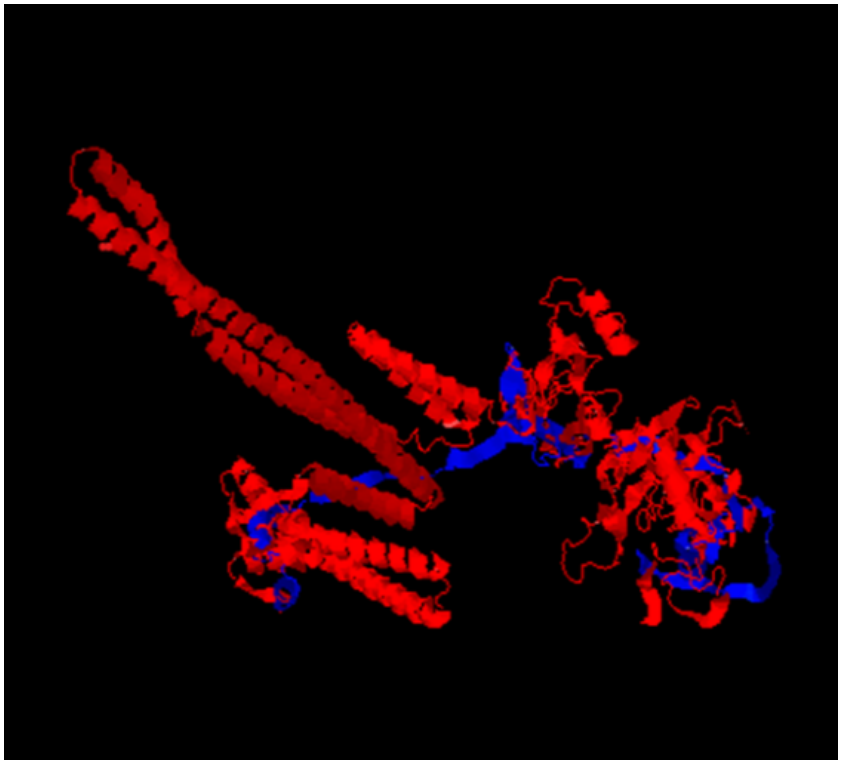


FIGURE 4.38: showing structure of NP001158027.1 shows aligned length of 84,RMSD score of5.22,aligned score of0.131,TM score of0.31866

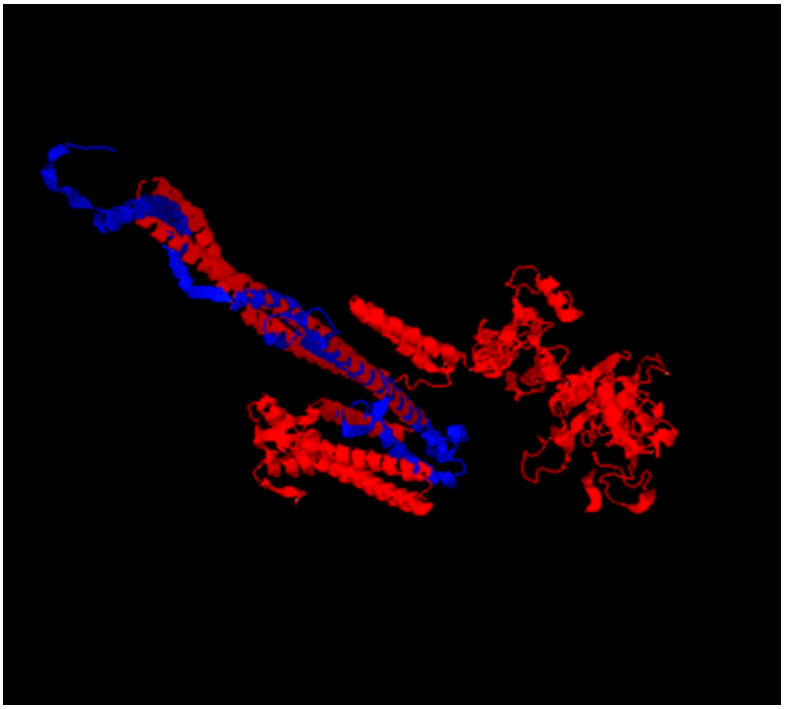


FIGURE 4.39: showing structure of NP-001158028.1 shows aligned length of 109 ,RMSD score, of 4.48aligned score of 0.073,TM score of 0.37535

In all figures(4.21-4.39) Blue colour represent mutated sequence while red colour represent wild sequence so by viewing all the figures, mutations and damages can be seen easily.

4.3 Identification of Significant Interactors of DISC1 Protein

4.3.1 Development of Protein Protein Interaction

After some time next page appears in Pickle 2.0 where identified interactors is present .And for finding protein protein interactions of DISC1 gene some parameters appeared on left side. Normalization level was set to protein, filtering method is cross checking and network setup is first neighbors, and then clicked on find interactions option.

After few moments list of interactions with DISC1 gene appears. Results are downloaded in Microsoft Excel sheet for further analysis. Downloaded results contains detail of PPI ID, detail of Interactors A&B, Sources, Cross checked confidence, Standard confidence, publications and number of publications [138].

When the result of Genemania comes, it shows interaction of DISC1 with other interacting genes. On the left side some information about the network is also given like percentage of physical interactions, co-expressions, predicted, co-localization, pathway, genetic interactions and shared protein domains. The resulting network can be viewed in many layouts.

I saved the interactions and networks of my gene. To save the results, I had clicked on the save option and from here I select the option of interactions data and networks data respectively and the result is downloaded in the form of notepad. Table 4.6 is showing such interactions. So, then I proceed further to perform the enrichment analysis of associated pathways in DISC1 gene [139]. Table 4.6 showing interacting genes, network group and network information

TABLE 4.6: showing interacting genes, network group and network information

Gene 1	Gene 2	Weight	Network group	Network
GPD2	DISC1	0.040339	Co-expression	Ramaswamy-Golub-2001
IMMT	ATF4	0.014595	Co-expression	Wang-Maris-2006
TRAIP	DISC1	0.009211	Co-expression	Wang-Maris-2006
SEMA7A	TCL1B	0.011372	Co-expression	Mallon-McKay-2013
GSK3B	TRIO	0.011653	Co-expression	Bild-Nevins-2006 B
GPD2	TRIO	0.013741	Co-expression	Rieger-Chu-2004
SEMA7A	MYH7	0.008499	Co-expression	Rieger-Chu-2004
SEMA7A	IL1RAPL2	0.002513	Co-expression	Dobbin-Giordano-2005
TCL1B	RASSF7	0.00868	Co-expression	Wu-Garvey-2007
CEP126	DISC1	0.016999	Co-expression	Chen-Brown-2002
SPTBN4	RASSF7	0.004684	Co-expression	Gysin-McMahon-2012
TRAIP	DISC1	0.015452	Co-expression	Sørlie-Børresen-Dale-2001
SRR	MYH7	0.015043	Co-expression	Cheok-Evans-2003
KCNQ5	TRIO	0.02208	Co-expression	Hegde-Luini-2015
SEMA7A	TRIO	0.01568	Co-expression	Hegde-Luini-2015
CEP63	GSK3B	0.006789	Co-expression	Singh-Celli-2011
CSF1R	ATF5	0.005804	Co-expression	Singh-Celli-2011
GPD2	CCDC141	0.003468	Co-expression	Finak-Park-2008
TRAIP	RASSF7	0.014763	Co-expression	Finak-Park-2008
SRR	IMMT	0.018719	Co-expression	Newman-Radeke-2012
CEP63	GSK3B	0.011232	Co-expression	Newman-Radeke-2012
GSK3B	IMMT	0.005881	Co-expression	Toedter-Baribaud-2011
TRAIP	IMMT	0.00615	Co-expression	Toedter-Baribaud-2011
NDEL1	IMMT	0.011606	Co-expression	Ross-Brown-2000
CSF1R	DISC1	0.008204	Co-expression	Ross-Brown-2000
SEMA7A	CCDC141	0.01256	Co-expression	Yu-Lee-2012
GPD2	CCDC141	0.004411	Co-expression	Scatolini-Chiorino-2010
CEP63	NDEL1	0.012055	Co-expression	Scatolini-Chiorino-2010

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
IMMT	DISC1	0.017628	Co-expression	Liu-Hsieh-2012
GPD2	GSK3B	0.015662	Co-expression	Liu-Hsieh-2012
TRAIP	SEMA7A	0.012496	Co-expression	Gómez-Abad-Piris-2011
TCL1B	DISC1	0.011324	Co-expression	Lee-Fine-2006
GSK3B	TRIO	0.011078	Co-expression	Yu-Tan-2008
SEMA7A	DISC1	0.004721	Co-expression	Yu-Tan-2008
TRAIP	ATF5	0.013551	Co-expression	Garber-Petersen-2001
GSK3B	TRIO	0.010483	Co-expression	Coelho-Hearing-2015
NDEL1	ATF4	0.008596	Co-expression	Holleman-Evans-2004
IL1RAPL2	RASSF7	0.01883	Co-expression	Salas-Chibon-2015
IL1RAPL2	MYH7	0.009675	Co-expression	Chowdary-Mazumder-2006
NDEL1	ATF4	0.008308	Co-expression	Lugthart-Evans-2005
ATF4	ATF5	0.021381	Co-expression	Postel-Vinay-Delattre-2012
ATF4	ATF5	0.006778	Co-expression	Den Boer-Pieters-2009 B
NDEL1	ATF4	0.005313	Co-expression	Den Boer-Pieters-2009 B
ATF4	ATF5	0.008647	Co-expression	Vallat-Bahram-2013
SEMA7A	RASSF7	0.008414	Co-expression	Kabbarah-Chin-2010
NDEL1	ATF4	0.004382	Co-expression	Jones-Libermann-2005
SEMA7A	IL1RAPL2	0.002853	Co-expression	Jones-Libermann-2005
TRAIP	TCL1B	0.016968	Co-expression	Lucas-Chute-2014
IL1RAPL2	DISC1	0.022986	Co-expression	Nakayama-Hasegawa-2007
MYH7	DISC1	0.014786	Co-expression	Minn-Massagué-2005
SEMA7A	IL1RAPL2	0.009935	Co-expression	Minn-Massagué-2005
CEP63	NDEL1	0.010347	Co-expression	Maertzdorf-Kaufmann-2011
IL1RAPL2	DISC1	0.007145	Co-expression	Shea-Musser-2010
TRAIP	TCL1B	0.009491	Co-expression	Shea-Musser-2010

