Homology Modeling, Docking and Protein analysis of L1 CAM 5 using Tetrahydrobiopetrin as a ligand

Dissertation

Submitted to the Orissa University of Agriculture & Technology, Bhubaneswar in partial fulfillment of the requirement for the award of degree of

MASTER OF SCIENCE IN BIOINFORMATICS

BY

SNEHALATA SAHU

Adm. No-29BI/08

DEPARTMENT OF BIOINFORMATICS

CENTRE FOR POST GRADUATE STUDIES

ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

BHUBANESWAR-751003

2010

Name of the Advisor **Mr. Surya Narayan Rath**

DEDICATED TO MY

BELOVED PARENTS

&

ADVISIOR

Orissa University Of Agriculture & Technology
Department Of Bioinformatics
Centre For Post Graduate Studies
Bhubaneswar

Mr. Surya Narayan Rath
Asst. professor of Bioinformatics

CERTIFICATE –I

This is to certify that thesis entitled **"Homology Modeling , Docking & protein analysis of L1 CAM 5 using Tetrahydrobiopterin as a Ligand"** submitted for award for the degree of **Master of Science** in the subject of **Bioinformatics** embodies a faithful bonafied research work carried out by **Snehalata sahu** (Adm. No. 29BI/08) under my guidance & supervision. No part of this thesis has been submitted by her 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 her.

Place- Bhubaneswar

Date- 20.8.10

SnRath
20.8.10

Mr. Surya Narayan Rath

Asst.Proff. cum Chairman
Advisory committee

CERTIFICATE –II

This is to certify that the dissertation entitled **"Homology Modeling ,Docking & protein analysis of L1 CAM 5 using Tetrahydrobiopterin as a Ligand"** submitted by **Snehalata sahu** to the Orissa University Of Agriculture & Technology, Bhubaneswar in the partial fulfillment of the requirements for the award of the degree of **Master of Science** in **Bioinformatics** has been approved by the students advisory committee after an oral examination of the same in collaboration with external examiner.

ADVISORY COMMITTEE

1. Mr. S.N. Rath — Chairman
   Asst.professor
   Department of Bioinformatics

2. Mr. S.K. Pradhan — Member
   Head
   Department of Bioinformatics

3. Miss S.R. Martha — Member
   Asst. professor
   Department of Bioinformatics

4. Mr. Abhimanyu Das — Member
   Head
   Department of CSA

External Examiner

ACKNOWLEDGEMENT

At a very outset, I express my profound gratitude to Mr.Surya Narayan Rath(Asst.professor), Dept Of Bioinformatics, OUAT, bhubaneswar, Orissa for encouraging me to carry out the dissertation work under his guidance.

I take this opportunity to express my deep sense of gratitude to Dr . N . Pradhan , prof &head of psychopharmacology dept. under whose excellent guidance , support this research project was undertaken.This training that I received in his laboratory will be extremely beneficial for me to purse research works in the field of computational biology.

I gratefully acknowledge Dr. D. Nagaraja ,Director of NIMHANS for giving me an opportunity to work in this institute. I thank Mr. K.Shankaracharaya system administrator for his timely help regarding softwares & network.

I am thankful to the member of the advisory committee Prof. Mr. S. K. Pradhan, Head , Dept. of Bioinformatics, O.U.A.T. & for his support ,help & encouragement. Miss S.R Martha Asst. Prof. dept Of Bioinformatics , Mr. A. Das Head of Dept. Of CSA for their support and encouragement to carry out this work successfully.

I would like to owe my sincere thanks to the Respected Ex-Dean,, Dr. B.K.Mishra & the present Dean Dr. H.K. Senapati for the encouragement.

I feel great pleasure in expressing deepest regards and cordial thanks to Miss S. Balbantray mam for continous guidance and support with out whom this would not be possible. Also I owe sincere thanks to for their guidance, & Miss Sudipta Mohanty for her support.

I would like to express my heartiest and cordial regards to my beloved Papa, Maa, my dearest bhai who are the real inspiration for me in every step of my life giving me unbound emotional support and my friend Jyoti, Sadashiv,Laxmipriya for unbound love and moral support. I feel great pleasure to express my love to all my sweetest friends for their strong mental support through out the project period & in these two years.

Last but not the least I express my gratitude to my beloved almighty, and lord Shiva for their invaluable inspiration for accomplishment of such a splendid work.

Snehalata Sahu

Snehalata sahu

Name of the Student	**: Snehalata sahu**
Admission No	**: 29BI/08**
Title Of Thesis	**:Homology Modelling,Docking & protein analysis of L1 CAM 5 using Tetrahydrobiopterin as a ligand.**
Degree for which thesis submitted	**:Master of Science in Bioinformatics**
Name of the Dept, College, & University	**:Department of Bioinformatics, Cetre for Post Graduate Studies, Orissa University Of Agriculture &Technology, Bhubaneswar, Orissa, 751003**
Year of submission	**: 2010**
Name of the advisor	**: Mr. Surya Narayan Rath.**

ABSTRACT

This thesis elucidates the attempt of Modeling and Docking the L1 CAM 5 protein & ligand Tetrahydrobiopterin using computational methods. Physical, chemical, biological properties, binding property and also evolutionary history of the protein is studied in order to validate & provide insights for the future drug development projects. The receptor protein L1 CAM 5 was taken as target, after blast search the protein 1U13pdb found 35% similarity with the target ,taken as template. Using the template structure ten different models were generated by Modeller9v7.The best model was chosen with highest verify 3D score. Then Drug Bank & KEGG database were searched and Tetrahydrobiopterin wih Acc. no. B9A061 was chosen as ligand molecule. The receptor and ligand structure were docked by using INSIGHT-II software. The ligand was placed into crude binding site representation via docking into averaged property fields derived from knowledge based potentials. The results were compared and the best docked structure was obtained. The predicted theoretical model of target receptor and the docking results would provide a boost in future for structure based drug design.

Mr. Surya Narayan Rath
Advisor

Snehalata Sahu
Author

CONTENTS

LIST OF FIGURE

LIST OF TABLE

ABBRIVIATIONS

SBDD	:	Structure Based Drug Design
MSA	:	Multiple Sequence Alignment
PDB	:	Protein Data Bank
CADD	:	Computer Aided Drug Design
FF	:	Force Field
SCRs	:	Structurally Conserved Regions
VRs	:	Variable Regions
ALI	:	Alignment File
MDF	:	Molecular Dynamics File
MSI	:	Molecular Simulation
PRO-CHECK	:	Protein check
CVFF	:	Consistence Valence Force Field
INP	:	Input File
SAVS	:	Structure Analysis & Verification Server
SOPMA	:	Self Optimized Prediction Method Alignmement
GOR	:	Garnier Osguthorpe & Robson

Chapter-1

INTRODUCTION

INTRODUCTION

In the recent years there has been growing interest in computer based screening. One of the driving forces 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 molecule modeling dealing with receptor & ligand has undergone a drastic transformation in the area of drug designing. In the series of molecular modeling experiments we addresses a number of fundamental questions that need exploration at a quantum chemical level of atomic bonding. At present computer aided drug designing (CADD) has replaced classical medicinal chemistry. Further molecular modeling & cellular automata model; have opened new avenues to understand the structure of biomolecules,drugs,protein-ligand interaction & de-novo modeling. The recent development focuses on our ability to access the information content in biological macromolecule. Information stored in the structure of the molecule is a function of their physical & chemical property. The more important path breaking development has been the ability to manipulate this information by virtue of understanding the structure of nuclic acid, polysaccharides & protein-ligand interaction.

Modeling the interaction of a drug with its receptor is a complex problem.The forces involved are the intermolecular association, hydrophobic, van der Waals interaction hydrogen bonding & electrostactic force of attraction. The major driving force for ligand building appears to be control by hydrogen bonding & electrostatic interaction. Modeling the intermolecular interaction in a ligand protein complex is termed difficult due to the involvement of many degress of freedom in a system & insufficient knowledge of the effect of solvent on the binding association. The interaction of the drug with the receptor site is stereo specific that receptor recognizes certain groups on the ligand, particularly the intermolecular distances & 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 & the

ligand, there 3D structure is essential. In the absence of experimentally determine structure, homology modeling plays an effective role in elucidating the structure.

FORMS OF PROTEIN

(1)PRIMARY STRUCTURE

The primary structure of protein is the amino acid sequence of polypeptide chain.

The amino acids are arranged in linear order with peptide bonds. The chain has a free –NH2 terminal & a free –COOH terminal.

(2)SECONDARY STRUCTURE

The polypeptide chain undergo folding to maintain a stability known as secondary structure of protein.

Two types of secondary structure i.e

(a)alpha helix

(b)beta helix

(a) alpha helix

- It is the spiral structure of protein.
- Alpha helix stabilized by extensive hydrogen bond.
- Right handed alpha helix is more stable than left handed helix.

(b) beta sheet

- Backbone of polypeptide chain is extended into a Zigzag rather than helical structure.
- Hydrogen bonds are formed between adjacent segments of polypeptide chain.
- Zigzag polypeptide chain arranged side by side.

(3)TERTIARY STRUCTURE

- 3-D arrangement of protein structure is reffered to as tertiary structure.
- It is stabilized by weak interatomic forces such as hydrogen bonds, ionic bonds, disulfide bonds, hydrophobic bonds, vanderwaals interaction, covalent bond.

  e.g: Myoglobin

(4)QUATERNARY STRUCTURE

- Some proteins contain 2 or more separate polypeptide chains. Arrangement of these protein subunits in 3-D complexes constitute quaternary structure.
  e.g.:Haemoglobin

SIGNIFICANCE OF PROTEIN STUDY

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 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 vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responces, cell adhesion, and the cell cycle. Proteins are also necessary in animals diets, since animals cannot synthesize all the amino acids they need and obtain essential amino acids from food. Through the process of digestion, animals break down ingested protein into amino acids that are then used in metabolism.

Proteins are fundamental components of living cells and include many substances, such as enzymes, hormones, and antibodies that are necessary for the proper functioning of an organism. They are essential in the diet of animals for the growth and repair of tissue. Proteins are essential building blocks for living systems hence their name, drawn from the Greek proteios, or "holding first place". Proteins are integral to the formation of DNA, a molecule that contains genetic codes for inheritances, and of hormones. Most of the dry weight of the body and the bodies of other animals is made of protein, as is a vast range of things with which we come into contact on a daily basis.

PROTEIN SYNTHESIS

- Proteins are the polymerase of amino acids. There are 20-varieties of amino acids which are taking part in protein synthesis.
- The amino acids are joined by peptide bonds in the specific sequence to form a polypeptide(protein).
- During protein synthesis the coded message of DNA is transcribed to m-RNA and then the information is translated into amino acid sequence of a polypeptide chain.Ribosome is the site of protein synthesis.

  Protein synthesis takes place following steps:

(a)Transcription of m-RNA on DNA.

(b)Activation of amino acids.

(c)Attachment of activated amino acids with tRNA.

(d)Translation.

FUNCTION OF PROTEIN

- Proteins act as enzymes, hormones and antibodies. They maintain fluid balance and acid base balance. They also transport substances such as oxygen, vitamins and minerals to target cells throughout the body.
- Structural proteins, such as collagen and keratin, are responsible for the formation of bones, hair and the outer layer of skin and they help maintain the structure of blood vessels and other tissues. In contrast motor proteins use energy and convert it into some form of mechanical work.
- Enzymes are proteins that facilitate chemical reactions without being changed in the process. The inactive form of an enzyme is called a proenzyme.
- Hormones (chemical messengers) are proteins that travel to one or more specific target tissues or organs and many have important regulatory functions. Insulin for example- plays a key role in regulating the amount of glucose in the blood.
- The body manufactures antibodies (gaint protein molecules) which combat invading antigens. Antigens are usually foreign substances such as bacteria and viruses that have entered the body and could potentially be harmful. Immunoproteins also called immunoglobulins or antibodies defend the body from

possible attack by these invaders by binding to the antigens and inactivating them. Proteins help to maintain the body's fluid and electrolyte balance.

- Proteins ensure that the proper types and amounts of fluid and minerals are present in each of the body's three fluid compartments. These fluid compartments are intracellular (contained within cells), extracellular (existing outside the cell) and intravascular(in the blood). Without this balance the body cannot function properly.
- Proteins also help to maintain balance between acids and bases within body fluids. The lower a fluid's pH the more acidic it is. Conversely the higher the pH the less acidic the fluid is. The body works hard to keep the pH of the blood near 7.4 (neutral).
- Proteins also act as carriers transporting many important substances in the bloodstream for delivery throughout the body. We have seen that each type of protein consists of a precise sequence of amino acids that allows it to fold up into a particular three-dimensional shape or conformation. But proteins are not rigid lumps of material.
- They can have precisely engineered moving parts whose mechanical actions are coupled to chemical events. It is this coupling of chemistry and movement that gives proteins the extraordinary capabilities that underlie the dynamic processes in living cells. Especially mitochondrial proteins play a vital role in living cells because it provides energy to the cells.

FUNCTIONAL AND STABLE STATE OF PROTEIN

The tertiary structure of protein is more functional because we found covalent, vander wall, ionic, hydrogen, hydrophobic, peptide and disulphide bond in this state.

PROTEIN STRUCTURE DETERMINATION IS NECESSARY

It essentially implements all the steps necessary to produce a high quality model of a protein. The whole process is fully automated and a potential user only submits the protein sequence.

Predicted protein structures can be used if very close homologs with known structure are available, but in most cases rational drug design requires iterative co-

crystallization of the protein–ligand complexes. Currently available structure prediction methods do not allow for high-quality predictions of the quaternary structure of protein complexes and for the prediction of interactions between proteins. Structure prediction methods have been used successfully in combination with sparse restrains obtained from nuclear Overhauser effects, residual dipolar couplings or backbone chemical shifts [96,97,98]. Recently, it was demonstrated that the PDB library is most likely complete for single domain protein structures at low to moderate resolution .

Determining protein 3D structure is one of the greatest challenges in computational biology. Nuclear magnetic resonance (NMR) spectroscopy is the second most popular method (after X-ray crystallography) for structure prediction. Given a molecule, NMR experiments provide upper and lower bounds on the interatomic distances. Thus, a number of heuristics are used in practice to generate 3D structures from NMR data. Because the quality of the computed structures and the time required to obtain them depend on the bound tightness, it is necessary to use efficient procedures for bound smoothing.

TRADITIONAL METHOD AVAILABLE FOR PROTEIN STRUCTURE DETERMINATION

Traditional methods, when applied to a family of proteins rather than a single sequence proved much more accurate at identifying core secondary structure elements. These predictions are actually much more useful than those for single sequence, since they tend to predict the core accurately. Probably the most famous early methods are those of Chou & Fasman, Garnier, Osguthorbe & Robson (GOR) and Lim. Although the authors originally claimed quite high accuracies (70-80 %).

LIMITATION

Finally, we sought to improve the biological utility of secondary structure prediction by identifying the subset of the predictions that are most likely to be correct.Using this approach, we found that the top 28% of the predictions were 860% accurate and the top 430% of the predictions were 81% accurate. These results indicate that, notwithstanding the limitations on overall accuracy of secondary structure prediction, a substantial proportion of a protein can be predicted with considerable accuracy.

PROTEIN STRUCTURE PREDICTION METHODOLOGY

Sequence–sequence,. profile–sequence, sequence–profile comparison methods represent a traditional evolutionary-based approach to predict structures of proteins. The simplest method aligns the sequence of the target with the sequence of the template using a substitution matrix. More sensitive methods define scores for aligning different amino acids separately for each position of the target sequence (PSI-BLAST) or the template sequence (RPS-BLAST). The scores are taken from the analysis of sequence variability in multiple alignments of the corresponding sequence families. Such position-specific scores are also called profiles. They are similar in format to the representation of sequence families used by prediction methods based on HMMs. Profile–profile comparison methods utilize the profiles generated by the sequence alignment methods. Instead of a lookup of a substitution score, they compare two vectors with each other when building the dynamic programming matrix used to draw the alignment. The comparison is usually conducted by calculating a dot product of the two positional vectors (as shown in the figure) or by multiplying one vector times a substitution matrix time the other vector. Depending on the choice of the comparison function the vectors are often rescaled before the operation. The sequence variability vectors are sometimes also augmented with meta information, such as predicted secondary structure as indicated . Threading or hybrid methods utilize the structure of the template protein in the comparison function. The position-specific alignment scores are computed for the template protein by replacing the side-chain of a residue with side-chains of all possible amino acids and by calculating the resulting substitution scores using statistically derived contact potentials. In addition, factors such as matching of predicted and observed secondary structure or burial preferences are also taken into account when aligning two positions. Most threading methods use frozen approximation where the sequence is threaded through the template structure and contacts are calculated between the target side-chain and side-chains of the residues of the template. In the much slower, defrosted threading template side-chains are replaced with side-chains of the target according to the alignment before calculating the contact scores. *Ab initio* methods represent a physical approach to predict the structure of the target protein. The methods are based on an energy function, which estimates the conformational energy of the chain of the modeled

protein. The energy can be calculated in a similar fashion as in the threading methods, i.e. utilizing contact potentials. The advantage of *ab initio* is that the database of folds does not constrain the set of possible results and theoretically any conformation can be generated and tested. *Ab initio* methods differ in employed energy functions and in the way conformational modifications are generated. Most common methods employ fragment insertion techniques or constrain the move set by placing the molecule on a lattice. Meta predictors represent statistical approaches to improve the accuracy of protein structure predictions. Simple meta predictors collect models from prediction servers, compare the models and select the one, which is most similar to other models. The consensus model corresponds to a model selected from the collected set and represents the final prediction. More advanced meta predictors are able to modify the set of collected models either by filing missing parts with *ab initio* or loop modeling or by creating hybrid models from segments of structures collected from prediction servers. Hybrid models have a higher chance to provide a more complete model but are sometimes unphysical in terms of chain connectivity.

