INSILICO MODELING, DOCKING AND ANALYSIS OF NEUREGULIN1 ISOFORM HRG-GAMMA PROTEIN IN SCHIZOPHRENIA USING INSIGHT II AND GOLD.

A
Thesis Submitted to The
Orissa University of Agriculture and Technology, Bhubaneswar
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Bioinformatics.

BY

GOPAL KRUSHNA BHOI
Adm. No. : 27BI/07

DEPARTMENT OF BIOINFORMATICS
CENTRE FOR POST GRADUATE STUDIES
ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY.
BHUBANESWAR,
2009

Thesis Advisor : Mr. Sukanta Kumar Pradhan

Dedicated to my Beloved Parents, Family and Friends

ORISSA UNIVERSITY OF AGRICULTURE & TECHNOLOGY
DEPARTMENT OF BIOINFORMATICS
CENTRE FOR POST GRADUATE STUDIES
BHUBANESWAR, ORISSA.

Mr. Sukanta Kumar Pradhan
Assistant Professor.

Date : 16.07.09

CERTIFICATE-I

This is to certify that the thesis entitled **"Insilico Modeling, Docking and Analysis of Neuregulin1 isoform HRG-gamma Protein in Schizophrenia using InsightII and GOLD"** submitted for the Degree of **Master of Science** in the subject for **Bioinformatics** embodies a faithful bonafide research work carried out by **Gopal Krushna Bhoi (Adm. No. : 27BI/07)** under my guidance and supervision and no part of this thesis has been submitted by him for any other degree or diploma.

I further certify that any help or information received during the course of investigation have been duly acknowledged by him.

SKpradhan
16/07/09

Place : Bhubaneswar
Date : 16.07.09

Mr. Sukanta Kumar Pradhan
Chairman, Advisory Committee

CERTIFICATE-II

This is to certify that the dissertation entitled **"Insilico Modeling, Docking and Analysis of Nuregulin1 isoform HRG-gamma Protein in Schizophrenia using InsightII and GOLD "** submitted by **Gopal Krushna Bhoi (Adm. No. : 27BI/07)** to the Orissa University of Agriculture and Technology, Bhubaneswar-751003 in the partial fulfillment of the requirements award for the degree of **Master of Science in Bioinformatics** has been approved by the student advisory committee after an oral examination of the same in collaboration with external examiner.

ADVISORY COMMITTEE

1. **Mr.Sukant Kumar Pradhan**
   Asst. Prof. Dept. of Bioinformatics.
   Chairman skpradhan 22/8/09

2. **Dr. P. N. Jagadev**
   Head, Dept. of Bioinformatics.
   Member 22/8/09

3. **Mr. R. R. Patra**
   Asst. Prof., Dept. of C.S.A.
   Member 22/8/09

External Examiner

Signature

Date : 22/8/09

ACKNOWLEDGEMENTS

I avail this priveledge to acknowledge the persons ,whose help & support has made me to complete this work succesfully.

It gives me great pleasure to express my sincere gratitude to my advisor **Mr. Sukanta Kumar Pradhan**, Asst. Professor, chairman-cum advisory committee, Dept. of Bioinformatics ,Orissa University of Agriculture and Technology, Bhubaneswar for his guidance, kind cooperation and good suggestions to complete my project.

Its my pleasure to thank member of the advisory committee **Dr. P. N. Jagadev**, H.O.D., Dept Of Bioinformatics, OUAT, Bhubaneswar who ever been source of inspiration and support in completing my project.

My hearty thanks to member of advisory committee **Mr. Rashmi Ranjan Patra**, Asst. Professor, dept. of Computer Science and Application.

My heartfelt thanks to my guide **Dr. N. Pradhan**, Professor, Dept. of Psychopharmacology, **National Institute of Mental Health and Neuro Science (NIMHANS)** Bangalore for providing me an opportunity to work under his guidance. I am grateful to him for all the encouragement, unrelenting support and fatherly concern during the period of my project.

I extend my hearty thanks and profound gratitude to **Dr. H. K. Senapati,** Dean PGF-cum-DRI for his support during my entire course of action.

I express my indebtedness to **Mr. K. Sankaracharya**, computer programmer for his timely help regarding software's and network.

I express my courteous gratitude to faculty members **Mr. Suryanarayan Rath, Ms. Sushma Rani Martha, Ms. Sucharita Balbantray, Ms. Sudipta Mohanty** and **Ms. Leepika Panda** for their kind support and valuable suggestions during my research work.

I am also obliged to the **Orissa University of Agriculture and Technology** for the kind help and support in providing the necessary instruments to carry out my research work.

I am sincerely indebted to all my **friends and co-trainees of NIMHANS** for their moral support, help and valuable suggestion through out my works.

I am very thankful to my friends **Bishwaranjan, Sabyasachi, Bikash, Yagyan and Asutosh** for their encouragement, company, valuable suggestions and support during the project.

At last, I take this opportunity to thank my **parents** who has shaped my life till now and family members for their silent sacrifices, constant encouragement, forbearance and blessings.

I thank once again all those who have helped and supported me directly or indirectly in completing this project work.

Gopal Krushna Bhoi

Gopal Krushna Bhoi

16.07.2009

Department of Bioinformatics
Centre for Post Graduate Studies.
O.U.A.T., Bhubaneswar.

Name of the student	**Gopal Krushna Bhoi.**
Admission No.	**27BI/07.**
Title of the Thesis	**Insilico Modeling, Docking and Analysis of Nuregulin1 isoform HRG-gamma Protein in Schizophrenia using InsightII and GOLD.**
Degree for which thesis is submitted	**Master of Science in Bioinformatics.**
Name of Dept., College and University	**Department of Bioinformatics Centre for Post Graduate Studies, Orissa University of Agriculture and Technology, Bhubaneswar.**
Year of submission	**2009.**
Name of Advisor	**Mr. Sukanta Kumar Pradhan.**

ABSTRACT

The field of Bioinformatics is growing at a phenomenal rate in the post-genomic era and the task in the early years of the millennium is to demonstrate how in-silico simulations facilitate experiments in the laboratories and how this knowledge can be applied in curing human diseases. This thesis elucidates the attempt of docking and modeling computationally the Neuregulin1 isoform HRG-gamma protein with ligand Ziprasidone using the softwares InsightII and GOLD. Due to abundant sequence information available from genome projects, an increasing number of structurally unknown proteins, homologous of known 3D structure will be discovered as new targets for drug design. Schizophrenia is a disease which affects the brain. Genetic variation in NRG1 isoform HRG gamma is associated with susceptible to Schizophrenia. Mutation in this protein contributes to Schizophrenia. Schizophrenia has been shown to affect the health-related quality of life to an extent of other chronic diseases such as Alzheimer's disease, Parkinson's disease and Epilepsy. It is essential to build the structure of this peptide and development of new drugs in order to control the disease. The Neuregulin1 isoform

HRG-gamma protein sequence for the receptor is taken and a template is searched using BLAST P. The three dimensional protein is constructed using the template by means of homology procedures of InsightII software. The models that were obtained validated to know which model was the best i.e. model with high score and minimum energy using Verify 3D tool. The energy of the model is further minimized by InsightII to reach its stable conformational state. The Docking analysis of this modeled structure was carried out by the Insight II and GOLD. A new set of homology models and docking results are obtained and these models can be used in future for structure based drug design.

Gopal Krushna Bhoi
16.07.2009

Mr. Sukanta Kumar Pradhan

(Advisor)

Gopal Krushna Bhoi

(Author)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

BLAST	Basic Local Alignment Search Tool.
CADD	Computer Added Drug Design.
DDBJ	DNA Data Base of Japan.
EMBL	European Molecular Biology Laboratory.
ExPasy	Expert Protein Analysis System.
FF	Force Field.
FDA	Food and Drug Administration.
GOLD	Genetic Optimization for Ligand Docking.
GOR IV	Garnier Osguthorpe and Robson-IV.
MSA	Multiple Sequence Alignment.
NRG1	Neuregulin1.
NCBI	National Centre for Biotechnology Information.
NLM	National Library of Medicine.
NMR	Nuclear Magnetic Resonance.
PDB	Protein Data Bank.
ProCheck	Protein Check.
RCSB	Research Collaboratory for Structural Bioinformatics.
SCR	Structurally Conserved Regions.
SNP	Single Nucleotide Polymorphism.
SOPMA	Self Optimized Prediction Method with Alignment.

CHAPTER-I

Introduction

INTRODUCTION

In the recent years there has been a growing interest in computer based screening. One of the driving force has been the increased efficiency of protein crystallography leading to the real possibility of using structure based design as a significant contributor to the discovery of novel ligands. The aspect of molecular modeling dealing with receptor and ligand has undergone a drastic transformation in the area of drug designing. In the series of molecular modeling experiments a number of fundamental questions that need exploration at a quantum chemical level of atomic binding. At present computer aided drug designing (CADD) has replaced classical medical chemistry. Further, molecular modeling has opened new avenues to understand the structure of bio-molecules, drugs, protein-ligand interaction and de-novo modeling. The recent development focuses on our ability to access the information content in biological macromolecules. Information stored in the structure of the molecules is a function of their physical and chemical properties. The more important path breaking development has been the ability to manipulate this information by virtue of understanding the structure of nucleic acids, polysaccharides and protein-ligand interaction.

Modeling the interaction of a drug with its receptor is a complex problem. The forces involved are intermolecular association, hydrophobic, Vander Walls interaction, hydrogen bonding and electrostatic forces. The major driving forces for ligand building appear to be controlled by hydrogen bonding and electrostatic interaction. Modeling the intermolecular interaction in the ligand-protein complex is termed difficult due to the involvement of many degrees of freedom in a system by insufficient knowledge of the effect of solvent on the binding association. The interaction of the drug at the receptor site is stereo specific in that the receptor recognizes certain groups on the ligand, particularly the intermolecular distances and molecular shape. The interaction is determined by the fit of the drug molecule to the receptor site that further induces a common biological response.

Thus the study of interaction between the receptor and the ligand, their 3D structures is essential. In the absence of experimentally determined structure, homology modeling plays an effective role in elucidating the structure.

1.1 Proteins

Proteins are large organic compounds made of amino acids arranged in a linear chain and joined together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. The sequence of amino acids in a protein is defined by a gene and encoded in the genetic code like other biological macromolecules such as polysaccharides and nucleic acids, proteins are essential parts of organisms and participate in every process within cells. Many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle. Proteins are also necessary in animal's diets, since animals cannot synthesize all the amino acids they need and must obtain essential amino acids from food. Through the process of Digestion, animals break down ingested protein into free amino acids that are then used in metabolism.

The word protein comes from the Greek word "prota", meaning "of primary importance". Proteins were first described and named by Swedish chemist Jons Jacob Berzelius in 1838. However, central role of proteins in living organisms was not fully appreciated until 1926, when James B Sumner showed that the enzyme urease was a protein. The first protein to be sequenced was insulin, by Frederick Sanger, who won the Nobel Prize for the achievement in 1958. The protein structure to be solved included hemoglobin and myoglobin, by Max Perutz and sir John Cowdery Kendrew respectively in 1958. The three dimensional structures of both proteins were first determined by X-ray diffraction analysis, Perutz and Kendrew shared the 1962 Nobel Prize for these discoveries.

Proteins are fundamental components all living cells and include many substances such as enzymes, hormones and antibodies that are necessary for proper

functioning of an organism. Proteins are essential building blocks for living system hence their name, drawn from Greek Proteios or "Holding first place".

Protein Structure

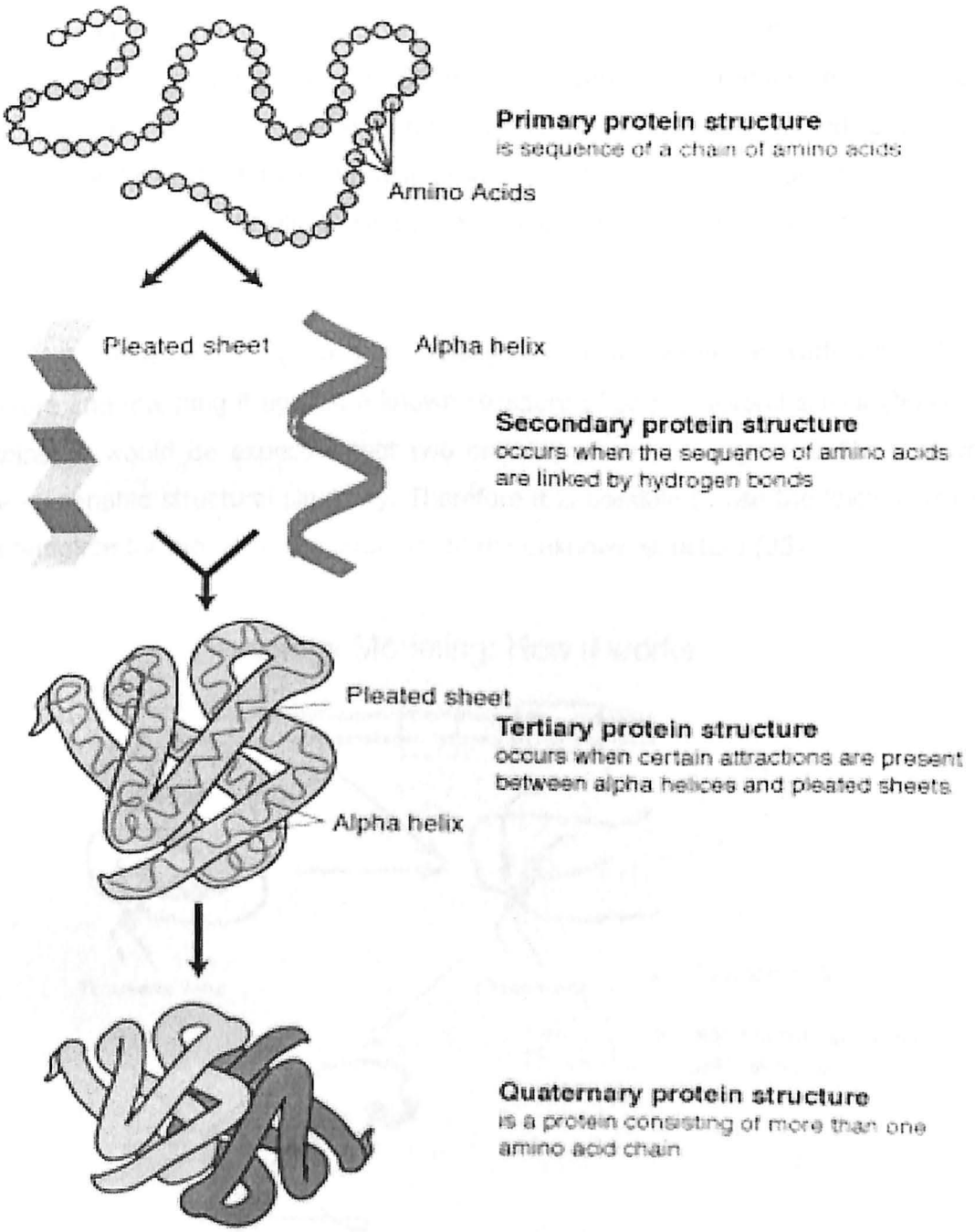


Fig.1 : Protein structure

1.2 Homology Modeling

Homology modeling, also known as comparative modeling, is a class of methods in protein structure prediction for constructing an atomic-resolution model of a protein from its amino acid sequence (the "query sequence" or "target"). Almost all homology modeling techniques rely on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity.

Homology modeling involves taking a known sequence with an unknown structure and mapping it against a known structure of one or several similar (homologs) proteins. It would be expected that two proteins of similar origin and function would have reasonable structural similarity. Therefore it is possible to use the known structure as a template for modeling the structure of the unknown structure.[25]

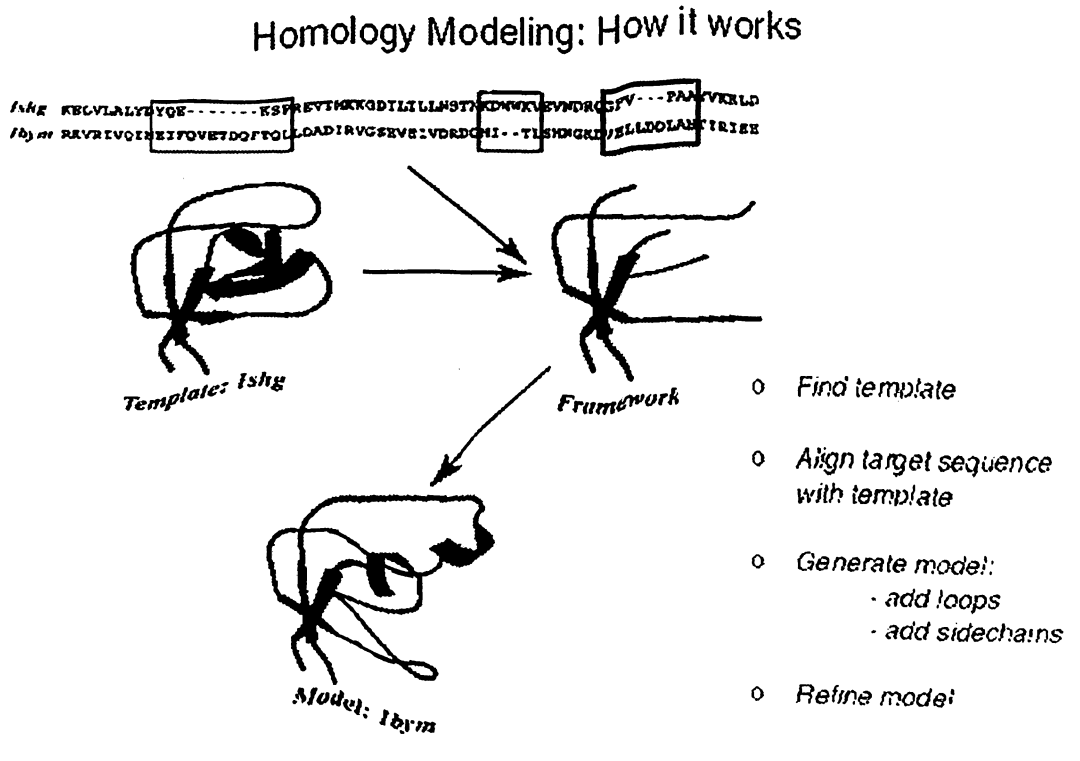


Fig 2 : Homology Modeling.

Modeling Overview

Differences between three-dimensional structures increase with decreasing sequence identity and accordingly the accuracy of models built by homology will also decreases. The errors in a model built on the basis of a structure with >90% sequence identity may be as low as the errors in crystallographically determined structures, except for a few individual side chains. If, as a test case, a known structure is built from another known structure, then in case of 50% sequence identity the RMS error in the modeled coordinates can be as large as 1.5 Angstrom, with considerably larger local errors. If the sequence identity is only around 25% the alignment is the main bottleneck for model building by homology, and large errors are often observed. With less than 25% sequence identity the homology often remains undetected. At present most model building by homology protocols start from the assumption that, except for the insertions and deletions, the backbone of the model is identical to the backbone of the structure. In practice, however, domain motions and 'bending' of parts of molecules with respect to each other is often seen. Even in case of significant bending short range interactions will not differ very much and the model will be perfectly adequate for rational protein engineering, etc. However, the prediction of local differences in the backbone between structures that are homologous in sequence still requires much research, some aspects of which will be described below.

General Procedures

The steps to creating a homology model are as follows:

1. Identify homolog protein and determine the extent of their sequence similarity with one another and the unknown.
2. Align the sequences.
3. Generate coordinates for core (structurally conserved) residues of the unknown structure from those of the known structure.
4. Generate conformations for the loops (structurally variable) in the unknown structure.
5. Refine and evaluate the unknown structure.

1.3 Molecular modeling

Molecular modeling is a collection of computer based techniques for deriving, representing and manipulating the structures and reactions of molecules, and those properties that are dependent on these three dimensional structures. In a natural way molecular modeling treats the molecule as a collection of weights connected with springs, where weights represent the nuclei and the springs represent the bonds. Using professional programs and super computers one can search for the most stable structure of a given molecule (Energy minimization) or stimulate its oscillations at any given temperature (Molecular dynamics). These techniques are useful for refinement of conformations of molecules determined by physical methods. It is far challenging to employ these methods to predict the conformation, Dynamics and properties of molecules whose structure is not yet determined by independent methods. Molecular modeling provides a possibility of screening for those modifications, which are most likely to be successful. Molecular modeling is very useful in predicting the alignment of molecules in complexes. It is most important in designing new drugs, which have to interact in desirable way with, for instance, cell receptors.

1.4 Molecular Docking

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions.

Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs. Given the biological and pharmaceutical significance of molecular docking, considerable efforts have been directed towards improving the methods used to predict docking.[21]

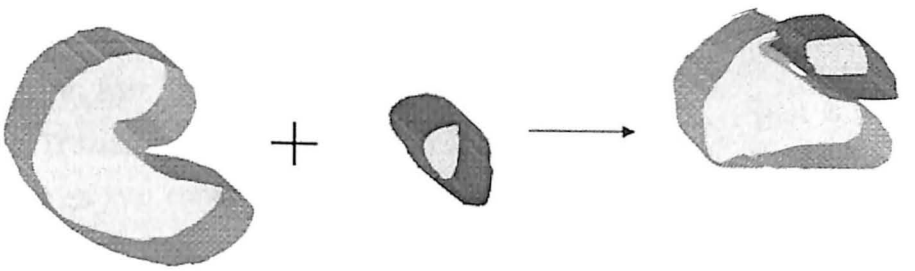


Fig 3 : Protein-Ligand Docking.