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
SPTBN4	SEMA7A	0.005895	Co-expression	Bonome-Birrer-2008
GPD2	CCDC141	0.005495	Co-expression	Finak-Park-2006
TRAIP	ATF5	0.005497	Co-expression	Finak-Park-2006
IL1RAPL2	MYH7	0.005825	Co-expression	Sorich-Evans-2008
SPTBN4	SEMA7A	0.005661	Co-expression	Sorich-Evans-2008
TCL1B	MYH7	0.003403	Co-expression	Chng-Fonseca-2007
IL1RAPL2	DISC1	0.009388	Co-expression	Einecke-Halloran-2010
IL1RAPL2	MYH7	0.013207	Co-expression	Freedman-Nevins-2011
NDEL1	ATF4	0.005981	Co-expression	Tomasson-Ley-2008
IL1RAPL2	DISC1	0.017643	Co-expression	Tomasson-Ley-2008
ATF4	ATF5	0.006408	Co-expression	Den Boer-Pieters-2009 A
NDEL1	ATF4	0.008041	Co-expression	Den Boer-Pieters-2009 A
CEP63	IMMT	0.009049	Co-expression	Den Boer-Pieters-2009 A
SEMA7A	DISC1	0.004326	Co-expression	Hatzis-Symmans-2011 B
IL1RAPL2	DISC1	0.009878	Co-expression	Radtke-Downing-2009
SEMA7A	MYH7	0.0023	Co-expression	Shahmanesh-Tomlinson-2015
GSK3B	TRIO	0.015738	Co-expression	Stratford-Yeh-2010
GPD2	DISC1	0.012029	Co-expression	Hannenhalli-Cappola-2006
SPTBN4	SEMA7A	0.004922	Co-expression	Hannenhalli-Cappola-2006
SRR	NDEL1	0.013571	Co-expression	Kogo-Mori-2011
CEP63	GSK3B	0.007857	Co-expression	Kogo-Mori-2011
ATF4	ATF5	0.018818	Co-expression	Agnelli-Neri-2009
IL1RAPL2	DISC1	0.010302	Co-expression	Agnelli-Neri-2009
CEP63	IMMT	0.010044	Co-expression	Kang-Willman-2010 B
CEP63	IMMT	0.010044	Co-expression	Kang-Willman-2010 A
IL1RAPL2	DISC1	0.011855	Co-expression	Homminga-Meijerink-2011
SRR	NDEL1	0.019113	Co-expression	Hummel-Siebert-2006

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
IL1RAPL2	DISC1	0.007631	Co-expression	Hummel-Siebert-2006
SPTBN4	SEMA7A	0.012026	Co-expression	Hummel-Siebert-2006
KCNQ5	MYH7	0.005571	Co-expression	Ong-Howell-2009
CEP63	IMMT	0.007715	Co-expression	Ong-Howell-2009
TCL1B	KCNQ5	0.010659	Co-expression	Ong-Howell-2009
IL1RAPL2	RASSF7	0.007919	Co-expression	Coustan-Smith-Campana-2011
SPTBN4	SEMA7A	0.007072	Co-expression	Coustan-Smith-Campana-2011
KCNQ5	TRIO	0.026016	Co-expression	Balgobind-den Boer-2011
SEMA7A	DISC1	0.007023	Co-expression	Zhang-Mullighan-2012
TCL1B	MYH7	0.027368	Co-expression	D'Alfonso-Shin-2013
SEMA7A	DISC1	0.007947	Co-expression	Li-Han-2009
SPTBN4	RASSF7	0.012779	Co-expression	Li-Han-2009
CEP63	IMMT	0.010071	Co-expression	Wang-McClelland-2010 A
CSF1R	TRIO	0.008671	Co-expression	Yang-Steele-2012
TRAIP	TCL1B	0.015643	Co-expression	Yao-Jallal-2008
NDEL1	GSK3B	0.002643	Co-expression	Elashoff-Topol-2011 A
CEP63	NDEL1	0.003616	Co-expression	Elashoff-Topol-2011 A
ATF4	ATF5	0.013616	Co-expression	Symmans-Pusztai-2010
GSK3B	TRIO	0.012794	Co-expression	Zhang-Foekens-2009
IL1RAPL2	DISC1	0.003784	Co-expression	Zhang-Foekens-2009
IL1RAPL2	MYH7	0.009491	Co-expression	Zhang-Foekens-2009
TCL1B	KCNQ5	0.014155	Co-expression	Watanabe-Nagawa-2007
SRR	IMMT	0.007399	Co-expression	Zhou-O'Keefe-2010
TRAIP	IMMT	0.006432	Co-expression	Zhou-O'Keefe-2010
CSF1R	ATF4	0.011305	Co-expression	Whitfield-Botstein-2002
CEP63	IMMT	0.025584	Co-expression	Wang-McClelland-2010 B

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
KCNQ5	TRIO	0.026809	Co-expression	Wouters-Delwel-2009
IL1RAPL2	DISC1	0.008181	Co-expression	Wouters-Delwel-2009
IL1RAPL2	DISC1	0.008134	Co-expression	Agnelli-Neri-2007
SEMA7A	IL1RAPL2	0.007819	Co-expression	Liang-Stephan-2007
SEMA7A	IL1RAPL2	0.00979	Co-expression	Taylor-Kwo-2008
IMMT	ATF4	0.008791	Co-expression	Bild-Nevins-2006 A
TCL1B	DISC1	0.008828	Co-expression	Barretina-Singer-2010
GSK3B	TRIO	0.010703	Co-expression	Blader-Boothroyd-2001
MYH7	GSK3B	0.028545	Co-expression	Kannan-Zweidler-McKay-2011
IL1RAPL2	DISC1	0.006228	Co-expression	Kannan-Zweidler-McKay-2011
NDEL1	ATF4	0.021408	Co-expression	Reeve-Halloran-2013
IL1RAPL2	DISC1	0.01009	Co-expression	Reeve-Halloran-2013
TCL1B	IL1RAPL2	0.020009	Co-expression	Taylor-Belle-2007
TRAIP	IMMT	0.015806	Co-expression	Chen-Zhao-2008
CSF1R	TRIO	0.009344	Co-expression	Fry-Samson-2008
MYH7	GSK3B	0.026043	Co-expression	Mullighan-Downing-2009
CSF1R	ATF5	0.01382	Co-expression	Elashoff-Topol-2011 C
KCNQ5	CCDC141	0.014356	Co-expression	Harms-Bichakjian-2013
SRR	IMMT	0.009396	Co-expression	Harms-Bichakjian-2013
IL1RAPL2	TRIO	0.004607	Co-expression	Harms-Bichakjian-2013
SEMA7A	CCDC141	0.019434	Co-expression	Harms-Bichakjian-2013
TRAIP	ATF5	0.012998	Co-expression	Harms-Bichakjian-2013
CSF1R	ATF5	0.007932	Co-expression	Levy-Hessner-2012 B
IMMT	ATF4	0.012705	Co-expression	Savola-Vakkila-2011
SEMA7A	IL1RAPL2	0.004483	Co-expression	Raponi-Beer-2006
CSF1R	ATF4	0.009356	Co-expression	Cuadras-Greenberg-2002

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
IL1RAPL2	DISC1	0.008591	Co-expression	Levy-Hessner-2012 A
IL1RAPL2	MYH7	0.007023	Co-expression	Levy-Hessner-2012 A
SEMA7A	RASSF7	0.005963	Co-expression	Levy-Hessner-2012 A
CSF1R	ATF5	0.00848	Co-expression	Levy-Hessner-2012 A
SEMA7A	TRIO	0.011028	Co-expression	Yoshihara-Tanaka-2010
TRAIP	IMMT	0.005203	Co-expression	Chuang-Kipps-2012
SRR	IMMT	0.009314	Co-expression	Scott-Rimsza-2014
SRR	MYH7	0.021291	Co-expression	Scott-Rimsza-2014
TCL1B	IL1RAPL2	0.024436	Co-expression	Niu-Wang-2010
GSK3B	TRIO	0.013243	Co-expression	Sheffer-Domany-2009 A
SEMA7A	MYH7	0.008007	Co-expression	Sheffer-Domany-2009 A
IMMT	ATF4	0.006981	Co-expression	Hessel-Tilley-2014
GPD2	IMMT	0.011648	Co-expression	Hessel-Tilley-2014
TCL1B	TRIO	0.017539	Co-expression	Wynn-Wong-2011
CSF1R	TRIO	0.01342	Co-expression	Wynn-Wong-2011
CSF1R	DISC1	0.008514	Co-expression	Miyake-Noguchi-2012
IMMT	ATF4	0.002365	Co-expression	Raman-Crystal-2009
CEP63	IMMT	0.003493	Co-expression	Bhojwani-Carroll-2006
TRAIP	IMMT	0.005422	Co-expression	Bhojwani-Carroll-2006
SPTBN4	SEMA7A	0.005155	Co-expression	Bhojwani-Carroll-2006
TCL1B	TRIO	0.012121	Co-expression	Kim-Green-2011 B
SPTBN4	SEMA7A	0.006937	Co-expression	Kim-Green-2011 B
IL1RAPL2	DISC1	0.011877	Co-expression	Liu-Bratslavsky-2013
TRAIP	IMMT	0.007558	Co-expression	Payton-Ley-2009
TCL1B	MYH7	0.01388	Co-expression	Klein-Dugas-2009
SEMA7A	DISC1	0.004577	Co-expression	Bild-Nevins-2006 C
SPTBN4	SEMA7A	0.005321	Co-expression	Bild-Nevins-2006 C
NDEL1	ATF4	0.009863	Co-expression	Zangrando-Basso-2009
IMMT	ATF4	0.004411	Co-expression	Wong-Shanley-2009 A