PROTEIN LIGAND INTERACTION

Biology functions via interaction between macromolecules such as nucleic acids, protein, carbohydrates & other biological molecules of all masses from inorganic ions to high molecular weight macromolecules. The very essence of this interaction is there exquisite sensivity & selectivity. Protein molecules function in there 3D state by recognition of another molecule, ligands. The specific recognition of ligands requires a rough geometrical match between the surface of protein & the ligand molecules. However the chemical complimentarity or fit between the binding site & the ligand is most important factor enabling high specific recognition of ligand molecules.

The classic view of the ligand protein binding is the "Lock and Key" concept (Fischer 1894) in which the ligand "Key" act as the complimentary to the binding site "Lock". Part of the interaction between the protein & ligand is simply the stearic fit the two pieces.i.e. similarly to a key fitting to a lock.

One method that can be applied to generate reasonable models of protein structure is Homology Modeling.

This procedure also termed comparative modeling or knowledge-based modeling , developes a three-dimensional model from a protein sequence based on the structure of homologous proteins. Several reviews on this topic have appeared. In the description that follows, some aspect of homology modeling that I may find useful in this course & in my research were discussed.

HOMOLOGY MODELING

In protein structure prediction, homology modeling also known as comparative modeling is a class of methods 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 (known as "templates") 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 protein sequences, detectable levels of sequence similarity usually imply significant structural similarity.

All Homology Modeling approaches consists of three steps:

- Finding homologus PDB files.
- Creation of the alignment using single or multiple alignments: Analysis of alignments; gap deletions and additions; secondary structure weighting.
- Structure calculation and model refinement . Given a correct alignment on a related template several methods can produce an accurate model, while without a correct alignment no method produce a good model.

GENERAL PROCEDURES TO CONSTRUCT HOMOLOGY MODELING

- Selection of target sequence
- Searching for one or more template
- Target-template alignment
- Structure building & refinement
- Model evalution
- Docking

APPROACHES OF HOMOLOGY MODELING

Consists of basic 3 steps:-

Step-1:- Finding homologous PDB files

Step-2:- Creation of alignment using multiple sequence alignment, Analysis of alignment Gap deletion & addition, secondary structure weighting etc.

Step-3:- Structure calculation & model refinement.

IMPORTANCE OF HOMOLOGY MODELING

Homology modeling plays a central role in determining protein structure in the structural genomics project. The importance of homology modeling has been steadily increasing because of the large gap that exists between the overwhelming number of available protein sequences and experimentally solved protein structures and also more importantly because of the increasing reliability and accuracy of the method. In fact, a protein sequence with over 30% identity to a known structure can often be predicted with an accuracy equivalent to a low-resolution X-ray structure. The recent advances in homology modeling, especially in detecting distance homologues, aligning sequences with template structures, modeling of loops and side chains as well as detecting errors in a model have contributed to reliable of protein structure which was not possible even several years ago. The ongoing efforts in solving protein structures, which can be time-consuming and often difficult will continue to spur the development of a host of new computational methods that can fill in the gap and further contribute to understanding the relationship between protein structure and function.

MOLECULAR MODELING

Molecular modeling is a collective term that refers to theoretical methods and computational techniques for deriving representing & manipulating the structures & reactions of molecules & those properties that are dependent on these 3-dimensional structures. In a natural way molecular modeling treats molecule as a collection of weights connected with springs where the weights represent the nucleic & the springs represent the bonds. Using professional programs & supercomputers one can search for the most stable structure of a given molecule (energy minimization) or simulate its oscillation at

any given temperature. These techniques are useful for the refinement of conformations of molecules determined by physical 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's important in designing new drugs, which have to interact in a desired way with for instance cell receptors.

COMPUTATIONAL APPROACH

Automated methods for structure based drug design are not mature enough to replace exhisting practices. De novo design entails the automatic design of novel ligands intended to bind with a given receptor with high affinity. Docking & scoring are complimentary methods which split the problem of ligand, docking involves finding the bound configuration of ligand & receptor assuming a specific interactions of reasonable affinity while scoring involves estimating the affinity of the ligand given a pericular bound configuration. Most docking & scoring methods are empirical approaches involving extrapolation from known ligand receptor complexes.

MODELING PROCESS

Modeling process divided into 9 stages:

- template recognition
- alignment
- alignment correction
- backbone generation
- generation of canonical loops (data based)
- side chain generation plus optimization
- ab initio loop building (energy based)
- overall model optimization (energy minimization)
- model verification with optional repeat of previous steps.

RASMOL

Rasmol is one of the most popular tools for protein structure visualization. It is a computer program writer for molecular graphics visualization used primarily for our

depiction and exploration of biological macromolecular structure such as those found in PDB. It was originally developed by Roger Sayle in the early 90s.

The various styles of representing a macromolecule are wire frame, space fill, cartoon, ribbon etc. It reads molecular structure files from protein data bank & recompute the information. It allows the user to display different NMR modules. It has proven to be a popular molecular modeling tool because:

- It is free.
- It runs on most computer system (including the Apple Mac, MS-windows, VMS, OS/2 & Unix/X-windows)
- Source code is freely available.
- Rasmol is readily integrated into the World-Wide Web.

Rasmol to be an important tool for research in structural biology.

TYPES OF DRUG DESIGNING

Drug design, also sometimes reffered to as rational drug design, is the inventive process of finding new medications based on the knowledge of the biological target.

The Drug is most commonly an organic small molecule which activates or inhibits the function of a biomolecule such as a protein which in turn result in a therapeutic benefit to the patient.

There are 2 major types of drug design.

Ligand based drug design

Structure based drug design.

(1) LIGAND BASED DRUG DESIGN

- Ligand based drug design (or indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest.
- These other molecules may be used to derive a pharmacophore which defines the minimum necessary structural characteristics a molecule must possess in order to bind to the target.
- Target built based on the knowledge of binds to it & it used to design new molecular entities that interact with the target.

(2) STRUCTURE BASED DRUG DESIGN

- Structure based drug design (or direct drug design) relies on knowledge of the three dimensional structure of the biological target obtained through methods such as X-ray crystallography or NMR spectroscopy.
- Due to X-ray crystallography or NMR spectroscopy structural dynamics & electronic properties about the ligand also increased.

(3) RATIONAL DRUG DESIGN

Rational drug design is a more focused approach , which uses information about the structure of a drug receptor or one of its natural ligands to identify or create candidate drugs. The 3-D structure of a protein can be determined using methods such as X-ray crystallography or nuclear magnetic resonance spectroscopy.

The first drug produced by rational design was "Relenza", which is used to treat influenza.

DOCKING

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 e.g. scoring function.

The association between biologically relevant molecules such as proteins , nucleic acids, carbohydrates & lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g. agonism vs antagonism). Therefore docking is useful for predicting both strength & type of signal produced.

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 & activity of the small molecule. Hence docking plays an important role in the rational desisn of drugs . Given the biological & pharmaceutical significance of molecular docking , considerable efforts have been directed towards improving the methods used to predict docking.

MOLECULAR DOCKING

Molecular 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. It is used to

predict the 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. The protein can be thought of as the "lock" and the ligand can be thought of as a "key". Molecular docking may be defind as an optimization problem, which would describe the "best-fit" orientation of a ligand that binds to a particular protein of intersest.

During the course of the process, the ligand and the protein adjust their conformation to achieve an overall "best-fit" and this kind of conformational adjustments resulting in the overall binding is referred to as "induced-fit".

APPROACHES OF MOLECULAR DOCKING

Two approaches are particularly popular within the molecular docking community. One approach used a matching technique that describes the protein and the ligand as a complementary surface. The second approach simulates the actual docking process in which the ligand protein pairwise interaction energies are calculated.

MECHANISM OF DOCKING

To perform a docking screen , the first requirement is a structure of the protein of interest. Usually the structure has been determined using a biophysical technique such as x-ray crystallography or less often, NMR spectroscopy. This protein structure and database of potential ligands serve as inputs to a docking program. The success of a docking program depends on two components: the search algorithm and the scoring function.

APPLICATION OF DOCKING

A binding interaction between a small molecule ligand & an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism. Docking is most commonly used in the field of drug design. Most drugs are small organic molecules & docking may be applied to:

Hit identification: Docking combined with a scoring function can be used to quickly screen large database of potential drugs in silico to identify molecules that are likely to bind to protein target of interest.

Lead optimization: Docking can be used to predict in where & in which relative orientation a ligand binds to a protein. This information may in turn be used to design more potent & selective analogs.

Bioremediation: Protein ligand docking can also be used to predict pollutants that can be degraded by enzymes.

OBJECTIVES OF INVESTIGATION

Main objectives of investigations are to explore the reality of protein ligand interaction & to learn about Homology Modeling & Docking, which includes:-

- Basic operation of InsightII graphics interface to display & manipulate modeled protein.
- Homology modeling of a structurally unknown protein.
- Setting up energy minimization.
- Studying the protein ligand interaction.
- Docking of ligand in modeled receptor protein molecule.

Chapter -2

REVIEW OF LITERATURE

REVIEW OF LITERATURE

Cell Adhesion Molecules (CAMs) are proteins located on the cell surface involved with the binding with other cells or with the extracellular matrix (ECM) in the process called cell adhesion.

These proteins are typically transmembrane receptors and are composed of three domains: an intracellular domain that interacts with the cytoskeleton, a transmembrane domain, and an extracellular domain that interacts either with other CAMs of the same kind (**homophilic** binding) or with other CAMs or the extracellular matrix (**heterophilic** binding).

The L1 adhesion molecule is a member of the immunoglobulin superfamily shared by neural and immune cells. In the nervous system L1 can mediate cell binding by a homophilic mechanism. To analyze its function on leukocytes we studied whether L1 could interact with integrins. Integrins are the family of heterophilic CAMs. Here we demonstrate that VLA-5, an RGD-specific fibronectin receptor on a wide variety of cell types, can bind to murine L1. ESb-MP cells(in mouse) expressing VLA-5 and L1 could be induced to aggregate in the presence of specific mAbs to CD24 (heat-stable antigen), a highly and heterogeneously glycosylated glycophosphatidylinositol-linked differentiation antigen of hematopoietic and neural cells. The aggregation was blocked by both mAbs to L1 and VLA-5, respectively. Aggregation was blocked also by a synthetic RGD-containing peptide derived from the Ig-domain VI of the L1 protein. ESb-MP subclones with low L1 expression could not aggregate. Also purified L1 coated to polystyrene beads could bind to platelets. The binding of L1-beads was again inhibited by mAbs to L1 and VLA-5, by soluble L1 and the L1-RGD peptide in a dose-dependent manner. The adhesion was strongly enhanced in the presence of Mn(2+)-ions and blocked by mAbs to VLA-5. We also demonstrate a direct L1-VLA-5 protein interaction. Our results suggest a novel binding pathway, in which the VLA-5 integrin binds to L1 on adjacent cells. Given its rapid downregulation on lymphocytes after induction of cell proliferation, L1 may be important in integrin-mediated and activation-regulated cell-cell interactions.

Families of CAMs

Most of the CAMs belong to 5 protein families: Ig (immunoglobulin) superfamily (IgSF CAMs), the integrins, the cadherins, the selectins and the lymphocyte homing receptors.

One classification system involves the distinction between calcium-independent CAMs and calcium-dependent CAMs.

Calcium-independent

Immunoglobulin superfamily

CAMs (IgSF CAMs) are either homophilic or heterophilic and bind integrins or different IgSF CAMs.

Here is a list of some molecules of this family:

SynCAMs Synaptic Cell Adhesion Molecules

NCAMs Neural Cell Adhesion Molecules

ICAM-1 Intercellular Cell Adhesion Molecule

VCAM-1 Vascular Cell Adhesion Molecule

PECAM-1 Platelet-endothelial Cell Adhesion Molecule

L1

CHL1

MAG

Nectins and nectin-like molecules

Integrins

The **Integrins** are a family of heterophilic CAMs that bind IgSF CAMs or the extracellular matrix. They are heterodimers, consisting of two noncovalently-linked subunits, called alpha and beta. Eighteen different alpha subunits that combine with 9 different beta subunits to form twenty-four known integrins; however not all combinations are observed.

Lymphocyte homing receptors

These are also known as addressins. Two well known examples are CD34 and GLYCAM-1.

Calcium-dependent

Cadherins

The **cadherins** are a family of homophilic CAMs, Ca^{2+}-dependent. The most important members of this family are E-cadherins (epithelial), P-cadherins (placental), and N-cadherins (neural).

Selectins

The **selectins** are a family of heterophilic CAMs that bind fucosylated carbohydrates, e.g., mucins. They are calcium-dependent. The three family members are E-selectin (endothelial), L-selectin (leukocyte), and P-selectin (platelet). The best-characterized ligand for the three selectins is P-selectin glycoprotein ligand-1 (PSGL-1), which is a mucin-type glycoprotein expressed on all white blood cells.

Autism is a behaviourally developmental disorder that is characterized by impaired development in communication, social interaction, and behavior.

Autism affects information processing in the brain by altering how nerve cells and their synapses connect and organize; how this occurs is not well understood.

Autism is classified as a category of disorders that is often described interchangeably with the broad spectrum of developmental disorders affecting young children and adults called the autistic spectrum disorders (ASD).

Additionally, autism can be found in association with other disorders such as mental retardation and certain medical conditions.

The number of children diagnosed as having autistic spectrum disorders is increasing for various reasons

A diagnosis of autism can be reliably made at between 2 and 3 years of age

Autism does not meet criteria for screening, but surveillance throughout the preschool years is recommended

Diagnosis is by history taking, focusing on the developmental story and systematically inquiring for core behaviours, and by observation in several settings

Features that may discriminate children with autism early in childhood

Lack of social smile, lack of appropriate facial expression, poor attention, impaired social interaction

Ignoring people, preference for aloneness, lack of eye contact, lack of appropriate gestures, lack of emotional expression, less looking at others, less pointing, less showing objects in the second year

Features that may discriminate children with autism in later childhood

In school age children, the following features should alert teachers and others to the possibility of autistic spectrum disorder and trigger discussion with parents and possible implementation of the local referral pathway

Communication impairments

Abnormalities in language development, including muteness and odd or inappropriate prosody

Persistent echolalia

Reference to self as "you," "she," or "he" beyond 3 years

Unusual vocabulary for child's age or social group

Limited use of language for communication or tendency to talk freely only about specific topics.

Social impairments

Inability to join in with the play of other children or inappropriate attempts at joint play (may manifest as aggressive or disruptive behaviour)

Lack of awareness of classroom "norms" (criticising teachers; overt unwillingness to cooperate in classroom activities).Easily overwhelmed by social and other stimulation

Failure to relate normally to adults .Showing extreme reactions to invasion of personal space and extreme resistance to being "hurried".

Causes

Researchers still have not reached agreement regarding its specific causes. Heritability contributes about 90% of the risk of a child developing autism, but the genetics of autism are complex and typically it is unclear which genes are responsible.

The large number of autistic individuals with unaffected family members may result from copy number variations—spontaneous deletions or duplications in genetic material during meiosis. Hence, a substantial fraction of autism cases may be traceable to genetic

causes that are highly heritable but not inherited: that is, the mutation that causes the autism is not present in the parental genome.

Environmental factors that have been claimed to contribute to or exacerbate autism, or may be important in future research, include certain foods, infectious disease, heavy metals, solvents, diesel exhaust, PCBs, phthalates and phenols used in plastic products, pesticides, brominated flame retardants, alcohol, smoking, illicit drugs, vaccines and prenatal stress. Among these factors, vaccines have attracted much attention, as parents may first become aware of autistic symptoms in their child around the time of a routine vaccination. Two theories link autism and vaccines. The first theory suggests that the MMR (Mumps-Measles-Rubella) vaccine may cause intestinal problems leading to the development of autism. The second theory suggests that a mercury-based preservative called thimerosal, used in some vaccines, could be connected to autism.

Symptoms of autism

The current Diagnosis and Statistical Manual of Mental Disorders-Fourth Edition, Treatment Revision (DSM-IV-TR) identifies three features that are associated with autism:

impairment in social interaction, communication, and behavior.

Impairment in social interaction

Patients with autism fail to develop normal personal interactions in virtually every setting. This means that affected persons fail to form the normal social contacts that are such an important part of human development.It is important to note that, contrary to popular belief, many, if not most, persons with this disorder are capable of showing affection, demonstrating affection bonding with their mothers or other caregivers. There is usually an inability to develop normal peer and sibling relationships and the child often seems isolated. There may be little or no joy or interest in normal age-appropriate activities.

Autistic infants show less attention to social stimuli, smile and look at others less often, and respond less to their own name. Autistic toddlers differ more strikingly from social norms; for example, they have less eye contact and turn taking, and are more likely to communicate by manipulating another person's hand.

Autistic children do not prefer being alone. Making and maintaining friendships often proves to be difficult for those with autism. For them, the quality of friendships, not the number of friends, predicts how lonely they feel.

Communication

Communication is usually severely impaired in persons with autism. What the individual understands (receptive language) as well as what is actually spoken by the individual (expressive language) are significantly delayed or nonexistent. Their speech may seem to lack the normal emotion and sound flat or monotonous.

Differences in communication may be present from the first year of life, and may include delayed onset of babbling, unusual gestures, diminished responsiveness, and vocal patterns that are not synchronized with the caregiver. In the second and third years, autistic children have less frequent and less diverse babbling, consonants, words, and word combinations; their gestures are less often integrated with words. Autistic children may have difficulty with imaginative play and with developing symbols into language.

Behaviors

Persons with autism often exhibit a variety of repetitive, abnormal behaviors. There may also be a hypersensitivity to sensory input through vision, hearing, or touch (tactile). As a result, there may be an extreme intolerance to loud noises or crowds, visual stimulation, or things that are felt.This person may head bang, scratch until blood is drawn, scream instead of speaking in a normal tone, or bring everything into close visual range.