A vast number of essential roles that protein plays require small molecules to bind to specific spots in the protein structure. Obtaining the atomic level details of the protein ligand interaction is a valuable tool in the development of novel pharmaceuticals. Three-dimensional molecular structure is one of the foundations of structure-based drug design. Often, data are available for the shape of a protein and a drug separately, but not for the two together. Docking is the process by which two molecules fit together in 3D space. [20]

Molecular docking is used to predict structure of the inter molecular complex formed between two or more molecules. The most interesting case is the protein-ligand interaction, because of its Application in Medicine. Ligand is a small molecule which interacts with protein binding sites. Binding sites are areas of protein known to be active of formatting of compounds. There are called binding modes. There are various forces involved in protein-ligand interaction. Protein ligand interaction involve one or more of the following types of bonding :

1. **Covalent Bonding.** The stabilities of these types of bonds hardly permit the formation of an easily reversible protein receptor complex. Only when the receptor is inactivated by an irreversible antagonist , there is the Formation of covalent bond.
2. **Hydrogen Bonding.** An important type of bonding between drug and receptor is a weak and easily Brocken H-bond. Since many proteins contain hydroxy1, amino, carboxy1 and carbonyl groups, they can form hydrogen Bonds with the receptor complex.
3. **Electrostatic Bonding.** The charged ions produced by the protein molecules may be attached to charged group within a receptor site.
4. **Dipole-Dipole Interactions.** Dipole forms a difference in the electro negativity of covalent bonded atoms. Alignment of dipoles is important. It contributes -1 to -5

Kcal/mole to protein ligand interaction. Amino acids with prominent dipole moments are Ser, Thr, Asn, Tyr, Cys, His.

5. **Charge Transfer Complexes**. When a molecule (or group) that is good electron donor comes into contact with a molecule or group, that is good electron acceptor, the donor may transfer some of its charge to its acceptor. This forms a charge transfer complex, which in effect is molecular dipole-dipole interaction.
6. **Hydrophobic Forces**. In presence of a non-polar molecule or a region of a molecule, the surrounding water molecules are around. When two non polar groups such as lipophilic group on a protein and non polar receptor group, each surrounded by ordered water molecules, become discovered in an attempt to associate with each other. This increase in entropy therefore results in a protein-ligand complex.
7. **Vander Walls Forces**. Vander Wall's bonds exist between all atoms, even those of noble gasses, and are based on polarisability or the induction of asymmetry in the electron cloud of an atom by a nucleus of a Neighboring atom.

1.5 Schizophrenia Disease

Schizophrenia is a chronic, severe, and disabling brain disease. Schizophrenia is a collection of mental and behavioral phenomena, a clinical syndrome. It is a serious mental illness that affects how a person thinks, feels, and behaves. The person finds it difficult to tell the difference between real and imagined experiences, to think logically, to express feelings, or to behave appropriately. Schizophrenia is a brain disorder that has been recognized throughout recorded history. It affects about 2% of Americans. People with schizophrenia may hear voices other people don't hear or they may believe that others are reading their minds, controlling their thoughts, or plotting to harm them. These experiences are terrifying and can cause fearfulness, withdrawal, or extreme agitation. People with schizophrenia may not make sense when they talk, may sit for hours without moving or talking much, or may seem perfectly fine until they talk about what they are really thinking. Because many people with schizophrenia have difficulty holding a job or caring for themselves, the burden on their families and society is significant as well.[16]

The History of Schizophrenia

The word "schizophrenia" is less than 100 years old. However the disease was first identified as a discrete mental illness by Dr. Emile Kraepelin in the 1887 and the illness itself is generally believed to have accompanied mankind through its history.

Written documents that identify Schizophrenia can be traced to the old Pharaonic Egypt, as far back as the second millennium before Christ. Depression, dementia, as well as thought disturbances that are typical in schizophrenia are described in detail in the Book of Hearts. The Heart and the mind seem to have been synonymous in ancient Egypt.

Cause

Experts now agree that schizophrenia develops as a result of interplay between biological predisposition (for example, inheriting certain genes) and the kind of environment a person is exposed to. These lines of research are converging: brain development disruption is now known to be the result of genetic predisposition and environmental stressors early in development (during pregnancy or early childhood), leading to subtle alterations in the brain that make a person susceptible to developing schizophrenia. Environmental factors later in life (during early childhood and adolescence) can either damage the brain further and thereby increase the risk of schizophrenia, or lessen the expression of genetic or Neuro developmental defects and decrease the risk of schizophrenia. In fact experts now say that schizophrenia (and all other mental illness) is caused by a combination of biological, psychological and social factors, and this understanding of mental illness is called the bio-psycho-social model.

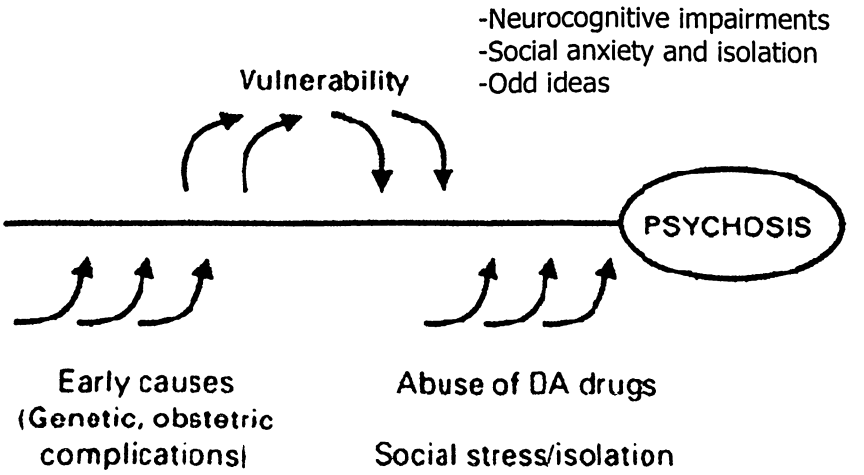


Fig 4 : Developmental Origins Of Schizophrenia.

The Path to Schizophrenia - The diagram above shows how biological, genetic and prenatal factors are believed to create a vulnerability to schizophrenia. Additional environmental exposures (for example, frequent or ongoing social stress and/or isolation

during childhood, drug abuse, etc.) then further increase the risk or trigger the onset of psychosis and schizophrenia.

Types of Schizophrenia

Paranoid schizophrenia : These persons are very suspicious of others and often have grand schemes of persecution at the root of their behavior. Hallucinations, and more frequently delusions, are a prominent and common part of the illness.

Disorganized schizophrenia : In this case the person is verbally incoherent and may have moods and emotions that are not appropriate to the situation. Hallucinations are not usually present.

Catatonic schizophrenia : In this case, the person is extremely withdrawn, negative and isolated, and has marked psychomotor disturbances.

Residual schizophrenia : In this case the person is not currently suffering from delusions, hallucinations, or disorganized speech and behavior, but lacks motivation and interest in day-to-day living.

Schizoaffective disorder : These people have symptoms of schizophrenia as well as mood disorder such as major depression, bipolar mania, or mixed mania.

Undifferentiated Schizophrenia : Conditions meeting the general diagnostic criteria for schizophrenia but not conforming to any of the above subtypes, or exhibiting the features of more than one of them without a clear predominance of a particular set of diagnostic characteristics.

Symptoms of Schizophrenia

The symptoms of schizophrenia fall into three broad categories:

- **Positive symptoms** are unusual thoughts or perceptions, including hallucinations, delusions, thought disorder, and disorders of movement.
- **Negative symptoms** represent a loss or a decrease in the ability to initiate plans, speak, express emotion, or find pleasure in everyday life. These symptoms are harder to recognize as part of the disorder and can be mistaken for laziness or depression.

- **Cognitive symptoms** (or cognitive deficits) are problems with attention, certain types of memory, and the executive functions that allow us to plan and organize. Cognitive deficits can also be difficult to recognize as part of the disorder but are the most disabling in terms of leading a normal life.

Diagnosis

The first step in getting treatment for schizophrenia is getting a correct diagnosis. Some of the symptoms that occur in schizophrenia also occur in other mental health conditions such as depression, mania, or after taking some 'street' drugs. Therefore, the diagnosis may not be clear at first. As a rule, the symptoms need to be present for several weeks before a doctor will make a firm diagnosis of schizophrenia.

Antipsychotic Medications

Antipsychotic medications have been available since the mid-1950s. They effectively alleviate the positive symptoms of schizophrenia. While these drugs have greatly improved the lives of many patients, they do not cure schizophrenia.

Everyone responds differently to antipsychotic medication. Sometimes several different drugs must be tried before the right one is found. People with schizophrenia should work in partnership with their doctors to find the medications that control their symptoms best with the fewest side effects. The older antipsychotic medications include chlorpromazine, haloperidol, perphenazine, and fluphenzine. The older medications can cause extrapyramidal side effects, such as rigidity, persistent muscle spasms, tremors, and restlessness. In the 1990s, new drugs, called atypical antipsychotics, were developed that rarely produced these side effects. The first of these new drugs was clozapine. It treats psychotic symptoms effectively even in people who do not respond to other medications, but it can produce a serious problem called agranulocytosis, a loss of the white blood cells that fight infection. Some of the drugs that were developed after clozapine was introduced such as risperidone, olanzapine, quietiapine, sertindole, and ziprasidone are effective and rarely produce extrapyramidal symptoms and do not cause agranulocytosis but they can cause weight gain and metabolic changes associated with an increased risk of diabetes and high cholesterol. No one can tell beforehand exactly how a medication will affect a particular individual, and sometimes several medications must be tried before the right one is found.[12]

Psychosocial Treatment

Numerous studies have found that psychosocial treatments can help patients who are already stabilized on antipsychotic medications deal with certain aspects of schizophrenia, such as difficulty with communication, motivation, self-care, work, and establishing and maintaining relationships with others. Learning and using coping mechanisms to address these problems allows people with schizophrenia to attend school, work, and socialize. Patients who receive regular psychosocial treatment also adhere better to their medication schedule and have fewer relapses and hospitalizations. A positive relationship with a therapist or a case manager gives the patient a reliable source of information, sympathy, encouragement, and hope, all of which are essential for for managing the disease. The therapist can help patients better understand and adjust to living with schizophrenia by educating them about the causes of the disorder, common symptoms or problems they may experience, and the importance of staying on medications.

Available treatments can relieve many of the disorder's symptoms, but most people who have schizophrenia must cope with some residual symptoms as long as they live. Nevertheless, this is a time of hope for people with schizophrenia and their families. Many people with the disorder now lead rewarding and meaningful lives in their communities. Researchers are developing more effective medications and using new research tools to understand the causes of schizophrenia and to find ways to prevent and treat it.[12]

OBJECTIVE

This study focuses on building molecular model of Neuregulin1 isoform HRG-gamma using molecular modeling techniques. This study also addresses the ligand-receptor interaction. This preliminary investigation of receptor-ligand can be exploited in the future theoretical designing on agonist and antagonist.

The objective of this work is :

- To model a 3-D structure of Neuregulin1 isoform HRG-gamma protein.
- Docking analysis of the protein using Ziprasidone ligand.

CHAPTER-II

Review of Literature

REVIEW OF LITERATURE

2.1 Neuregulin1

Neuregulin1 (NRG1) is a leading schizophrenia susceptibility gene. Neuregulin1 is one of four proteins in the neuregulin family that act on the EGFR family of receptors. Neuregulin1 is produced in numerous isoforms by alternative splicing, which allows it to perform a wide variety of functions. It is essential for the normal development of the nervous system and the heart.

A series of genetic mapping, gene expression and animal studies implicate Neuregulin1 (NRG1) mediated signaling in the neuro-pathogenesis of the brain disorder schizophrenia (SZ). Understanding the molecular disease pathways involved is a major research focus in Schizophrenia. A recent report by Sei and colleagues (2008) identifies an interesting method of investigating the role of the NRG1 gene in vivo using transformed B lymphoblasts from patients with Schizophrenia. NRG1 gene plays critical roles involving many aspects of neurodevelopment and this study focuses on a readily measurable phenotype, cell migration. This is of interest as a substantial, if controversial, literature suggests that early defects in cell migration may be relevant to Schizophrenia susceptibility.[15]

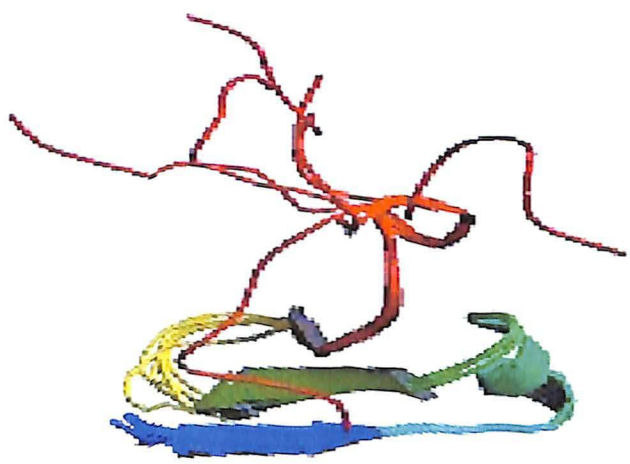


Fig 5 : NRG1 Gene.

Function

Neuregulin1 (NRG1) was originally identified as a 44-kD glycoprotein that interacts with the NEU/ERBB2 receptor tyrosine kinase to increase its phosphorylation on tyrosine residues. It is known that an extraordinary variety of different isoforms are produced from the NRG1 gene by alternative splicing. These isoforms include heregulins (HRGs), glial growth factors (GGFs) and sensory and motor neuron-derived factor (SMDF). They are tissue-specifically expressed and differ significantly in their structure. The HRG isoforms all contain immunoglobulin (Ig) and epidermal growth factor-like (EGF-like) domains. GGF and GGF2 isoforms contain a kringle-like sequence plus Ig and EGF-like domains; and the SMDF isoform shares only the EGF-like domain with other isoforms. The receptors for all NRG1 isoforms are the ERBB family of tyrosine kinase transmembrane receptors. Through interaction with ERBB receptors, NRG1 isoforms induce the growth and differentiation of epithelial, neuronal, glial, and other types of cells.

Neuregulin1 (NRG1) is a leading schizophrenia susceptibility gene. The NRG1 locus on chromosome 8p shows linkage to the disorder, and genetic association has been found between schizophrenia and various non-coding polymorphisms and haplotypes, especially at the 5′ end of the NRG1 gene, in many but not all case-control and family studies. NRG1 is a pleiotropic growth factor, important in nervous system development and functioning; roles include the modulation of neuronal migration, synaptogenesis, gliogenesis, neuron-glia communication and neurotransmission. Understanding the neurobiology of NRG1 and its involvement in schizophrenia is challenged by the complexity of the gene, which gives rise to multiple functionally distinct isoforms, including six "types" of NRG1 defined by 5′ exon usage. Type IV and type I NRG1 may be particularly relevant to schizophrenia, with initial data showing altered expression of these isoforms in the disorder or in association with NRG1 risk alleles. We review the structure and functions of NRG1, consider the evidence for and against it being a schizophrenia susceptibility gene, and discuss mechanisms that might underlie the contribution of NRG1 to disease pathophysiology.

Clinical significance

Neuregulin1 interactions are thought to play a role in the pathological mechanism of schizophrenia. A high-risk deCODE (Icelandic) haplotype was discovered in 2002 on

the 5'-end of the gene. The SNP8NRG243177 allele from this haplotype was associated in 2006 with a heightened expression of the Type IV NRG1 in the brains of people suffering from schizophrenia. The protein also has the putative ability to protect the brain from damage induced by stroke.

2.1.1 Neuregulin1: A target for schizophrenia

Approximately 2% of the population develops schizophrenia during their lifetime, making this disease more prevalent than Alzheimer's disease, diabetes, or multiple sclerosis. Patients report a variety of symptoms such as hearing internal voices not heard by others, or believing that other people are reading their minds, controlling their thoughts, or plotting to harm them. These symptoms result in fear and withdrawal. In the USA, patients with schizophrenia occupy about 25% of all hospital beds and account for about 20% of all social security disability days. 10% of patients take their own lives. There is a significant genetic component to schizophrenia and genomic analysis may offer one approach to the identification of new treatment options. DeCODE Genetics have adopted this approach and as a result neuregulin 1 (NRG1) was identified as a candidate gene for schizophrenia. NRG1 is expressed at central nervous system synapses and has a clear role in the expression and activation of neurotransmitter receptors, including glutamate receptors that are implicated in schizophrenia. NRG1 and its receptor therefore deserve further attention in relation to the treatment of schizophrenia.[1]

2.1.2 Neuregulin1 and susceptibility to schizophrenia

The cause of schizophrenia is unknown, but it has a significant genetic component. Pharmacologic studies, studies of gene expression in man, and studies of mouse mutants suggest involvement of glutamate and dopamine neurotransmitter systems. However, so far, strong association has not been found between schizophrenia and variants of the genes encoding components of these systems. NRG1 is expressed at central nervous system synapses and has a clear role in the expression and activation of neurotransmitter receptors, including glutamate receptors. Mutant mice heterozygous for either NRG1 or its receptor, ErbB4, show a behavioral phenotype that overlaps with mouse models for schizophrenia. Furthermore, NRG1 hypomorphs have fewer functional NMDA receptors than wild-type mice. Here it demonstrate that the behavioral

phenotypes of the NRG1 hypomorphs are partially reversible with clozapine, an atypical antipsychotic drug used to treat schizophrenia.[2]

2.1.3 Neuregulin1 Genotype and Schizophrenia

The neuregulin1 (NRG1) gene has been the subject of considerable excitement within the psychiatric genetics literature since it was originally identified as a potential susceptibility locus for schizophrenia. Case-control and family-based genetic association studies of the NRG1 gene in healthy control groups and clinically diagnosed schizophrenia patients were included. The evidence for association of the SNP8NRG221533 polymorphism continued to be non-significant. It discuss a number of problems in the interpretation of a disparate and inconsistent gene-disease association literature, including the difficulties associated with determining what constitutes replication across studies which vary in their methods, marker sets employed, phenotype definition, and other study characteristics.[3]

2.1.4 Neuregulin1 transcripts are differentially expressed in Schizophrenia and regulated by 5' SNPs associated with the disease

Genetic variation in neuregulin1 (NRG1) is associated with schizophrenia. The disease-associated SNPs are noncoding, and their functional implications remain unknown. We hypothesized that differential expression of the NRG1 gene explains its association to the disease. Here examined four of the disease-associated SNPs that make up the original risk haplotype in the 5' upstream region of the gene for their effects on mRNA abundance of NRG1 types I–IV in human postmortem hippocampus. Diagnostic comparisons revealed a 34% increase in typeI mRNA in schizophrenia and an interaction of diagnosis and genotype (SNP8NRG221132) on this transcript. Of potentially greater interest, a single SNP within the risk haplotype (SNP8NRG243177) and a 22-kb block of this core haplotype are associated with mRNA expression for the novel type IV isoform in patients and controls. Bioinformatics promoter analyses indicate that both SNPs lead to a gain/loss of putative binding sites for three transcription factors, serum response factor, myelin transcription factor-1, and High Mobility Group Box Protein-1. These data implicate variation in isoform expression as a molecular mechanism for the genetic association of NRG1 with schizophrenia.[4]

2.1.5 Neuregulin1-induced cell migration is impaired in schizophrenia

Neuregulin1 (NRG1), a candidate susceptibility gene for schizophrenia, plays a critical role in neuronal migration and central nervous system development. However, its relation to schizophrenia pathogenesis is unknown. B lymphoblasts migrate to NRG1 through the ErbB-signaling system as observed in neuronal cells. It assessed NRG1-induced cell migration in B lymphoblasts from patients with schizophrenia and found that NRG1-induced migration is significantly decreased compared with control individuals in two independent cohorts. This impaired migration is related at least in part to reduced phosphorylation in the patients. Moreover, the magnitude of NRG1-induced migration is associated with polymorphisms of the NRG1 and catechol-o-methyltransferase genes and with an epistatic interaction of these genes. This study demonstrates that the migratory response of schizophrenia-derived cells to NRG1 is impaired and is associated with genetic variations in more than one schizophrenia susceptibility gene, providing a novel insight into potential neurodevelopmental mechanisms of schizophrenia.[5]

2.1.6 Neuregulin1 in neural development, synaptic plasticity and schizophrenia

Schizophrenia is a highly debilitating mental disorder that affects 1% of the general population, yet it continues to be poorly understood. Recent studies have identified variations in several genes that are associated with this disorder in diverse populations, including those that encode neuregulin 1 (NRG1) and its receptor ErbB4. The past few years have witnessed exciting progress in our knowledge of NRG1 and ErbB4 functions and the biological basis of the increased risk for schizophrenia that is potentially conferred by polymorphisms in the two genes. An improved understanding of the mechanisms by which altered function of NRG1 and ErbB4 contributes to schizophrenia might eventually lead to the development of more effective therapeutics.[6]

2.1.7 Impact of neuregulin1 on the pathophysiology of schizophrenia in human post-mortem studies

To a large extend schizophrenia has been shown to be heritable, with neuregulin1 (NRG1) one of the candidate genes considered to play a role in the pathophysiology of the disorder. While several polymorphisms within this gene have

been reported to be associated with schizophrenia, the impact of NRG1 risk genotypes on disturbed brain function and symptoms of the disease is unknown and might be elucidated using post-mortem studies. Neuregulins are signalling proteins and the NRG1 family encodes at least 15 different splice variants, classified into four isoforms. They play an important role in cell differentiation, migration, myelination and proliferation of oligodendrocytes and neurons. Dysfunction in these processes may be related to neurodevelopmental disturbances in schizophrenia. NRG1· isoforms are differentially expressed in relevant brain regions of schizophrenia patients such as the prefrontal cortex and hippocampus and may contribute to pathophysiological processes. Different NRG1 genotypes have been shown to influence gene expression of isoforms and the risk-associated variants are in primarily non-coding and promoter regions, probably operating by altering gene expression or splicing.[8]

2.1.8 Genetic variation in the schizophrenia-risk gene neuregulin1 correlates with brain activation and impaired speech production in a verbal fluency task in healthy individuals

Impaired performance in verbal fluency tasks is an often replicated finding in schizophrenia. In functional neuroimaging studies, this dysfunction has been linked to signal changes in prefrontal and temporal areas. Since schizophrenia has a high heritability, it is of interest whether susceptibility genes for the disorder, such as NRG1, modulate verbal fluency performance and its neural correlates. Four hundred twenty-nine healthy individuals performed a semantic and a lexical verbal fluency task. A sub sample of 85 subjects performed an overt semantic verbal fluency task while brain activation was measured with functional magnetic resonance imaging (MRI). NRG1 status was determined and correlated with verbal fluency performance and brain activation. For the behavioral measure, there was a linear effect of NRG1 status on semantic but not on lexical verbal fluency. Performance decreased with number of risk-alleles. NRG1 genotype does influence language production on a semantic level in conjunction with the underlying neural systems. These findings are in line with results of studies in schizophrenia and may explain some of the cognitive and brain activation variation found in the disorder.[9]

2.2 Ziprasidone Ligand

Ziprasidone is an antipsychotic medication. It works by changing the effects of chemicals in the brain. Ziprasidone is used to treat schizophrenia and the manic symptoms of bipolar disorder (manic depression). Ziprasidone may also be used for other purposes not listed in this medication guide.