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
TCL1B	RASSF7	0.021494	Co-expression	Wong-Shanley-2009 A
CSF1R	TRIO	0.011451	Co-expression	Wong-Shanley-2009 A
IL1RAPL2	SRR	0.007977	Co-expression	Driscoll-Munshi-2010
CSF1R	TRIO	0.015567	Co-expression	Wong-Shanley-2009 B
IMMT	ATF4	0.018878	Co-expression	Suárez-Fariñas-Krueger-2012
SRR	IMMT	0.011257	Co-expression	Lenz-Staudt-2008
IL1RAPL2	SRR	0.008139	Co-expression	Hanamura-Shaughnessy-2006
IL1RAPL2	DISC1	0.00619	Co-expression	Ioannidis-Flaño-2012
SEMA7A	IL1RAPL2	0.006412	Co-expression	Ioannidis-Flaño-2012
GPD2	GSK3B	0.00756	Co-expression	Desmedt-Sotiriou-2007
IL1RAPL2	DISC1	0.019756	Co-expression	Desmedt-Sotiriou-2007
IL1RAPL2	DISC1	0.016745	Co-localization	Johnson-Shoemaker-2003
SEMA7A	DISC1	0.025791	Co-localization	Johnson-Shoemaker-2003
CSF1R	DISC1	0.028151	Co-localization	Johnson-Shoemaker-2003
CSF1R	SEMA7A	0.025212	Co-localization	Johnson-Shoemaker-2003
ATF5	DISC1	0.711687	Pathway	Wu-Stein-2010
ATF4	DISC1	0.164889	Pathway	Wu-Stein-2010
GSK3B	ATF4	0.013889	Pathway	Wu-Stein-2010
TRIO	DISC1	0.544643	Physical Interactions	IREF-DIP
IMMT	DISC1	0.431157	Physical Interactions	IREF-DIP

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
NDEL1	DISC1	0.212005	Physical Interactions	IREF-DIP
NDEL1	IMMT	0.267992	Physical Interactions	IREF-DIP
ATF5	DISC1	0.102376	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
ATF4	DISC1	0.028206	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
IMMT	DISC1	0.055388	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
NDEL1	DISC1	0.040931	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
NDEL1	IMMT	0.101212	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
CCDC141	DISC1	0.33126	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
CEP126	DISC1	0.193353	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
RASSF7	DISC1	0.189766	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
MYH7	DISC1	0.14837	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
KCNQ5	DISC1	0.131839	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
SRR	DISC1	0.108499	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
CEP63	DISC1	0.102592	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
SPTBN4	DISC1	0.077629	Physical Interactions	BIOGRID-SMALL-SCALE-STUDIES
TCL1B	DISC1	0.185946	Physical Interactions	Huttlin-Gygi-2015
ATF5	DISC1	0.046231	Physical Interactions	IREF-INTACT
ATF4	DISC1	0.007135	Physical Interactions	IREF-INTACT
TRIO	DISC1	0.037171	Physical Interactions	IREF-INTACT
IMMT	DISC1	0.005407	Physical Interactions	IREF-INTACT
GSK3B	DISC1	0.00193	Physical Interactions	IREF-INTACT
NDEL1	DISC1	0.00504	Physical Interactions	IREF-INTACT
CCDC141	DISC1	0.055905	Physical Interactions	IREF-INTACT
CEP126	DISC1	0.003754	Physical Interactions	IREF-INTACT
CEP126	IMMT	0.006786	Physical Interactions	IREF-INTACT
RASSF7	DISC1	0.063889	Physical Interactions	IREF-INTACT
MYH7	DISC1	0.027677	Physical Interactions	IREF-INTACT
CEP63	DISC1	0.007948	Physical Interactions	IREF-INTACT

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
SPTBN4	DISC1	0.020204	Physical Interactions	IREF-INTACT
SPTBN4	GSK3B	0.013032	Physical Interactions	IREF-INTACT
SPTBN4	GSK3B	0.02159	Physical Interactions	Vinayagam-Wanker-2011
ATF5	DISC1	0.097224	Physical Interactions	IREF-HPRD
NDEL1	DISC1	0.098202	Physical Interactions	IREF-HPRD
CCDC141	DISC1	0.413135	Physical Interactions	IREF-HPRD
KCNQ5	DISC1	0.154744	Physical Interactions	IREF-HPRD
SPTBN4	DISC1	0.105465	Physical Interactions	IREF-HPRD
GSK3B	DISC1	0.622719	Predicted	I2D-IntAct-Mouse2Human
NDEL1	DISC1	1	Predicted	I2D-MGI-Mouse2Human
ATF4	ATF5	0.020905	Shared protein domains	INTERPRO
CCDC141	TRIO	0.018942	Shared protein domains	INTERPRO
TRAIP	DISC1	0.046213	Shared protein domains	INTERPRO
SPTBN4	TRIO	0.013242	Shared protein domains	INTERPRO

Table 4.6 continued from previous page

Gene 1	Gene 2	Weight	Network group	Network
SPTBN4	CCDC141	0.016388	Shared protein domains	INTERPRO
ATF4	ATF5	0.025182	Shared protein domains	PFAM
CCDC141	TRIO	0.018971	Shared protein domains	PFAM
SPTBN4	TRIO	0.008561	Shared protein domains	PFAM
SPTBN4	CCDC141	0.03105	Shared protein domains	PFAM

The given results of Genemania shows interactions of DISC1 with other interacting genes.Most interacting genes include TRIO,ATF5,KCNQ5, SPTBN4, CCDC141, GSK3B and NDEL1.Weight of interacting genes network ranges from 0.00193(GSK3B-DISC1 network)to 1(NDEL1-DISC1 network).Network groups include Co-expression, co-localization, pathway, Physical Interactions, predicted and shared protein domains.The results of Genemania also include different networks.

4.3.2 Functional Module Analysis within the Network

The result of Gephi is basically a network of genes which are connected together by nodes and edges and can be visualized better here. The obtained network in Gephi is saved along with the network interactions values and percentage of nodes modularity in the MS-Word file. The figure 4.40 given below shows network interactions of DISC1 with other genes. In Figure 4.40 the network overview it is observed that average degree is 0.952, average weighted degree is also 0.952, graph density is 0.048 and connected components is 1.Edges are 95.24% and nodes are 4.76% in the obtained network as shown in the figure above [141].

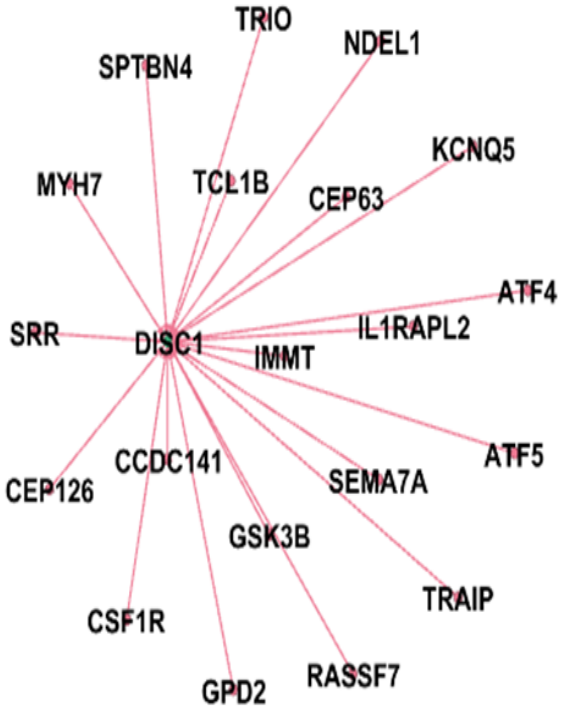


FIGURE 4.40: showing network interaction of DISC1 with other genes as predicted by Gephi software

4.3.3 GO Pathway and Enrichment Analysis

On the result page of EnrichNet tool, five different options are present. From which I select the option 1 which was gene similarity ranking, then results appear in the form of table showing different values.

Columns with annotation (pathway/process), significance of network distance distribution (XD-score), significance of overlap (Fisher test, q-value),dataset size (overlap) were saved in MS-Word file [142].

Then I move towards option 2 of result which was a graph plotted between XD score and significance of overlap (fisher test, q-value). So, I saved the regression plot and relevant threshold and score values.

This was done for saving results of KEGG,GO gene ontology (Molecular function, GO gene ontology (cellular component), GO ontology (biological process) respectively [143]. Table 4.7 to 4.10 shows the obtained results.

TABLE 4.7: showing results of KEGG

Annotation (pathway/ process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Glycine,serine and threonine metabolism	0.268	1.00	SRR
Prostate cancer	0.184	0.75	GSK3B ATF4
Endometrial cancer	0.151	1.00	GSK3B
Basal cell carcinoma	0.141	1.00	GSK3B
Hedgehog signaling pathway	0.138	1.00	GSK3B
Neurotrophin signaling pathway	0.123	0.75	GSK3B ATF4
Colorectal cancer	0.123	1.00	GSK3B
Axon guidance	0.118	0.75	GSK3B SEMA7A
Long-term potentiation	0.110	1.00	ATF4
Viral myocarditis	0.106	1.00	MYH7
B cell receptor signaling pathway	0.099	1.00	GSK3B

Table 4.7 continued from previous page

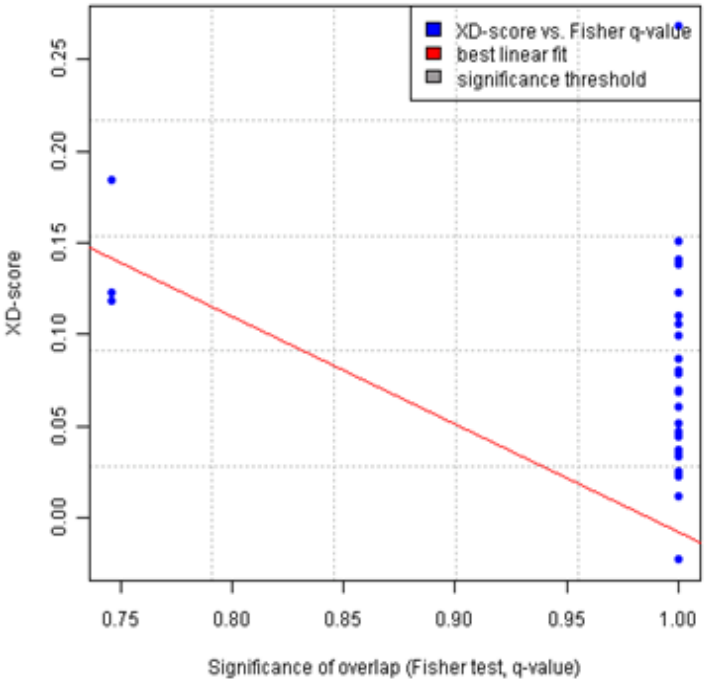
Annotation (pathway/ process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Cardiac muscle contraction	0.099	1.00	MYH7
Glycerophospholipid metabolism	0.099	1.00	GPD2
Hypertrophic cardiomyopathy (HCM)	0.087	1.00	MYH7
ErbB signaling pathway	0.081	1.00	GSK3B
Hematopoietic cell lineage	0.081	1.00	CSF1R
Dilated cardiomyopathy	0.079	1.00	MYH7
GnRH signaling pathway	0.069	1.00	ATF4
Melanogenesis	0.068	1.00	GSK3B
T cell receptor signaling pathway	0.061	1.00	GSK3B
Cell cycle	0.051	1.00	GSK3B
Tight junction	0.047	1.00	MYH7
Insulin signaling pathway	0.045	1.00	GSK3B
Wnt signaling pathway	0.037	1.00	GSK3B