A person with this disorder may sit for hours turning off and on a light switch, twirling a spinning toy, or stacking nesting objects.

Treatments for Autism

There is no known cure for autism, but it is treatable. Many people with autism become more responsive as they come to better understand the world. Some children's symptoms may improve significantly or resolve altogether. The goals of treatment include:

- Stopping inappropriate behaviors so the child can relate better to others.
- Teaching the child to attend to purposeful activity. This can help the child succeed in educational settings.

- Helping the child learn self-care skills.
- Providing opportunities for the child to socialize with others.
- Improving the child's communication skills.
- Teaching parents how to provide helpful educational and social experiences for their child.

In most cases, treatment is provided in an individualized program that focuses on behavior modification and skills development. Treatment also may involve medication to help control specific symptoms. Usually a team of specialists is involved. The team may include a psychologist, a special education teacher, a speech therapist, an occupational therapist, a child development specialist, and trained aides.

In general, treatment programs tend to be more effective if they build on the child's unique interests; if they engage the child in highly structured activities according to a predictable schedule; and if they provide regular rewards for desired behavior. Also, the parents' involvement is very important to the development of autistic children.

Diagnosis

There are no specific medical laboratory tests to diagnose autism. The diagnosis of autism is made by taking into account the child's complete medical and behavioral history, lengthy observation of the child's behavior, and ruling out other problems that may cause some of the same symptoms.

The first diagnosis :Leo Kanner (1943) was the first person to describe and name a pattern of behaviour he observed in a small group of young children, which he termed early infantile autism. Asperger (1944), one year after Kanner's original paper, wrote about another behaviour pattern in older children and adolescents, which, though different in detail, clearly overlapped with Kanner's accounts. Asperger also used the term 'autistic' in relation to the behaviour he saw.

Mechanism

Autism's symptoms result from maturation-related changes in various systems of the brain. How autism occurs is not well understood. Its mechanism can be divided into two

areas: the pathophysiology of brain structures and processes associated with autism, and the neuropsychological linkages between brain structures and behaviors.The behaviors appear to have multiple pathophysiologies.

Autism affects the amygdala, cerebellum, and many other parts of the brain. Unlike many other brain disorders such as Parkinson's, autism does not have a clear unifying mechanism at either the molecular, cellular, or systems level; it is not known whether autism is a few disorders caused by mutations converging on a few common molecular pathways, a large set of disorders with diverse mechanisms. Autism appears to result from developmental factors that affect many or all functional brain systems and to disturb the timing of brain development more than the final product. Neuroanatomical studies and the associations with teratogens strongly suggest that autism's mechanism includes alteration of brain development soon after conception. This anomaly appears to start a cascade of pathological events in the brain that are significantly influenced by environmental factors. Just after birth, the brains of autistic children tend to grow faster than usual, followed by normal or relatively slower growth in childhood. It seems to be most prominent in brain areas underlying the development of higher cognitive specialization. Hypotheses for the cellular and molecular bases of pathological early overgrowth include the following:

An excess of neurons that causes local overconnectivity in key brain regions.

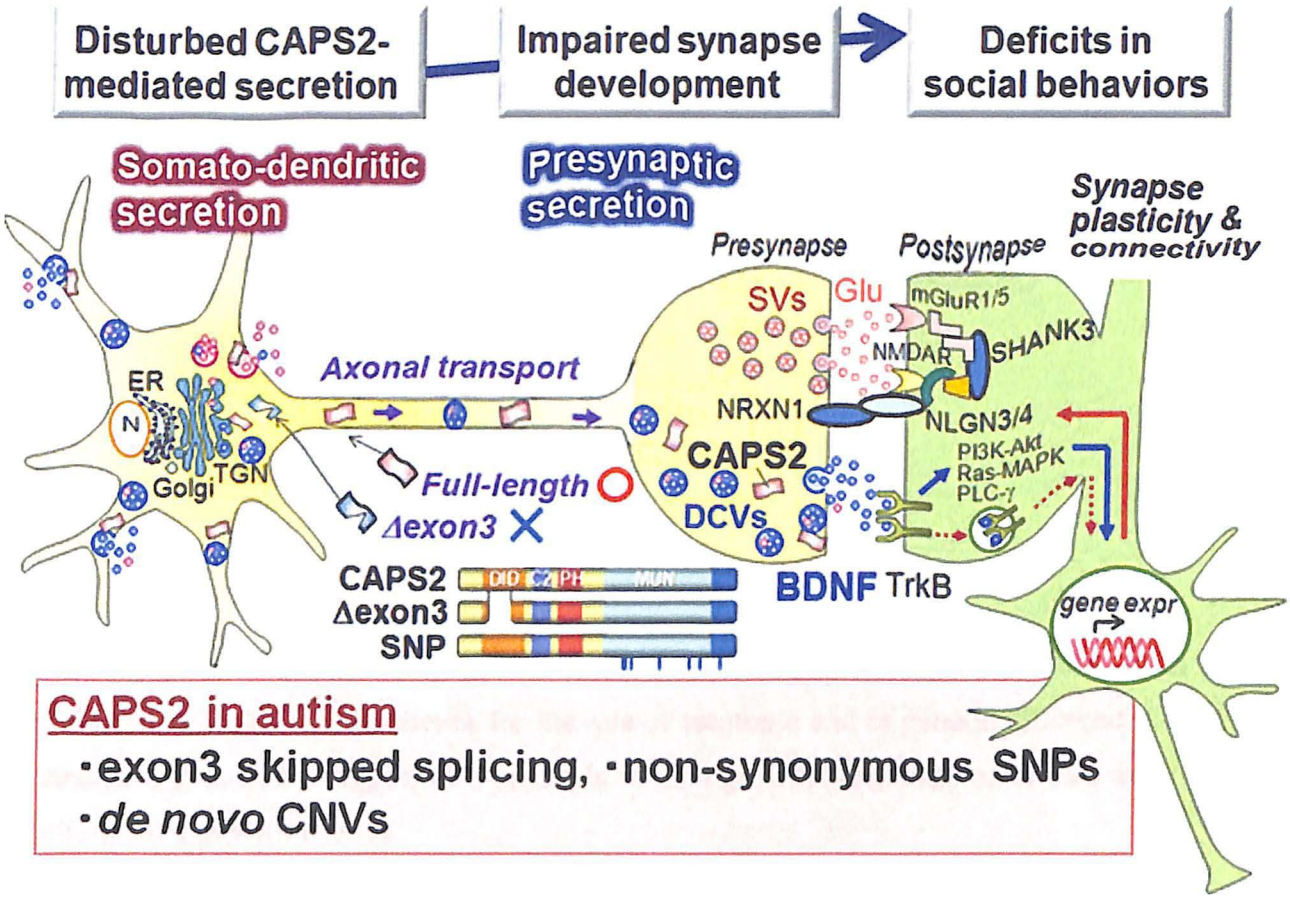
Disturbed neuronal migration during early gestation.

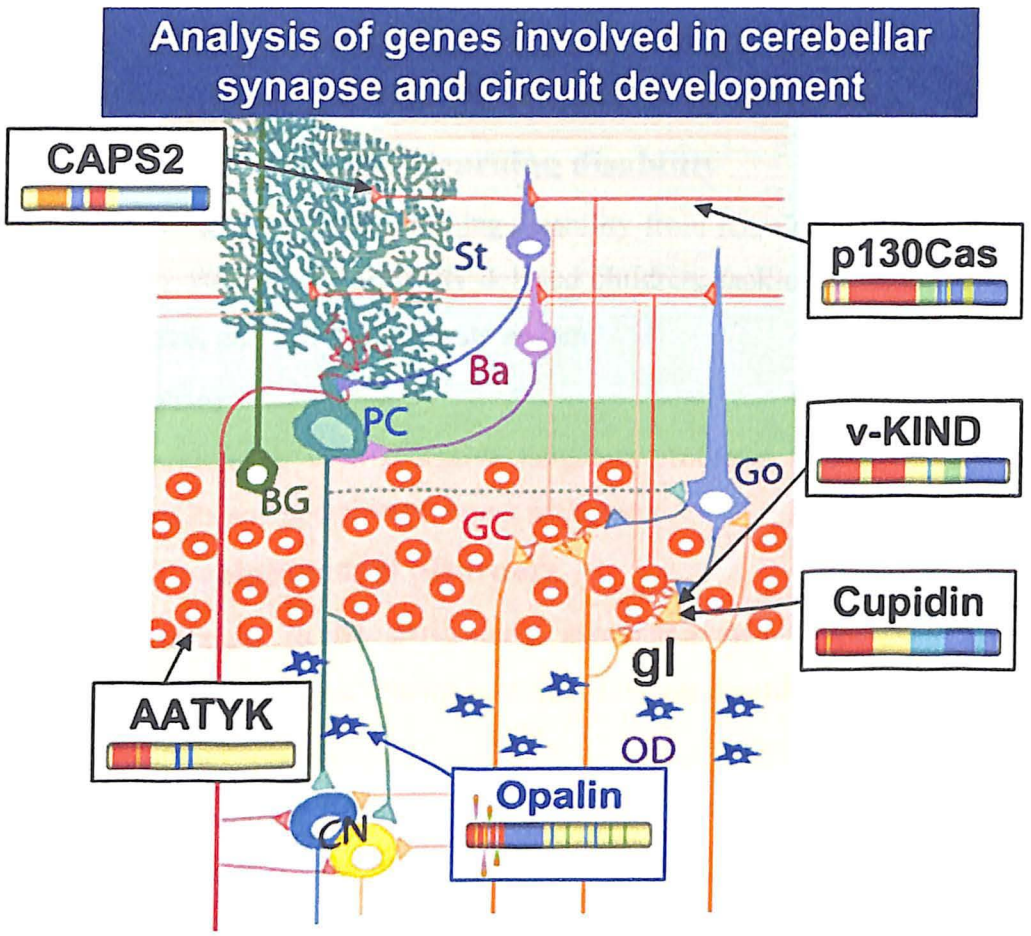
Unbalanced excitatory–inhibitory networks.

Abnormal formation of synapses and dendritic spines for example, by modulation of the neurexin–neuroligin cell-adhesion system or by poorly regulated synthesis of synaptic protein. It is possible that aberrant immune activity during critical periods of neurodevelopment is part of the mechanism of some forms of ASD.

- CAPS2 is the gene name of B9A061_HUMAN.
- CAPS2 protein associated with secretory vesicles containing Brain Derived Neurotrophic Factor (BDNF) and NeuroTrophin-3 (NT-3).
- CAPS2 enhances release of BDNF and NT-3 both are essential for normal cerebellar development.

- CAPS2 reduce secretion of BDNF and NT-3, increased cell death , fewer branched dendrites on purkinje cells.
- CAPS2 regulates neurotrophin release from granules cells leading to regulate cell differentiation and survival during cerebellar development.





The relationship of neurochemicals to autism is not well understood; several have been investigated, with the most evidence for the role of serotonin and of genetic differences in its transport. Some data suggest an increase in several growth hormones; other data argue for diminished growth factors.

Regression

In approximately 25-30% of affected children, obvious stasis and sometimes clear regression of development occurs between 15 and 21 months of age. This can involve loss of word use (usually within the first 10 word stage) accompanied by social withdrawal, loss of eye contact and play interests, and sometimes change of sleep and eating habits. Onset of unusual behaviours such as overfocused looking at objects and flapping or other mannerisms may also be noted. Regression of skills needs careful medical assessment, as neurodegenerative conditions, Rett syndrome, and epilepsy may present in this way. No evidence exists to show that regression is on the increase.

Common differential diagnoses

Mental retardation or general learning disability

Differentiating autism plus learning disability from learning disability alone may be difficult in very young or profoundly delayed children; lack of imitation, poor social relatedness, eye gaze, and gestures indicate autism.

Language disorder

Children presenting with receptive language problems may also have limited imaginative play skills and social impairment with peers .

Other specific developmental disorders

Although thought to be particularly associated with Asperger's syndrome, clumsiness or incoordination (developmental coordination disorder) can occur in any of the autistic spectrum disorders.

Other disorders causing diagnostic difficulty

Reactive attachment disorders of childhood after early emotionally depriving experiences, such as those in Romanian adoptees, may present overlapping behaviours but also differences from autism[w3].

Early onset epilepsy, Rett syndrome, or neurodegenerative disorders.

The L1 CAM 5 is selected as target because yet there is no any 3D structure available and L1 CAM 5 causes the Autism disease directly or indirectly.

RECENT PUBLICATIONS ON L1 CAM 5 PROTEIN

- L1 cell adhesion molecule (L1CAM) overexpression is often associated with bad prognosis in various human carcinomas. Recent studies also suggest a role of L1CAM in pancreatic ductal adenocarcinomas (PDAC). To further address its contribution, we expressed functional domains of L1CAM in PT45-P1 PDAC cells. We found that L1CAM that is full length (L1-FL), but neither the soluble ectodomain (L1ecto) nor the cytoplasmic part (L1cyt), could enhance cell proliferation or tumour growth in mice. Expression of L1-FL resulted in constitutive activation of NF-κB, which was abolished by L1CAM knockdown. We showed that the expression of IL-1β was selectively upregulated by L1-FL,

and increased IL-1β levels were instrumental for sustained NF-κB activation. IL-1β production and NF-κB activation were abolished by knockdown of α5-integrin and integrin-linked kinase, but insensitive to depletion of L1CAM cleavage proteinases. Supporting these data, PT45-P1 cells transduced with an L1CAM mutant deficient in integrin binding (L1-RGE) did not support the described L1-FL functions. Our results suggest that membranous L1CAM interacts with RGD&!minus;-binding integrins, leading to sustained NF-κB activation by IL-1β production and autocrine/paracrine signalling. The unravelling of this novel mechanism sheds new light on the important role of L1CAM expression in PDAC cells.

- Recent work has identified L1CAM (CD171) as a novel marker for human carcinoma progression. Functionally, L1CAM promotes tumor cell invasion and motility, augments tumor growth in nude mice, and facilitates experimental tumor metastasis. These functional features qualify L1 as an interesting target molecule for tumor therapy. Here, we generated a series of novel monoclonal antibodies (mAb) to the L1CAM ectodomain that were characterized by biochemical and functional means. All novel mAbs reacted specifically with L1CAM and not with the closely related molecule CHL1, whereas antibodies to the COOH terminal part of L1CAM (mAb2C2, mAb745H7, pcytL1) showed cross-reactivity. Among the novel mAbs, L1-9.3 was selected and its therapeutic potential was analyzed in various isotype variants in a model of SKOV3ip cells growing i.p. in CD1 nude mice. Only therapy with the IgG2a variant efficiently prolonged survival and reduced tumor burden. This was accompanied by an increased infiltration of F4/80-positive monocytic cells. Clodronate pretreatment of tumor-bearing animals led to the depletion of monocytes and abolished the therapeutic effect of L1-9.3/IgG2a. Expression profiling of tumor-derived mRNA revealed that L1-9.3/IgG2a therapy induced altered expression of cellular genes associated with apoptosis and tumor growth. Our results establish that anti-L1 mAb therapy acts via immunologic and nonimmunologic effector mechanism to block tumor growth. The novel antibodies to L1CAM could become helpful tools for the therapy of L1-positive human carcinomas.

- The L1 syndrome is an X-linked recessive disease caused by mutations in the *L1CAM* gene. To date more than 200 different mutations have been reported, scattered over the entire gene, about 35% being missense mutations. Although it is tempting to consider these missense mutations as being disease-causing.we have updated and upgraded our *LICAM* mutation database with more pathogenicity data and clinical information collected from the literature or generated by our own research. As a result, the renewed database offers condensed scientific information, allowing conclusions to be drawn on the pathogenicity and severity of *LICAM* mutations based on multiple factors.
- Mental health is an essential ingredient in the quality of life. Recent studies carried out in countries like Germany, USA, France, England and Belgium have provided evidence for the involvement of L1 (CAM) mutations in various X-linked mental retardation syndromes. L1 CAM is a neural cell adhesion molecule belonging to the superfamily of the immunoglobulins and is critical for proper CNS development in humans.
- This study was aimed to screen idiopathic mental retardation cases for L1 CAM mutations.In the present study, a deletion was observed in two cases with idiopathic mental retardation in exon 26-27. Hence it is worthwhile to screen for L1CAM mutations in cases of idiopathic mental retardation. Detection of mutations will be very helpful for prenatal diagnosis and to impart genetic counseling to the parents and families who are at risk.
- A recent publication from the Felsenfeld laboratory http:// www. mssm. edu/ labs/ felsenfeld) fills a gap in our knowledge of how the interaction of L1-type CAMs with the membrane skeleton adaptor protein ankyrin is severed by phosphorylation and suggests a feedback mechanism whereby the neurite-stimulating activity of L1-CAM is inversely connected to its cytoskeleton binding.
- Five novel mutations have been identified in the gene encoding L1CAM, a neural cell adhesion protein, in families with X linked hydrocephalus (XHC). Interestingly, all five mutations are in the evolutionarily highly conserved Ig-like domains of the protein. The two frameshift mutations (52insC and 955delG) and

the nonsense mutation (Trp276Ter) most probably result in functional null alleles and complete absence of L1CAM at the cell surface. The two missense mutations (Tyr194Cys and Pro240Leu) may considerably alter the structure of the L1CAM protein. These data provide convincing evidence that XHC is genetically extremely heterogeneous.

- We have examined the effects of 25 missense mutations on binding to homophilic (L1) and heterophilic (TAX-1) ligands as well as on intracellular trafficking. All but three of these result in reduced ligand binding or impaired movement to the surface of COS and CHO cells. Therefore, we demonstrate for the first time that most missense mutations found in affected families have functional consequences. Furthermore, mutations that are predicted to affect the structure of individual extracellular domains are more likely to affect intracellular processing and/or ligand binding than those mutations affecting surface properties of the molecule.