Ziprasidone was the fifth atypical antipsychotic to gain FDA approval (February 2001). In the United States, Ziprasidone is Food and Drug Administration (FDA) approved for the treatment of schizophrenia, and the intramuscular injection form of ziprasidone is approved for acute agitation in schizophrenic patients. Ziprasidone has also received approval for acute treatment of mania and mixed states associated with bipolar disorder. The brand name Geodon has been suggested to bring to mind the phrase 'down (don) to earth (geo)' referring to the goals of the medication.

The oral form of ziprasidone is the hydrochloride salt, ziprasidone hydrochloride. The intramuscular form, on the other hand, is the mesylate salt, ziprasidone mesylate trihydrate, and is provided as a lyophilized powder.[23]

2.2.1 Mechanism of Action

The mechanism of action of Ziprasidone, as with other drugs used to treat schizophrenia, is unknown. Ziprasidone exhibited high in vitro binding affinity for the dopamine D2 and D3, the serotonin 5HT2A, 5HT2C, 5HT1A, 5HT1D and alpha 1-adrenergic receptors, and moderate affinity for the histamine H1 receptor. However, it has been proposed that the drug's therapeutic activity in schizophrenia is mediated through a combination of dopamine Type 2 (D2) and serotonin Type 2 ($5HT_2$) receptor antagonism.

2.2.2 Pharmacology

Ziprasidone is a psychotropic agent belonging to the chemical class of benzisoxazole derivatives and is indicated for the treatment of schizophrenia. Ziprasidone is a selective monoaminergic antagonist with high affinity for the serotonin Type 2 (5HT2), dopamine Type 2 (D2), 1 and 2 adrenergic, and H1 histaminergic receptors. Ziprasidone acts as an antagonist at other receptors, but with lower potency. Antagonism at receptors other than dopamine and 5HT2 with similar receptor affinities

may explain some of the other therapeutic and side effects of Ziprasidone. Ziprasidone's antagonism of muscarinic M1-5 receptors may explain its anticholinergic effects. Ziprasidone's antagonism of histamine H1 receptors may explain the somnolence observed with this drug. Ziprasidone's antagonism of adrenergic a1 receptors may explain the orthostatic hypotension observed with this drug.

It has a high affinity for dopamine, serotonin, and alpha-adrenergic receptors and a moderate affinity for histamine receptors. Ziprasidone also displays some inhibition of synaptic reuptake of serotonin and norepinephrine, although the clinical significance of this is unknown.[24]

2.2.3 Pharmacokinetics

The systemic bioavailability of ziprasidone administered intramuscularly is 100%, or 60%, administered orally with food. After a single dose intramuscular administration, the peak serum concentration typically occurs at about 60 minutes after the dose is administered, or earlier. Steady state plasma concentrations are achieved within one to three days. The mean half-life ranges from two to five hours. Exposure increases in a dose-related manner and following three days of intramuscular dosing, little accumulation is observed.

2.2.4 Adverse effects

Ziprasidone received a black box warning due to increased mortality in elderly patients with dementia-related psychosis. It also slightly increases the QTc interval in some patients and increases the risk of a potentially lethal type of heart arrythmia known as torsades de pointes. Ziprasidone should be used cautiously in patients taking other medications likely to interact with ziprasidone or increase the QTc interval. Ziprasidone is known to cause activation into mania in some bipolar patients. This medication can cause birth defects, according to animal studies, although this side effect has not been confirmed in humans.

Adverse events reported for ziprasidone include severe chest pains, sedation, insomnia, orthostasis, life-threatening neuroleptic malignant syndrome, akathisia, and the development of permanent neurological disorder tardive dyskinesia. Rarely, temporary speech disorders may result.[24]

CHAPTER-III

Materials & Methods

MATERIALS AND METHODS

3.1 Materials Required

The following materials were used for achieve the objective.

A. Searching of protein and ligand
   - NCBI
   - Uniprot
   - PDB
   - BLAST
   - PubChem
   - Drug Bank

B. Composition of Protein
   - Prot Param

C. Secondary Structure Prediction
   - SOPMA
   - GOR4

D. Multiple Sequence Alignment
   - Clustal W

E. Modeling Protein
   - Insight II

F. Validation of Model
   - Verify 3D
   - Pro Check
   - Rampage Server

G. Docking of Protein
   - Cast P Server
   - GOLD
   - Insight II
   - Patch Dock
   - Rasmol

3.1.1 Searching a Protein and Ligand

NCBI

The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI is located in Bethesda, Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper. The NCBI houses genome sequencing data in Gen Bank and an index of biomedical research articles in Pub Med Central and Pub Med, as well as other information relevant to biotechnology. All these databases are available online through the Entrez search engine.

The NCBI is directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in Bioinformatics. He also leads an intramural research program, including groups led by Stephen Altschul (another BLAST co-author), David Landsman, and Eugene Koonin (a prolific author on comparative genomics.

UniProt

The Universal Protein Resource (UniProt) is the world's most comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR. UniProt is comprised of three components, each optimized for different uses. The UniProt Knowledgebase (UniProtKB) is the central access point for extensive curated protein information, including function, classification, and cross-reference. The UniProt Reference Clusters (UniRef) databases combine closely related sequences into a single record to speed searches. The UniProt Archive (UniParc) is a comprehensive repository, reflecting the history of all protein sequences.

PDB

The Protein Data Bank (PDB) is the single worldwide depository of information about the three-dimensional structures of large biological molecules, including proteins and nucleic acids. These are the molecules of life that are found in all organisms including bacteria, yeast, plants, flies and mice and in healthy as well as diseased humans. Understanding the shape of a molecule helps to understand how it works.

The data, typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world, are released into the public domain, and can be accessed at no charge on the internet. The PDB is overseen by an organization called the Worldwide Protein Data Bank, ww PDB. The RCSB PDB is a portal for information about these molecules, and as such enables research and education about the molecular basis of life. The PDB is available at no cost to all users.

The PDB was established in 1971 at Brookhaven National Laboratory and originally contained 7 structures. In 1998, the Research Collaboratory for Structural Bioinformatics (RCSB) became responsible for the management of the PDB.

BLAST

In Bioinformatics, Basic Local Alignment Search Tool, or BLAST is an algorithm for computing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

PubChem

PubChem is a database of chemical molecules. The system is maintained by the National Center for Biotechnology Information (NCBI), a component of the National Library of Medicine, which is part of the United States National Institutes of Health (NIH). PubChem can be accessed for free through a web user interface. Millions of compound structures and descriptive datasets can be freely downloaded via FTP. PubChem contains substance descriptions and small molecules with fewer than 1000 atoms and 1000 bonds. The American Chemical Society tried to get the U.S. Congress to restrict the operation of PubChem, because they claim it competes with their Chemical Abstracts Service. More than 80 database vendors contribute to the growing PubChem database

PubChem consists of three dynamically growing primary databases.

- Compounds, 19 million entries, contain pure and characterized chemical compounds.
- Substances, 41 million entries, contain also mixtures, extracts, complexes and uncharacterized substances.

- BioAssay, bioactivity results from 1156 high-throughput screening programs with several million values

It has properties like chemical structure, name fragments, chemical formula, molecular weight, XLogP, and hydrogen bond donor and acceptor count. It contains its own online molecule editor with SMILES/SMARTS and InChI support that allows the import and export of all common chemical file formats to search for structures and fragments.

Drug Bank

The DrugBank database available at the University of Alberta is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, pathway) information. The database contains nearly 4800 drug entries including >1480 FDA-approved small molecule drugs, 128 FDA-approved biotech (protein/peptide) drugs, >71 nutraceuticals and >3200 experimental drugs. Additionally, more than 2500 protein (i.e. drug target, non-redundant) sequences are linked to these drug entries. Each DrugCard entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data. It is maintained by David Wishart and Craig Knox. The simple text query (above) supports general text queries of the entire textual component of the database. Clicking on a given DrugCard button brings up the full data content for the corresponding drug. A complete explanation of all the DrugCard fields and sources is given here. The PharmaBrowse button allows users to browse through drugs as grouped by their indication. This is particularly useful for pharmacists and physicians, but also for pharmaceutical researchers looking for potential drug leads. The ChemQuery button allows users to draw (using MarvinSketch applet or a ChemSketch applet) or write a chemical compound and to search DrugBank for chemicals similar or identical to the query compound. The TextQuery button supports a more sophisticated text search the text portion of DrugBank. The SeqSearch button allows users to conduct BLASTP (protein) sequence searches of the 18,000 sequences contained in DrugBank. Both single and multiple sequence (i.e. whole proteome) BLAST queries are supported. The Data Extractor button opens an easy-to-use relational query search tool that allows users to select or search over various combinations of subfields. The Data Extractor is the most sophisticated search tool for Drug Bank.

3.1.2 Composition of Protein

Prot Param

ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity.

ProtParam computes various physico-chemical properties that can be deduced from a protein sequence. No additional information is required about the protein under consideration. The protein can either be specified as a Swiss-Prot/TrEMBL accession number or ID, or in form of a raw sequence. White space and numbers are ignored. If we provide the accession number of a Swiss-Prot/TrEMBL entry, we will be prompted with an intermediary page that allows us to select the portion of the sequence on which we would like to perform the analysis. The choice includes a selection of mature chains or peptides and domains from the Swiss-Prot feature table (which can be chosen by clicking on the positions), as well as the possibility to enter start and end position in two boxes. By default the complete sequence will be analyzed.

3.1.3 Secondary Structure Prediction

SOPMA

SOPMA is a secondary structure prediction method. SOPMA (Self-Optimized Prediction Method with Alignment) is an improvement of SOPM method. These methods are based on the homologue method of Levin et al (1986). SOPMA correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins. Joint prediction with SOPMA and a neural networks method (PHD) correctly predicts 82.2% of residues for 74% of co-predicted amino acids. The improvement takes place in the fact that SOPMA takes into account information from an alignment of sequences belonging to the same family (Geourjon and Deleage, 1995). So it helps in significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments.

GOR4

GOR IV (GARNIER OSGUTHORPE and ROBSON-IV) is the fourth version of GOR secondary structure prediction methods based on the information theory. It was developed by J.Garnier, D.Osguthorpe and B.Robson (J.Mol.Biol. 120,97, 1978). There is no defined decision constant. The GOR method is based on information theory and was developed by J.Garnier, D.Osguthorpe and B.Robson (J.Mol.Biol. 120,97, 1978). The present version, GOR IV, uses all possible pair frequencies within a window of 17 amino acid residues and is reported by J. Garnier. J.F. Gibrat and B.Robson in Methods in Enzymology, vol 266, p 540-553 (1996). After cross validation on a data base of 267 proteins, the version IV of GOR has a mean accuracy of 64.4% for a three state prediction (Q3). The program gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil ;the second output gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one of highest probability compatible with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues. This program was written by Jean-Francois Gibrat and was modified by Stephen Pheiffer in order to increase it's efficiency. These scores are corrected by decision constants and the highest one gives the conformational state for the residue in the center of the window.

3.1.4 Multiple Sequence Alignment

ClustalW

Clustal W is a general purpose multiple sequence alignment program for DNA and Protein. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing cladogram or phylogram. Multiple alignments of protein sequences are important tools in studying sequences. The basic information they provide is the identification of conserved sequence regions. This is very useful in designing experiments to test and modify the function of specific proteins, in predicting the function and structure of proteins and in identifying new members of protein families.

It can be used to produce a multiple sequence alignment. Using the web form the user need only input or upload a file of the sequences that they want to align in an

accepted format. The other options on the form are set to the default values for producing a multiple alignment.

3.1.5 Modeling Protein

InsightII

InsightII is a sophisticated molecular modeling environment that provides a powerful graphical interface to best-of-breed algorithms for molecular dynamics, homology modeling, de novo design, and electrostatics making it the perfect for protein modelers, computational chemists, and structural biologists. InsightII is a comprehensive graphic molecular modeling program. In conjunction with moleculare maechanics/dynamics programs such a Discover or CHARMm, you can use the InsightII program to build and manipulate virtually any class of molecule or molecular system. Inconjunction with other Accelrys products, you can study molecular properties. Building, displaying and studying Molecules are done by issuing commands to the InsightII program.

Most commands create and affect objects. Molecular modeling is a general term that covers a wide range of molecular graphics and computational chemistry techniques used to build, display, manipulate, simulate, and analyze molecular structures and to calculate properties of these structures. Molecular modeling is used in several different research areas, and thereforethe term does not have a rigid definition. To a chemical physicist, molecular modeling might imply performing a high quality quantum mechanical calculation using a supercomputer on a structure with 4 or 5 atoms, to an organic chemost, molecular modeling might mean displaying and modifying a candidate drug molecule on a desktop computer. The criterion for successful modeling experiment should not be how accurately the calculations are performed, but whether they are useful in rationalizing the behavior of the molecule or in enhancing the creativity of the chemist in the design of novel compounds.

Modules in InsightII

The Module pull down contains commands that you use to switch from one module to another. This pull down is accessed by picking the MSI logo in the top corner of the Insight II program's window, and contains the names of available modules.

Builder

The Builder command activates the Builder module. This module allows constructing new molecules from molecular fragments or individual atoms. It also allows modifying such properties as atom type, hybridization, potential function parameters, bond order, and geometry of existing molecules.

Biopolymer

The Biopolymer command activates the Biopolymer module. The commands in this module facilitate the building and modification of peptides, proteins, polynucleic acids, and carbohydrates. In particular the peptide commands can be used to: build up peptide sequences while imposing secondary structure, delete or replace residues in peptides and proteins, impose secondary structure on existing peptides and proteins, and change N and C terminal capping groups.. The nucleic acid commands can be used to: build single, double, or triple-stranded poly nucleotides in A, B, or Z form; delete or replace nucleotides in strands, measure angles and distances between bases, and cap a prime or ligate strand. The carbohydrate commands can be used to link monosaccharides in a number of ways and to change the enantiomer or anomer of a monosaccharide within a molecule. This module includes access to many of the commands from the Builder module, allow to perform builder functionality while in the Biopolymer module.

Delphi

The Delphi module has commands to define Delphi calculation parameters (Setup), perform a Delphi calculation (Run_ Delphi), analyze the results (Potential), and create charge and radius templates (Templates). Delphi itself is a software package which calculates the electrostatic potential in and around macro-molecules, using a finite difference solution to the Poisson-Boltzmann equation. The program allows specification of ionic strength as well as dielectric constants of both the solvent and the molecule of interest. In addition, periodic boundary conditions may be used for molecules or systems with repetitive portions, such as nucleic acids.

Solvation

The Solvation module has been implemented in order to simplify and speed up calculating the solvation energy using Delphi. It involves one or two Delphi runs (based on the parameter set--CFF91, PARSE--used), a Discover run for calculating the intramolecular energy, and a calculation of the total accessible solvent area. In addition,

this module provides several other simple solvation models. Finally, a separate option to calculate the total solvent accessible surface area, using current Insight II atomic radii, is also provided.

Discover

The Discover module provides an interface to the Discover program. The interface allows for the definition of minimization and dynamics calculations. It can perform simulations using various forms of constraints and restraints, including template forcing, torsion forcing, tethering, and NOE. This module also enables to do free energy calculations, and provides options for querying ongoing jobs.

Discover_3

The Discover_3 module provides an interface to the Discover 3.0 program by enabling to specify molecular mechanics simulations using that program. It also provides basic options for querying ongoing jobs. The interface allows for the definition of minimization and dynamics calculations.

Docking

The Docking module provides facilities for calculating the no bond energy between two molecules using explicit vander Waal's energy, explicit electrostatic (Coulombic) energy, or the combination of vander Waal's and electrostatic energies. The number of atoms included in the calculation can be limited by specifying a monomer- or residue-based cutoff. Alternatively, the computation can be done approximately, using a pre computed energy grid.

Homology

The Homology module contains commands to help build a model of a protein given only its amino acid sequence and the three-dimensional structure of at least one other protein. Facilities are provided to find related proteins, to find regions of structural conservation among related proteins, to align amino acid sequences, and to assign coordinates based on these alignments.

Binding Site

The Binding Site module contains methods to characterize a protein binding site of active site.

Analysis

The Analysis module is used to perform molecular conformation analysis. This module provides functionality for analyzing the trajectory data output by the molecular mechanics program Discover (separately licensed from MSI). This functionality may also be used to analyze conformational data produced by other molecular mechanics

programs, or by any other method, provided the data is formatted in a file type recognizable by InsightII. The analysis is performed by defining properties of interest, identifying the atoms that uniquely define the property, and then graphing or tabulating those properties against each other. Four types of properties are available for analysis: total energy, time, distance (including point-plane), and periodic (angles, dihedral angles, and plane-plane angles).

DeCipher

The DeCipher command activates the DeCipher module, which is used to perform molecular conformation analysis. This module provides functionality for analyzing the configuration data output by the molecular mechanics program Discover, separately licensed from MSI. This functionality may also be used to analyze conformational data produced by other molecular mechanics programs, or by any other method, provided the data is formatted in a file type recognizable by InsightII. The analysis is performed by defining properties of interest, identifying the atoms that uniquely define the property, and then graphing or tabulating those properties against each other.

3.1.6 Validation of Model

Programs like Verify 3D, ProCheck and Rampage server are used to analyze the final structure created through homology modeling. Criteria for analysis of correctness include:

1. Dihedral angles
2. Peptide bonds
3. Side Chain conformations
4. Hydrogen bonding
5. Hydrophobic residue

All models built by homology modeling are prone to errors induced during many stages of model building. Models should be checked for the normality of the torsion angles, bond angles and bond length.

Verify 3D

Verify 3D provides a visual analysis of the quality of the putative crystal structure for a protein. An effective test of the accuracy of the 3D protein model is a comparison of the model to its own amino acid sequence, using a 3D profile computed from the

atomic coordinates of the structure. The accuracy of the model can be assumed by its 3D profile, regardless of whether the model has been derived by the X-ray, NMR or Computational methods. Verify 3D works best with proteins with at least 100 residues. Verify 3D expects this crystal structure to be submitted in PDB format ie like : target.B99990005.pdb.

ProCheck

A program to check the stereo chemical quality of the protein structure. ProCheck suite of programs for assessing the "Stereo chemical quality" of a given protein structure. The only input required for ProCheck is the PDB file holding the coordinates of the structure of interest. The aim of ProCheck is to assess how normal, or conversely how unusual, the geometry of the residues in a given protein structure is, as compared with stereo chemical parameters derived from well-refined, high-resolution structures.

Rampage Server

Ramachandran plot, developed by Gopalasamundram Narayana Ramchandran's graduate student V.Sasisekharan, is a way to visualize dihedral angles φ against ψ of amino acid residues in protein structure. It shows the possible conformations of φ and ψ angles for a polypeptide. One would expect that larger side chains would result in more restrictions and consequently a smaller allowable region in the Ramachandran plot. In practice this does not appear the case; only the methylene group at the β position has an influence.

Glycine has a hydrogen atom, with a smaller van der Waals radius, instead of a methyl group at the β position .Hence it is least restricted and this is apparent in the Ramachandran Plot for Glycine for which the allowable area is considerably larger. In contrast, The Ramachandran Plot for proline shows only a very limited number of possible combinations of φ and ψ. The amino acid in allowed and disallowed region and also percentage was calculated. If greater than 90% in allowed region then it is a good model.

3.1.7 Docking of Protein

Cast P server

Binding sites and active sites of proteins and DNAs are often associated with structural pockets and cavities. Cast P server uses the weighted Delaunay triangulation and the alpha complex for shape measurements. It provides identification and

measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules. It measures analytically the area and volume of each pocket and cavity, both in solvent accessible surface and molecular surface.

This server is developed and maintained by Joe Dundas and Zheng Ouyang as an update to the Cast P server developed by Andrew Binkowski and Shapor Naghibzadeh, under the guidance of Prof. Jie Liang, who developed the Cast P program and the original server. The computation is based on the pocket algorithm of the alpha shape theory, and its core is the alpha shape API developed in Edelsbrunner's group and the NCSA. We would like to thank Dr. Herbert Edelsbrunner for the alpha shape API and Dr. Clare K. Woodward for encouragement.

GOLD 2.1

GOLD stands for Genetic Optimization for Ligand Docking. It is a program for calculating the docking modes of small molecules in protein binding sites and is provided as part of the GOLD Suite, a package of programs for structure visualisation and manipulation (Hermes), for protein-ligand docking (GOLD) and for post-processing (GoldMine) and visualisation of docking results. The product of a collaboration between the University of Sheffield, GlaxoSmithKline plc and CCDC, GOLD is very highly regarded within the molecular modelling community for its accuracy and reliability.

GOLD features:

- A genetic algorithm (GA) for protein-ligand docking.
- An easy to use interface with interactive docking set-up.
- A comprehensive docking set-up wizard.
- Full ligand flexibility.
- Partial protein flexibility, including protein side chain and backbone flexibility for up to ten user-defined residues.
- Energy functions partly based on conformational and non-bonded contact information.
- A variety of constraint options.
- Automatic consideration of cavity bound water molecules.
- Improved handling and control of metal coordination geometries.
- Options for generating diverse solutions, based on RMSD.