Table 4.7 continued from previous page

Annotation (pathway/ process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Protein processing in endoplasmic reticulum	0.035	1.00	ATF4
Alzheimer's disease	0.034	1.00	GSK3B
Pathways in cancer	0.033	1.00	CSF1R GSK3B
Chemokine signaling pathway	0.026	1.00	GSK3B
Endocytosis	0.024	1.00	CSF1R
Focal adhesion	0.023	1.00	GSK3B
Cytokine -cytokine receptor interaction	0.012	1.00	CSF1R
MAPK signaling pathway	0.012	1.00	ATF4

Table 4.7 contain results of KEGG obtained by EnrichNet(Network based Enrichment Analysis)and here I get annotation pathways and processes, along with values of Significance of network distance distribution (XD-Score) , Significance of overlap (Fisher-test, q-value) Dataset size known as overlap. XD-score refers to similarity score of the network and finds network connectivity of pathway and gene set. Fisher test is performed in order to measure the significance of gene set

over of the pathway and the gene set. For example Glycine, serine and threonine metabolism has the highest XD-score of 0.268 and significance overlap value is 1 and overlap is SRR gene.



Absolute Pearson correlation between XD-scores and Fisher q-values:	**0.41**
XD-score significance threshold: (reqression fit equivalent to Fisher q-value of 0.05 + upper bound of 95% confidence interval for linear fitting)	**0.73**

FIGURE 4.41: Graph drawn between XD-score and Significance of overlap(Fisher test,q-value)

Graph in figure 4.41 shows the XD-score relation to the significance score; for the classical overlap-based Fisher test. The plot between the Xd-scores and Fisher q-values pearson correlation enables the user to determine the goodness of the linear fit between the two scoring lists, which can be used to choose a significance threshold for the XD-score.

TABLE 4.8: showing results of Gene Ontology(Molecular Function)

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Tau-Protein Kinase Activity	0.80068	1	GSK3B
Glycine Binding	0.67481	1	SRR
Ionotropic Glutamate Receptor Binding	0.62536	1	GSK3B
Ankyrin Binding	0.58250	1	SPTBN4
Alpha-Tubulin Binding	0.54500	1	NDEL1
Cytokine Binding	0.51191	1	CSF1R
Microfilament Motor Activity	0.51191	1	MYH7
Inward Rectifier Potassium Channel Activity	0.45618	1	KCNQ5
NF-Kappab Binding	0.39159	1	GSK3B
Beta-Tubulin Binding	0.35750	1	NDEL1
RNA Polymerase II Transcription Factor Binding	0.34250	1	GSK3B
Integrin Binding	0.21934	1	GSK3B SEMA7A

Table 4.8 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Receptor Signaling Protein Activity	0.20201	1	TRAIP
Structural Constituent Of Muscle	0.19180	1	MYH7
Transmembrane Receptor Protein Tyrosine Kinase Activity	0.18705	1	CSF1R
Voltage-Gated Potassium Channel Activity	0.16617	1	KCNQ5
P53 Binding	0.15231	1	GSK3B
Protein Dimerization Activity	0.14179	1	ATF4 ATF5
Pyridoxal Phosphate Binding	0.14039	1	SRR
Motor Activity	0.13004	1	MYH7
Beta-Catenin Binding	0.13004	1	GSK3B

Table 4.8 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Rho Guanyl-Nucleotide Exchange Factor Activity	0.11683	1	TRIO
Structural Constituent Of Cytoskeleton	0.09226	1	SPTBN4
Guanyl-Nucleotide Exchange Factor Activity	0.08477	1	TRIO
Phosphotransferase Activity, Alcohol Group As Acceptor	0.08477	1	TRIO
PDZ Domain Binding	0.08250	1	SRR
Microtubule Binding	0.08140	1	NDEL1
Actin Binding	0.04331	1	SPTBN4 MYH7
Protein Tyrosine Kinase Activity	0.04263	1	TRIO CSF1R GSK3B
Ubiquitin Protein Ligase Binding	0.04250	1	GSK3B
Ion Channel Activity	0.04210	1	KCNQ5

Table 4.8 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Phospholipid Binding	0.04001	1	SPTBN4 TRIO
Atpase Activity	0.03982	1	MYH7
Calmodulin Binding	0.03910	1	MYH7
Transmembrane Signaling Receptor Activity	0.03672	1	IL1RAPL2

Gene Ontology (GO) is the largest source of information on the molecular functions of genes. Table 4.8 contain results of Gene Ontology (Molecular Function)obtained by EnrichNet(Network based Enrichment Analysis)and here I get annotation pathways and processes, along with values of Significance of network distance distribution (XD-Score) , Significance of overlap (Fisher-test, q-value) and Dataset size known as overlap. Here Tau-Protein Kinase Activity has the highest XD-score of 0.80068 and significance overlap value is 1 and overlap is GSK3B gene.

TABLE 4.9: showing results of Gene Ontology (Cellular Components)

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Myosin Filament	0.6273	0.834	MYH7
Contractile Fiber	0.6273	0.834	MYH7
Muscle Myosin Complex	0.6273	0.834	MYH7

Table 4.9 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Kinesin Complex	0.4844	0.084	NDEL1 DISC1
Receptor Complex	0.3178	0.834	CSF1R
Sarcomere	0.3059	0.834	MYH7
Cell Leading Edge	0.2948	0.834	NDEL1
Cell Body	0.2747	0.084	NDEL1 DISC1
Microtubule Associated Complex	0.2657	0.834	NDEL1
Dendritic Shaft	0.2572	0.834	GSK3B
Myofibril	0.2491	0.834	MYH7
Myosin Complex	0.1801	0.981	MYH7
Stress Fiber	0.1759	0.981	MYH7
Kinetochore	0.1344	1.000	NDEL1
Voltage-Gated Potassium Channel Complex	0.1251	1.000	KCNQ5
Condensed Chromosome Kinetochore	0.1229	1.000	NDEL1
Z Disc	0.1130	1.000	MYH7
PML Body	0.1060	1.000	SPTBN4
Nuclear Matrix	0.1028	1.000	SPTBN4

Table 4.9 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Spindle Pole	0.1013	1.000	CEP63
Dendritic Spine	0.0998	1.000	GSK3B
Growth Cone	0.0955	1.000	GSK3B
Apical Part Of Cell	0.0929	1.000	SRR
Microtubule Organizing Center	0.0916	1.000	ATF4
Axon	0.0792	0.834	SPTBN4 NDEL1
Anchored To \ Membrane	0.0710	1.000	SEMA7A
Centrosome	0.0688	0.084	NDEL1 CEP63 GSK3B DISC1
Spindle	0.0670	1.000	NDEL1
Nuclear Envelope	0.0641	1.000	NDEL1
Transcription Factor Complex	0.0610	0.834	ATF4 ATF5
Focal Adhesion	0.0559	1.000	MYH7
Ribonucleoprotein Complex	0.0547	1.000	GSK3B
Neuronal Cell Body	0.0539	0.834	SRR GSK3B
Membrane Raft	0.0444	1.000	GSK3B

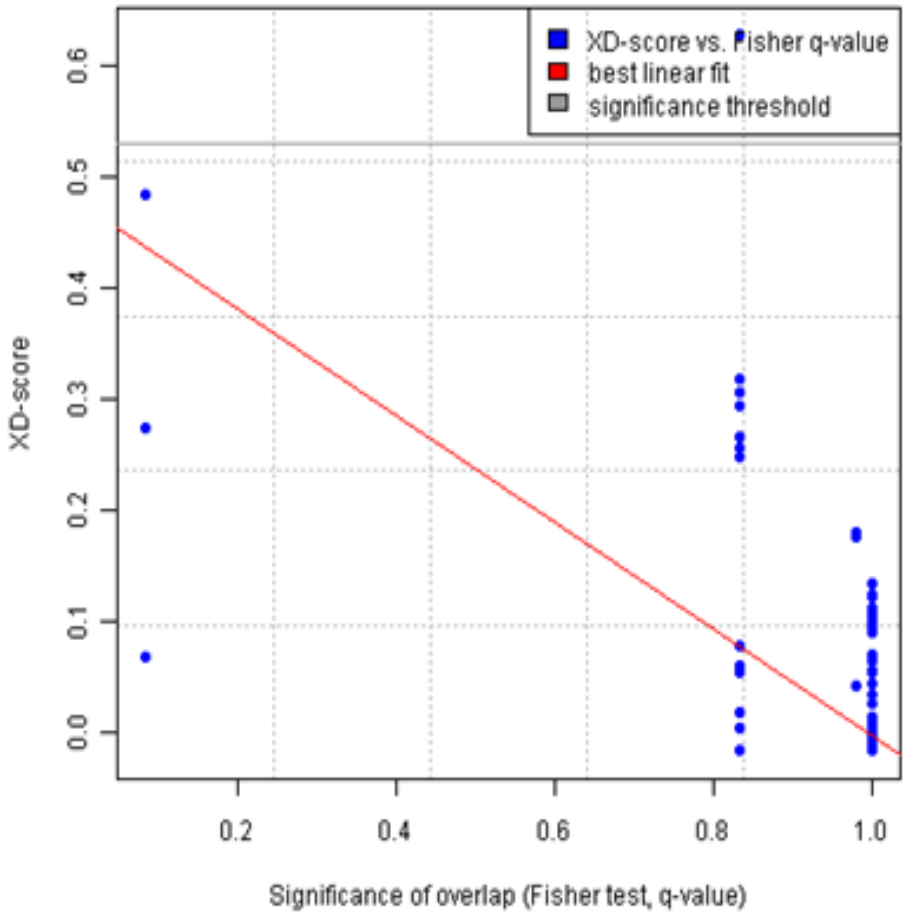
Table 4.9 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Mitochondrial Inner Membrane	0.0429	0.981	GPD2 IMMT
External Side Of Plasma Membrane	0.0353	1.000	SEMA7A
Soluble Fraction	0.0274	1.000	SRR GSK3B
Microtubule	0.0188	0.834	NDEL1 DISC1
Protein Complex	0.0149	1.000	GSK3B
Cell Surface	0.0107	1.000	CSF1R TRIO
Membrane	0.0089	1.000	CSF1R GSK3B SEMA7A KCNQ5
Perinuclear Region Of Cytoplasm	0.0046	1.000	TRAIP
Cytosol	0.0043	1.000	SPTBN4 NDEL1 TRIO CEP63 GSK3B