- We detected a transition c2308G→A in exon 18 that caused an amino acid change in codon 770. The patient's mother and two sisters were heterozygous for the same mutation. This newly described mutation predicts the substitution of an aspartate by asparagine (D770N) in the second fibronectin (Fn2) domain of the extracellular portion of the mature L1 protein. Even if amino acid substitution does not significantly change the physico-chemical properties of the Fn2 domain, it seems clear that the integrity of this domain is required to maintain the biological functions of the protein. The feature peculiar to this patient is the decelerated head growth post-natally, leading to microcephaly. Mutations of *L1CAM* associated with prolonged survival may hamper post-natal brain and head growth.

Chapter-3

MATERIALS & METHODS

MATERIALS & METHODS

FREE SOFTWARE / DATABASES USED

NCBI

The Entrez Global Query Cross-Database Search System is a powerful federated search engine or web portal that allows users to search many discrete health sciences databases at the National Center for Biotechnology Information (NCBI) website. NCBI is a part of the National Library of Medicine (NLM) itself a department of the National Institutes of Health (NIH) of the United States government. Enterz also happens to be the French second person plural form of the verb "to enter", meaning literally "come in".Enterz Global Query is an integrated search and retrieval system that provides access to all databases simultaneously with a single query string and user interface. Enterz can efficiently retrieve related sequences, structures and references. The Enterz system can provide views of gene and protein sequences and chromosome maps. Some textbooks are also available online through the Enterz system.

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.

Blast

In Bioinformatics Basic Local Alignment Search Tool or BLAST is an algorithm for comparing 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.

PDB

The PDB archive contains information about experimentally determined structures of proteins, nucleic acids and complex assemblies. As a member of the WWPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized , download and analyzed by users.

EXPASY

The **ExPASy** (**Ex**pert **P**rotein **A**nalysis **Sy**stem) is a proteomics server of the Swiss Institute of Bioinformatics (SIB) which analyzes protein sequences and structures and two-dimensional gel electrophoresis (2-D Page electrophoresis).The server functions in collaboration with the European Bioinformatics Institute. ExPASy also produces the protein sequence knowledgebase, UniProtKB/Swiss-Prot, and its computer annotated supplement, UniProtKB/Trembl.

RCSB

The **Research Collaboratory for Structural Bioinformatics (RCSB)** is dedicated to improving our understanding of the function of biological systems through the study of the 3-D structure of biological macromolecules. RCSB members work cooperatively and equally through joint grants and subsequently provide free public resources and publications to assist others and further the fields of bioinformatics and biology.

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 a part of the United states National Institutes of Health (NIH).
- Pubchem can be accessed for free through a web user interface.
- Pubchem contains substance descriptions & small molecules which fewer than 1000 atoms & 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 Services.
- More than 80 database vendors contribute to the growing pubchem database.

SOPMA

Recently a new method called the self-optimized prediction method (SOPMA) has been described to improve the success rate in the prediction of the secondary structure of proteins. In this paper we report improvements brought about by predicting all the sequences of a set of aligned proteins belonging to the same family. It is based on homologue method of Levin etal (1986). This can be accessed via NDSA (Network Protein Sequence Analysis) server (http://www.npsabil.ibcp.frl) SOPMA takes into the account the information from an alignment of sequence belonging to the same family.

GOR IV

GOR IV is the fourth version of GOR secondary structure prediction methods based on the information theory. There is no defined decision constant. GOR IV uses all possible pair frequencies within the window of 17 amino acid residues.

MODELLER

Modeller (Sali and Blundell 1993) is perhaps the most frequently used homology modeling program. It is one of the first fully automated programs and it is also relatively fast, making it suitable for whole genome modeling (Marti-Renom et al. 2000; Pieper et al. 2004). It is a tool that predicts protein structure by comparative modeling methods. The program requires a sequence alignment of the query and a template to be

submitted as input. The sequence is required in the non-reduent data base using BLAST. The top hit structural homologue with at least 30% sequence identy is selected as template. Pair wise alignment of the query and the template sequence is carried out and this alignment is used as input to the MODELLER program. The structure of the template sequence is downloaded from the PDB and used as the modeling template. The models with the lowest energy are then submitted to the PROCHECK program that checks the stereochemical properties of the protein model. Models are obtained by satisfying spatial restraints derived from the alignment and expressed as probability density functions (pdfs) for the different types of restraints. The pdfs restrain CA-CA and backbone N-O distances, backbone and side-chain dihedral angles for different residue types. The generated model violates these restraints as little as possible. A new version of Modeller was recently released and both the new 7v7 and old 6v2 have been tested here. In addition a third Modeller version, Modeller 6v2-10 was also tested. Here 10 models are created for each alignment using different initial random seeds and the model with the lowest RMSD to the template structure is chosen. The reason for including this program was that Modeller sometimes has a problem with convergence, i.e., producing models with extended structures.

VARIFY 3D

The quality of the models was assessed by using the structure verification program Verify 3D (Luthy et al. 1992) which tests the compatibility of a protein structure with its amino acid sequence. Verify 3D constructs a profile for the three-dimensional model in which each residue position is characterized by its environmental score. These scores were derived from a statistical analysis of high-resolution protein structures from the PDB . The Verify 3D profile is graphically represented by the numerical scores as a function of the residue number in the structure or model. For high-resolution, experimentally determined structures, the Verify 3D scores are positive and consistently high (>0.2), indicating that they provide a reliable means to assess the quality of a protein structure. We show that an effective test of the accuracy of a 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 3D profile of correct protein structures match their own sequences with high scores. The accuracy of a protein model can be assessed

by its 3D profile, regadless of whether the model has been derived by X-ray, NMR or computational procedures.

PROCHECK

The stereo-chemical quality of the module was checked using RAMPAGE. The model proteins were submitted to the RAMPAGE server. The stereo-chemical quality of the protein was assessed by plotting the Ramachandran plot.

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.

INSIGHT II PACKAGES

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.

MODULES IN INSIGHT II

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 therefore the 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 chemist, 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.

Builder

The Builder command activates the Builder module. This module allows you to construct new molecules from molecular fragments or individual atoms. It also allows you to modify 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, change N and C terminal capping groups. Other commands in this module allow you to search a database for, and display regions of, proteins which meet a given geometric criterion. The nucleic acid commands can be used to: build single, double, or triple-stranded polynucleotides in A, B, or Z form; delete or replace nucleotides in strands, measure angles and distances between bases, and cap a prime or legate 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, allowing you to perform builder functionality while in the Biopolymer module.

Discover

The Discover module provides an interface to the Discover program. The interface allows for the definition of minimization and dynamics calculations. We can perform simulations using various forms of constraints and restraints, including template

forcing, torsion forcing, tethering and NOE. This module also enables to be 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 options for querying ongoing jobs. The interface allows for the definition of minimization dynamics calculations.

Docking

The Docking module provides facilities for calculating the nonbond energy between two molecules using explicit Van der Waals energy, explicit electrostatic (Coulombic) energy or the combination of Van der Waals 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 precomputed 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.

FILE FORMATS IN INSIGHTII

Molecular Topology File (.top)

The molecular topology file is generated by the translated pdb file of any peptide or protein to a molecular topology file. This topology file contains the complete descriptions of all the interactions in the peptide or a protein molecule.

Molecular structure files (.gro & .pdb)

When molecular topology is generated , it also translates the structure file (.pdb) to a grooms structure file (.gro). The main different between a pdb file & a groms file is their format & that a gromos file can hold velocities. However if we do not need the velocities it can use the pdb file in all programs.

Molecular Dynamics Parameter File (.mdb)

The molecular dynamics parameter file (.mdb) file contains all informations about the molecular Dynamics simulation itself i.e., time-step, no. of steps, temperature, pressure etc. The easiest way of handling such a file is by adapting a sample .mdb file. A sample mdb file can be found online.

Run Input File (.tpr)

The next steps to combine molecular structures (.gro file) , topology (.top file), MD- parameters file (.mdb files) to generate a run input file (.tpr file). This file contains all information needed to start a simulation with gromacs. The grompp program processes all input files & generates the run input file.

Trajectory File (.trr)

Once the run input file is available, we can start simulation. The program which starts the simulation is called mdrun. The only input of mdrun we usually need to start run is the run input file (.tpr file). The output files of mdrun are the trajectory files (.trrfile) & a log file (.log).

Index File (.ndx)

Sometimes we may need an index file to specify actions on group of atoms (i.e. , temperature, acceleration, freezing). Usually the default index grps. Will be sufficient.

Gromacs Trajectory Viewer (.ngmx file)

This program reads a trajectory file, a run input file & an index file & plots a 3D structure of the molecule on the standard window screen. The features like 3D views, rotation, translation & scaling of the molecules labels on atoms, animation of trajectories user friendly menus options to show computational box have been implemented.

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 Insight II.

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 Insight II.

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.

RAMACHANDRAN PLOT

The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure (expect those at the chain termini). Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other side chain types. The colouring/shading on the plot represents the different regions described in Morris et al. (1992): the darkest areas correspond to the "core" regions representing the most favourable combinations of phi-psi values. Ideally, one would hope to have over 90% of the residues in these "core" regions. The percentage of residues in the "core" region is one of the better guides to stereochemical quality. Note that additional Ramachandran plots can also be generated, as follows:

Separate plots for each of the 20 different amino acid types

Separate plots for just the Gly & Pro residues

Options

The main options for the Ramachandran plot are:-

Labelling of residues in disallowed regions can be switched off, or alternatively can be extended into the other regions.

Shading/colouring of the different regions can be switched off.

The plot can be in colour or black and white.

A "publication version" of the plot (without the outer border and statistics) can be generated.

PATCHDOCK

PatchDock algorithm is inspired by object recognition and image segmentation techniques used in Computer Vision. Docking can be compared to assembling a jigsaw puzzle. When solving the puzzle we try to match two 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. PathDock employes a similar technique. Given two molecules, their surfaces are divided into patches according to the surface shape. These patches correspond to patterens that visually distinguish between puzzle pieces. Once the patches are identified they can be superimposed using shape matching algorithms.

METHODOLOGY

Alignment

The target protein could generally be aligned either by pairwise or multiple sequence aligned in insightII package, both the methods were available & the similarity matters in aligning the target.

HOMOLOGY MODELLING USING INSIGHTII

Homology is an application within InsightII. Most InsightII commands can be used on molecules created with Homology and vice versa. All rules and conventions for InsightII apply when I am working with Homology. For example, commands in

Homology can be typed in or selected from the pulldowns and parameter blocks, just like commands in InsightII.

General steps involved in InsightII

First we have to generate a sequence alignment between reference proteins. Then model protein sequence with reference protein sequences. Finally creating a modeler input and launching the background jobs.

STEPS

- Select Homology from the module
- Load the reference molecules

  Go to molecules

  Get reference pdb files.
- Extracting sequences from the protein structures

  Go to sequences/Extract

  Sequence window will appear.
- Creating a sequence alignment between reference proteins

  Go to alignment

  Pairwise alignment/Multiple alignment
- Read the sequence (for modeled protein)

  Go to sequences

  Get sequence
- Aligning the model protein with reference proteins

  Go to alignment

  Pairwise alignment/Multiple alignment
- Creating the modeler input files and launching the jobs

  Go to modeler

  Build model.

Verify 3D

As methods for determining protein three-dimensional (3D) structure develop, a continuing problem is how to verify that the final protein model is correct. The revision

of several protein models to correct errors has prompted the development of new criteria for judging the validity of X-ray and NMR structures, as well as the formation of energetic and empirical methods to evaluate the correctness of protein models. Verify 3D Structure evaluation Sever is a tool designed to help in the refinement of crystallographic structures. It will provide with a visual analysis of the quality of a crystal structure for a protein. Verify 3D expects this crystal structure to be submitted in PDB format. Verify 3D works best on proteins with at least 100 residues.

Energy Minimization

The energy minimization process would be carried out in Discover 3 model. The system could be set up with the described parameters level. The output fill to be generated after minimization would be analyzed and changed according to the state. The minimization or Discover Run could be carried out within 15 to 20 minutes. The process of minimization should be carried out with different output set up different iteration and algorithm. Finally dynamics calculation could also been performed in complete validation of the protein to be modeled. Graphs could be generated using the Analysis module to visualize the change in Energy levels with the total and current energy average.

STEPS

- We get modeled protein pdb file (target)
- Go to molecule
- Get pdb for protein
- Add Hydrogens
- Go to module and select Biopolymer
- **Modify/Hydrogens** command
- Set Potential
- **Select** command allows to determine the current forcefield
- The Insight II program supports these forcefields: CFF, CVFF, AMBER, CHARMm and ESFF which are specified through the **Forcefield/Select** command.
- Go to Discover3 module

- Select setup/system
- Select calculate/minimize
- D_Run/Run

RAMACHANDRAN PLOT IN INSIGHTII

- Create a InsightII table with the data that you want to graph (e.g., measure Phi and Psi angles using the Homology/ProStat/Residue_Dihedral menu). Manually edit the table if necessary (edit blank cells, insert new columns etc).
- Click the LEFT mouse button on the top cell of the column that contains X-axis data (e.g., Phi)
- Hold down the <Ctrl> button and click the LEFT mouse button on the top cell of the column that contains Y-axis data (e.g., Psi).
- Click on the **Graph icon at the bottom of the table window** (the middle icon). A graph will appear.
- Click on the **Graph icon on the left of the InsightII window** (two icons below the blue & white "Biuosym module" icon).
- Change the **Label** of the **Graph** to "Ramachandran plot"
- Change the **CharSize** of the **Axes** and **Graph** to 0.04.
- Change the **Color** of the **X axis, Y axis, Box_segment 1** and **Box_Segment 2** to light blue.
- Change the **Color** of the **Title** to green.
- Change the **Color** of the **plot** to yellow.
- Use **Modify_Display** into modify the display of graph. Turn off **bar, line, graph** and turn on **point.** Select a **point Symbol** and change **Symbol Scale** to 10.0 . Set **Graph** to **graph_name:1**
- Set the **Threshold** of the X axis and Y axis to a **Min Value** of -180 and a **Max Value** of 180. Turn on **Zoom_Axis.**
- Divide plot point into four quadrants: Change the **Tick Mark** so that **Extended** tick marks **Start** at -400 and **End** at 400 with a **Mark Step** of 200. Use a **Label Display** of **none**. Add these marks to the X axis and Y axis.

Target validation

The target could be validated using SAVS server, by analyzing the Ramachandran plot, the residue in the core region & the residues in allowed region be viewed. The 3D verification could be describe the percentage of efficacy of the protein.

DOCKING

Docking Using Patchdock

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. We have developed PatchDock , very efficient algorithm for protein-small ligand and protein-protein docking. PatchDock 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 complementarity criteria. It is a suite of freely available web services for protein structural analysis. PatchDock performs prediction of protein-protein and protein small molecule interactions. The input to all services is either protein PDB codes or protein structures upload to the server. Once the docking request is submitted, the patchDock algorithm starts the prediction process. The user is notified when the results are ready by an email message that contains a link to a web page where the predictions are presented.

STEPS

- Load the Receptor molecule
- Load the Ligand molecule
- Set the Rmsd value
- Run the server
- Get the results
- Go to Rasmol and view structure of docking.

DOCKING USING INSIGHTII

GRID DOCKING

- Build a receptor molecule
- Molecule/Get pdb
- Build a ligand molecule
- Module/Builder
- Draw ligand/optimize
- Define a Subset for Binding site
- Subset/Define
- Add Hydrogen Bond
- Measure/Hydrogen Bond
- Create Ligand Receptor Assembly
- Assembly/Associate
- Assign Potential for Ligand/Receptor molecules
- Module/Force Field
- Select/potential/CVFF
- Select the Docking Module
- Click on Docking_Grid/Creats, and create an enclosure-style grid about RECEPTOR i.e (EM1) with a border space of 5 angstroms and a 2 angstrom grid step.
- Click on Docking_Grid/Comput, and compute a docking grid of em2 with Van der waals and Coulomb energies and a cutoff of 10 angstroms. This take several minutes to computer. Make a visible grid by turning on make_vis_grid with a grid name of EM1_VGRD.
- Click on Grid/Display, and display all points of EM1_VGRD.
- Click on Grid/color and color all grid points using the charge_spectrum (turn "use_spectrum" ON)
- Click on Grid/Display, and turn off the display of the grid points.
- Evalute grid slices:
- Click on Grid/Slice

- Set the Scalar grid name to EM1_VGRD
- Set the Plane Number to 1
- Set the Plane Direction to Z
- Set the Plane Height to 59.5221
- Set the H intervals to 20
- Set the Plane Width to 59.5221
- Set the W intervals to 20
- Set the Plane Type to Map plane
- Set the Spectrum Name to CHARGE_SPECTRUM
- Set the Plane Style to Filled
- Click execute. Move the plane by moving the slide bar in the Parameters menu. When the plane bisects the molecule, the center of 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 canter 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 set to -2 and Displacement set to -6
- Add a Plane Number 2 at a Contour level of -3.
- Add a Plane Number 3 at a Contour level of -4.
- Add a Plane Number 4 at a Contour level of -5.
- Add a Plane Number 5 at a Contour level of -6.
- Add a Plane Number 6 at a Contour level of -7.

The interaction energy becomes lower (more negative) nearer the active site. Try other planes and parameters. When I done,set Plane Type to OFF for all planes.

- Evaluate Grid Contours

  Click on Grid/Contour

  Set Contour_Name_Root to contour1

Set Level Specification to Single

Set Contour Level to -15

Set Display Style to Solid

Set Color to yellow

Note where the best interaction energies lie. Then delete all contour objects.