- Automatic derivation of GA settings for particular ligands.
- A choice of GoldScore, ChemScore or Astex Statistical Potential (ASP) scoring functions.
- Extensive options for customising or implementing new scoring functions through a Scoring Function Application Programming Interface, allowing users to modify the GOLD scoring-function mechanism.
- Links between GOLD and GoldMine, enabling ligands to be received from or sent to a GoldMine database before or after having been docked in GOLD respectively.

GOLD has been fully validated against 305 diverse and extensively checked protein-ligand complexes from the PDB. 72% of GOLD's top-ranked solutions were found to be accurate using stringent success criteria. A further 85 diverse, high quality drug-like complexes have been validated; GOLD reproduces the observed binding mode within 2.0 Angstroms for 81% of the structures.

Patch Dock

Patch Dock is an algorithm for molecular docking. The Patch Dock method performs structure prediction of protein-protein and protein-small complexes The input is two molecules of any type: proteins, DNA, peptides, drugs. The output is a list of potential complexes sorted by shape complimentarily criteria.

Patch Dock algorithm is inspired by object recognition and image segmentation techniques used in computer version. Docking can be compared to assembling a jigsaw puzzle. When the puzzle we try to much pieces by picking one piece and searching for the complementary one. We concentrate on the patterns that are unique for the puzzle element and look for the matching patterns in the rest of the pieces. Patch Dock employs a similar technique. Given two molecules, their surface is divided into patches according to the surface shape. These patches correspond to patterns that visually distinguish between puzzle pieces. Once the patches are identified, they can be superimposed using shape matching algorithms.

The algorithm has three major stages:

- Molecular Shape Representation – in this step it computes the molecular surface of the molecule. Next, it applies a segmentation algorithm for detection of

geometric patches (concave, convex and flat surface pieces). The patches are filtered, so that only patches with 'hot spot' residues are retained.

- Surface Patch Matching – it applies a hybrid of the Geometric Hashing and Pose-Clustering matching techniques to match the patches detected in the previous step. Concave patches are matched with convex and flat patches with any type of patches.
- Filtering and Scoring – the candidate complexes from the previous step are examined. It discards all complexes with unacceptable penetrations of the atoms of the receptor to the atoms of the ligand. Finally, the remaining candidates are ranked according to a geometric shape complementarily score.

Automated prediction of protein-protein interactions and protein-small molecule interactions is one of the most challenging problems in structural biology. Many biological studies, both in academia and in industry, may benefit from credible high-accuracy interaction predictions, In the docking problem the goal is to find the correct association of two interacting molecules given a structural representation for each molecule separately, In the case of protein-protein docking an accurate prediction will point out most of the residue-residue contacts involved in the target interaction.

Rasmol

Rasmol is a molecular graphics program intended for the visualisation of proteins, nucleic acids and small molecules. The program reads in molecular co-ordinate files and interactively displays the molecule on the screen in a variety of representations and colour schemes. Supported input file formats include Brookhaven Protein Databank (PDB), Tripos Associates' Alchemy and Sybyl Mol2 formats, Molecular Design Limited's (MDL) Mol file format, Minnesota Supercomputer Centre's (MSC) XYZ (XMol) format and CHARMm format files. If connectivity information is not contained in the file this is calculated automatically. The loaded molecule can be shown as wireframe bonds, cylinder 'Dreiding' stick bonds, alpha-carbon trace, space-filling (CPK) spheres, macromolecular ribbons, hydrogen bonding and dot surface representations. Different parts of the molecule may be represented and coloured independently of the rest of the molecule or displayed in several representations simultaneously. The displayed molecule may be rotated, translated, zoomed and z-clipped (slabbed) interactively using either the mouse, the scroll bars, the command line or an attached dial box. RasMol can read a

prepared list of commands from a 'script' file (or via inter-process communication) to allow a given image or viewpoint to be restored quickly. RasMol has been developed at the University of Edinburgh's Biocomputing Research Unit and the Bimolecular Structure Department, Glaxo Research and Development, Greenford, U.K.

3.2 Methodology

The following steps were followed to achieve the objective.

- **Step1 :** Target Identification
- **Step2 :** Identification of homolog using BLAST P
- **Step3 :** secondary structure prediction
    - SOPMA
    - GOR4
- **Step4 :** Homology Modeling using insight II
- **Step5 :** Energy Minimization
- **Step6 :** Finding Ramachandran Plot in Insight II
- **Step7 :** Docking in GOLD
- **Step8 :** Docking in Insight II
    - Grid Docking
    - Delphi Docking
- **Step9 :** Patch Docking

3.2.1 Target Identification

Before starting homology modeling the requirement was to select a protein sequence whose structure is going to modeled. For that following steps were followed:

- The Neuregulin1 isoform HRG-gamma protein was retrieved from NCBI (National Centre for Biotechnology Information) homepage.
- In ExPASY the name of protein was given and various results were found all the results were gone thoroughly and finally the sequence which does not have any structure in protein data bank (PDB) and uniprot was taken for further analysis.

3.2.2 Identification of Homolog

In homology modeling, the sequence of a protein to be modeled is searched to identify its homolog, that has (>=35%) sequence similarity and its three-dimensional structures are available. Steps are :

- The selected protein was then entered in to BLAST P and the results were obtained.
- From the results the sequence having more than 35% sequence identity and 0.0% of gaps with the target sequence was taken as the template.
- The template 2PND was downloaded from Protein data Bank.
- After taking the target and template sequences homology modeling was done with InsightII.

3.2.3 Secondary Structure Prediction

Combining methods using different approaches to predict secondary structure is useful to validate the prediction. The first step is to select the secondary structure prediction methods we want to use. Then, the secondary structure consensus prediction program generates a secondary consensus. To predict the secondary structure of Neuregulin1 isoform HRG-gamma protein various tools like SOPMA and GOR IV.

1. SOPMA

SOPMA is a secondary structure prediction method. SOPMA (Self-Optimized Prediction Method with Alignment) is an improvement of SOPM method. These methods are based on the homologue method of Levin et al (1986). SOPMA correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins. Joint prediction with SOPMA and a neural networks method (PHD) correctly predicts 82.2% of residues for 74% of co-predicted amino acids. The improvement takes place in the fact that SOPMA takes into account information from an alignment of sequences belonging to the same family (Geourjon and Deleage, 1995).

So it helps in significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments.

Steps

- The SOPMA homepage was retrieved from the URL. This is available at expasy server.
- After getting the homepage of SOPMA there is a box provided for inputting a sequence.
- The FASTA format of the Neuregulin 1 isoform HRG gamma protein sequence which was obtained from NCBI was pasted here.
- The name of the sequence was also given in the provided box and the output width was keeping same as constant i.e 70.
- Finally "SUBMIT" option was clicked.

2. GORIV

GOR IV (Garnier Osguthorpe and Robson-IV) is the fourth version of GOR secondary structure prediction methods based on the information theory. It was developed by J.Garnier, D.Osguthorpe and B.Robson (J.Mol.Biol. 120,97, 1978). There is no defined decision constant. The GOR method is based on information theory and was developed by J.Garnier, D.Osguthorpe and B.Robson (J.Mol.Biol. 120,97, 1978). The present version, GOR IV, uses all possible pair frequencies within a window of 17 amino

acid residues and is reported by J. Garnier. J.F. Gibrat and B.Robson in Methods in Enzymology, vol 266, p 540-553 (1996). After cross validation on a data base of 267 proteins, the version IV of GOR has a mean accuracy of 64.4% for a three state prediction (Q3). The program gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil ;the second output gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one of highest probability compatible with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues. This program was written by Jean-Francois Gibrat and was modified by Stephen Pheiffer in order to increase it's efficiency. These scores are corrected by decision constants and the highest one gives the conformational state for the residue in the center of the window.

Steps

- The GORIV homepage was retrieved from the URL. This is available at expasy server.
- After getting the homepage of GORIV there is a box provided for inputting a sequence.
- The FASTA format of the Neuregulin 1 isoform HRG gamma protein sequence which was obtained from NCBI was pasted here.
- The name of the sequence was also given in the provided box and the output width was keeping same as constant i.e 70.
- Finally "SUBMIT" option was clicked.

3.2.4 Homology Modeling using Insight II

As mentioned before, homology is an application within InsightII. Most Insight II commands can be used on molecules created with Homology and vice. All rules and conventions for Insight II apply when you are working with homology. Example. Just like commands in Insight II.

Steps

- From the module Homology was selected.
- The template protein molecule was loaded. For this molecules/ get/ Gopal/ template.pdb/ execute was clicked.

- Then to extract sequence from the protein structures Sequence/ extract was clicked. Sequence window will appear on screen.
- The sequence option was selected and get sequence was clicked for getting target sequence.
- Alignment/ pair wise alignment was selected to create a sequence alignment between proteins. Here sequence1 is the target sequence and sequence2 is the template sequence and other value was set to default and executed.
- Then box/ create boxes/ freeze boxes was selected for box creation and freeze the boxes.
- For assigning of coordinates *sequence/ assign cords* was clicked.
- For generation of loops Loops/ generate was clicked.
- Model refinement was done. For this Refine/end repair/execute was selected.
- The modeler input files were created.
- For building of models modeler/ build model was clicked and no. of models given as 10. This will take some times depending on the sequence length. Finally after completion 10 models are generated on my folder.

3.2.5 Energy Minimization

It presents a detailed study of the performance and reliability of design procedures based on energy minimization. The analysis is carried out for model protein where exact results can be obtained through exhaustive enumeration. The efficiency of design techniques is assessed as a function of protein lengths and number of class into which amino acids are coarse grained. It turns out that, while energy minimization strategies can identify correct solutions in most circumstances, it may be impossible for numerical implementations of design algorithms to meet the efficiency required to yield correct solutions in realistic contexts. We also investigated how the design efficiency varies when putative solutions are required to obey some external constraints and found that a restriction of the sequence space impairs the design performance rather than boosting it. Finally some alternate design strategies based on a correct treatment of the free energy are discussed.

Steps

- To get the target protein molecule/ get/ gopal/ neuregulin1.pdb was clicked.
- Module/ Biopolymer was selected
- Addition of hydrogen was done using modify/ hydrogen/ pH set as 7/ execute.

- Potential was set. FF/ select/ execute. Select command allows us to determine the current force field.
- FF/ potential/ fix all was selected and executed.
- The Insight II program supports these force fields. CFF .CVFF. AMBER CHARMm, and ESFF, which are specified through the force field/ select command
- Discover_3 module was opened from module.
- Setup/ system was selected, object name was given as neuregulin1 and executed.
- For energy minimization calculate/ minimize/ maximum steps was set to 1000 and executed.
- Finally D_run is used for running the program. This is done by D_run/ run/ execute.
- Then measure/ energy/ execute was selected. It will show the final energy value on the screen after completion.

3.2.6 Finding Ramachandran Plot in InsightII

- The selected protein in .pdb format was retrieved and homology module was selected.
- An InsightII table was created for graph (e.g., measure phi and psi angles using the Homology / Prostat/ Residue Dihedral menu). This is done by prostat/ residue_dihederal/ turn on phi, psi/ residue table name was given as residue and executed. A table will appear. Manually edit the table if necessary (edit blank cells, insert new columns, etc).
- The LEFT mouse button was clicked on the top cell of the column that contains X–axis data (e.g., phi).
- Hold down the <Ctrl> button and the LEFT mouse button was clicked on the top cell of the column that contains Y-axis data (e.g., psi).
- The Graph icon was clicked at the bottom of the table window (the middle icon). A graph will appear.
- Then analysis module was selected. Graph/ label was clicked. Label text was given as "Ramachandran Plot" and was executed.
- The Char Size of the Axes and Graph was set to 0.04.
- Color of the X-axis, Y-axis, Box_segment 1 and Box_ segment 2 was set as light blue.

- The color of the plot was changed to yellow.
- Graph/ modify_display was selected. Bar, line and graph were set as "turn off" and point as "turn on". Point symbol was given as "X", symbol scale=10.0 and was executed.
- Graph option was turn on, plot spec was given as Graph 1 and executed.
- Threshold of the X axis and Y axis was set to a Min Value of -180 and a Max Value of 180. Zoom_Axis was turn on and executed.
- Plot was divided into four quadrants. Tick Mark was changed so that extended tick marks Start at -400 and End at 400 with a Mark Step of 200. These marks were added to the X axis and Y axis respectively.

3.2.7 Docking in GOLD 2.1

GOLD is the Genetic Optimization for Ligand Docking. It has a good hit rate and fitness value overall. Gold uses genetic algorithm to provide docking of flexible ligand and a protein with flexible hydroxyl groups. Otherwise the protein is considered to be rigid. This makes it a good choice when the binding pocket contains amino acids that form hydrogen bonds with the ligand. Gold uses a scoring function that is based on favorable conformations found in Cambridge Structural Database and on empirical results on weak chemical interactions.

Steps

- GOLD software was opened by typing "GOLD" in the terminal window.
- Ligands are added by clicking on Add/Delete button and all the 8 Ligand molecules were added.
- Then the Protein molecule was added by clicking on Protein button.
- Define Active site was clicked and atom number which are obtained from Cast P server was given.
- Active site radius was set to 10.0.
- Parameters was selected and the output parameter file was changed.
- All other parameters was leave to their default values.
- GOLD docking was run.

- After the process got over many files in the folder were generated. In that just check the "best ranking list file". That's very important file where ligand ranking will be given.

3.2.8 Docking in InsightII

I. Grid Docking

- For build a receptor molecule : Molecule/get pdb was selected and the target protein was retrieved. Similarly for ligand molecule : Module/Builder eas selected and retrieved.
- Draw ligand/optimize.
- Subset/Define was used for define a Subset for Binding site.
- Measure/Hydrogen Bond was selected for adding Hydrogen bond.
- Then to create Ligand Receptor Assembly Assembly/Associate was clicked.
- For assigning Potential for Ligand/Receptor molecules Module/Force Field/Select/Potential/CVFF.
- Docking Module was selected.
- Docking Grid/Create was clicked and create an enclosure-style grid about receptor i.e (Neuregulin1) was create with a border space of 5 angstroms and a 2 angstrom grid step.
- Docking Grid/Compute was clicked and a docking grid of em2with Van der waals and Coulomb energies and a cutoff of 10 angstroms was computed. A visible grid was made by turning on make_vis_grid with a grid name of Neuregulin1_VGRD.
- Then Grid/Display was selected and display all points of Neuregulin1_VGRD.
- Grid/color was clicked and color all grid points using the charge spectrum (turn "use spectrum" ON)
- Again Grid/Display was chosen and the grid points was turned off.
- Grid slice was evaluated. For this Grid/Slice was clicked. The Scalar grid name was set to Neuregulin1_VGRD, Plane Number was set to 1, Plane Direction to Z, Plane Height to 59.5221, H intervals to 20, Plane Width to 59.5221, W intervals to 20, Plane Type to Map plane, Spectrum Name to Charge_Spectrum, Plane Style to Filled and was executed.
- The plane was moved by moving the slide bar in the Parameters menu. (when the plane bisects the molecule, the center of the plane is blue: the interaction

energy inside the molecule is high (positive, or Bad). When the plane is moved towards the ligand and outside the protein, the plane turns red: the interaction energy is low (negative, or Good). Position the plane near the ligand; notice that the plane is red in the center, and blue near the edges: the interaction energy is good in the center and bad at the edges, so the ligand is attracted towards the center, the active site ! This model may actually be useful).

- Try a Contour plane in the X direction with a Slice_Spectrum, Contour Levels was set to -2, and Displacement was set to -6
  Plane Number 2 was added at a contour level of -3. Similarly -4, -5, -6, -7. The interaction energy becomes lower (more negative) nearer the active site. After finishing, Plane Type was set to "OFF" for all planes.

- Grid Contours was evaluated. This is done by selecting Grid/Contour. Contour_Name_Root was set to contour1, Level Specification to Single, Contour Level to -15, Display Style to Solid, Color to yellow. (Note where the best interaction energies lie. Then delete all contour objects).
- For evaluate the Intermolecular Energies Evaluate/Intermolecular was clicked and a monitor energy was added between receptor and ligand via the grid.
- The ligand position was set so that the VdW and Elect energies are < 0, and the Total energy is < -10.
- The system position was set for best visualization of the ligand-receptor interaction.
- Then the file was saved in grid dock .psv.
- Evaluate/Intermolecular was clicked and a monitor energy was added between receptor and ligand.
- And finally the docked structure was seen.

II. Delphi Docking

- Delphi module in InsightII was opened, receptors folder was selected and restored it.
- Addition of hydrogen was done from modify.
- Setup/initialize was selected. parameters options was set to be default param and was executed.
- Select the inhibitor i.e. the Ligand. Go to molecule and archive the .car file of the ligand.

- Reference object was set to on. Set reference name to the receptor and was executed.
- Then subset/interface was selected. em1subset was typed for subset name. Centre of subset was clicked and ligand name (Ziprasidone) was typed. Search domain was clicked and typed as Neuregulin1. Collection level to monomer/residue was set. Then radius of subset was set as 5.00 and executed.
- Subset/list was selected and subset/list was set to em1subset by choosing it.
- The text port off box at the bottom of the InsightII window was set. Then molecule/display was selected and molecule pick level was set to subset by picking it. Molecule spec was clicked and em1subset was selected.
- Setup/boundary was selected. Boundary condition was set to approx_coulombic and executed.
- The grid was set up for Delphi calculation.
- Setup/list_setup was clicked to ensure that parameters are set correctly.
- Setup/files was selected and Delphi potential output file format was set to ".phi".
- Delphi/run was used to run. assembly/molecule should be set to em1 and the job name was written as lesson and was executed.
- Setup/boundary. set boundary condition to focussing. Set focusing file to lesson8.phi.execute
- Setup/grid was selected. Grid center should be set to molecule_region. Assem/mol was set to leave subset by selecting it. Molecule region was set to em1subset and set grid size to extent. Extent button was clicked and typed as 90.0. Then setup/list setup was used for setting the correct parameters.
- Assembly/mol level was set to molecule and assemble/molecule to em1 and executed.
- Docking grid/create was selected. Set from_file was on. Molecule spec was clicked and em1subset was typed. Sampling rate was set to 1 and executed.
- Docking grid/compute was selected. Vander waals to on and coulomb to off was clicked, also molecule name and finally the receptor was given and executed.
- Evaluate/intermolecular was select. Monitor was set to on, grids to on. Executed. Transform/connect was selected. Object name was set to ligand. Finally .psv folder was saved to see the docked structure.

3.2.9 Patch Docking

Automated prediction of protein-protein interactions and protein-small molecule interactions is one of the most challenging problems in structural biology. Many biological studies, both in academia and in industry, may benefit from credible high-accuracy interaction predictions. In the docking problem the goal is to find the correct association of two interacting molecules given a structural representation for each molecule separately. W have developed patch Dock, a very efficient algorithm for protein-small ligand and protein- protein docking, patch dock is an algorithm for molecular docking. The input is two molecules of any type: proteins, DNA, peptides, drugs. The output is a list of potential complexes sorted by shape complementarily criteria. It is a suite of freely available web services for protein structural analysis. Patch Dock performs prediction of protein- protein and protein-small molecule interactions. The input to all services is either protein PDB codes or protein structures uploaded to the sever. Once the docking request is submitted. The patch dock algorithm starts the prediction process. The user is notified when results are ready by an email message that contains a link to a web page where the predictions are presented.

Steps

- Ligand and Receptor molecule was loaded.
- RMSD value was set.
- The server was run and got the results.
- In the result page .pdb file was downloaded.
- Rasmol software was opened and docking structure was viewed.

CHAPTER-IV

Results

RESULTS

After the complete study about the disease it was found that schizophrenia is caused by protein **Neuregulin1 isoform HRG gamma,** the target. The target information and sequence was retrieved from NCBI.

5.1 Target sequence

>gi|116006965|ref|NP_004486.2| neuregulin 1 isoform HRG-gamma [Homo sapiens]
MSERKEGRGKGKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCETSSEYSSLRFKWFK
NGNELNRKNKPQNIKIQKKPGKSELRINKASLADSGEYMCKVISKLGNDSASANITIVESNEIITGMPASTEGA
YVSSESPIRISVSTEGANTSSSTSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLCK

BLAST P Result

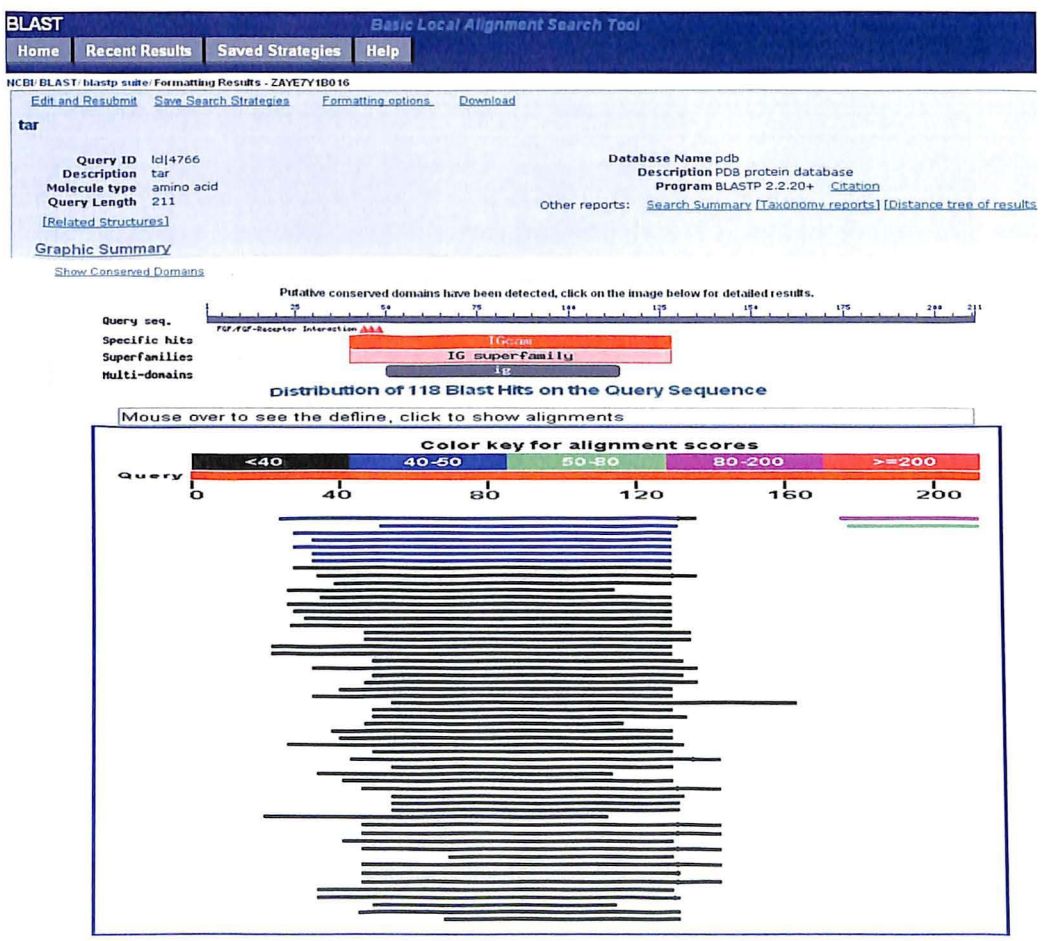


Fig 6 : Protein BLAST Result.