Table 4.9 continued from previous page

Annotation (Pathway/ Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Plasma Membrane	0.0039	0.834	SRR ATF4 TRAIP CSF1R GSK3B SEMA7A DISC1 KCNQ5
Cytoplasm	0.0012	1.000	SPTBN4 SRR NDEL1 ATF4 TRAIP ATF5 MYH7 GSK3B

Graph in figure 4.42 drawn between XD-score and significance of overlap(Fisher test,q-value)



Absolute Pearson correlation between XD-scores and Fisher q-values:	**0.54**
XD-score significance threshold: (reqression fit equivalent to Fisher q-value of 0.05 + upper bound of 95% confidence interval for linear fitting)	**0.53**

FIGURE 4.42: shows results for the sequence NP001158010.1

Table 4.10: showing results of Gene Ontology (Biological Process)

Annotation (Pathway/Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Positive Regulation Of Axon Extension	0.92978	0.50	NDEL1 SEMA7A
Cytoskeletal Anchoring At Plasma Membrane	0.88241	1.00	SPTBN4
Positive Regulation Of Cell Motility	0.88241	1.00	CSF1R
Myotube Differentiation	0.80059	1.00	GSK3B
Response To Lithium Ion	0.80059	1.00	GSK3B
Positive Regulation Of Protein Export From Nucleus	0.80059	1.00	GSK3B
Regulation Of Microtubule-Based Process	0.80059	1.00	GSK3B
Negative Regulation Of Astrocyte Differentiation	0.73241	1.00	ATF5
Cellular Response To Cytokine Stimulus	0.73241	1.00	CSF1R
Inner Cell Mass Cell Proliferation	0.67472	1.00	NDEL1
Macrophage Differentiation	0.67472	1.00	CSF1R

Table 4.10 continued from previous page

Annotation (Pathway/Process)	Significance of network distance distribution (XD-Score)	Significance of overlap (Fisher-test, q-value)	Dataset size (overlap)
Transmission Of Nerve Impulse	0.62527	1.00	SPTBN4
Regulation Of Heart Rate	0.62527	1.00	MYH7
Adult Heart Development	0.62527	1.00	MYH7
Ruffle Organization	0.62527	1.00	CSF1R
Positive Regulation Of Protein Serine/Threonine Kinase Activity	0.62527	1.00	CSF1R
Response To Endoplasmic Reticulum Stress	0.58241	1.00	ATF4
Positive Regulation Of Peptidyl-Threonine Phosphorylation	0.58241	1.00	GSK3B
Myoblast Fusion	0.58241	1.00	GSK3B
Adult Behavior	0.58241	1.00	SPTBN4
Regulation Of Gene Expression By Genetic Imprinting	0.58241	1.00	GSK3B
Positive Regulation Of Cell-Matrix Adhesion	0.58241	1.00	GSK3B
Negative Regulation Of Neurogenesis	0.54491	1.00	ATF4

Table 4.10 continued from previous page

Annotation (Pathway/Process)	Significance of network distance distribution (XD-Score)	Significance of overlap (Fisher-test, q-value)	Dataset size (overlap)
Positive Regulation Of Protein Complex Assembly	0.54491	1.00	GSK3B
Response To Reactive Oxygen Species	0.51182	1.00	MYH7
Actin Filament Capping	0.51182	1.00	SPTBN4
Positive Regulation Of Protein Tyrosine Kinase Activity	0.51182	1.00	CSF1R
Monocyte Differentiation	0.51182	1.00	CSF1R
Regulation Of Neuronal Synaptic Plasticity	0.45610	1.00	GSK3B
Establishment Of Cell Polarity	0.43241	1.00	GSK3B
Osteoclast Differentiation	0.43241	1.00	CSF1R
Positive Regulation Of Rac Gtpase Activity	0.41098	1.00	GSK3B
Gluconeogenesis	0.40102	0.87	GPD2 ATF4
Negative Regulation Of Protein Binding	0.39150	1.00	GSK3B

Table 4.10 continued from previous page

Annotation (Pathway/Process)	Significance of network distance distribution (XD-Score)	Significance of overlap (Fisher-test, q-value)	Dataset size (overlap)
Phosphatidylinositol Metabolic Process	0.39150	1.00	CSF1R
Protein Export From Nucleus	0.39150	1.00	GSK3B
Positive Regulation Of Tyrosine Phosphorylation Of Stat3 Protein	0.37372	1.00	CSF1R
Response To Morphine	0.35741	1.00	SRR
Establishment Or Maintenance Of Cell Polarity	0.35741	1.00	GSK3B
Ventricular Cardiac Muscle Tissue Morphogenesis	0.32857	1.00	MYH7
Spindle Assembly	0.32857	1.00	CEP63
DNA Damage Checkpoint	0.31575	1.00	CEP63
Glycogen Metabolic Process	0.31575	1.00	GSK3B
Positive Regulation Of Multicellular Organism Growth	0.29276	1.00	SPTBN4
Adult Walking Behavior	0.28241	1.00	SPTBN4
Epithelial To Mesenchymal Transition	0.28241	1.00	GSK3B

Table 4.10 continued from previous page

Annotation (Pathway/Process)	Significance of network distance distribution (XD-Score)	Significance of overlap (Fisher-test, q-value)	Dataset size (overlap)
Fertilization	0.28241	1.00	SPTBN4
Positive Regulation Of Protein Catabolic Process	0.28241	1.00	GSK3B
Negative Regulation Of MAP Kinase Activity	0.26366	1.00	GSK3B
Regulation Of Inflammatory Response	0.24712	1.00	SEMA7A
Positive Regulation Of Protein Binding	0.24712	1.00	GSK3B
Muscle Filament Sliding	0.23956	1.00	MYH7
Phosphatidylinositol-Mediated Signaling	0.20741	1.00	CSF1R GSK3B
Positive Regulation Of ERK1 And ERK2 Cascade	0.19928	1.00	CSF1R SEMA7A
Hippocampus Development	0.18696	1.00	GSK3B
Positive Regulation Of Peptidyl-Serine Phosphorylation	0.18241	1.00	GSK3B

Table 4.10 continued from previous page

Annotation (Pathway/Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Phosphorylation	0.17806	1.00	GSK3B
Lipopolysaccharide Biosynthetic Process	0.16991	1.00	TRIO
Fat Cell Differentiation	0.15888	1.00	GSK3B
Protein Homotetramerization	0.15549	1.00	SRR
Axonogenesis	0.15384	1.00	SPTBN4 GSK3B
Cellular Amino Acid Metabolic Process	0.14031	1.00	ATF4
Peptidyl-Tyrosine Phosphorylation	0.13759	1.00	CSF1R
Peptidyl-Serine Phosphorylation	0.13496	1.00	GSK3B
Cell Migration	0.12995	1.00	NDEL1 GSK3B
Chromosome Segregation	0.12757	1.00	NDEL1

Table 4.10 continued from previous page

Annotation (Pathway/Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Microtubule Cytoskeleton Organization	0.12757	1.00	NDEL1 DISC1
Activation Of Signaling Protein Activity Involved In Unfolded Protein Response	0.12087	1.00	ATF4
Integrin-Mediated Signaling Pathway	0.10917	1.00	SEMA7A
Negative Regulation Of Canonical Wnt Receptor Signaling Pathway	0.10917	1.00	GSK3B
Regulation Of Rho Protein Signal Transduction	0.10917	1.00	TRIO
Hemopoiesis	0.10741	1.00	CSF1R
Neuron Projection Development	0.10083	1.00	NDEL1
Canonical Wnt Receptor Signaling Pathway	0.10083	1.00	GSK3B DISC1

Table 4.10 continued from previous page

Annotation (Pathway/Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Axon Guidance	0.09491	0.63	SPTBN4 TRIO GSK3B SEMA7A
Endoplasmic Reticulum Unfolded Protein Response	0.09352	1.00	ATF4
ATP Catabolic Process	0.09085	1.00	MYH7
Mitotic Prometaphase	0.09085	1.00	NDEL1
Transmembrane Receptor Protein Tyrosine Kinase Signaling Pathway	0.08830	1.00	CSF1R
M Phase Of Mitotic Cell Cycle	0.07919	1.00	NDEL1
Neuron Migration	0.07616	1.00	NDEL1 DISC1
Potassium Ion Transport	0.07332	1.00	KCNQ5

Table 4.10 continued from previous page

Annotation (Pathway/Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Muscle Contraction	0.07241	1.00	MYH7
Regulation Of Cell Shape	0.07065	1.00	CSF1R
Positive Regulation Of Protein Phosphorylation	0.06813	1.00	CSF1R
Induction Of Apoptosis By Extracellular Signals	0.06732	1.00	TRIO
Anti-Apoptosis	0.06692	1.00	ATF5 GSK3B
Positive Regulation Of Cell Migration	0.06575	1.00	CSF1R
Nerve Growth Factor Receptor Signaling Pathway	0.06536	1.00	TRIO GSK3B
Protein Complex Assembly	0.06000	1.00	KCNQ5
G2/M Transition Of Mitotic Cell Cycle	0.05934	1.00	CEP63
Fibroblast Growth Factor Receptor Signaling Pathway	0.05868	1.00	GSK3B

Table 4.10 continued from previous page

Annotation (Pathway/Process)	**Significance of network distance distribution (XD-Score)**	**Significance of overlap (Fisher-test, q-value)**	**Dataset size (overlap)**
Sensory Perception Of Sound	0.05804	1.00	SPTBN4
Organ Morphogenesis	0.05618	1.00	GSK3B
Aging	0.05558	1.00	SRR
Central Nervous System Development	0.05218	1.00	IL1RAPL2
Epidermal Growth Factor Receptor Signaling Pathway	0.05008	1.00	GSK3B

Chapter 5

Conclusions and Recommendations

Schizophrenia is worldwide present and approximately effects 1 percent of the human population. It is very harmful and effects mental health of the individual. It is generally associated with emotional impairment, social dysfunction and cognitive deficits. So, it is very important to know about the genes, pathways which are involved in the schizophrenia. Disrupted in schizophrenia 1 (DISC1) is a scaffold and multifunctional large protein which consists of 854 amino acids that has a 93,611 kDa molecular weight which occurs in humans and is encoded by the important candidate DISC1 gene. Early studies provided linkage evidences between schizophrenia and a balanced translocation involving chromosomes 1 and 11. It is a protein that performs various functions including cell proliferation, regulation, differentiation, migration, and cell to cell adhesion, neurogenesis, and provides diverse understanding regarding schizophrenia pathophysiology.