- _dock.psv
- Click on Evaluate/Intermolecular and ADD a MONITOR energy between RECEPTOR and LIGAND. Evaluate the Intermolecular Energies

  Click on Evaluate/Intermolecular and ADD a MONITOR energy between RECEPTOR and LIGAND via the GRID.

  Position the ligand so that the VdW and Elect energies are <0 and Total energy is <-10.
- Position the system for best visualization of the ligand-receptor interaction.
- Save a folder of the model using the name manual
- Delete all objects.

TOXICITY PREDICTION

- First the homepage of Pharma algorithms was opened by typing "Pharma algorithm" in the google homepage.
- The homepage of Pharma algorithms was displayed in which clicked on the green colored "ADME/TOX WEB" option.
- Next clicked on continue button.
- The SMILES notation from Drugbank and submitted in the TOXBOX package & "submit smiles" option was clicked.
- Selected the "ADME BOXES" option to predict various parameters values . Like bioavailability, absorption, active transport, protein binding, physical chemical properties, log d values etc.

ToxBox

Pharma Algorithms announces ADME/Tox WEB, a workgroup version of ADME and Tox Boxes. ADME/Tox WEB has been specifically designed to fit into departmental

and corporate collaborative networks. Users can access the prediction tools at any time and from any computer on the organization's intranet without additional installation. ADME/Tox WEB offers its owners powerful interface customization capabilities and direct programmatic access to its modules.

Benefits

- Run Tox Boxes on any computer, no installation required
- Keep the same elegant, hassle-free user interface and world-class predictive power of and Tox Boxes
- Instantly display a look-up of the most similar structures from the training set with literature references and experimental property values
- Use its meticulously designed and dependable algorithms to your advantage
- Design proprietary user interfaces depending on the task at hand
- Access its prediction engine directly from within your scripts or intranet applications

Features

- Contains all the ADME Boxes and Tox Boxes predictive functionality

ADME overall oral bioavailability, pKa, logD, P-gp substrate and inhibitor specificity, solubility in pure water and in buffer, Abraham solvation parameters, active transport properties, absorption, physicochemical properties.

TOXICITY acute toxicity (mouse and rat), genotoxicity, health effects (blood,cardiovascular, gastrointestinal, kidney, liver, lungs).

- Has the standard ADME Boxes and Tox Boxes user interface
- Supports user clipboard copy-and-paste, interactive dictionary lookup of compounds with literature references and experimental property values
- Reads SMILES, *.mol, *.skc, *.sk2, *.cdx files from chemical drawing programs (e.g. Isis/Draw, ChemDraw, ChemSketch)
- Batch calculations can be run from within user scripts or custom graphical interfaces by directly communicating with the modules via an open protocol
- Compatible with Windows, Mac, Linux/UNIX

Chapter -4

RESULTS

RESULTS

Sequence Analysis

The chemistry and composition of the protein was studied using Protparam tool

Number of amino acids: 475

>tr|B9A061|B9A061_HUMAN Putative uncharacterized protein CAPS2 OS=Homo sapiens GN=CAPS2 PE=4 SV=1
MINTGTQMDLEVKGVAATSRSQIQPFFGRKKPLQQRWTSESWTNQNSCPPVVPRLDLGSL
VDSDDEDQNIIPENLPAPTDKCKLKYQQCKTEIKEGYKQYSQRNAENTKSNVTHKQSPRN
KIDEKCVQDEEANTDDLTTLDRKAILQQGYADNSCDKQQRARKLDAEIVAAEKKKQIVAE
QVMIDHLSRAVISDPEQNLAIEQKESDHILPDSKMTPLRFRKRTLHETKIRTHSTLTENV
LSHKLQFDGRIVSRTNVLPFIQKSIYSHQCGRRKGKQYRLGDFYVGATLTFLSSDHLSLP
ESIKENTLLKLRITNIDQIALDSLKTASMEQEDDIIIQETNDRLVFKAIQDVLKEKLHKR
GVRILTGLGKYFQQLDKEGNGLLDKADFKQALKVFHLEVSEKDFESAWLILNDNGNGKVD
YGEFKRGIIGEMNEYRKSYVRKAFMKLDFNKSGSVPIINIRKCYCAKKHSQVISG

Number of amino acids: 475

Molecular weight: 54374.8

Theoretical pI: 8.97

Amino acid composition:

Ala (A)	23	4.8%
Arg (R)	27	5.7%
Asn (N)	24	5.1%
Asp (D)	35	7.4%
Cys (C)	8	1.7%
Gln (Q)	33	6.9%
Glu (E)	31	6.5%
Gly (G)	23	4.8%
His (H)	11	2.3%
Ile (I)	33	6.9%
Leu (L)	43	9.1%
Lys (K)	49	10.3%
Met (M)	7	1.5%
Phe (F)	15	3.2%
Pro (P)	15	3.2%
Ser (S)	33	6.9%
Thr (T)	26	5.5%
Trp (W)	3	0.6%
Tyr (Y)	12	2.5%
Val (V)	24	5.1%
Pyl (O)	0	0.0%

Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 66
Total number of positively charged residues (Arg + Lys): 76

Atomic composition

Carbon	C	2382
Hydrogen	H	3855
Nitrogen	N	687
Oxygen	O	736
Sulfur	S	15

Formula: $C_{2382}H_{3855}N_{687}O_{736}S_{15}$
Total number of atoms: 7675

Extinction coefficients

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

Ext. coefficient 34880
Abs 0.1% (=1 g/l) 0.641, assuming all pairs of Cys residues form cystines

Ext. coefficient 34380
Abs 0.1% (=1 g/l) 0.632, assuming all Cys residues are reduced

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 40.85
This classifies the protein as unstable.

Aliphatic index: 81.89

Grand average of hydropathicity (GRAVY): -0.726

BLAST RESULT

For selection of appropriate template, blast p was done against target protein sequence. The template was selected with high similarity.

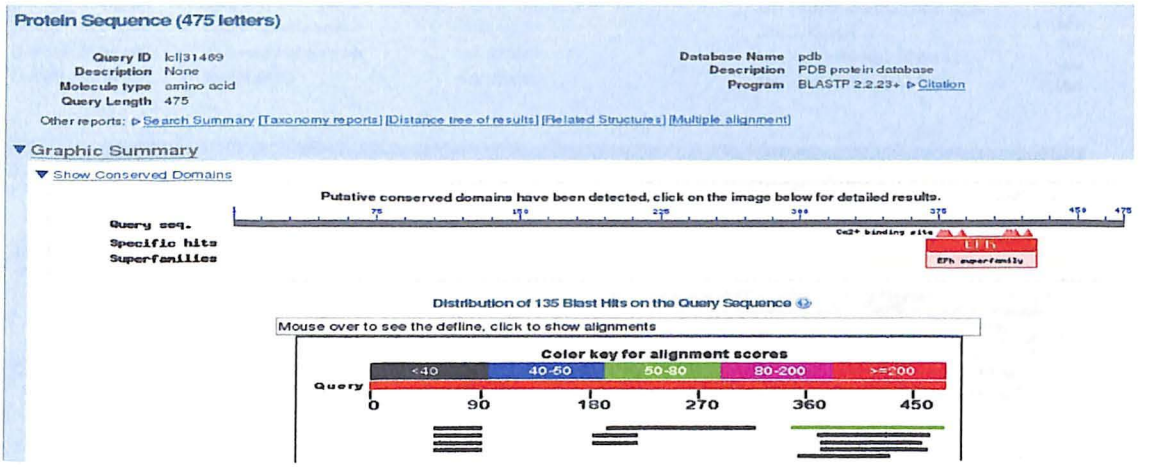


Fig 1:Blast result page

pdb|1U13|A Chain A, Crystal Structure Analysis Of The C37lC151TC442A-Triple Mutant Of Cyp51 From Mycobacterium Tuberculosis
Length=455

Score = 30.8 bits (68), Expect = 1.3, Method: Compositional matrix adjust.
Identities = 14/39 (35%), Positives = 20/39 (51%), Gaps = 0/39 (0%)

```
Query  54   RLDLGSLVDSDDEDQNIIPENLPAPTDKCKLKYQQCKTE  92
            R+ G LV +    N IPE+ P P D   +Y+Q + E
Sbjct  338  RIHEGDLVAASPAISNRIPEDFPDPHDFVPARYEQPRQE  376
```

Template sequence

>gi|51247801|pdb|1U13|A Chain A, Crystal Structure Analysis Of The C371C151TC442A-Triple Mutant Of Cyp51 From Mycobacterium Tuberculosis
MSAVALPRVSGGHDEHGHLEEFRTDPIGLMQRVRDELGDVGTFQLAGKQVVLLSGSHANEFFFRAGDDDL
DQAKAYPFMTPIFGEGVVFDASPERRKEMLHNAALRGEQMKGHAATIEDQVRRMIADWGEAGEIDLLDFF
AELTIYTSSATLIGKKFRDQLDGRFAKLYHELERGTDPLAYVDPYLPIESFRRRDEARNGLVALVADIMN
GRIANPPTDKSDRDMLDVLIAVKAETGTPRFSADEITGMFISMMFAGHHTSSGTASWTLIELMRHRDAYA
AVIDELDELYGDGRSVSFHALRQIPQLENVLKETLRLHPPLIILMRVAKGEFEVQGHRIHEGDLVAASPA
ISNRIPEDFPDPHDFVPARYEQPRQEDLLNRWTWIPFGAGRHRCVGAAFAIMQIKAIFSVLLREYEFEMA
QPPESYRNDHSKMVVQLAQPAAVRYRRRTGVHHHH

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
O94868	O94868_HUMAN		Putative uncharacterized protein MAPKAPK5	MAPKAPK5	Homo sapiens (Human)	467
A2EEY0	A2EEY0_HUMAN		Valyl-tRNA synthetase	VARS DASS-324023 5-001	Homo sapiens (Human)	1,182
B9A061	B9A061_HUMAN		Putative uncharacterized protein CAPS2	CAPS2	Homo sapiens (Human)	475
Q16521	Q16521_AEDAE		Calcium-calmodulin dependent protein kinase I...	AAEL013823 AAEL013824 AaeL_AAEL013823 AaeL_AAEL013824	Aedes aegypti (Yellowfever mosquito) (Culex aegypti)	490
A4HXT8	A4HXT8_LEIIN		60S ribosomal protein L10a, putative	LinJ18.0630 LinJ06.2990	Leishmania infantum	214
B7QS02	B7QS02_IXOSC		Putative uncharacterized protein	IscW_ISCW011594	Ixodes scapularis (Black-legged tick) (Deer tick)	699
B7QMB9	B7QMB9_IXOSC		Myosin IA, putative	IscW_ISCW014356	Ixodes scapularis (Black-legged tick) (Deer tick)	690
B7PJP7	B7PJP7_IXOSC		Phosphoinositide 3-kinase regulatory subunit,...	IscW_ISCW017661	Ixodes scapularis (Black-legged tick) (Deer tick)	1,351
B7PNH0	B7PNH0_IXOSC		Serine threonine protein kinase, putative	IscW_ISCW006840	Ixodes scapularis (Black-legged tick) (Deer tick)	173
B7QNP1	B7QNP1_IXOSC		MAP kinase-activated protein kinase, putative	IscW_ISCW015270	Ixodes scapularis (Black-legged tick) (Deer tick)	364
Q17L96	Q17L96_AEDAE		Myosin heavy chain, nonmuscle or smooth muscl...	AAEL001411 AaeL_AAEL001411	Aedes aegypti (Yellowfever mosquito) (Culex aegypti)	1,888
A4HEP7	A4HEP7_LEIBR		Valyl-tRNA synthetase, putative	LbrM30_V2.3120	Leishmania braziliensis	967
B7PUV0	B7PUV0_IXOSC		Putative uncharacterized protein	IscW_ISCW007674	Ixodes scapularis (Black-legged tick) (Deer tick)	580
B7QDV0	B7QDV0_IXOSC		Myosin IA, putative	IscW_ISCW009241	Ixodes scapularis (Black-legged tick) (Deer tick)	1,096

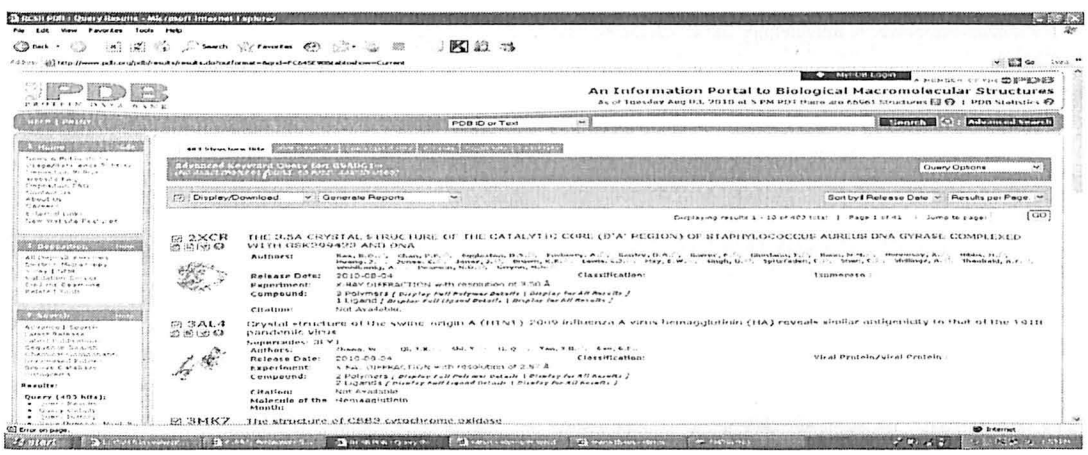


Fig 2: Showing Acc.no & PDB page.

SECONDARY STRUCTURE PREDICTION

Various secondary structure prediction servers like GOR4 and SOPMA were used and the secondary structure of the target protein was predicted.

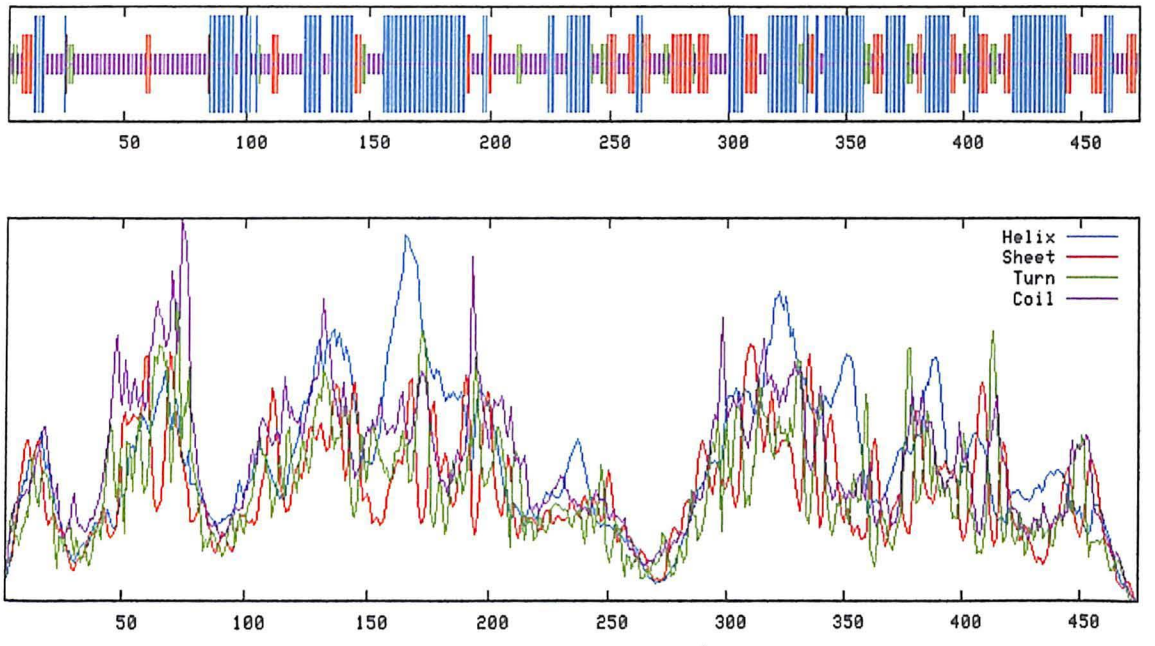
SOPMA result for : target

Abstract Geourjon, C. & Deléage, G., SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments., Cabios (1995) 11, 681-684

View SOPMA in: [AnTheProt (PC) , Download...] [HELP]