After the running it was found that the template 2PND have 38% of sequence identity with the target sequence and having gap 0.0%. the complete information about template is :

> pdb|2PND|A Chain A, Structure Or Murine Crig
Length=119, Score = 26.6 bits (57), Expect = 6.6, Method: Compositional matrix adjust, Identities = 10/26 (38%), Positives = 13/26 (50%), Gaps = 0/26 (0%)

Query 89 QKKPGKSELRINKASLADSGEYMCKV 114
K PG L+IN + D Y C+V
Sbjct 72 HKVPGDVSLQINTLQMDDRNHYTCEV 97

5.2 Template Sequence

The template sequence was retrieved from Protein Data Bank which is available at http://www.rcsb .org/pdb/explore/explore.do?structureid=2PND

>2PND:A|PDBID|CHAIN|Structure of Murine CRIg|SEQUENCE|
GHPTLKTPESVTGTWKGDVKIQCIYDPLRGYRQVLVKWLVRHGSDSVTIFLRDSTGDHIQQAKYRGRLKVSH
KVPGDVSLQINTLQMDDRNHYTCEVTWQTPDGNQVIRDKIIELRVRK

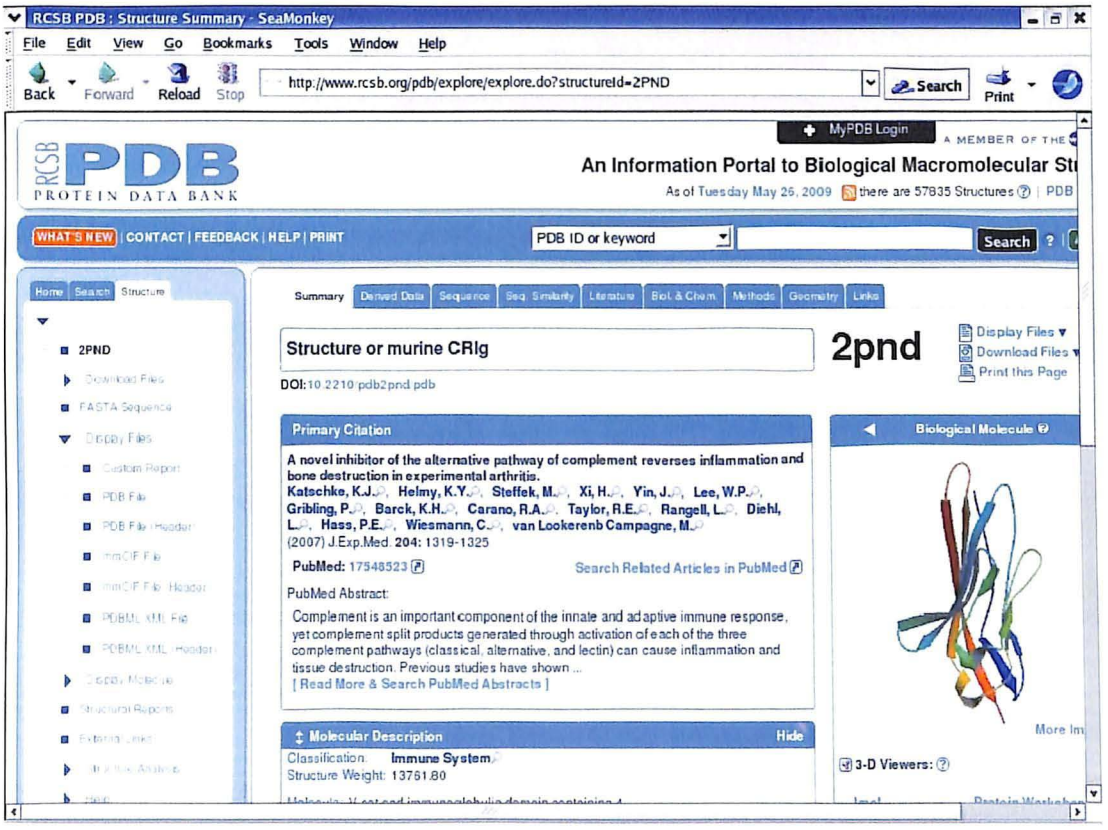


Fig 7 : Template 2PND from Protein Data Bank.

5.3 Prot Param

User-provided sequence

MSERKEGRGK GKGKKKERGS GKKPESAAGS QSPALPPRLK

EMKSQESAAG SKLVLRCETS SEYSSLRFKW FKNGNELNRK

NKPQNIKIQK KPGKSELRIN KASLADSGEY MCKVISKLGN

DSASANITIV ESNEIITGMP ASTEGAYVSS ESPIRISVST

EGANTSSSTS TSTTGTSHLV KCAEKEKTFC VNGGECFMVK

DLSNPSRYLC K

Number of amino acids: 211
Molecular weight: 22754.8
Theoretical pI: 9.62

Amino acid composition:

Amino Acid	Number	In (%)
Ala (A)	13	6.2
Arg (R)	10	4.7
Asn (N)	12	5.7
Asp (D)	3	1.4
Cys (C)	6	2.8
Gln (Q)	4	1.9
Glu (E)	19	9.0
Gly (G)	18	1.5
His (H)	1	0.5
Ile (I)	10	4.7
Leu (L)	12	5.7
Lys (K)	27	12.8
Met (M)	5	2.4
Phe (F)	4	1.9
Pro (P)	9	4.3
Ser (S)	33	15.6
Thr (T)	12	5.7
Trp (W)	1	0.5
Tyr (Y)	4	1.9
Val (V)	8	3.8
Pyl (O)	0	0.0
Sec (U)	0	0.0

Table 1 : Amino acid Compositions

Total number of negatively charged residues (Asp + Glu): 22

Total number of positively charged residues (Arg + Lys): 37

Atomic composition

Carbon	C	968
Hydrogen	H	1607
Nitrogen	N	287
Oxygen	O	321
Sulfur	S	11

Formula : $C_{968}H_{1607}N_{287}O_{321}S_{11}$
Total number of atoms : 3194

Extinction coefficients

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 11835
Abs 0.1% (=1 g/l) 0.520, assuming ALL Cys residues appear as half cystines

Ext. coefficient 11460
Abs 0.1% (=1 g/l) 0.504, assuming NO Cys residues appear as half cystines

Estimated half-life

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index

The instability index (II) is computed to be 45.05
This classifies the protein as unstable.

Aliphatic index: 57.82

Grand average of hydropathicity (GRAVY): -0.785

5.4 ClustalW Results

Number of sequences	2
Alignment score	48
Sequence format	Pearson
Sequence type	aa
Output file	clustalw2-20090528-0708533472.output
Alignment file	clustalw2-20090528-0708533472.aln
Guide tree file	clustalw2-20090528-0708533472.dnd
Your input file	clustalw2-20090528-0708533472.input

Table 2 : ClustalW Results

Scores Table

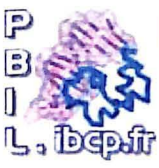
SeqA	Name	Len(aa)	SeqB	Name	Len(aa)
1	gi|116006965|ref|NP_004486.2|	211	2	2PND_A|PDBID|CHAIN|SEQUENCE	119

Alignment

```
gi|116006965|ref|NP_004486.2|      MSERKEGRGKGKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAG 50
>2PND_A|PDBID|CHAIN|SEQUENCE       ----------------GHPTLKTPESVTGTWKGDV--KIQCIYDPLRGYR 32
                                                 : : *.***.:*: .  :  ::: : .   .

gi|116006965|ref|NP_004486.2|      SKLVLRCETSSEYSSLRFKWFKNGNELNRKNKPQNIKIQ-KKPGKSELRI 99
>2PND_A|PDBID|CHAIN|SEQUENCE       QVLVKWLVRHGSDSVTIFLRDSTGDHIQQAKYRGRLKVSHKVPGDVSLQI 82
                                   . **       .. *   *    ..*:.::: :   .:*:. * **. .*:*

gi|116006965|ref|NP_004486.2|      NKASLADSGEYMCKVISKLGNDSASANITIVESNEIITGMPASTEGAYVS 149
>2PND_A|PDBID|CHAIN|SEQUENCE       NTLQMDDRNHYTCEVTWQTP-----------------------DGNQVI 108
                                   *. .: * ..* *:*  :                          :*  *

gi|116006965|ref|NP_004486.2|      SESPIRISVSTEGANTSSSTSTSTTGTSHLVKCAEKEKTFCVNGGECFMV 199
>2PND_A|PDBID|CHAIN|SEQUENCE       RDKIIELRVRK---------------------------------------- 119
                                    :. *.: * .

gi|116006965|ref|NP_004486.2|      KDLSNPSRYLCK 211
>2PND_A|PDBID|CHAIN|SEQUENCE       ------------
```

ClustalW is used for Multiple Sequence Alignment. It align the target sequence with the template. The result shows detailed information about no. of sequences, scores, sequence format and type and 4 types of file like output file, alignment file, guide tree file and input file. Here the alignment shows the colour output which helps us to identify the amino acids distinctly.

5.5 Secondary Structure Prediction



Pôle Bioinformatique Lyonnais
Network Protein Sequence Analysis

NPS@ is the IBCP contribution to PBIL in Lyon, France

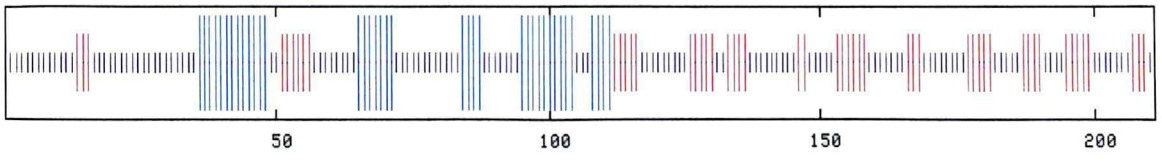
GOR4 result for : Neuregulin1 isoform HRG-gamma

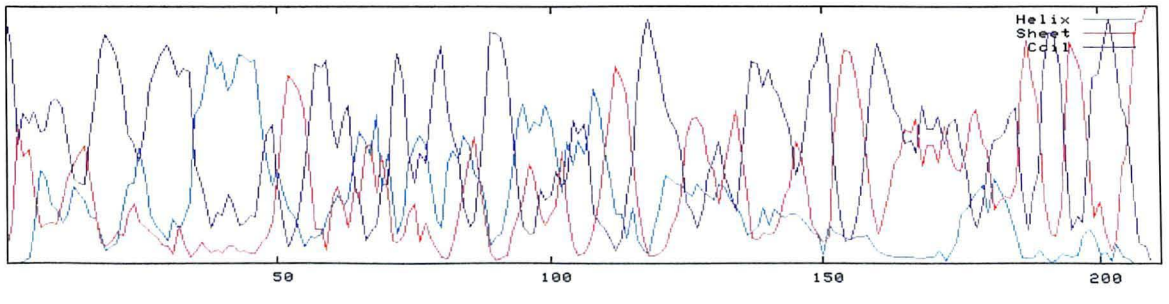
```
          10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MSERKEGRGKGKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCETSSEYSSLRFKW
ccccccccccccccceeecccccccccccccccccchhhhhhhhhhhhhcceeeeeecccccccchhhhh
FKNGNELNRKNKPQNIKIQKKPGKSELRINKASLADSGEYMCKVISKLGNDSASANITIVESNEIITGMP
hhccccccccccchhhhccccccchhhhhhhhhhccchhhheeeeeccccccccceeeeecceeeecccc
ASTEGAYVSSESPIRISVSTEGANTSSSTSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLC
cccccceeccccceeeeeecccccccceeecccccccceeeeecccccceeeecccceeeeeccccccceee
K
c
```

Sequence length : 211

GOR4 :

Alpha helix	(Hh)	38 is 18.01%
3_{10} helix	(Gg)	0 is 0.00%
Pi helix	(Ii)	0 is 0.00%
Beta bridge	(Bb)	0 is 0.00%
Extended strand	(Ee)	51 is 24.17%
Beta turn	(Tt)	0 is 0.00%
Bend region	(Ss)	0 is 0.00%
Random coil	(Cc)	122 is 57.82%
Ambigous states	(?)	0 is 0.00%
Other states		0 is 0.00%





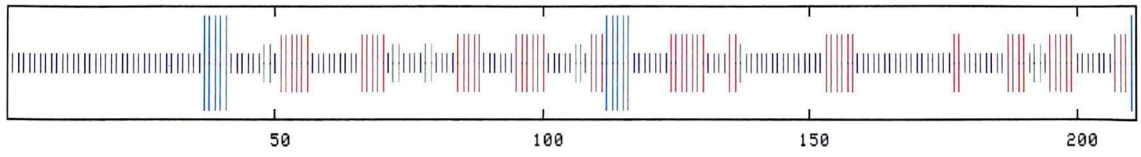
SOPMA result for : Neuregulin1 Isoform HRG-gamma

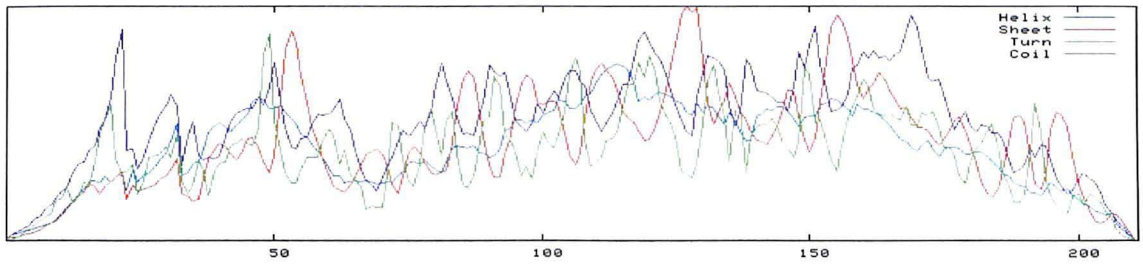
```
         10         20         30         40         50         60         70
         |          |          |          |          |          |          |
MSERKEGRGKGKGKKKERGSGKKPESAAGSQSPALPPRLKEMKSQESAAGSKLVLRCETSSEYSSLRFKW
hccccccccccccccccccccccccccccccccccccccchhhhhccccccttceeeeeeccccccccceeee
FKNGNELNRKNKPQNIKIQKKPGKSELRINKASLADSGEYMCKVISKLGNDSASANITIVESNEIITGMP
ecttccccttcccceeeeecccccceeeeeeccccctteeehhhhhccccccceeeeeeecccceetcc
ASTEGAYVSSESPIRISVSTEGANTSSSTSTSTTGTSHLVKCAEKEKTFCVNGGECFMVKDLSNPSRYLC
ccccccccccccceeeeeeccccccccccccccccccceecccccccceeeeecttceeeeeccccccceee
K
H
```

Sequence length : 211

SOPMA :

Alpha helix	(Hh)	12 is 5.69%
3_{10} helix	(Gg)	0 is 0.00%
Pi helix	(Ii)	0 is 0.00%
Beta bridge	(Bb)	0 is 0.00%
Extended strand	(Ee)	54 is 25.59%
Beta turn	(Tt)	11 is 5.21%
Bend region	(Ss)	0 is 0.00%
Random coil	(Cc)	134 is 63.51%
Ambigous states	(?)	0 is 0.00%
Other states		0 is 0.00%





Parameters :

Window width	17
Similarity threshold	8
Number of states	4

5.6 Homology Modeling

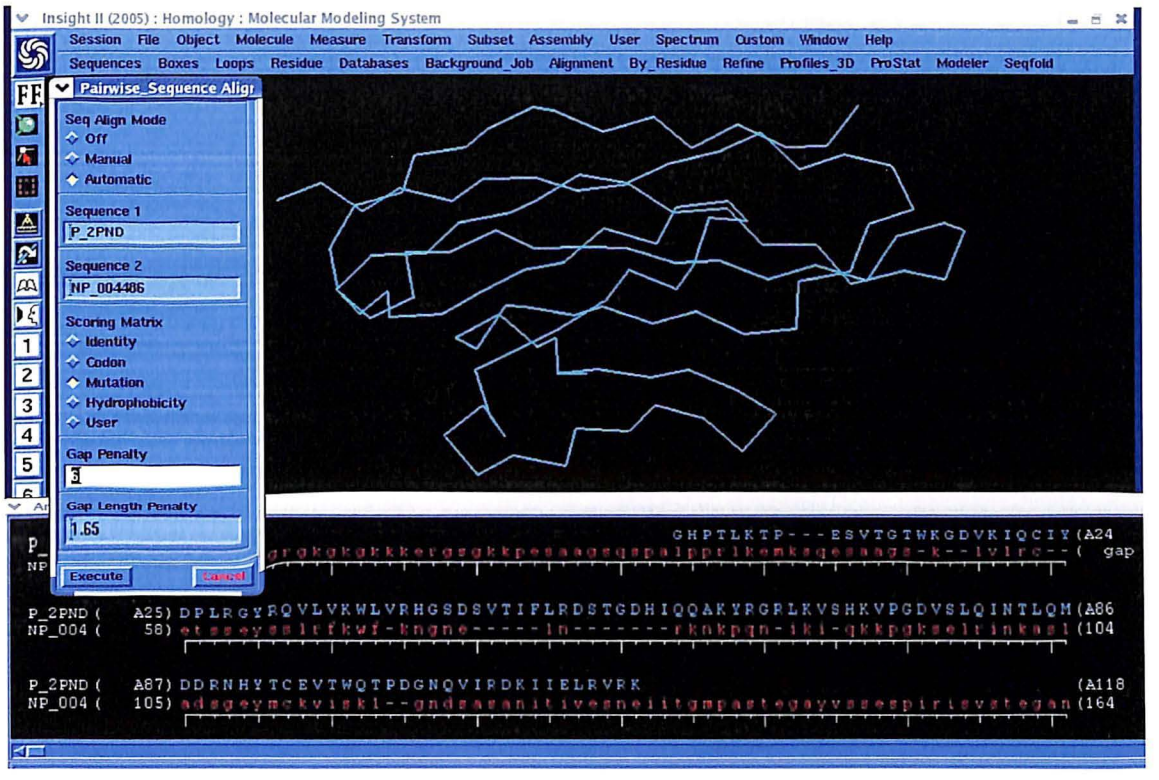


Fig. 8 : Pair wise Alignment of target with template.

In homology modeling, the sequence of a protein to be modeled (its three-dimensional structure is unknown) is searched to identify its homolog, The template has 38% sequence identity and its three-dimensional structures are available. A critical step in the development of a homology model is the alignment of the unknown sequence with the homologues. Many methods are available for sequence alignment, and sometimes the most perplexing task is deciding which methods to apply. The above figure (Fig.8) shows the pair wise alignment between target with template. Here blue sequence indicates the template sequence and red as target.

Factors to be considered when performing an alignment are, algorithm to use for sequence alignment, scoring method to apply, and whether and how to assign gap penalties.

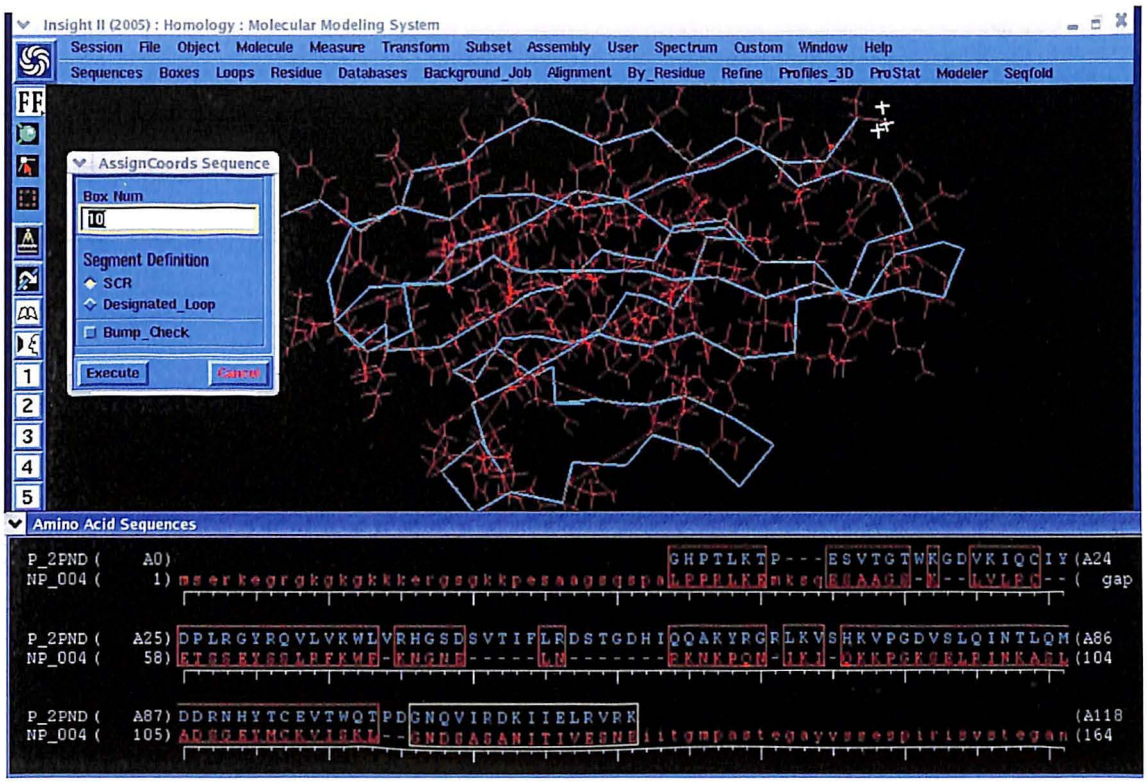


Fig. 9 : Assigning coordinates.