So, first objective of my study is to investigate deleterious and damaging regions in DISC1 protein variants. So, to achieve my objective I analysis 23 sequences of DISC1 protein by the use of different tools which indicate the deleterious and damaging regions of DISC1 protein. disEMBL tool is used to predict the disordered protein regions in the sequence. Then the obtained results are verified by

PrDOS server which confirms the disordered regions in the protein sequence. Moving forward, functional maps of protein sequences and prediction of the functional variations in the amino acid sequence is done by fuNTRP tool, such damaged amino acids are highlighted and in the next step, insertions, deletions and multiple substitutions in the structure of protein sequence are predicted by the use of PROVEAN tool and the obtained results are verified by the Polyphen 2 tool. It is unique tool because of predictive features, alignment and classification methods. Furthermore, this tool also predicts about impact of substitution of amino acid on the function, stability and structure of human proteins. So, by using above mentioned tools and soft wares, deleterious and damaging regions in the protein sequences were predicted and analyzed. Then moving towards next objective of the study, some more steps are performed.

Next objective of the study is to identify structural variations in DISC1 and associated pathways involved in schizophrenia. To, identify the structural variations in the protein sequences; firstly MutPred 2 is effectively used in order to detect functional and structural properties of amino acids. MutPred is a web based tool which is used to identify amino acid substitutions whether they are benign or damaging. It almost predict 50 different properties which may include altered disordered interface, altered DNA binding, gain of helix, loss of strand and loss of phosphorylation site etc. So, that molecular mechanism of pathogenicity can be interpreted.

After the identification of structural variations in the amino acid sequences, I proceed further and move ahead to the I-Stable 2.0 tool which is basically used to predict the thermal stability of the protein sequences by using many characteristic modules. I-Stable can be operated by the use of two input types: one is on the base of structure and other input method is on the base of the sequence. After analyzing stability of the sequences, I opted to perform the multiple sequence alignment of the DISC1 sequences with the relevant sequences from other organisms. This is done by the use of Consurf server. Consurf server server is used to predict functional regions present in the protein sequence. It also checks and find evolutionary conservative positions of amino acid in the protein sequence. The

evolutionary conservation of amino acid depends upon the structural and functional importance.After performing multiple sequence alignment of the sequences, construction of the phylogenetic trees is the next step. Phylogenetic tree is constructed by using Neighbor joining with maximum likelihood distance method. Method of calculation was Bayesian and Best fit was model which was substitution for proteins. Then Swiss model expasy server is used to predict and validate the structure of obtained protein sequences. Moreover Phyre2 tool is used to detect the 3D protein models of mutant sequences. Raptor X is used for the prediction of secondary structure and tertiary structure modeling of the templates. It is basically used to validate the structures which were earlier detected by Phyre2 server. This server also predicts binding sites, disordered regions, solvent accessibility along with secondary and tertiary structure prediction. Saves 5.0 is used to confirm the protein structure stability and by providing graph in the result, compares expected and observed value. TM-align server is specific algorithm used to sequence independent protein structure and compare the disordered and wild protein sequence. It first aligns the structure on the base of similarity in the structure. After identification of structural and functional variations then it is important to identify the pathways involved in the schizophrenia and also to find the interactors of DISC1 gene so to predict pathways I proceed to the next objective of my study. Last objective of my study is to identify significant interactors of DISC1 protein and to perform pathway analysis to elucidate DISC1 and its variant in the pathophysiology of schizophrenia.

Firstly PICKLE 2.0 is selected as it can find protein protein interactions of the mouse and the human beings directly and mine is human gene so PICKLE 2.0 is good enough option for me to proceed easily. Moreover it is meta-database which is used to detect direct protein-protein interaction network in humans and also uses the reviewed human complete proteome of Uniprot as a standard. Then in the next step GENEMANIA is used to predict the function of genes and set of genes. It is user friendly and flexible web interface which can generate functions of genes and select genes for functional assay. It can also operate for single gene queries, multiple gene queries and for searching network. Genemania has high accuracy

rate algorithm, large database so it is very useful for analyzing function genes. Afterwards I proceed to the Gephi 0.9.2 is software which is used to visualize and analyze the graphs and networks freely. It helps the user to explore and manipulate the interacting networks. It can deal with 20,000 nodes at a time. Lastly I used EnrichNet tool which is used for enrichment analysis of the network. It is web based tool which evaluate function, components, processes and pathways among the proteins and genes.

Overall it is concluded that in this study the structural, functional analysis of DISC1 translocation and other sequence variants provide us a path to explore the role of DISC1 in schizophrenia. DISC1 is a multifactorial complex protein with various molecular interactors. This unique property of multiple interactors makes it a suitable candidate for find new therapeutic targets for schizophrenia .In the futher the predicted analysis and conclusion can be validated in the wet lab so that role of DISC1 can be further investigated and it may help in the cure and prevention of the schizophrenia.

Bibliography

[1]. Xu, Y., J. Ren, and H. Ye, Association between variations in the disrupted in schizophrenia 1 gene and schizophrenia: A meta-analysis. Gene, 2018. 651: p. 94-99.

[2]. Luis Sanchez-Pulido, Chris P. Ponting, Structure and evolutionary history of DISC1, Human Molecular Genetics, Volume 20, Issue R2, 15 October 2011, Pages R175–R181, https://doi.org/10.1093/hmg/ddr374.

[3]. Sawa A, Snyder SH. Genetics. Two genes link two distinct psychoses. Science. 2005 Nov 18;310(5751):1128-9. doi: 10.1126/science.1121114. PMID: 16293746.

[4]. Soares, D.C., et al., DISC1: structure, function, and therapeutic potential for major mental illness. ACS chemical neuroscience, 2011. 2(11): p. 609-632.

[5]. Nehring, R.B., et al., The metabotropic GABAB receptor directly interacts with the activating transcription factor 4. Journal of Biological Chemistry, 2000. 275(45): p. 35185-35191.

[6]. Donnell, A.F., et al., Identification of pyridazino [4, 5-b] indolizines as selective PDE4B inhibitors. Bioorganic & medicinal chemistry letters, 2010. 20(7): p. 2163-2167.

[7]. Bord, L., et al., Primate disrupted-in-schizophrenia-1 (DISC1): high divergence of a gene for major mental illnesses in recent evolutionary history. Neuroscience research, 2006. 56(3): p. 286-293.

[8]. Bradshaw, N.J. and D.J. Porteous, DISC1-binding proteins in neural development, signalling and schizophrenia. Neuropharmacology, 2012. 62(3): p. 1230-1241.

[9]. Wang, H.-Y., et al., Gene polymorphisms of DISC1 is associated with schizophrenia: Evidence from a meta-analysis. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2018. 81: p. 64-73.

[10]. Hennah, W., et al., A haplotype within the DISC1 gene is associated with visual memory functions in families with a high density of schizophrenia. Molecular psychiatry, 2005. 10(12): p. 1097-1103.

[11]. Brandon, N.J., et al., Understanding the role of DISC1 in psychiatric disease and during normal development. Journal of Neuroscience, 2009. 29(41): p. 12768-12775.

[12]. Pletnikov, M., et al., Inducible expression of mutant human DISC1 in mice is associated with brain and behavioral abnormalities reminiscent of schizophrenia. Molecular psychiatry, 2008. 13(2): p. 173-186.

[13]. Austin, C., et al., Expression of Disrupted-In-Schizophrenia-1, a schizophrenia-associated gene, is prominent in the mouse hippocampus throughout brain development. Neuroscience, 2004. 124(1): p. 3-10.

[14]. Roberts RC. Schizophrenia in translation: disrupted in schizophrenia (DISC1): integrating clinical and basic findings. Schizophr Bull. 2007 Jan;33(1):11-5. doi: 10.1093/schbul/sbl063. Epub 2006 Nov 30. PMID: 17138582; PMCID: PMC2632285.

[15]. Bradshaw, N.J. and C. Korth, A Structural Organization for Disrupted in Schizophrenia 1, Identified by High-Throughput Screening, Reveals Distinctly Folded Regions Which Are Bisected by Mental Illness-Related Mutations Antony SK Yerabham, Philippe J. Mas, Christina Decker, Dinesh C. Soares, Oliver H. Weiergräber, Luitgard Nagel-Steger, Dieter Willbold. Darren J. Hart.

[16]. Morris, J.A., et al., DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. Human molecular genetics, 2003. 12(13): p. 1591-1608.

[17]. Lipina, T.V., et al., Synergistic interactions between PDE4B and GSK-3: DISC1 mutant mice. Neuropharmacology, 2012. 62(3): p. 1252-1262.

[18]. Porteous, D.J., et al., DISC1 as a genetic risk factor for schizophrenia and related major mental illness: response to Sullivan. Molecular psychiatry, 2014. 19(2): p. 141-143.

[19]. Porteous DJ, Millar JK, Brandon NJ, Sawa A. DISC1 at 10: connecting psychiatric genetics and neuroscience. Trends Mol Med. 2011 Dec;17(12):699-706. doi: 10.1016/j.molmed.2011.09.002. Epub 2011 Oct 19. PMID: 22015021; PMCID: PMC3253483.

[20]. Mouaffak, F., et al., Association of Disrupted in Schizophrenia 1 (DISC1) missense variants with ultra-resistant schizophrenia. The pharmacogenomics journal, 2011. 11(4): p. 267-273.

[21]. McGlashan, T.H. and R.E. Hoffman, Schizophrenia as a disorder of developmentally reduced synaptic connectivity. Archives of general psychiatry, 2000. 57(7): p. 637-648.

[22]. Ozeki, Y., et al., Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. Proceedings of the National Academy of Sciences, 2003. 100(1): p. 289-294.

[23]. Sullivan, P.F., K.S. Kendler, and M.C. Neale, Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of general psychiatry, 2003. 60(12): p. 1187-1192.