10 20 30 40 50 60
70

```
         |          |          |          |          |          |
|
MINTGTQMDLEVKGVAATSRSQIQPFFGRKKPLQQRWTSESWTNQNSCPPVVPRLDLGSLVDSDD
EDQNI
eecttcceeeeehhhhhccccccccchettccccccccccccccccccccccccccccccccceecccc
ccccc.
IPENLPAPTDKCKLKYQQCKTEIKEGYKQYSQRNAENTKSNVTHKQSPRNKIDEKCVQDEEANTD
DLTTL
ccccccccccccccehhhhhhhhhhcchhhhhcchtcccceeecccccccccchhhhhhhhcccch
hhhhh
DRKAILQQGYADNSCDKQQRARKLDAEIVAAEKKKQIVAEQVMIDHLSRAVISDPEQNLAIEQKE
SDHIL
hhhheeettcccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheecccccchheecccc
ccccc
PDSKMTPLRFRKRTLHETKIRTHSTLTENVLSHKLQFDGRIVSRTNVLPFIQKSIYSHQCGRRKG
KQYRL
cttccccccccccchhhccccchhhhhhhhhhhtcccttteeeecccccceeehhheeecccccctt
ceeee
GDFYVGATLTFLSSDHLSLPESIKENTLLKLRITNIDQIALDSLKTASMEQEDDIIIQETNDRLV
FKAIQ
eeeeetceeeeecccccccchhhhhhheeeeeeccccchhhhhhhhhhhhhttthheeehhcchhhh
hhhhh
DVLKEKLHKRGVRILTGLGKYFQQLDKEGNGLLDKADFKQALKVFHLEVSEKDFESAWLILNDNG
NGKVD
hhhhhhhhtttceeeechhhhhhhhhtttcceechhhhhhhhhhheecccttchhhheeeecttt
cccee
YGEFKRGIIGEMNEYRKSYVRKAFMKLDFNKSGSVPIINIRKCYCAKKHSQVISG
ehhhhhhhhhhhhhhhhhhhhhhheeecccccccceeeeehhhhcccccceeeec
Sequence length :   475
SOPMA :
   Alpha helix      (Hh) :   184 is  38.74%
   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) :    81 is  17.05%
   Beta turn        (Tt) :    29 is   6.11%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   181 is  38.11%
   Ambigous states  (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```



Parameters :
Window width : 17
Similarity threshold : 8
Number of states : 4

Fig 3: Showing SOPMA result page

TABLE NO 1

Alpha helix (Hh): 38.74%

Beta helix (Bb): 0.00%

Extended strand (Ee): 17.05%

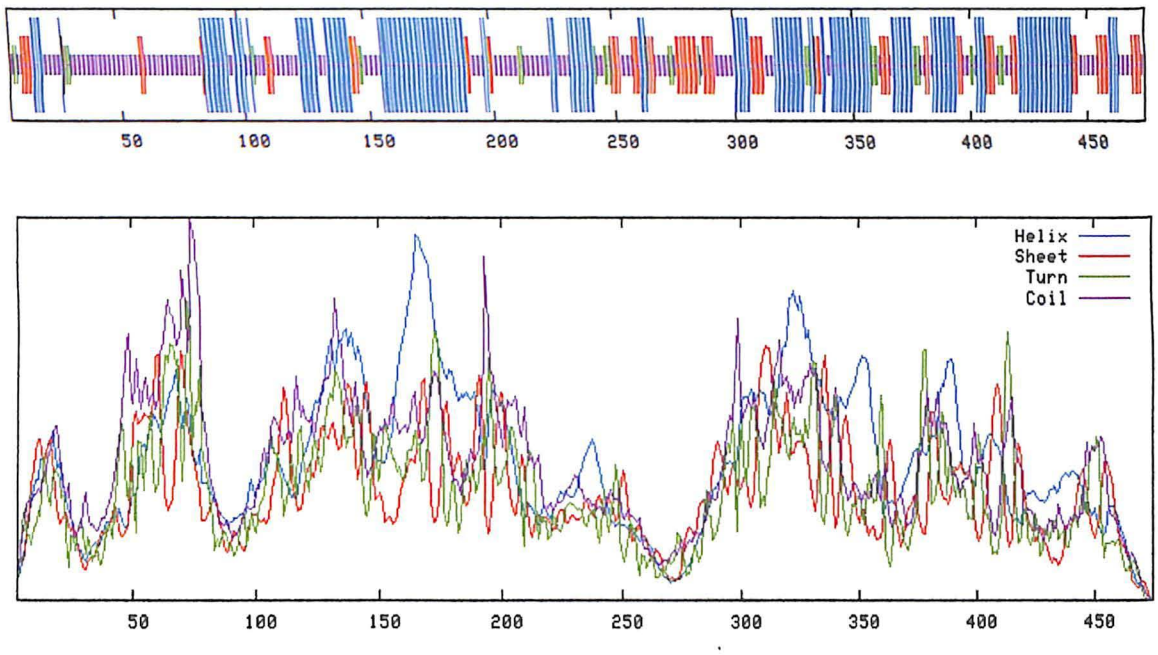
Beta turn (Tt): 6.11%

Bend region (Ss): 0.00%

Random coil (Cc) : 38.11%

GOR4 result for : ptnxxx0

Abstract GOR secondary structure prediction method version IV, J. Garnier, J.-F. Gibrat, B. Robson, Methods in Enzymology,R.F. Doolittle Ed., vol 266, 540-553, (1996)



Parameters :
Window width : 17
Similarity threshold : 8
Number of states : 4

Fig 3: Showing SOPMA result page

TABLE NO 1

Alpha helix (Hh): 38.74%

Beta helix (Bb): 0.00%

Extended strand (Ee): 17.05%

Beta turn (Tt): 6.11%

Bend region (Ss): 0.00%

Random coil (Cc) : 38.11%

GOR4 result for : ptnxxx0

Abstract GOR secondary structure prediction method version IV, J. Garnier, J.-F. Gibrat, B. Robson, Methods in Enzymology,R.F. Doolittle Ed., vol 266, 540-553, (1996)

```
         10         20         30         40         50         60
70
         |          |          |          |          |          |
|
MINTGTQMDLEVKGVAATSRSQIQPFFGRKKPLQQRWTSESWTNQNSCPPVVPRLDLGSLVDSDD
EDQNI
ccccccchhhhccccccccccccccccccccceeeeeeeeeccccccccccccccccceeccccc
ccccc
IPENLPAPTDKCKLKYQQCKTEIKEGYKQYSQRNAENTKSNVTHKQSPRNKIDEKCVQDEEANTD
DLTTL
ecccccccccccceeecccceecccchhhhhhccccccccceecccccccccccccccccccccc
hhhhh
DRKAILQQGYADNSCDKQQRARKLDAEIVAAEKKKQIVAEQVMIDHLSRAVISDPEQNLAIEQKE
SDHIL
hhhhhhhcccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhc
ccceee
PDSKMTPLRFRKRTLHETKIRTHSTLTENVLSHKLQFDGRIVSRTNVLPFIQKSIYSHQCGRRKG
KQYRL
ccccchhhhhhhhhhhcccccccccchhhhhhhhccccceeecccccccccccceeeeccccccccc
cceee
GDFYVGATLTFLSSDHLSLPESIKENTLLKLRITNIDQIALDSLKTASMEQEDDIIIQETNDRLV
FKAIQ
cceeccceeeeeccccccccchhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhchhhhhhhhhhh
hhhhh
DVLKEKLHKRGVRILTGLGKYFQQLDKEGNGLLDKADFKQALKVFHLEVSEKDFESAWLILNDNG
NGKVD
hhhhhhhhhcceeeeecccchhhhhhhhcccchhhhhhhhhhhhhhcccccchhhhhhhheeecccc
cccee
YGEFKRGIIGEMNEYRKSYVRKAFMKLDFNKSGSVPIINIRKCYCAKKHSQVISG
ccccccchhhhcchhhhhhhhhhhhhhhcccccccceeeeeecccccccccccceeec
```

Sequence length : 475

GOR4 :

Alpha helix	(Hh) :	191	is	40.21%
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) :	58	is	12.21%
Beta turn	(Tt) :	0	is	0.00%
Bend region	(Ss) :	0	is	0.00%
Random coil	(Cc) :	226	is	47.58%
Ambigous states	(?) :	0	is	0.00%
Other states	:	0	is	0.00%

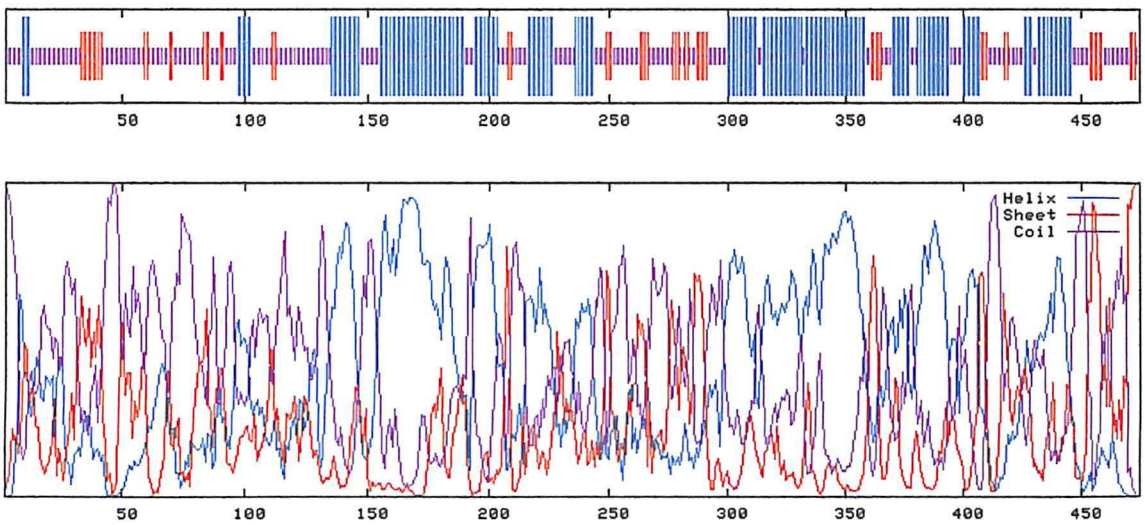


Fig 4: Showing GOR4 result page

TABLE NO 2

Alpha helix (Hh):40.21%

Beta bridge (Bb): 0.00%

Extended strand (Ee): 12.21%

Beta turn (Tt): 0.00%

Bend region (Ss): 0.00%

Random coil (Cc) : 47.58%

Homology Modeling in Insight II

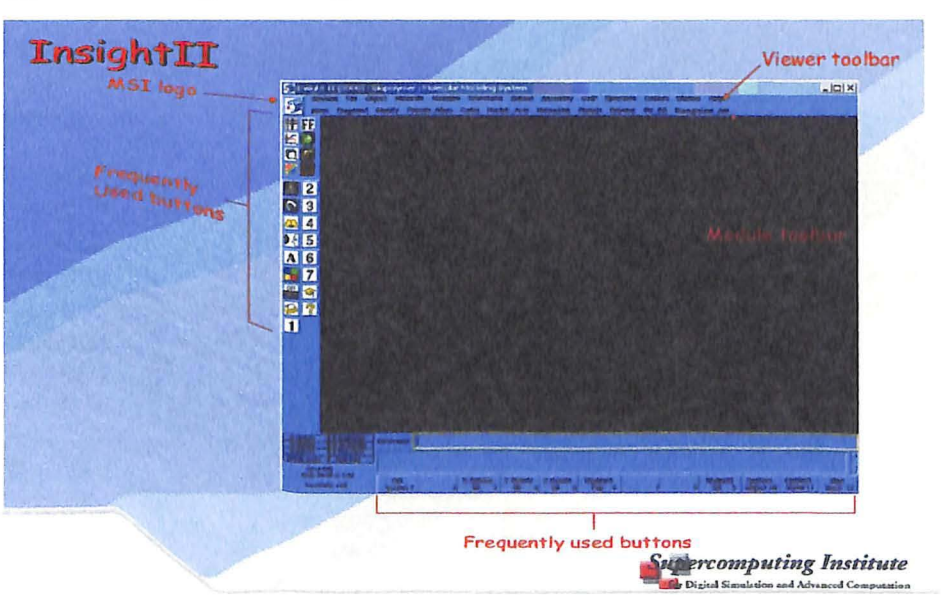


Fig 5: InsightII Homepage

Protein Having Unknown Structure is Modeled Using its template with the help of InsightII

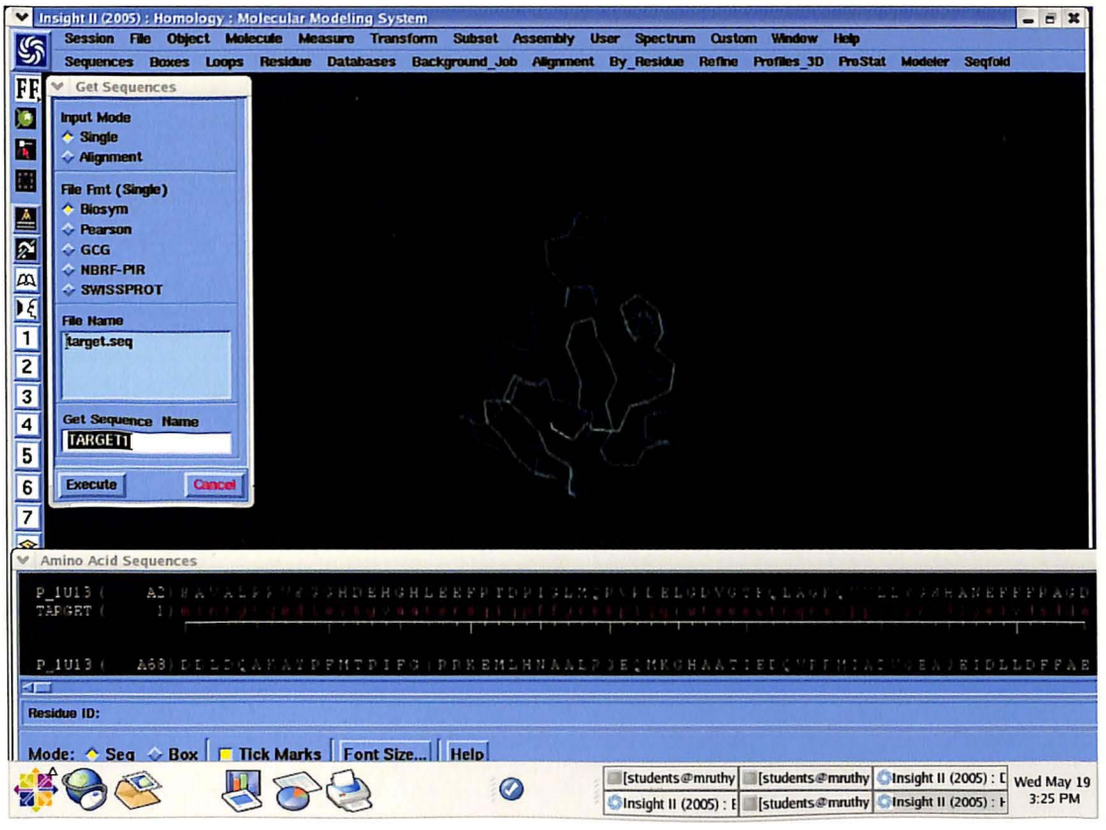


Fig 6: Model protein of 1U13

TABLE NO 3

Table Of Ten Predicted Model Protein:

Model protein	Score
TAR.B99990001.pdb	0.49
TAR.B99990002.pdb	0.47
TAR.B99990003.pdb	0.46
TAR.B99990004.pdb	0.46
TAR.B99990005.pdb	0.51
TAR.B99990006.pdb	0.53
TAR.B99990007.pdb	0.57
TAR.B99990008.pdb	0.46
TAR.B99990009.pdb	0.43
TAR.B99990010.pdb	0.49

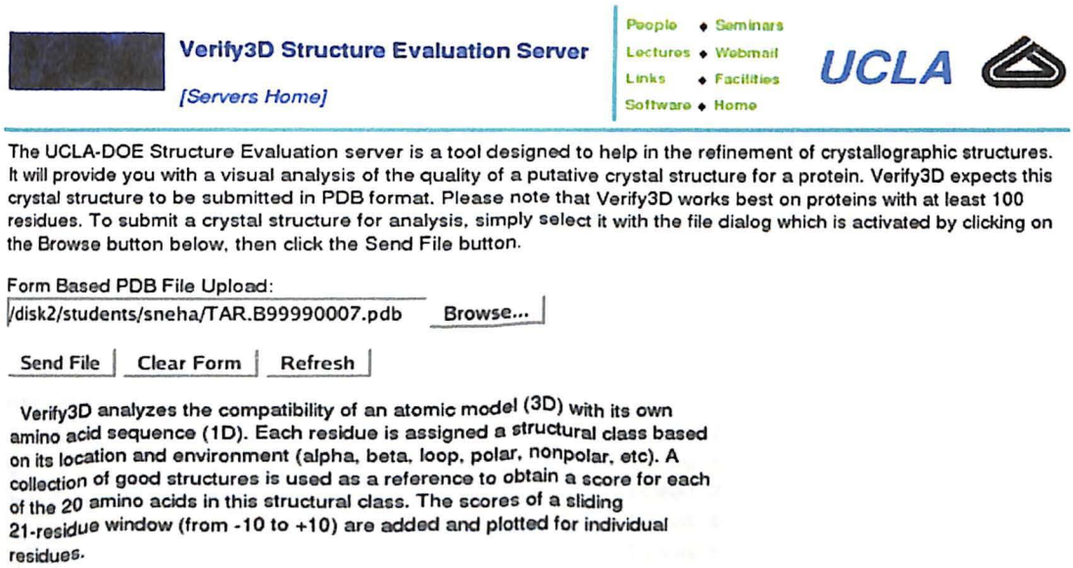


Fig 7:Verify 3D Homepage

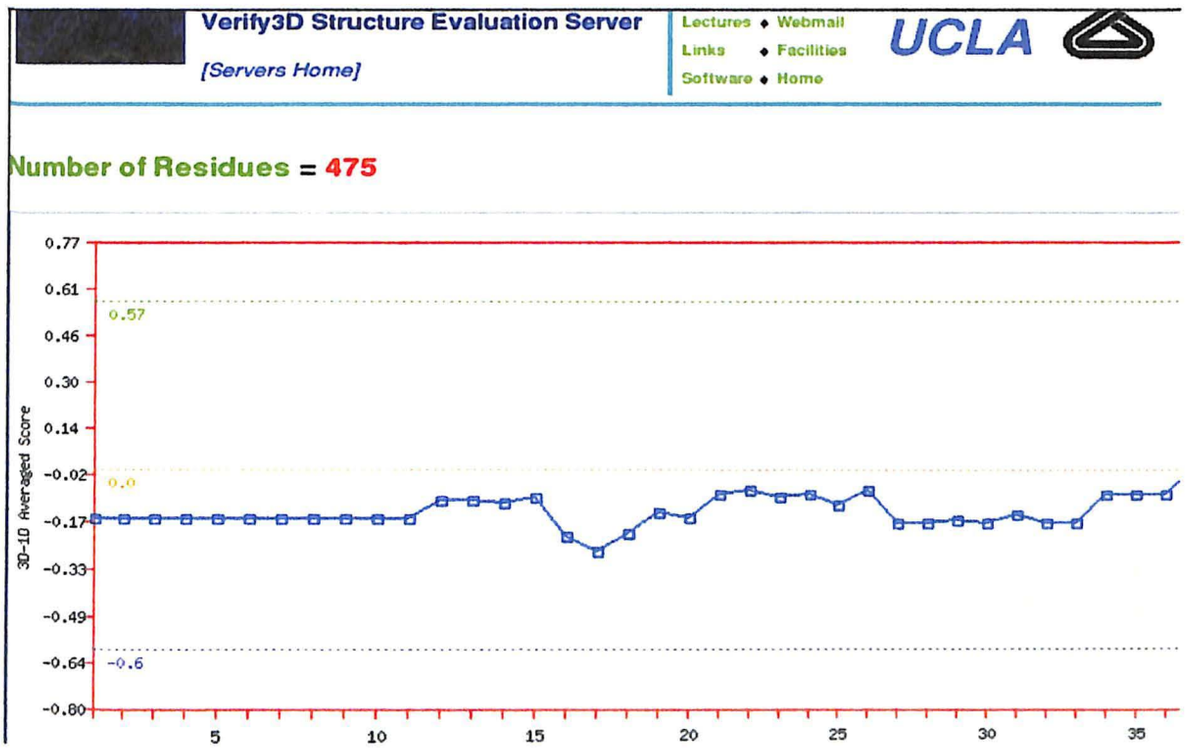


Fig 8: Varify_3D graph showing Highest Value:

RAMPAGE: Assessment of the Ramachandran Plot

The model developed in Insight II was further assessed using online server of Ramachandran Plot i.e.,RAMPAGE SERVER and it was found that 90.7% of residues were found in favoured region , 7.6% of residues were found in allowed region and 1.7% of residues in outlier region.

TABLE NO 4

Evaluation of residues:

```
Residue [  271 :GLY] (  49.18,-107.90) in Allowed region
Residue [  280 :LEU] (-137.45,  -93.65) in Allowed region
Residue [  289 :LEU] (-120.62,   62.88) in Allowed region
Residue [  307 :THR] (-102.74,   39.08) in Allowed region
Residue [  347 :LYS] (-112.54,   44.41) in Allowed region
Residue [  351 :ASP] (-158.62, -146.91) in Allowed region
Residue [  373 :GLN] ( -54.49,  102.43) in Allowed region
Residue [  374 :GLN] ( 179.35, -176.75) in Allowed region
Residue [  381 :GLY] (-155.19,  112.22) in Allowed region
Residue [  382 :LEU] (-145.86,   88.41) in Allowed region
Residue [  384 :ASP] ( 173.12,  151.38) in Allowed region
Residue [  395 :PHE] (-151.23, -163.72) in Allowed region
Residue [  396 :HIS] ( -91.43, -156.44) in Allowed region
Residue [  397 :LEU] (-135.46, -153.92) in Allowed region
Residue [  407 :ALA] (  63.79, -161.91) in Allowed region
Residue [  422 :GLY] ( 122.33,  -56.35) in Allowed region
Residue [  457 :ILE] (-122.07,   72.19) in Allowed region
Residue [  460 :ILE] (-159.66, -168.46) in Allowed region
Residue [  464 :TYR] (-111.78,   65.20) in Allowed region
Residue [  206 :SER] (-166.29, -135.02) in Outlier region
Residue [  212 :ASP] (  -6.50,   73.24) in Outlier region
Residue [  213 :SER] (  13.48,  103.83) in Outlier region
Residue [  219 :ARG] ( 116.45,  173.39) in Outlier region
Residue [  312 :ARG] (-162.72, -138.42) in Outlier region
Residue [  336 :ILE] (  41.76,  110.18) in Outlier region
Residue [  442 :LYS] (  91.95,  130.59) in Outlier region
Residue [  458 :ILE] (  54.88,  150.49) in Outlier region
Number of residues in favoured region   (~98.0% expected)  :  429 ( 90.7%)