When generating coordinates for the unknown structure, one needs to model main chain atoms and side chain atoms. It is straightforward to generate the coordinates of the main chain atoms of the unknown structure from those of the known structures. Side chain coordinates are copied if the residue type in the unknown is identical or very similar to that in the known homologues. For other side chain coordinates one can apply a side chain rotamer library in a systematic approach to explore possible side chain conformations. It may be desirable to weight the contribution of each homologue in each Structurally Conserved Regions based on the extent of similarity with the unknown. In the above figure (Fig.9) it indicates the assigning the co-ordinates in the structurally conserved regions.

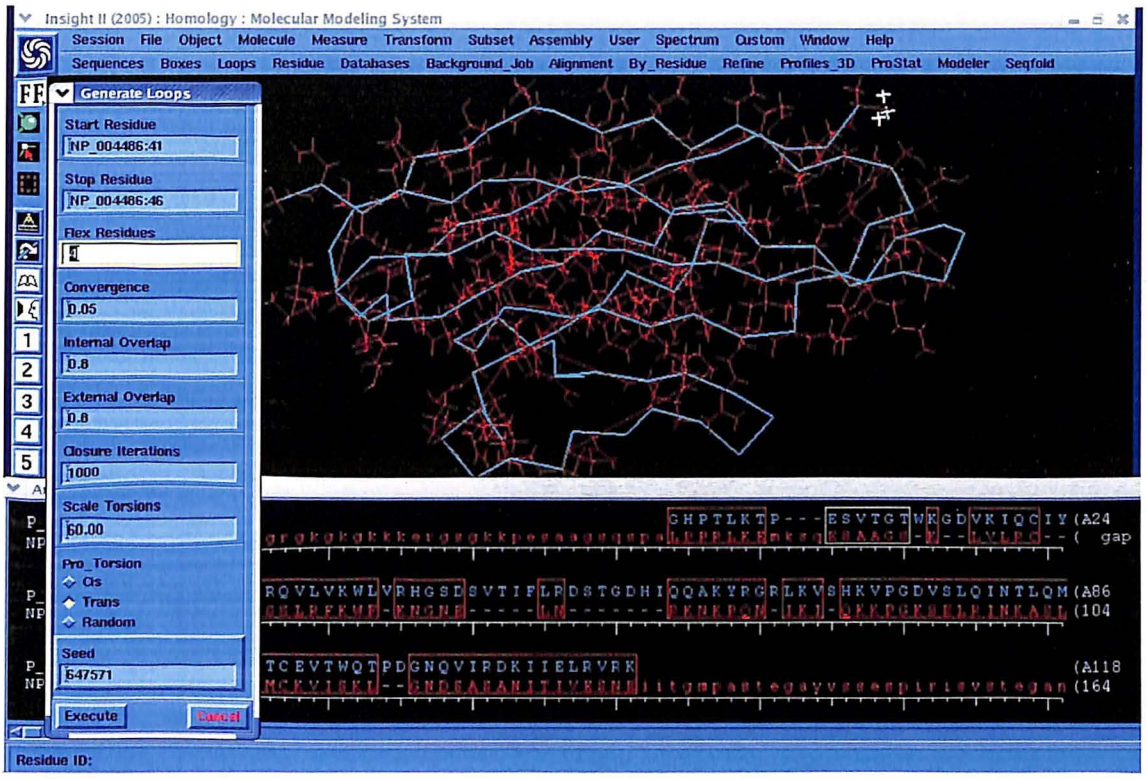


Fig. 10 : Loop Generation.

Regions of the target sequence that are not aligned to a template are modeled by loop modeling, they are the most susceptible to major modeling errors and occur with higher frequency when the target and template have low sequence identity. The coordinates of unmatched sections determined by loop modeling programs are generally much less accurate than those obtained from simply copying the coordinates of a known structure, particularly if the loop is longer than 10 residues. The above figure (Fig.10) indicates the generation of loop for the better modeling.

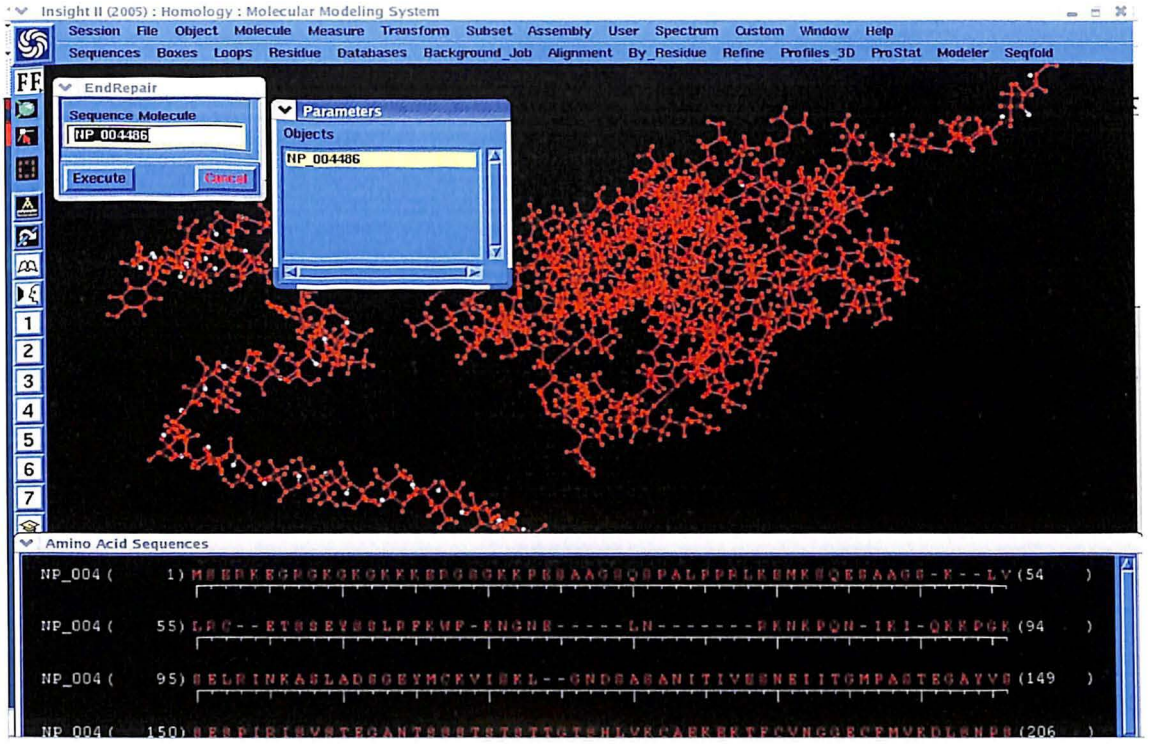


Fig. 11 : Model Refinement.

For a homology model from any source, it is important to demonstrate that the structural features of the model are reasonable in terms of what is know about protein structures in general. That is, it analyzed three-dimensional structures of proteins from which basic principles of protein structure and folding have been developed. Above figure (Fig.11) shows the model refinement. The white spot indicates the hydrogen atoms.

The criteria for analysis of correctness include:

- main chain conformations in acceptable regions of the Ramachandran map
- planar peptide bonds
- side chain conformations that correspond to those in the rotamer library
- hydrogen-bonding of polar atoms if they are buried
- proper environments for hydrophobic and hydrophilic residues
- no bad atom-atom contacts
- no holes inside the structure.

5.7 Validation of Model

Programs that provide structure analysis along with output that is useful for publication include Verify 3D, Rampage server and ProCheck.

A) Verify 3D

Verify3D Structure Evaluation Server is a tool designed to help in the refinement of crystallographic structures. It will provide us with a visual analysis of the quality of a putative crystal structure for a protein. Verify3D expects this crystal structure to be submitted in PDB format.

Sl. No.	Modeling Pdb	Values [verify3D]
1	NP_004486.B99990001.pdb	0.51
2	NP_004486.B99990002.pdb	0.51
3	NP_004486.B99990003.pdb	0.49
4	NP_004486.B99990004.pdb	0.44
5	NP_004486.B99990005.pdb	0.51
6	NP_004486.B99990006.pdb	0.54
7	NP_004486.B99990007.pdb	0.46
8	NP_004486.B99990008.pdb	0.41
9	NP_004486.B99990009.pdb	0.38
10	NP_004486.B99990010.pdb	0.38

Table 3 : Verify 3D Results

From the above results , 6th model has the highest value. So it has been considered as a best model and **NP_004486.B99990006.pdb** (0.54) was taken for further work.

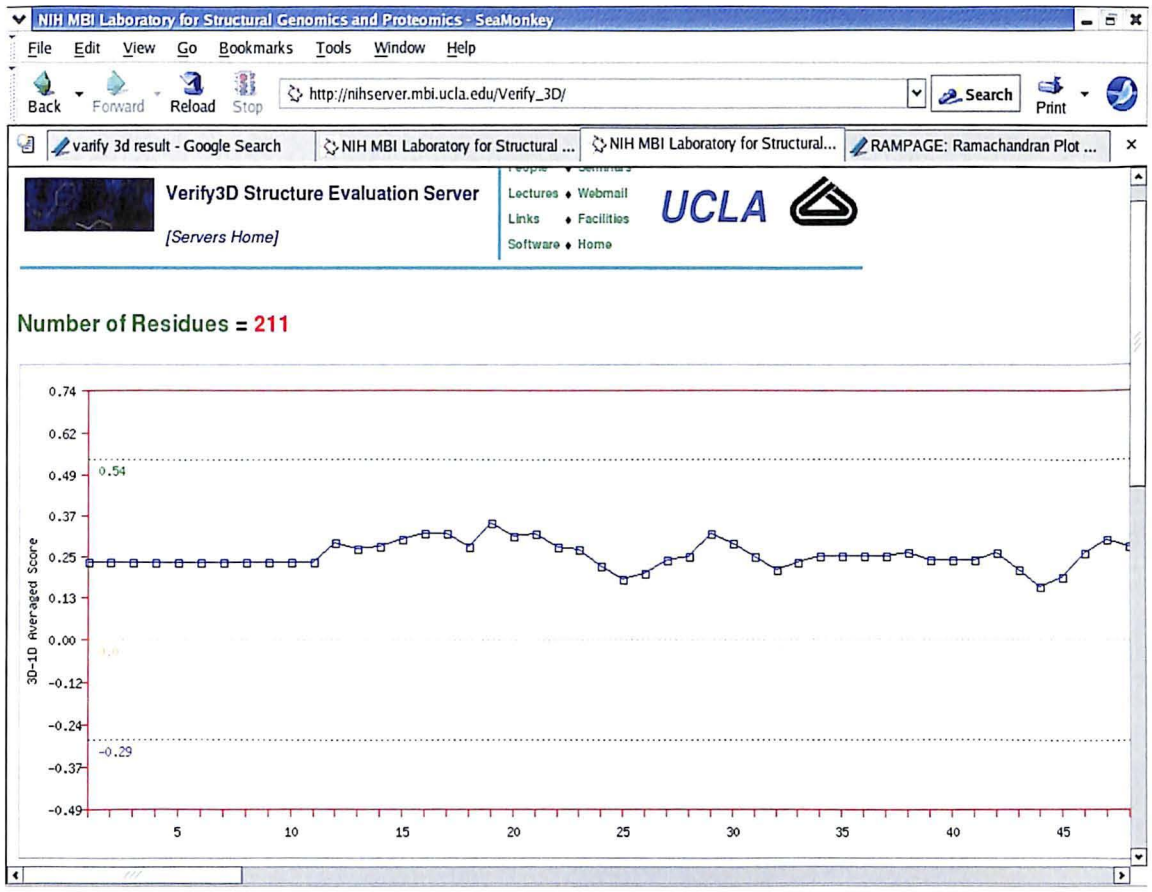


Fig. 12 : Verify 3D Graph.

The least energy was found with the model NP_004486.B99990006.pdb. In the above plot, the vertical axis represents the average 3D-1D profile score for residues in a 21-residue sliding window, the center of which is at the sequence position indicated by the horizontal axis. Scores for the first 9 and the final 9 sequence positions have no meaning. A window length of 21 residues strikes a useful. The model NP_004486.B99990006.pdb was found to be with least energy with highest value i.e. 0.54 .

B) RAMPAGE Server: Assessment of the Ramachandran Plot

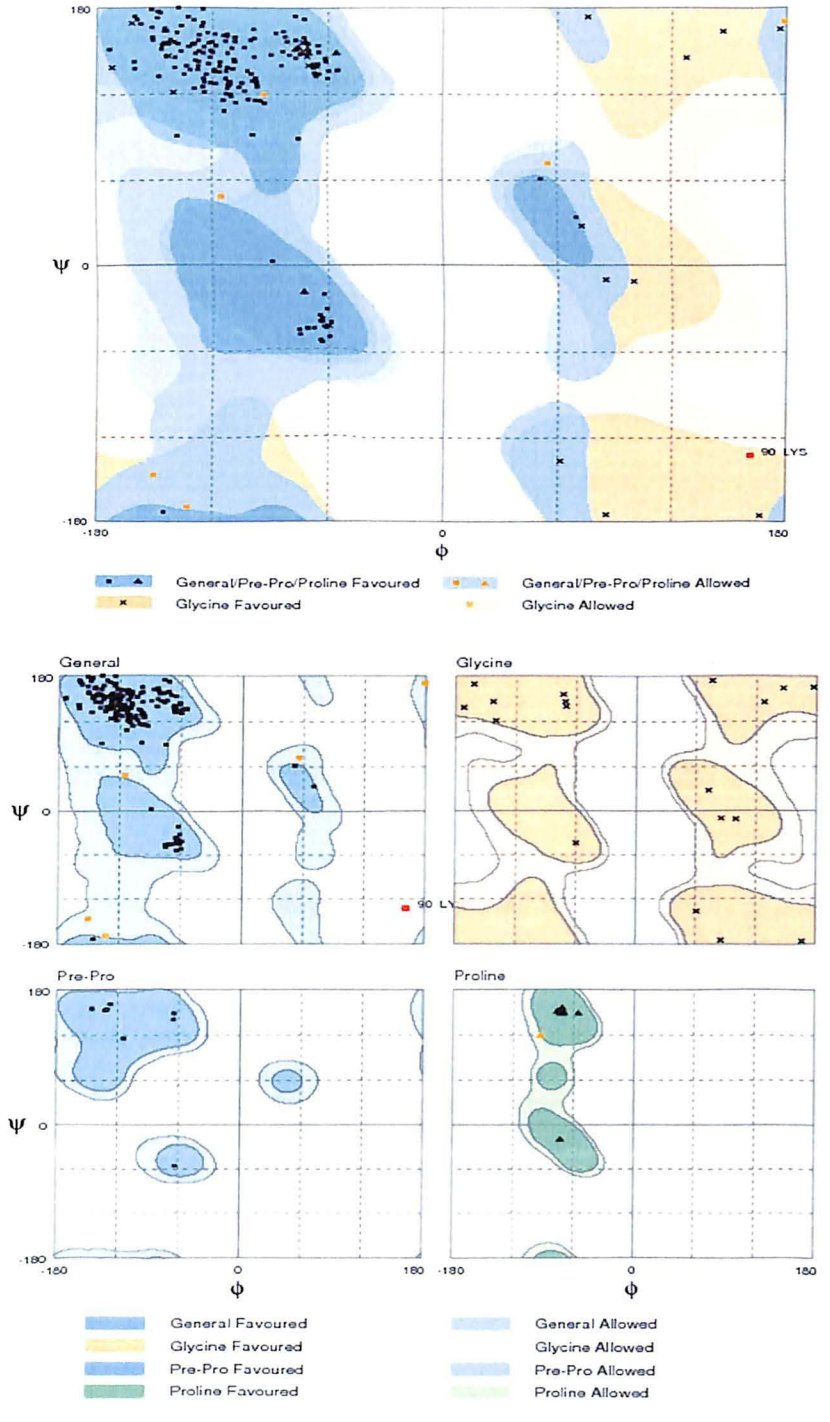


Fig. 13 : Rampage Server Result.

Evaluation of residues

Number of residues in favoured region	(~98.0% expected):	202 (96.7%)
Number of residues in allowed region	(~2.0% expected):	6 (2.9%)
Number of residues in outlier region:		1 (0.5%)

C) ProCheck

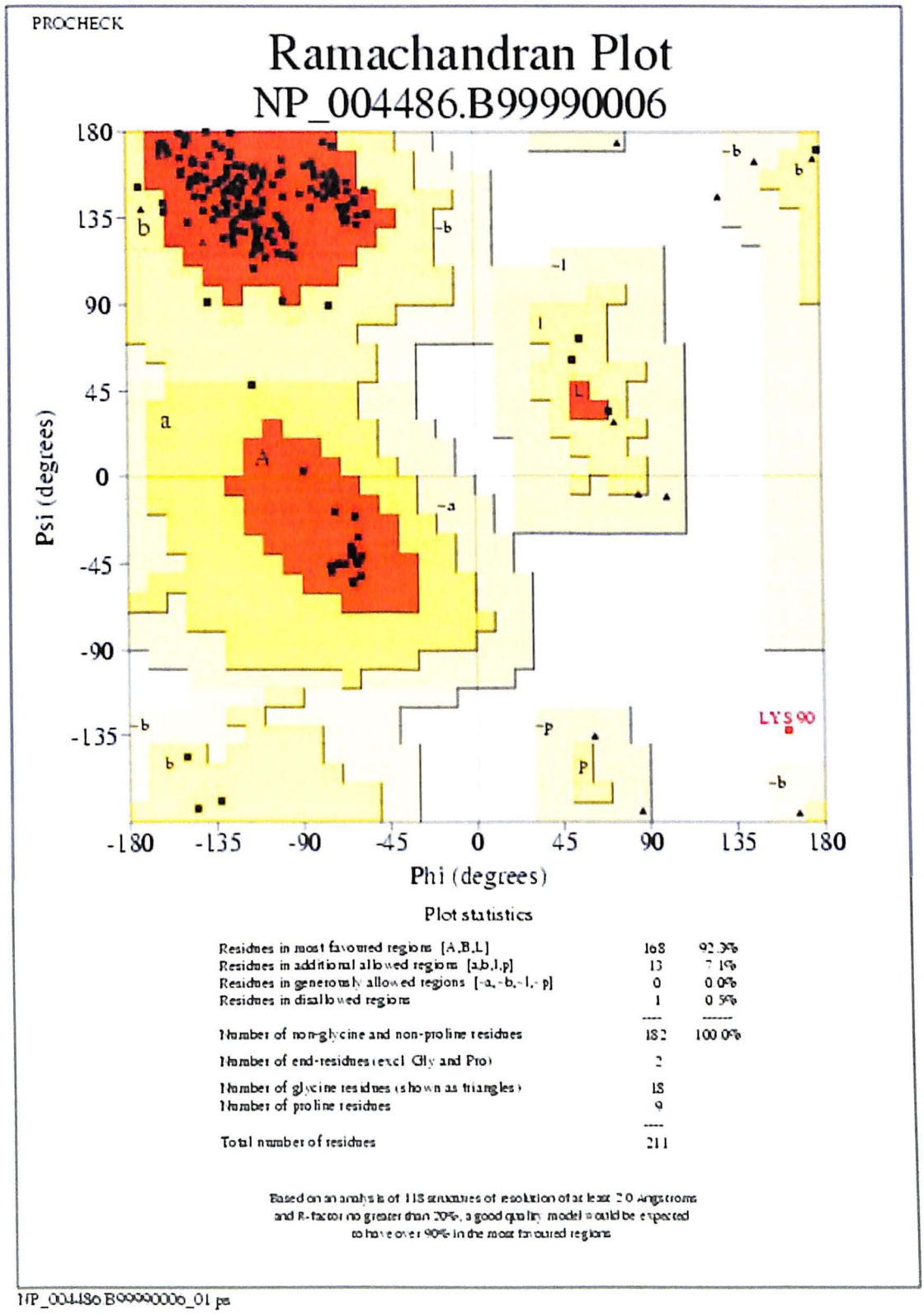


Fig. 14 : Ramachandran Plot.