[24]. JACKSON, J.H., A dynamic interplay between positive and negative factors in insanity. Levels of dissolution and mental symptoms. Dialogues in Philosophy, Mental & Neuro Sciences, 2020. 13(1).

[25]. Fatani, B.Z., et al., Schizophrenia: etiology, pathophysiology and management-a review. The Egyptian Journal of Hospital Medicine, 2017. 69(6): p. 2640-2646.

[26]. Need, A.C., et al., Jessica M. Maia, Kevin V. Shianna, Min He, Elizabeth T. Cirulli, Curtis E. Gumbs, Qian Zhao, C. Ryan Campbell, Linda Hong, Peter Rosenquist, 6 Anu Putkonen, 7 Tero Hallikainen, 7 Eila Repo-Tiihonen, 7 Jari Tiihonen, 7, 8 Deborah L. Levy, 9 Herbert Y. Meltzer, 10 and David B. Goldstein1, 11.

[27]. Tam, G.W., et al., The role of DNA copy number variation in schizophrenia. Biological psychiatry, 2009. 66(11): p. 1005-1012.

[28]. Faraone, S.V., L. Taylor, and M. Tsuang, The molecular genetics of schizophrenia: an emerging consensus. Expert reviews in molecular medicine, 2002. 4(14): p. 1.

[29]. Brandon, N.J. and A. Sawa, Linking neurodevelopmental and synaptic theories of mental illness through DISC1. Nature Reviews Neuroscience, 2011. 12(12): p. 707-722.

[30]. Angst, J., et al., Bipolar manic-depressive psychoses: results of a genetic investigation. Human Genetics, 1980. 55(2): p. 237-254.

[31]. Tsuang, M.T., S.V. Faraone, and J.A. Fleming, Familial transmission of major affective disorders: is there evidence supporting the distinction between unipolar and bipolar disorders? The British Journal of Psychiatry, 1985. 146(3): p. 268-271.

[32]. Gershon, E.S., et al., A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. Archives of General psychiatry, 1982. 39(10): p. 1157-1167.

[33]. Möller H. J. (2003). Bipolar disorder and schizophrenia: distinct illnesses or a continuum. The Journal of clinical psychiatry, 64 Suppl 6, 23–28.

[34]. Ishizuka, K., et al., Functional characterization of rare NRXN1 variants identified in autism spectrum disorders and schizophrenia. Journal of Neurodevelopmental Disorders, 2020. 12(1): p. 1-16.

[35]. De Crescenzo, F., et al., Autistic symptoms in schizophrenia spectrum disorders: a systematic review and meta-analysis. Frontiers in psychiatry, 2019. 10: p. 78.

[36]. Kendler, K.S., et al., The Clinical Characteristics of Major Depression as Indices of the. British Journal of Psychiatry, 1994. 165: p. 66-72.

[37]. Kety, S.S., et al., Mental illness in the biological and adoptive relatives of schizophrenic adoptees: replication of the Copenhagen study in the rest of Denmark. Archives of general psychiatry, 1994. 51(6): p. 442-455.

[38]. Pulver, A.E., K.Y. Liang, and G. Vogler, Estimating effects of proband characteristics on familial risk: II. The association between age at onset and familial risk in the Maryland schizophrenia sample. Genetic Epidemiology, 1991. 8(5): p. 339-350.

[39]. Liang, K.Y. and G. Vogler, Estimating effects of probands' characteristics on familial risk: I. Adjustment for censoring and correlated ages at onset. Genetic epidemiology, 1991. 8(5): p. 329-338.

[40]. Spitzer, R.L., E. Robins, and J. Endicott, Research diagnostic criteria (RDC) for a selected group of functional disorders. 1989: Research Assessment and Training Unit, New York State Psychiatric Institute.

[41]. Wender, P.H., et al., Psychiatric disorders in the biological and adoptive families of adopted individuals with affective disorders. Archives of General Psychiatry, 1986. 43(10): p. 923-929.

[42]. Trost, S., et al., DISC1 (disrupted-in-schizophrenia 1) is associated with cortical grey matter volumes in the human brain: a voxel-based morphometry (VBM) study. Journal of psychiatric research, 2013. 47(2): p. 188-196.

[43]. Shao, L., et al., Disrupted-in-Schizophrenia-1 (DISC1) protein disturbs neural function in multiple disease-risk pathways. Human molecular genetics, 2017. 26(14): p. 2634-2648.

[44]. Blackwood, D., et al., Schizophrenia and affective disorders—cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. The American Journal of Human Genetics, 2001. 69(2): p. 428-433.

[45]. Eykelenboom, J.E., et al., A t (1; 11) translocation linked to schizophrenia and affective disorders gives rise to aberrant chimeric DISC1 transcripts that encode structurally altered, deleterious mitochondrial proteins. Human molecular genetics, 2012. 21(15): p. 3374-3386.

[46]. Malavasi, E.L., et al., DISC1 regulates N-methyl-D-aspartate receptor dynamics: abnormalities induced by a Disc1 mutation modelling a translocation linked to major mental illness. Translational psychiatry, 2018. 8(1): p. 1-16.

[47]. Millar, J.K., et al., DISC1 and DISC2: discovering and dissecting molecular mechanisms underlying psychiatric illness. Annals of medicine, 2004. 36(5): p. 367-378.

[48]. Whalley, H.C., et al., Effects of a balanced translocation between chromosomes 1 and 11 disrupting the DISC1 locus on white matter integrity. PLoS One, 2015. 10(6): p. e0130900.

[49]. Zhou, X., et al., Insoluble DISC1-Boymaw Fusion Proteins Generated by the DISC1 Translocation. Molecular psychiatry, 2010. 15(7): p. 670.

[50]. Marshall, C.R., et al., Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nature genetics, 2017. 49(1): p. 27-35.

[51]. Sachs, N., et al., A frameshift mutation in Disrupted in Schizophrenia 1 in an American family with schizophrenia and schizoaffective disorder. Molecular psychiatry, 2005. 10(8): p. 758-764.

[52]. Ma, J.-H., et al., Association on DISC1 SNPs with schizophrenia risk: a meta-analysis. Psychiatry research, 2018. 270: p. 306-309.

[53]. Liu, C.-M., et al., Genetic associations and expression of extra-short isoforms of disrupted-in-schizophrenia 1 in a neurocognitive subgroup of schizophrenia. Journal of human genetics, 2019. 64(7): p. 653-663.

[54]. Tiihonen, J., et al., 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). The Lancet, 2009. 374(9690): p. 620-627.

[55]. Green, M.F., et al., Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? Schizophrenia bulletin, 2000. 26(1): p. 119-136.

[56]. Ma, L., et al., Cloning and characterization of Disc1, the mouse ortholog of DISC1 (Disrupted-in-Schizophrenia 1). Genomics, 2002. 80(6): p. 662-672.

[57]. Kamiya, A., et al., A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. Nature cell biology, 2005. 7(12): p. 1167-1178.

[58]. Narayanan, S., et al., Molecular characterization of disrupted in schizophrenia-1 risk variant S704C reveals the formation of altered oligomeric assembly. Journal of Biological Chemistry, 2011. 286(51): p. 44266-44276.

[59]. Kvajo, M., et al., A mutation in mouse Disc1 that models a schizophrenia risk allele leads to specific alterations in neuronal architecture and cognition. Proceedings of the National Academy of Sciences, 2008. 105(19): p. 7076-7081.

[60]. Meltzer, H.Y., et al., Clozapine treatment for suicidality in schizophrenia: international suicide prevention trial (InterSePT). Archives of general psychiatry, 2003. 60(1): p. 82-91.

[61]. Chubb, J., et al., The DISC locus in psychiatric illness. Molecular psychiatry, 2008. 13(1): p. 36-64.

[62]. Yerabham, A.S., et al., A structural organization for the Disrupted in Schizophrenia 1 protein, identified by high-throughput screening, reveals distinctly folded regions, which are bisected by mental illness-related mutations. Journal of Biological Chemistry, 2017. 292(16): p. 6468-6477.

[63]. Millar, J.K., et al., Disruption of two novel genes by a translocation co-segregating with schizophrenia. Human molecular genetics, 2000. 9(9): p. 1415-1423.

[64]. Thomson, P.A., et al., DISC1 genetics, biology and psychiatric illness. Frontiers in biology, 2013. 8(1): p. 1-31.

[65]. Nakata, K., et al., DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. Proceedings of the National Academy of Sciences, 2009. 106(37): p. 15873-15878.

[66]. Leliveld, S.R., et al., Insolubility of disrupted-in-schizophrenia 1 disrupts oligomer-dependent interactions with nuclear distribution element 1 and is associated with sporadic mental disease. Journal of Neuroscience, 2008. 28(15): p. 3839-3845.

[67]. Burdick, K.E., et al., Elucidating the relationship between DISC1, NDEL1 and NDE1 and the risk for schizophrenia: evidence of epistasis and competitive binding. Human molecular genetics, 2008. 17(16): p. 2462-2473.

[68]. Song, W., et al., Identification of high risk DISC1 structural variants with a 2% attributable risk for schizophrenia. Biochemical and biophysical research communications, 2008. 367(3): p. 700-706.

[69]. Song, W., et al., Identification of high risk DISC1 protein structural variants in patients with bipolar spectrum disorder. Neuroscience letters, 2010. 486(3): p. 136-140.

[70]. Szeszko, P.R., et al., DISC1 is associated with prefrontal cortical gray matter and positive symptoms in schizophrenia. Biological psychology, 2008. 79(1): p. 103-110.

[71]. Hashimoto, R., et al., Impact of the DISC1 Ser704Cys polymorphism on risk for major depression, brain morphology and ERK signaling. Human molecular genetics, 2006. 15(20): p. 3024-3033.

[72]. Eastwood, S., C. Hodgkinson, and P. Harrison, DISC-1 Leu607Phe alleles differentially affect centrosomal PCM1 localization and neurotransmitter release. Molecular psychiatry, 2009. 14(6): p. 556-557.

[73]. Atkin, T., et al., Disrupted in Schizophrenia-1 regulates intracellular trafficking of mitochondria in neurons. Molecular psychiatry, 2011. 16(2): p. 122-124.

[74]. Takahashi, T., et al., The Disrupted-in-Schizophrenia-1 Ser704Cys polymorphism and brain morphology in schizophrenia. Psychiatry Research: Neuroimaging, 2009. 172(2): p. 128-135.

[75]. Callicott, J.H., et al., Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. Proceedings of the National Academy of Sciences, 2005. 102(24): p. 8627-8632.