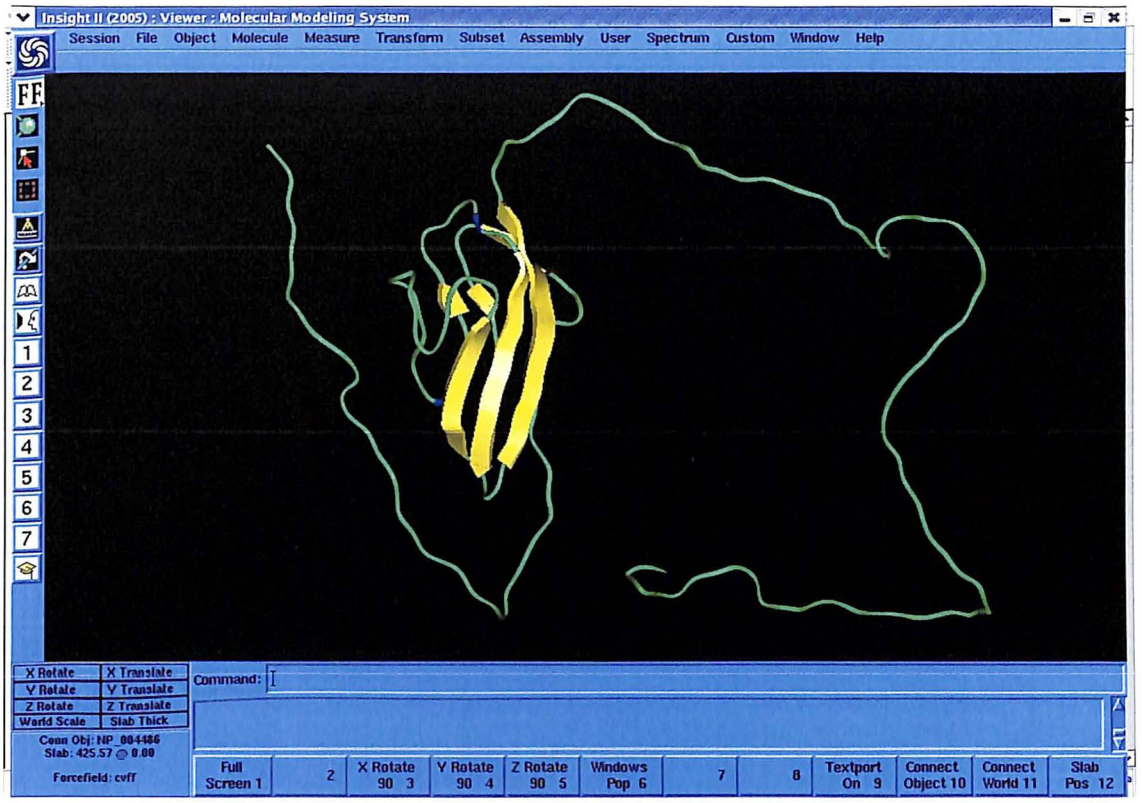


Fig. 15 : Screenshot of Modeled Protein.

So procedure of Homology Modeling starts with the target identification to model refinement. These steps include pair wise sequence alignment of target i.e Neuregulin1 isoform HRG-gamma with the template 2PND, assigning co-ordinates, loop generation and refinement of model. After doing the complete study and procedure of Homology Modeling, finally 10 models were generated in InsightII. The models having high score and list energy i.e 6^{th} model was selected for further analysis for docking. The above figure (fig.15) indicates the final modeled protein having least energy.

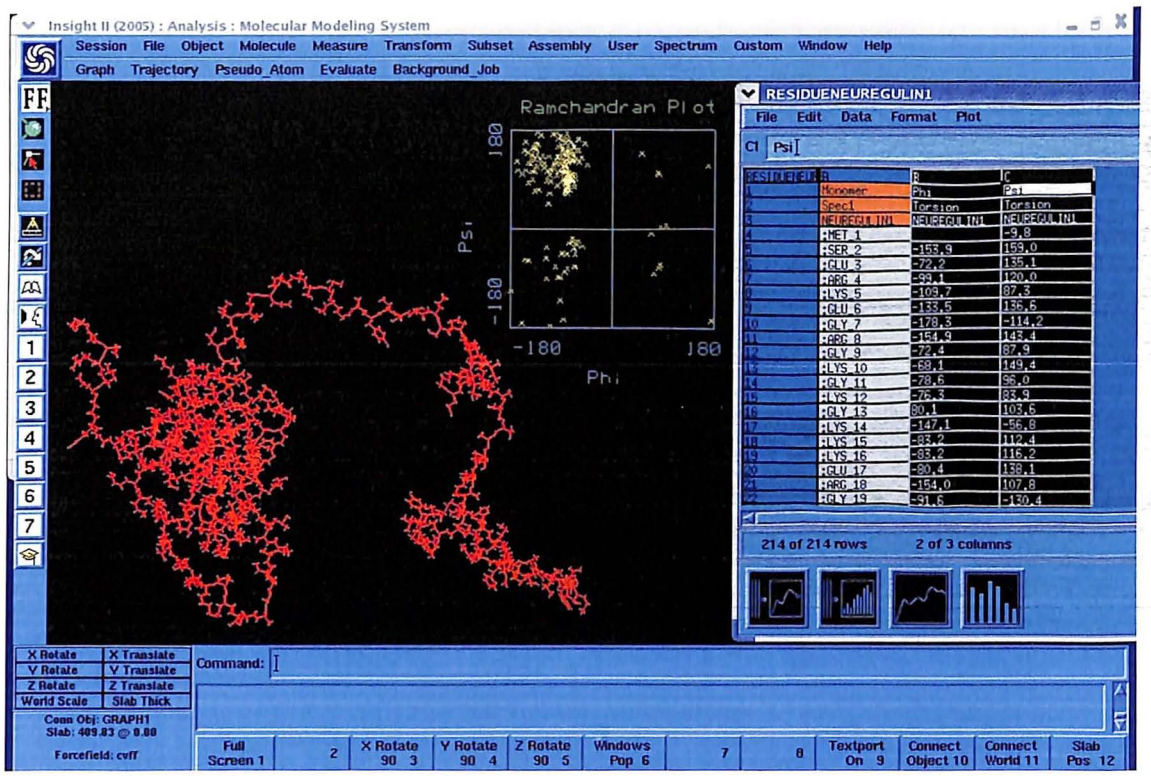


Fig. 16 : Screenshot of Ramachandran Plot on InsightII.

After selection of a good model, its better to analyse the Ramachandran Plot. Here in the above figure (Fig.16) it shows the Ramachandran Plot in InsightII. It indicates phi, psi angles, peptide bond planarity, bond lengths, bond angles, hydrogen-bond geometry, and side-chain conformations of this protein structures as a function of atomic resolution. Thus, the expected values of these parameters are known and can be compared to a modeled structure based on the atomic resolution of the structures from which the model was developed.

5.8 Energy Minimization

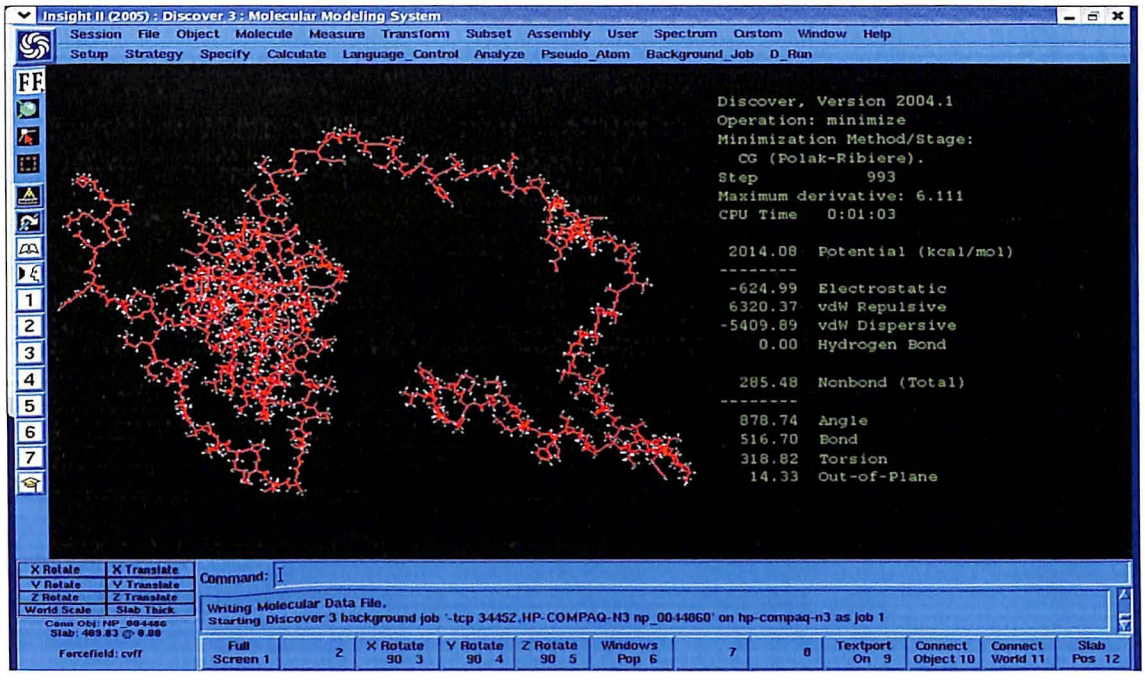


Fig. 17 : Energy minimization.

Energy	Initial	Final
Total Potential Energy	5797.79	2014.08
Electrostatic	2314.23	-624.99
VdW_Repulsive	7131.80	6320.37
VdW Dispersive	-5531.31	-5409.89
Hydrogen Bond	0.00	0.00
Nonbond	3914.73	285.48
Angle	970.55	878.74
Bond	473.61	516.70
Torsion	407.12	318.82
Out of Plane	20.87	14.33

Table 4 : Energy Minimization Summary

Various types of energies and forces are affects the protein molecule. These includes potential energy, electrostatic forces, Vander Waals repulsive, dispersive, hydrogen bond torsion angle etc. Once any irregularities have been resolved, the entire structure may then be subjected to further refinement. This process may consist of energy minimization with restraints, especially for the Structurally Conserved Regions. The restraints then may be gradually removed for subsequent minimizations. It also may be advantageous to apply molecular dynamics in conjunction with energy minimization. For any of these refinement procedures, the structure should be solvated, using for example crystallographic waters from the known homologues, a solvent shell, or a periodic box of pre-equilibrated water molecules.

In the above figure (Fig.17) it is the running mode of energy minimization process. This is done using Discover module of InsightII. Also the steps for iteration was taken as 1000 for better energy minimization and least energy.

So energy minimization helps for the stable conformation of the molecule. Modeling method will produce unfavorable contacts and bonds. Energy minimization is used to regularize local bond and angle geometry Relax close contacts and geometric strain extensive energy minimization will move coordinates away from real structure . so it was kept to a minimum.

5.9 Docking

5.9.1 Docking in GOLD 2.1

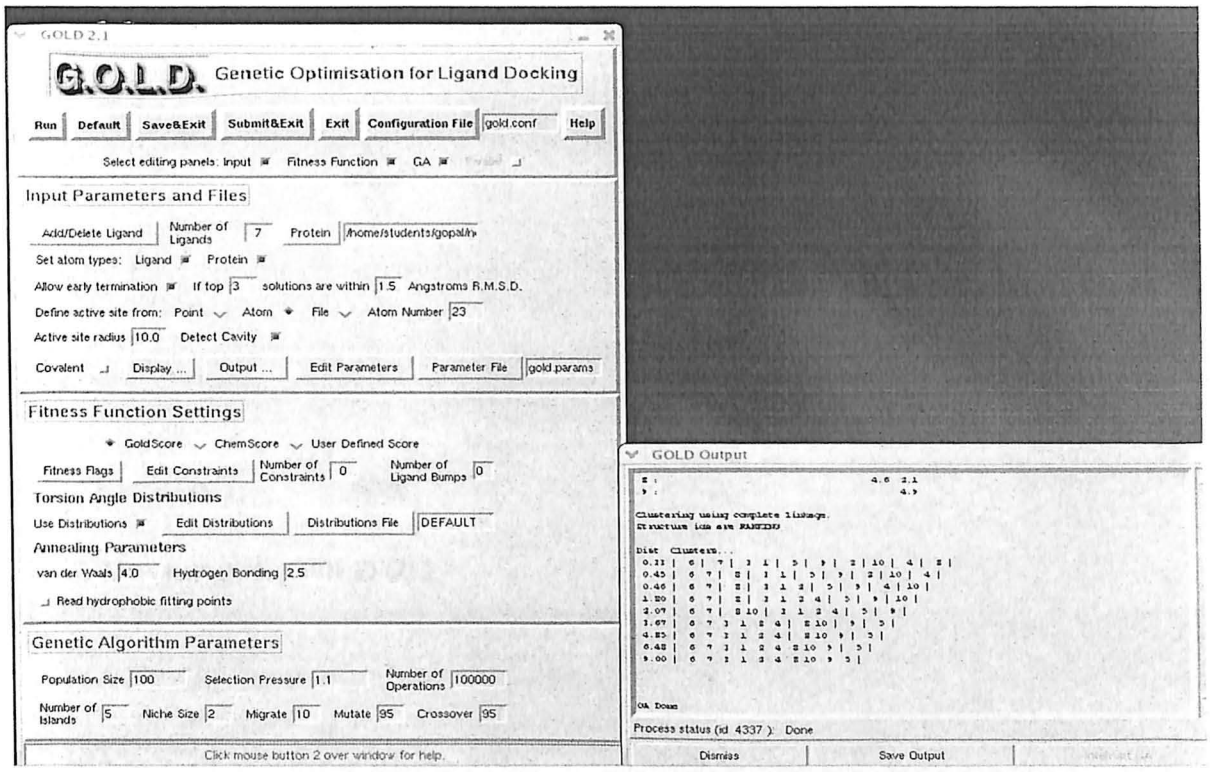


Fig. 18 : Docking in GOLD.

File name	Fitness	S(hb ext)	S(vdw ext)	S(vdw int)
aripiprazol_m1_2.mol2	40.57	4.91	26.19	-0.35
clozapine_m1_2.mol2	37.03	2.98	24.12	0.89
olanzapine_m1_3.mol2	36.67	0.17	25.92	0.86
paliperidon_m1_5.mol2	42.28	9.98	26.39	-3.99
quetiapine_m1_4.mol2	38.33	0.00	30.16	-3.15
risperidon_m1_9.mol2	45.57	19.24	21.23	-2.86
ziprasidone_m1_6.mol2	**50.15**	**21.85**	**21.43**	**-1.17**

Table 5 : Best ranking Results in GOLD.

From the above best ranking file generated by taking seven drug in GOLD Dock, It was Found that Ziprasidone has more fitness score, so Ziprasidone was taken as a ligand, which can successfully bind to receptor Neuregulin 1 isoform HRG-gamma to suppress disease Schizophrenia.

Binding Site of Neuregulin1 isoform HRG-gamma

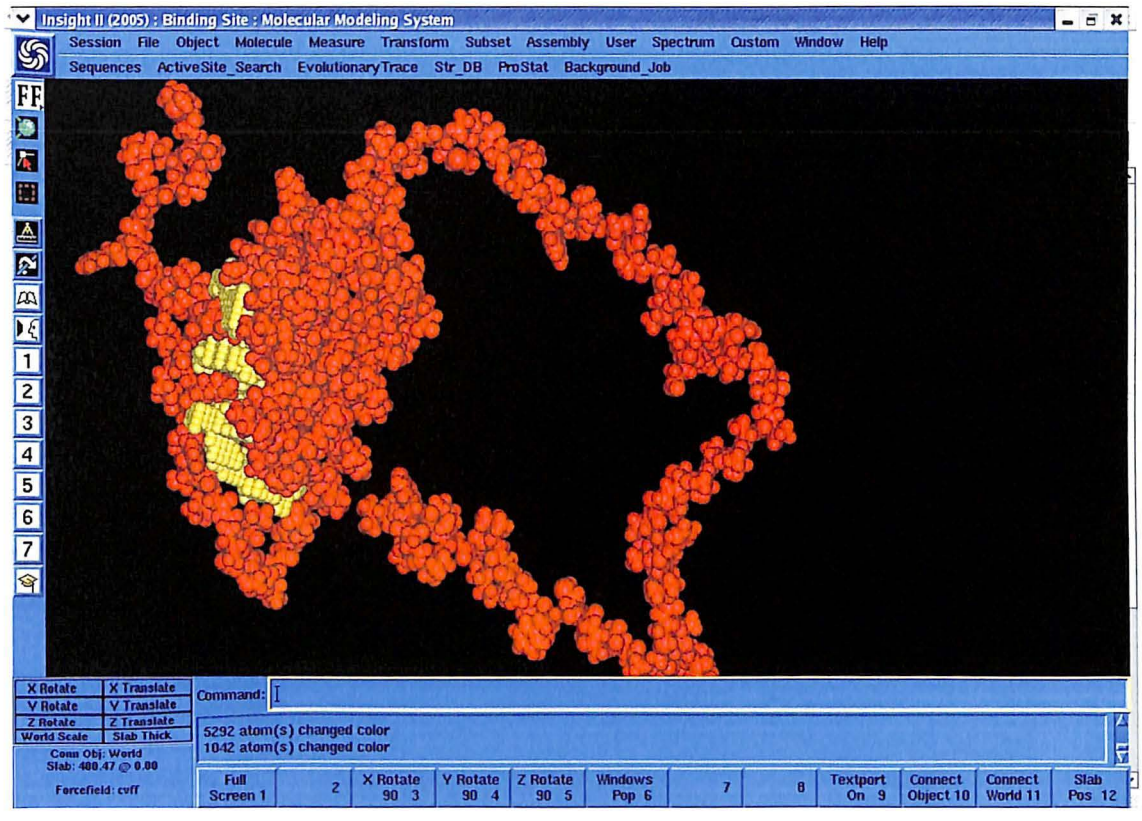


Fig. 19 : Binding Site of neuregulin1 isoform HRG-gamma protein.

The above picture (Fig.19) shows the binding site of the modeled protein i.e Neuregulin1 isoform HRG-gamma. The binding site indicates the appropriate space at which a ligand is perfectly binds to form a receptor-liland complex. Here red colour shows the protein molecule and yellow colour shows the binding site of the protein.

Ziprasidone Ligand

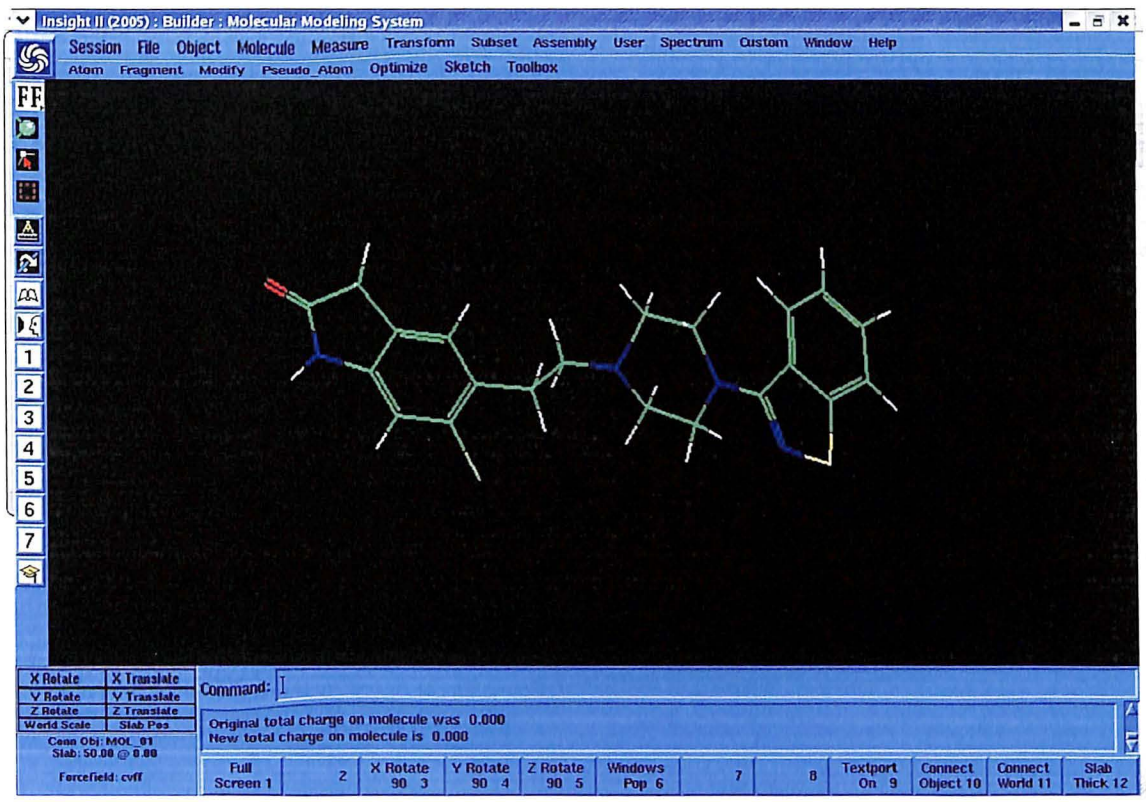


Fig. 20 : Ligand - Ziprasidone in InsightII.

The above figure (Fig.20) indicates the three dimensional structure of Ziprasidone ligand. Originally the ligand is in 2D structure when downloaded from Drug Bank. It was then converted to 3 dimensional structure using InsightII. After this the energy minimization was also carried out for least energy and stable conformation. It also carried out in 1000 iteration steps like the receptor.

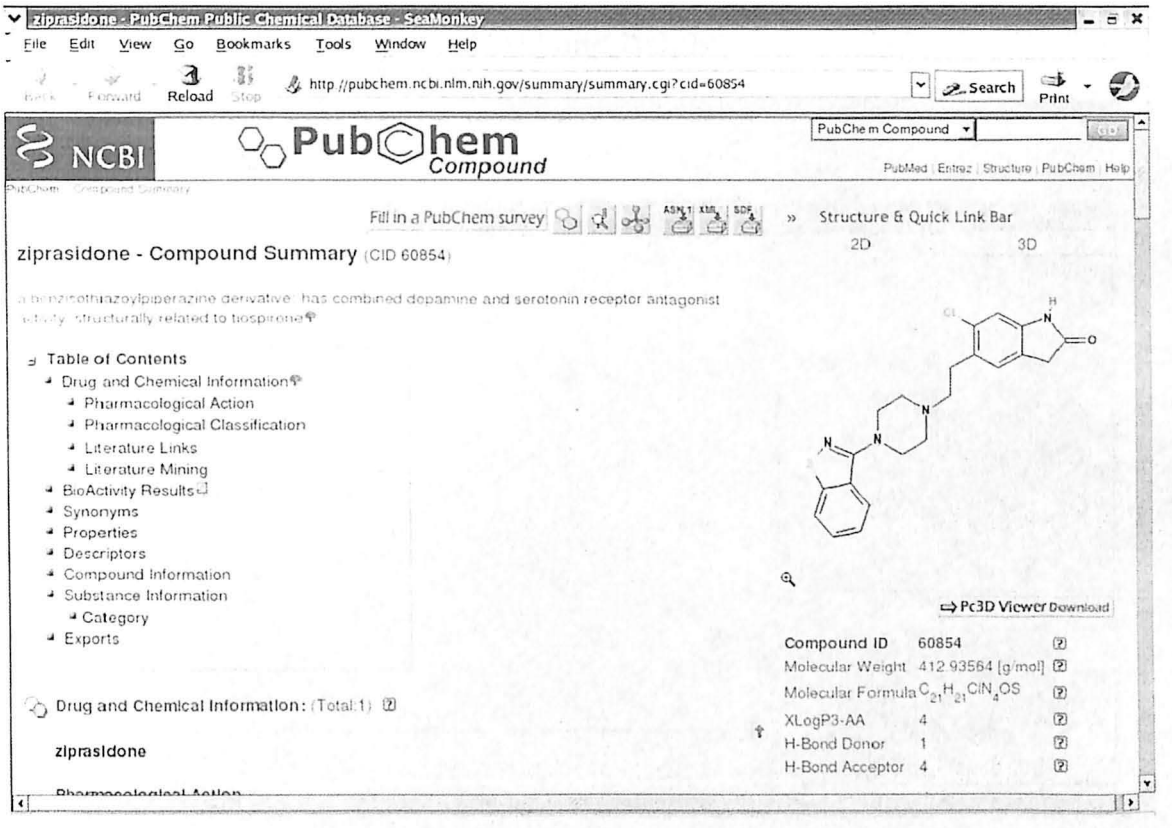


Fig. 21: The ligand Ziprasidone retrieved from PubChem.

Molecular Weight	**412.93564 [g/mol]**
Molecular Formula	**$C_{21}H_{21}ClN_4OS$**
XLogP3-AA	**4**
H-Bond Donor	**1**
H-Bond Acceptor	**4**
Rotatable Bond Count	**4**
Tautomer Count	**3**
Exact Mass	**412.11246**
Mono Isotopic Mass	**412.11246**
Topological Polar Surface Area	**48.5**
Heavy Atom Count	**28**
Formal Charge	**0**
Complexity	**573**
Isotope Atom Count	**0**
Defined Atom Stereo Center Count	**0**
Undefined Atom Stereo Center Count	**0**
Defined Bond Stereo Centre Count	**0**
Undefined Bond Stereo Center Count	**0**
Covalently-Bonded Unit Count	**1**

Table 6 : Properties Computed from the Structure of Ziprasidone.

5.9.2 Docking in InsightII (Grid and Delphi)

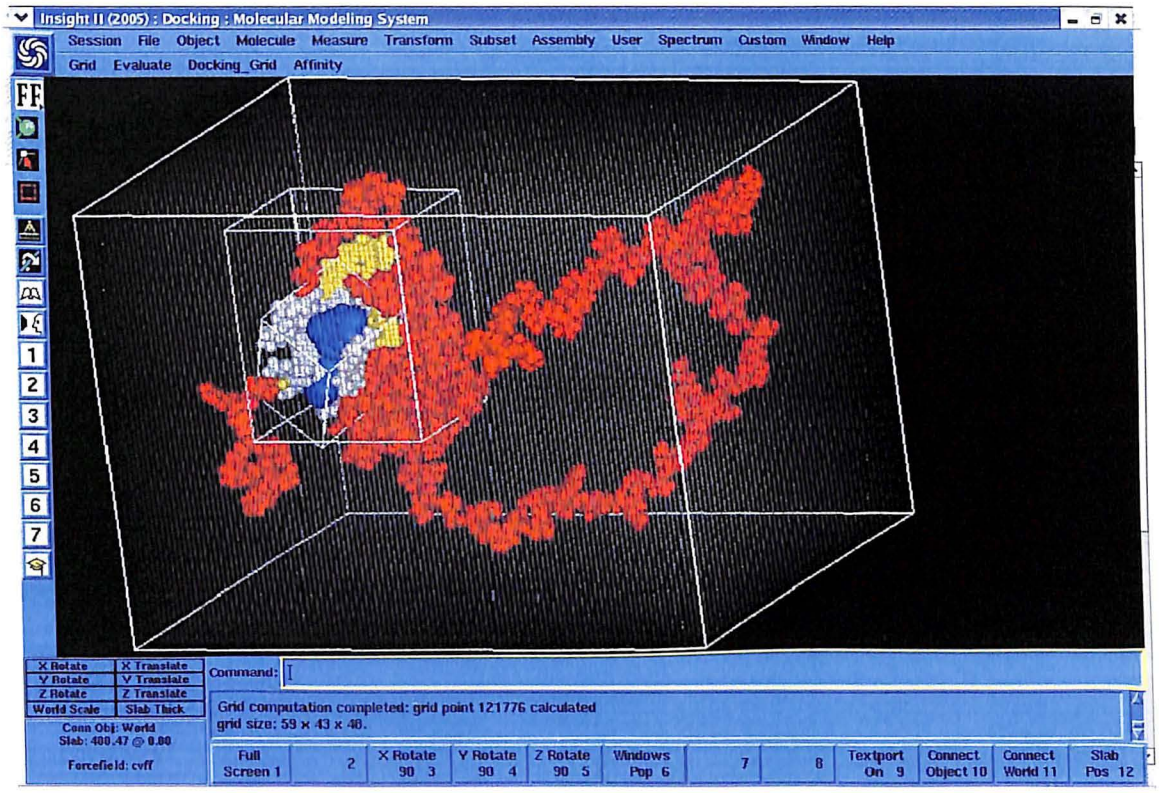


Fig. 22 : Grid Point Display.

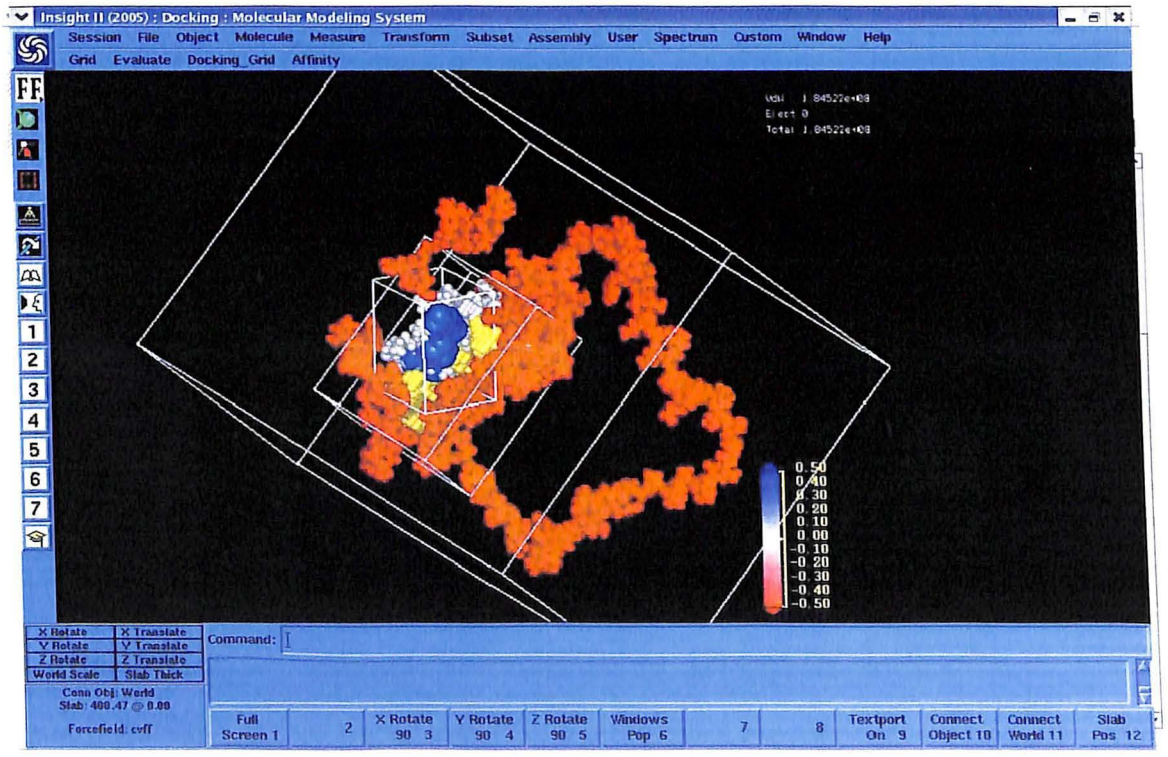


Fig. 23 : Docked Structure in InsightII (Grid Docking).

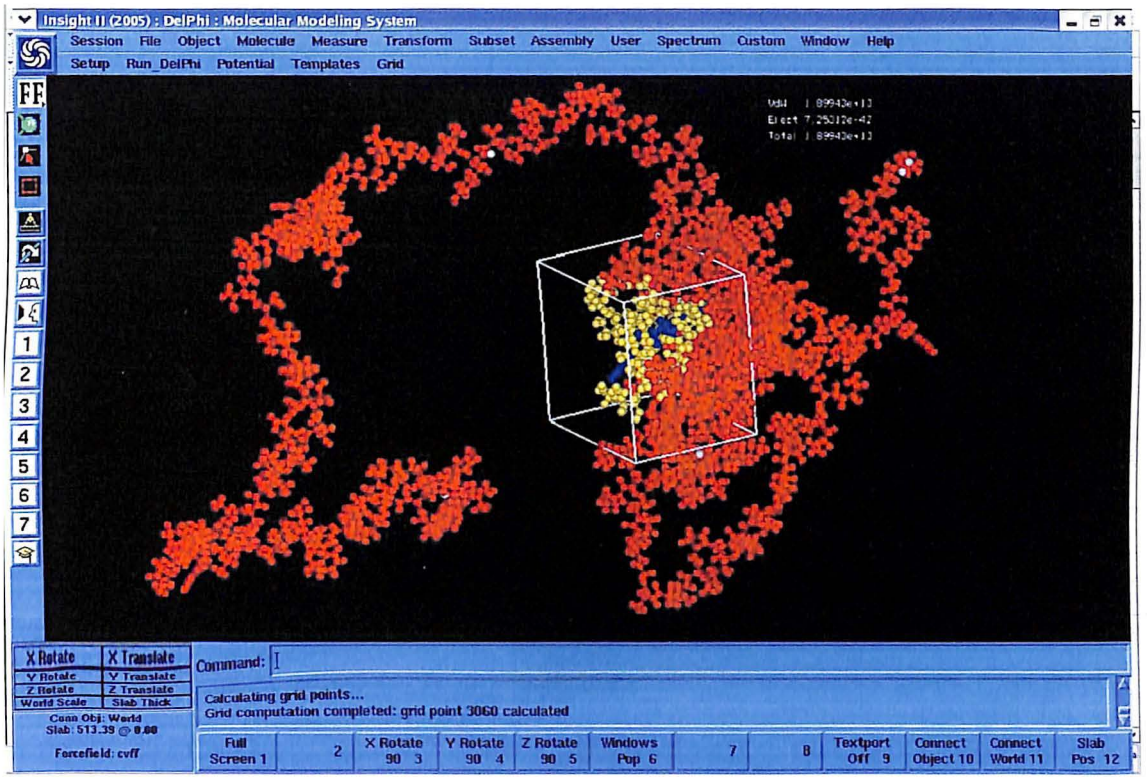


Fig. 24 : Docked Structure in InsightII (Delphi Docking).

Docking provides the binding of a drug (ligand) at the binding site of the desired protein molecule. In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. A ligand is simply a molecule with interacts with a protein, by specifically binding to the protein. In fig.22, it describes about the grid docking and computed the grid points. There are 3 nos. of grids are given. One is for ligand, one is for binding site and the final big one is for receptor. The red colour indicates the receptor, blue is for ligand, yellow is for binding site and white is for assembly. Assembly region is the part of a binding site and is most attractive for ligand binding. Fig.23 shows the final docked structure in Grid Docking. Fig.24 describe about the Delphi Docking in InsightII. It also measures the minimal energy value for stable complex.

5.9.3 Patch Docking

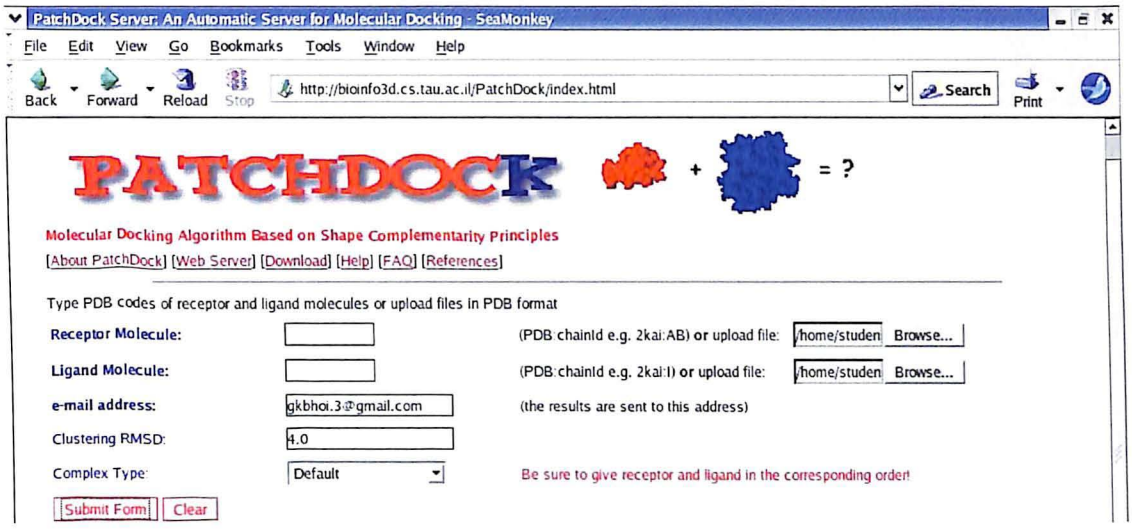


Fig. 25 : Patch Dock Server Page.

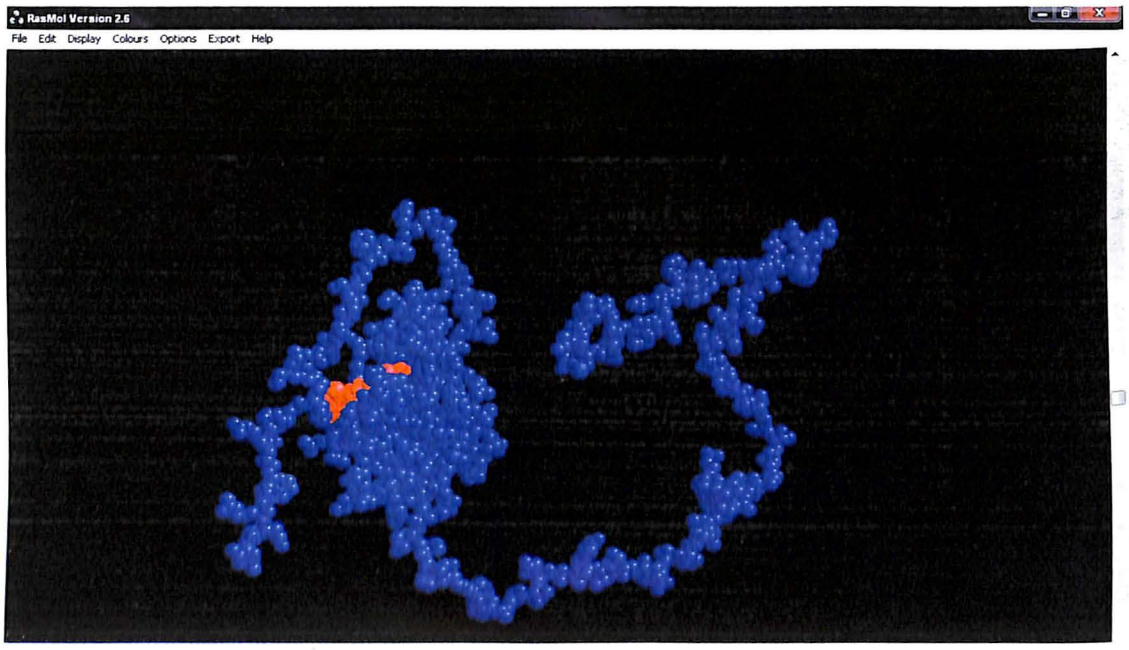


Fig. 26 : Docked Structure viewed in Rasmol.

After submitted the query in Patch Dock, the results was obtained. The .pdb files were downloaded and viewed in Rasmol viewer (Fig.26). The red colour shows the ligand Ziprasidone and blue colour indicates the receptor protein neuregulin1 isoform HRG-gamma.

CHAPTER-V

Discussion

DISCUSSION

The present study on schizophrenia the chronic disease seen round the globe. NRG1 isoform HRG-gamma was found to be the genetic component involved, as the structure of the protein (was essential to know the biological functions as well as to design novel drugs) was not available, homology modeling is usually the method of choice when clear relationship of homology between the sequence of target protein and at least one known structure found. This approach would give reasonable results based on the assumptions that are tertiary structures of two protein will be similar if the sequences are related. The target sequence was first studied about the amino acid composition, atomic composition using the Prot Param which is a primary structure prediction tool. It was found that Neuregulin1 isoform HRG-gamma protein have the no. of amino acids 211, the molecular weight is 22754.8 and theoretical pI is 9.62. It also found each amino acid composition in percentage. The serine having 33 aa shows the highest amino acid composition. Neuregulin1 isoform HRG-gamma contains 3194 nos. of atoms. It also describes the extinction coefficient, estimated half life and instability index. The secondary structure prediction was done using SOPMA and GORIV to know about the α-helix and β-turns. GOR IV shows the α-helix in position 38 is 18.01% while SOPMA shows at the position 12 is 5.69%. Similarly GOR IV shows the no β-turns while SOPMA in position 11 is 5.21%.

In part of modeling the structure, the target sequence was searched for the templates using BLAST P. The template 2PND (structure of Murin Crig) was selected having sequence identity 38% and gap 0.0%. Using ClustalW The multiple sequence alignment has been done by taking the template and target sequences as the parameters. It shows in a distinct colour for identifying the amino acid. The Homology Modeling was done using InsightII software package. Then it goes pair wise sequence alignment between target and template to the model generation. It involved :

1. Identify homolog protein and determine the extent of their sequence similarity with one another and the unknown.
2. Sequence alignment.

3. Generation of co-ordinates of the unknown structure from those of the known structure.
4. Generate conformations for the loops in the unknown structure.
5. Refinement, model building and evaluate the unknown structure.

The target-template file was provided as an input and 10 models were generate. The structural validation was then done to find the best structure using Verify 3D tool. The best model among 10 homologous models were chosen on the basis of the best score acquired by Verify 3D. It shows 6^{th} model have the highest score i.e. 0.54 having least energy. The Ramachandran Plot was then performed to find the phi-psi dihedral angles. It showed a normal distribution of points with phi values and psi value clustered in a few distinct regions with 92.3% and 7.1% of residues occupying most favoured and allowed regions respectively. 0.5% of residues were in disallowed regions. The energy of the modeled structure was minimized to reach stable conformational state. The energy minimization was carried out using Discover_3 module of InsightII. Steepest and Conjugate are the algorithms that were used to minimize the energy of the molecule. Also the Ramachandran plot analysis was carried out in InsightII.

To perform docking the binding site of the protein molecule was found by active site search in Binding site module of InsightII. The ligand (Drug) was searched in Drug Bank, NCBI Pub Chem etc. The ligands Aripiprazol, Clozapine, Olanzapine, Paliperidon, Quetiapine, Risperidon and Ziprasidone were selected as the inhibitors which helps for treatment of schizophrenia. For find out the suitable drug GOLD 2.1 was performed which shows the results of Best ranking files. In that file it contains the fitness value. The highest value was found in Ziprasidone having 50.15 fitness. I choose Ziprasidone as best inhibitor because it exert best fitness score comparison with other. The structure of Ziprasidone was built using Builder domain and optimized using InsightII to proceed with docking. Three types of docking analysis were carried out. Out of which two are (Grid Docking and Delphi Docking) using InsightII and the other i.e. Patch Dock which is a on line tool. The results of Patch Docking was viewed in Rasmol viewer tool.

CHAPTER-VI

Summary

SUMMARY

After complete studied in literature review it was found that Schizophrenia is caused by Neuregunin1 protein. The NRG1 gene encodes neuregulin1, a signaling protein that mediates cell-cell interactions and plays a critical role in the growth of organ systems. Neuregulin1 is a protein which in humans is encoded by the NRG1 gene. Genetic linkage and association have implicated neuregulin1 as a schizophrenia susceptibility gene. NRG1 is one of the neuregulin family of proteins, which have a broad range of bioactivities in the central nervous system and contain an epidermal growth factor.

Schizophrenia is a genetically brain disorder in its clinical expression both genetic and environmental factors are thought to contribute to the pathogenesis of The Schizophrenia. Predisposing genetic influences include associations with NRG1 isoform HRG-gamma. It is essential to know the structure of the protein in order to know its functioning which is in turn important to design novel drugs. The secondary structure of the protein has been carried out. The 3D structure of the protein was modeled using InsightII. This modeled structure showed a minor but important structural difference to the template protein. The model was validated and docking has been performed. The basic computational and experimental strategies used in this project will be of general utility for the analysis of other protein ligand interactions. As the structure of NRG1 isoform HRG-gamma was modeled it will be very useful in designing other inhibitors that can suppress the disease more effectively than the present.

Scientists worldwide are studying schizophrenia so they will be able to develop new ways to prevent and treat the disorder. The only way it can be understood is for researchers to study the illness as it presents itself in those who suffer from it.