[76]. Di Giorgio, A., et al., Association of the Ser704Cys DISC1 polymorphism with human hippocampal formation gray matter and function during memory encoding. European Journal of Neuroscience, 2008. 28(10): p. 2129-2136.

[77]. Sprooten, E., et al., Association of white matter integrity with genetic variation in an exonic DISC1 SNP. Molecular psychiatry, 2011. 16(7): p. 688-689.

[78]. Prata, D., et al., Effect of disrupted-in-schizophrenia-1 on pre-frontal cortical function. Molecular psychiatry, 2008. 13(10): p. 915-917.

[79]. DeRosse, P., et al., Disrupted in schizophrenia 1 genotype and positive symptoms in schizophrenia. Biological psychiatry, 2007. 61(10): p. 1208-1210.

[80]. Ayhan, Y., et al., Differential effects of prenatal and postnatal expressions of mutant human DISC1 on neurobehavioral phenotypes in transgenic mice:

evidence for neurodevelopmental origin of major psychiatric disorders. Molecular psychiatry, 2011. 16(3): p. 293-306.

[81]. Taylor, M.S., et al., Evolutionary constraints on the Disrupted in Schizophrenia locus. Genomics, 2003. 81(1): p. 67-77.

[82]. Tsien, J.Z., P.T. Huerta, and S. Tonegawa, The essential role of hippocampal CA1 NMDA receptor–dependent synaptic plasticity in spatial memory. Cell, 1996. 87(7): p. 1327-1338.

[83]. Olney, J.W. and N.B. Farber, Glutamate receptor dysfunction and schizophrenia. Archives of general psychiatry, 1995. 52(12): p. 998-1007.

[84]. Coyle, J.T., Glutamate and schizophrenia: beyond the dopamine hypothesis. Cellular and molecular neurobiology, 2006. 26(4-6): p. 363-382.

[85]. Merritt, K. and A. Egerton, 165. Nature of Glutamate Alterations in Schizophrenia: A Meta-Analysis of Proton Magnetic Resonance Spectroscopy Studies. Schizophrenia Bulletin, 2017. 43(Suppl 1): p. S84.

[86]. Hall, J., et al., Associative learning and the genetics of schizophrenia. Trends in neurosciences, 2009. 32(6): p. 359-365.

[87]. Wu, Q., Y. Li, and B. Xiao, DISC1-related signaling pathways in adult neurogenesis of the hippocampus. Gene, 2013. 518(2): p. 223-230.

[88]. Grimes, C.A. and R.S. Jope, The multifaceted roles of glycogen synthase kinase 3β in cellular signaling. Progress in neurobiology, 2001. 65(4): p. 391-426.

[89]. Ali, A., K.P. Hoeflich, and J.R. Woodgett, Glycogen synthase kinase-3: properties, functions, and regulation. Chemical reviews, 2001. 101(8): p. 2527-2540.

[90]. Mao, Y., et al., Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3β/β-catenin signaling. Cell, 2009. 136(6): p. 1017-1031.

[91]. Pruitt, K.D., T. Tatusova, and D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research, 2007. 35(suppl_1): p. D61-D65.

[92]. Iakoucheva, L.M. and A.K. Dunker, Order, disorder, and flexibility: prediction from protein sequence. Structure, 2003. 11(11): p. 1316-1317.

[93]. Kabsch, W. and C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers: Original Research on Biomolecules, 1983. 22(12): p. 2577-2637.

[94]. Smith, D.K., et al., Improved amino acid flexibility parameters. Protein Science, 2003. 12(5): p. 1060-1072.

[95]. Li, X., et al., Comparing predictors of disordered protein. Genome Informatics, 2000. 11: p. 172-184.

[96]. Lieutaud, P., B. Canard, and S. Longhi, MeDor: a metaserver for predicting protein disorder. BMC genomics, 2008. 9(S2): p. S25.

[97]. Miller, M., et al., funtrp: identifying protein positions for variation driven functional tuning. Nucleic acids research, 2019. 47(21): p. e142-e142.

[98]. Choi, Y., et al., Predicting the functional effect of amino acid substitutions and indels. PloS one, 2012. 7(10): p. e46688.

[99]. Choi, Y. and A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics, 2015. 31(16): p. 2745-2747.

[100]. PROVEAN, S. and S. GO, Praveen P. Balgir* and Suman Rani. 2017.

[101]. Adzhubei, I.A., et al., A method and server for predicting damaging missense mutations. Nature methods, 2010. 7(4): p. 248-249.

[102]. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics, 2013. 76(1): p. 7.20. 1-7.20. 41.

[103]. Li, B., et al., Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics, 2009. 25(21): p. 2744-2750.

[104]. Chen, C.-W., et al., iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules. Computational and structural biotechnology journal, 2020.

[105]. Ashkenazy, H., et al., ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic acids research, 2010. 38(suppl_2): p. W529-W533.

[106]. Glaser, F., et al., ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics, 2003. 19(1): p. 163-164.

[107]. Celniker, G., et al., ConSurf: using evolutionary data to raise testable hypotheses about protein function. Israel Journal of Chemistry, 2013. 53(3-4): p. 199-206.

[108]. Bienert, S., et al., The SWISS-MODEL Repository—new features and functionality. Nucleic acids research, 2017. 45(D1): p. D313-D319.

[109]. Schwede, T., et al., SWISS-MODEL: an automated protein homology-modeling server. Nucleic acids research, 2003. 31(13): p. 3381-3385.

[110]. Waterhouse, A., et al., SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research, 2018. 46(W1): p. W296-W303.

[111]. Biasini, M., et al., SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research, 2014. 42(W1): p. W252-W258.

[112]. Kelley, L.A., et al., The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols, 2015. 10(6): p. 845-858.

[113]. Basyuni, M., et al. Protein modelling of triterpene synthase genes from mangrove plants using Phyre2 and Swiss-model. in Journal of Physics: Conference Series. 2018. IOP Publishing.

[114]. Källberg, M., et al., Template-based protein structure modeling using the RaptorX web server. Nature protocols, 2012. 7(8): p. 1511-1522.

[115]. Gao, Y., et al., RaptorX-Angle: real-value prediction of protein backbone dihedral angles through a hybrid method of clustering and deep learning. BMC bioinformatics, 2018. 19(4): p. 100.

[116]. Swain, S.S., In silico Approach in the Prediction and Analysis of the Three-dimensional Structure of Maleylacetate reductase: A Biodegrading Protein. International Journal Bioautomation, 2013. 17(4): p. 217-226.

[117]. Sashidhara, K.V., et al., Designing, synthesis of selective and high-affinity chalcone-benzothiazole hybrids as Brugia malayi thymidylate kinase inhibitors: in vitro validation and docking studies. European Journal of Medicinal Chemistry, 2015. 103: p. 418-428.

[118]. Mukherjee, S. and Y. Zhang, MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. Nucleic acids research, 2009. 37(11): p. e83-e83.

[119]. Malod-Dognin, N. and N. Pržulj, GR-Align: fast and flexible alignment of protein 3D structures using graphlet degree similarity. Bioinformatics, 2014. 30(9): p. 1259-1265.

[120]. Zhang, Y. and J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic acids research, 2005. 33(7): p. 2302-2309.

[121]. Gioutlakis, A., M.I. Klapa, and N.K. Moschonas, PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. PLoS One, 2017. 12(10): p. e0186039.

[122]. Mostafavi, S., et al., GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome biology, 2008. 9(S1): p. S4.

[123]. Franz, M., et al., GeneMANIA update 2018. Nucleic acids research, 2018. 46(W1): p. W60-W64.

[124]. Jacomy, M., et al., ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one, 2014. 9(6): p. e98679.

[125]. Bastian, M., S. Heymann, and M. Jacomy. Gephi: an open source software for exploring and manipulating networks. in Proceedings of the International AAAI Conference on Web and Social Media. 2009.

[126]. Glaab, E., et al., EnrichNet: network-based gene set enrichment analysis. Bioinformatics, 2012. 28(18): p. i451-i457.

[127]. Gupta, B. and R.B. Mishra, Protein Network for Associating Genes with Dementia. International Journal of Computer Applications, 2013. 83(10).

[128]. Meng, F., V. Uversky, and L. Kurgan, Computational prediction of intrinsic disorder in proteins. Current protocols in protein science, 2017. 88(1): p. 2.16. 1-2.16. 14.

[129]. Itan, Y., et al., The mutation significance cutoff: gene-level thresholds for variant predictions. Nature methods, 2016. 13(2): p. 109-110.

[130]. Vikas Pejaver, Jorge Urresti, Jose Lugo-Martinez, Kymberleigh A. Pagel, Guan Ning Lin, Hyun-Jun Nam, Matthew Mort, David N. Cooper, Jonathan Sebat, Lilia M. Iakoucheva, Sean D. Mooney, Predrag Radivojac bioRxiv 134981; doi: https://doi.org/10.1101/134981.

[131]. Chen, C.-W., J. Lin, and Y.-W. Chu. iStable: off-the-shelf predictor integration for predicting protein stability changes. in BMC bioinformatics. 2013. Springer.

[132]. Ashkenazy, H., et al., ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic acids research, 2016. 44(W1): p. W344-W350.

[133]. Veidenberg, A., A. Medlar, and A. Löytynoja, Wasabi: an integrated platform for evolutionary sequence analysis and data visualization. Molecular biology and evolution, 2016. 33(4): p. 1126-1130.

[134]. Kopp, J. and T. Schwede, The SWISS-MODEL Repository: new features and functionalities. Nucleic acids research, 2006. 34(suppl_1): p. D315-D318.

[135]. Wang, S., et al., RaptorX-Property: a web server for protein structure property prediction. Nucleic acids research, 2016. 44(W1): p. W430-W435.

[136]. Pramanik, K., et al., An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of Pseudomonas aeruginosa. Journal of Genetic Engineering and Biotechnology, 2017. 15(2): p. 527-537.

[137]. Dong, R., et al., mTM-align: an algorithm for fast and accurate multiple protein structure alignment. Bioinformatics, 2018. 34(10): p. 1719-1725.

[138]. Dimitrakopoulos, G.N., M.I. Klapa, and N.K. Moschonas, PICKLE 3.0: Enriching the human Meta-database with the mouse protein interactome extended via mouse-human orthology. Bioinformatics, 2020.

[139]. Warde-Farley, D., et al., The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research, 2010. 38(suppl_2): p. W214-W220.