Number of residues in allowed region    ( ~2.0% expected)  :   36 (  7.6%)
Number of residues in outlier region                       :    8 (  1.7%)
```

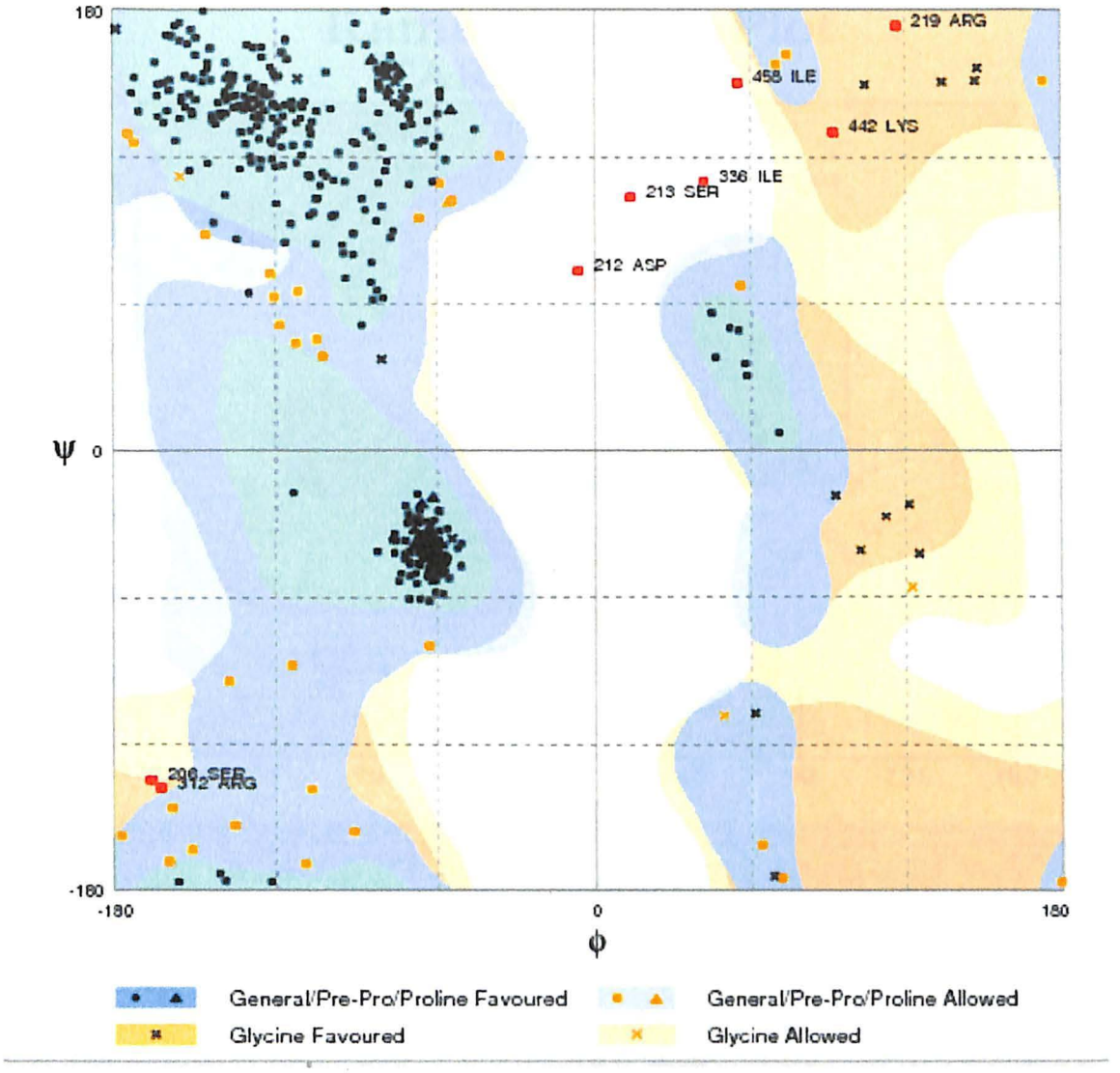
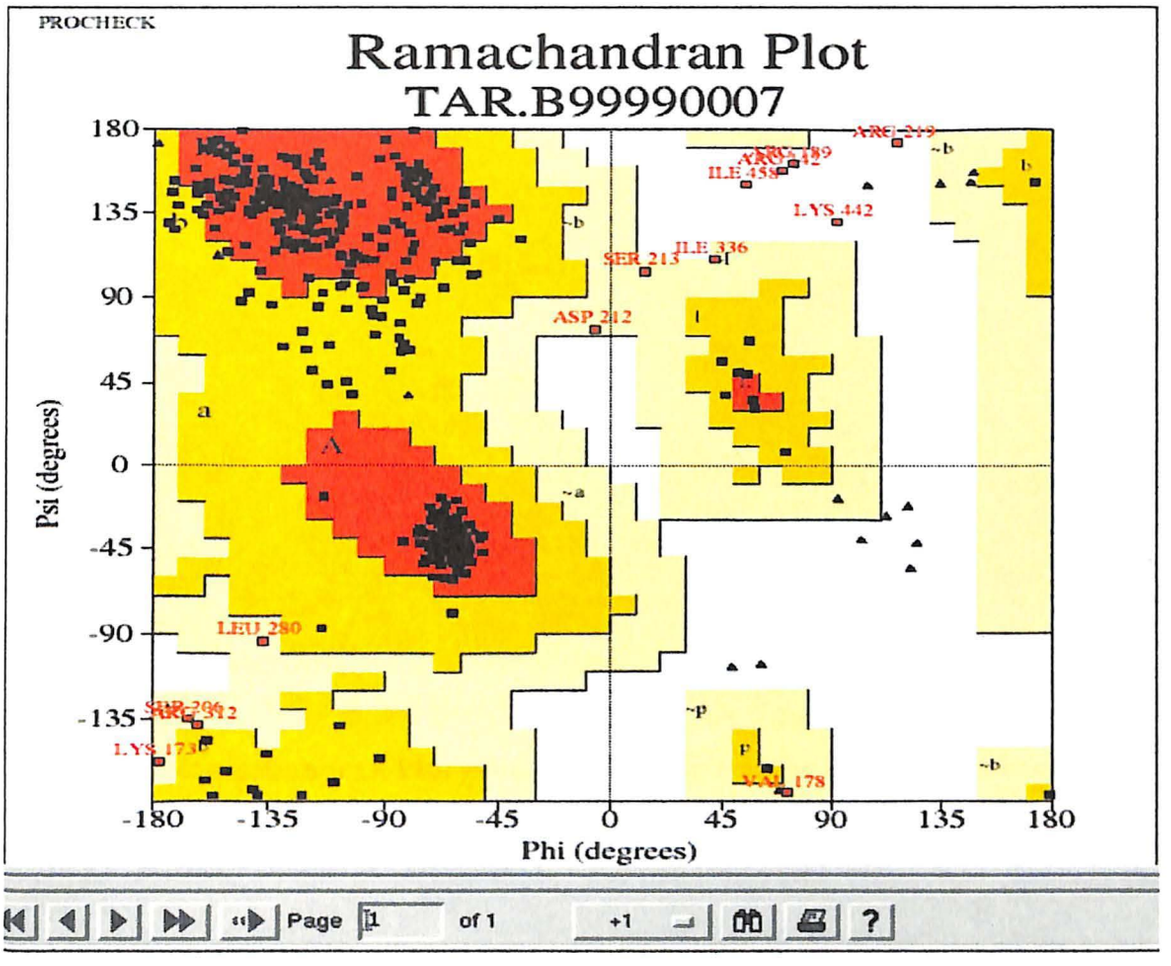


Fig 9:Ramachandran Plot generated for the model protein.



Residues in most favoured regions [A,B,L]	354	81.2%
Residues in additional allowed regions [a,b,l,p]	69	15.8%
Residues in generously allowed regions [~a,~b,~l,~p]	8	1.8%
Residues in disallowed regions	5	1.1%
Number of non-glycine and non-proline residues	436	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	23	
Number of proline residues	15	
Total number of residues	475	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig 10:showing Rampage in PROCHECK.

Ramachandran plot in InsightII

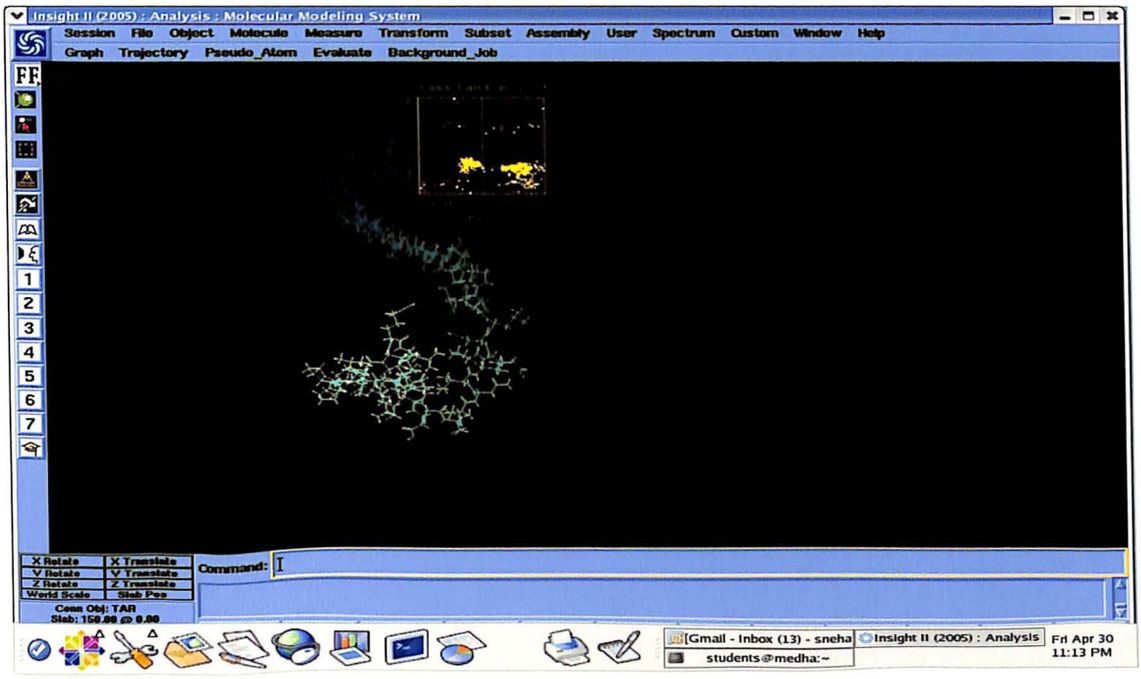


Fig 11: Ramachandran Plot generated in Insight II for the Modeled Protein

Energy Minimization in InsightII

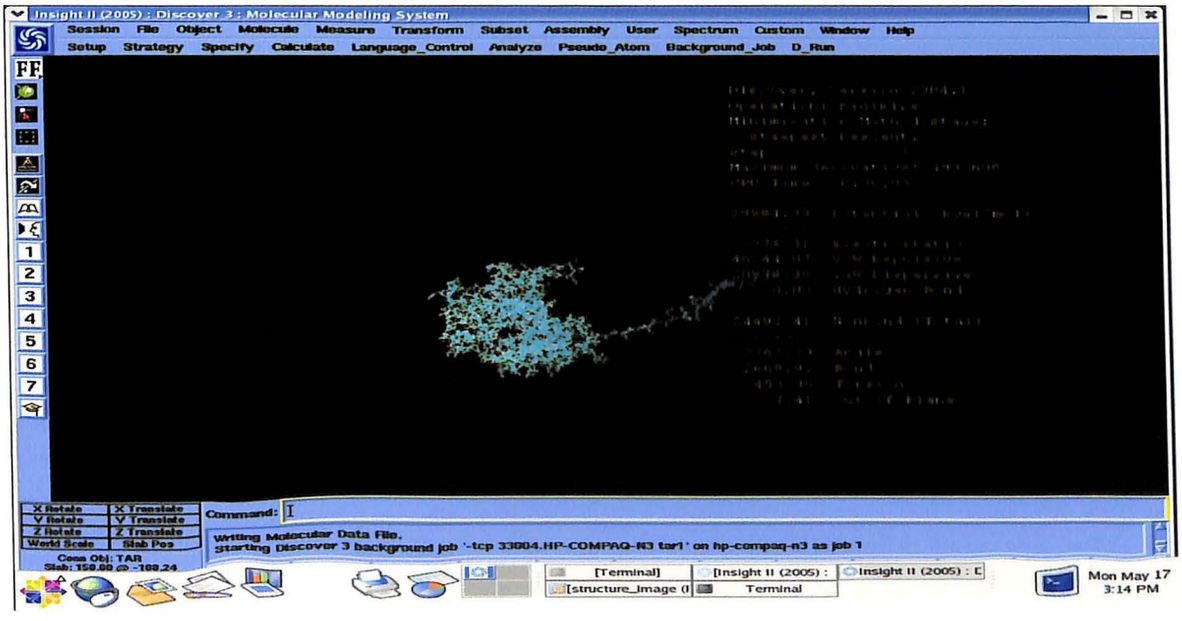


Fig 12: Energy minimization in 3 step.

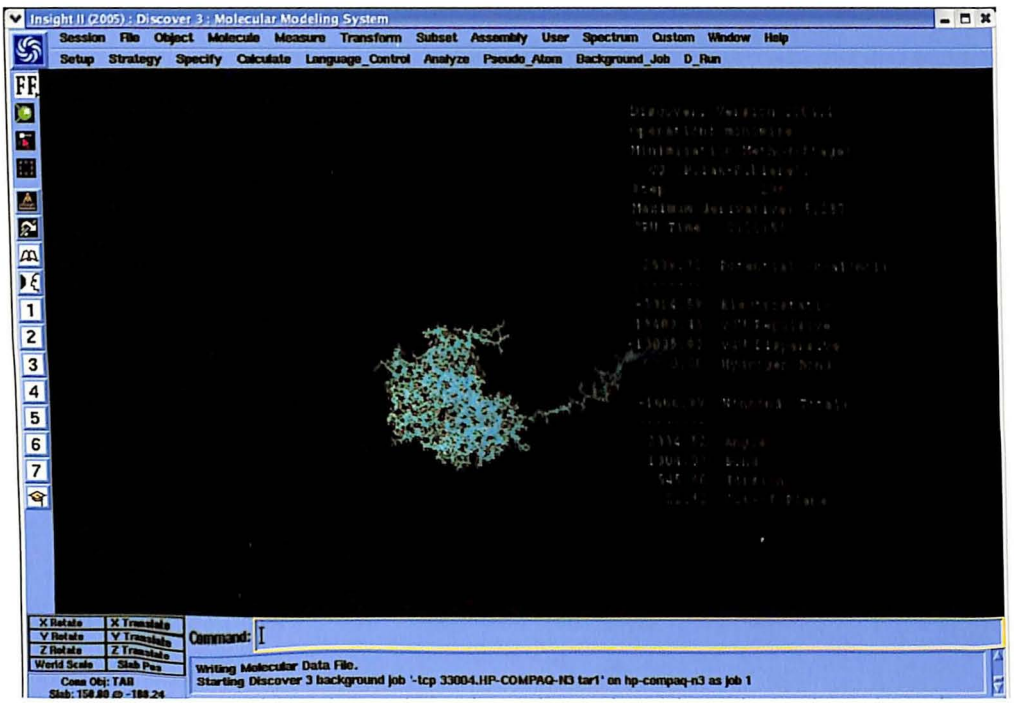


Fig 13: Energy minimization in 296 step.

DOCKING RESULTS

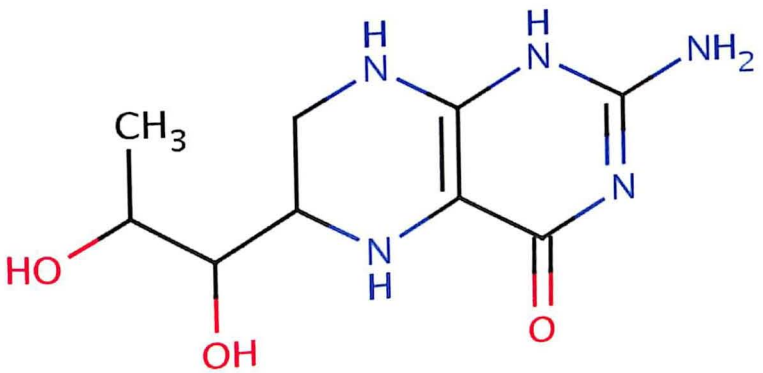


Fig 14 : Ligand structure

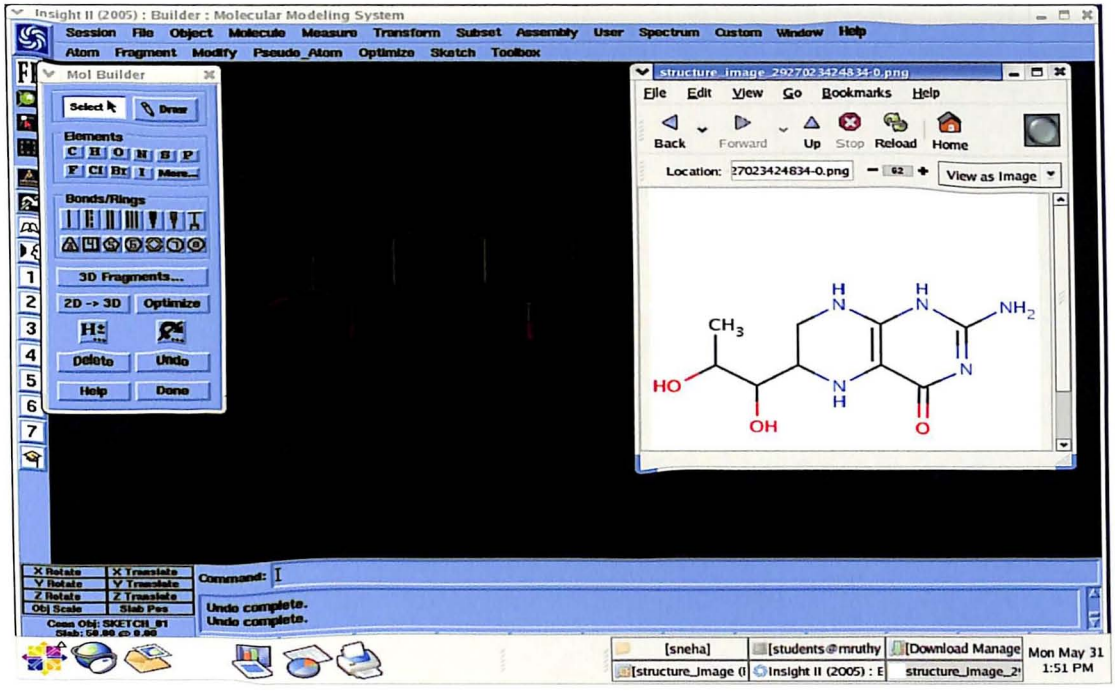


Fig 15: Building Ligand in InsightII

Docking in InsightII

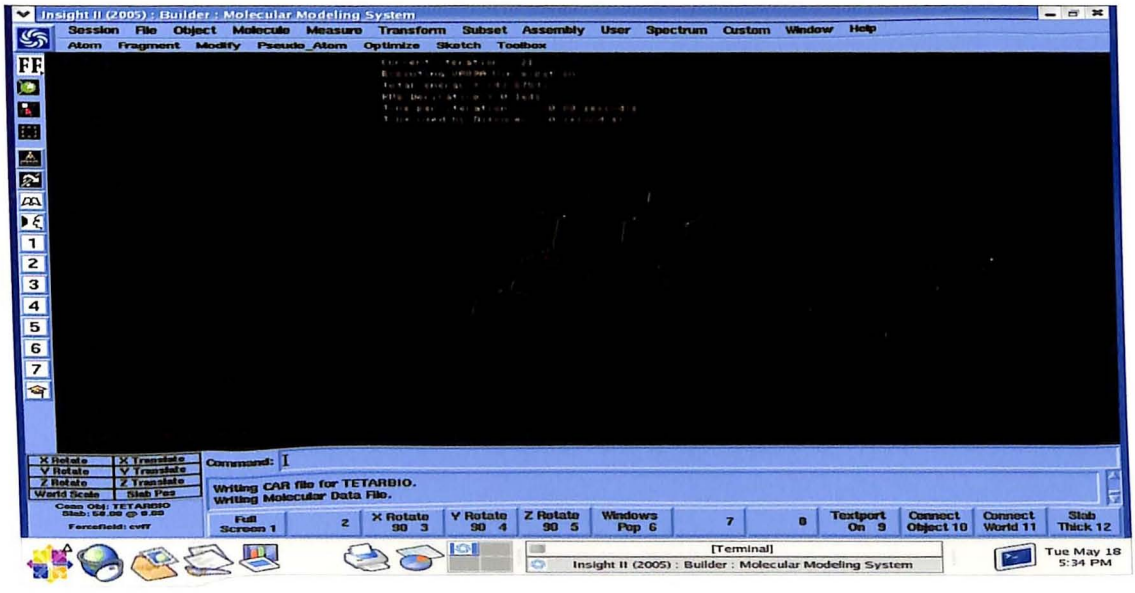


Fig 16: Optimization of Ligand

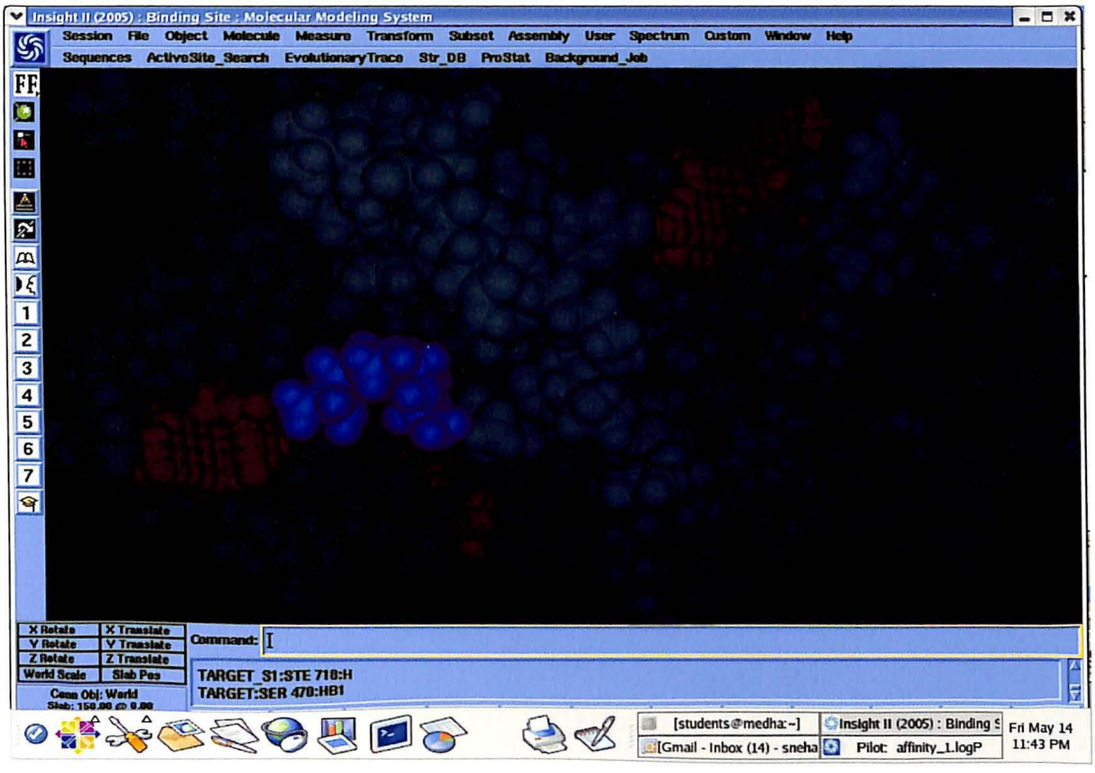


Fig 17: Ligand and receptor bind in Active site

Patch Dock

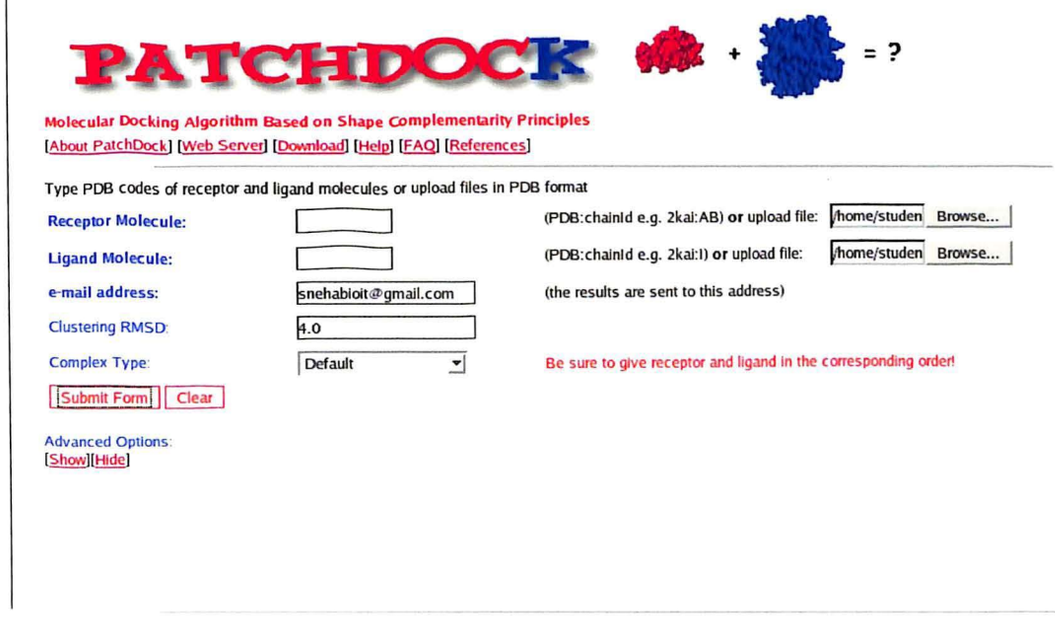


Fig 18: PatchDock Homepage

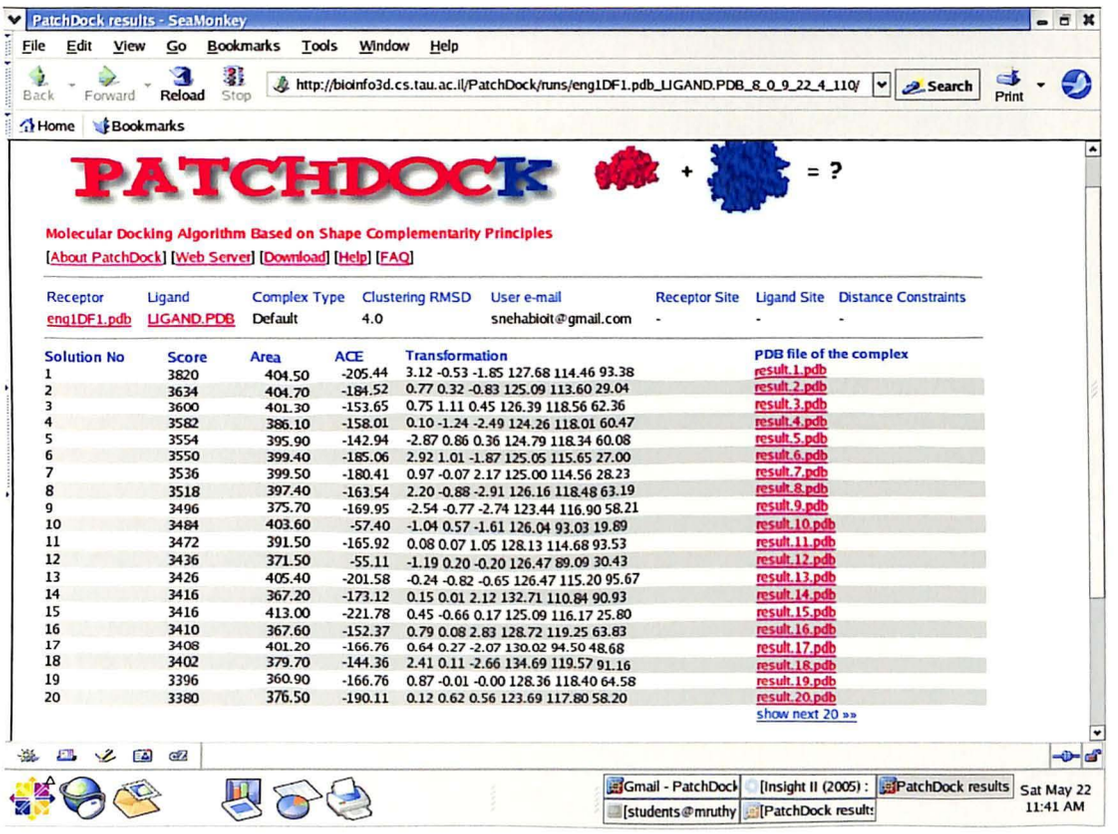


Fig 19: PatchDock PDB result

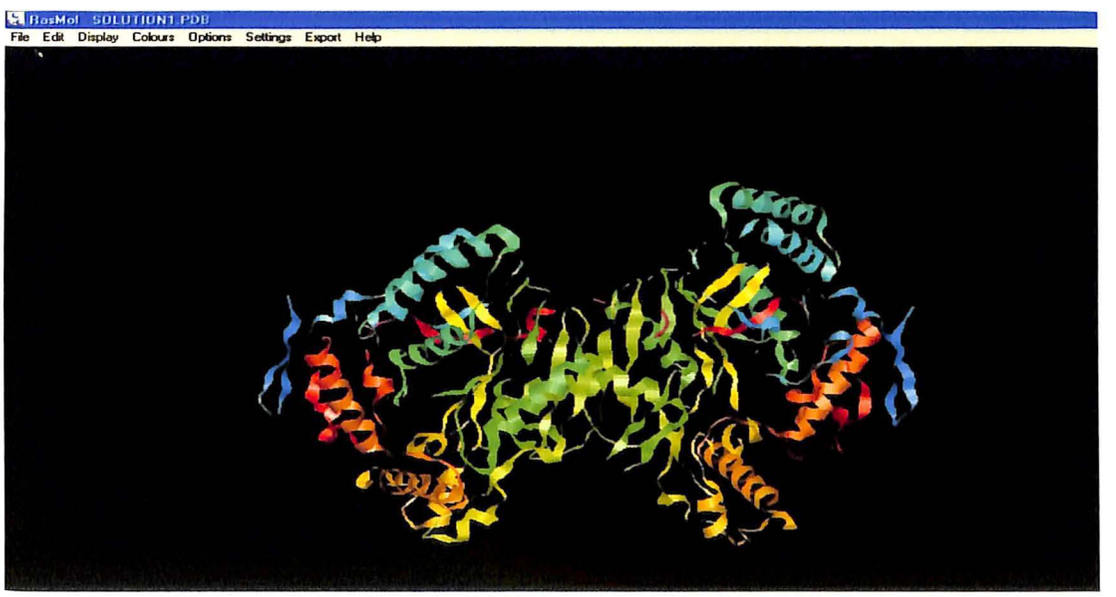


Fig 20:PatchDock result

GRID DOCKING

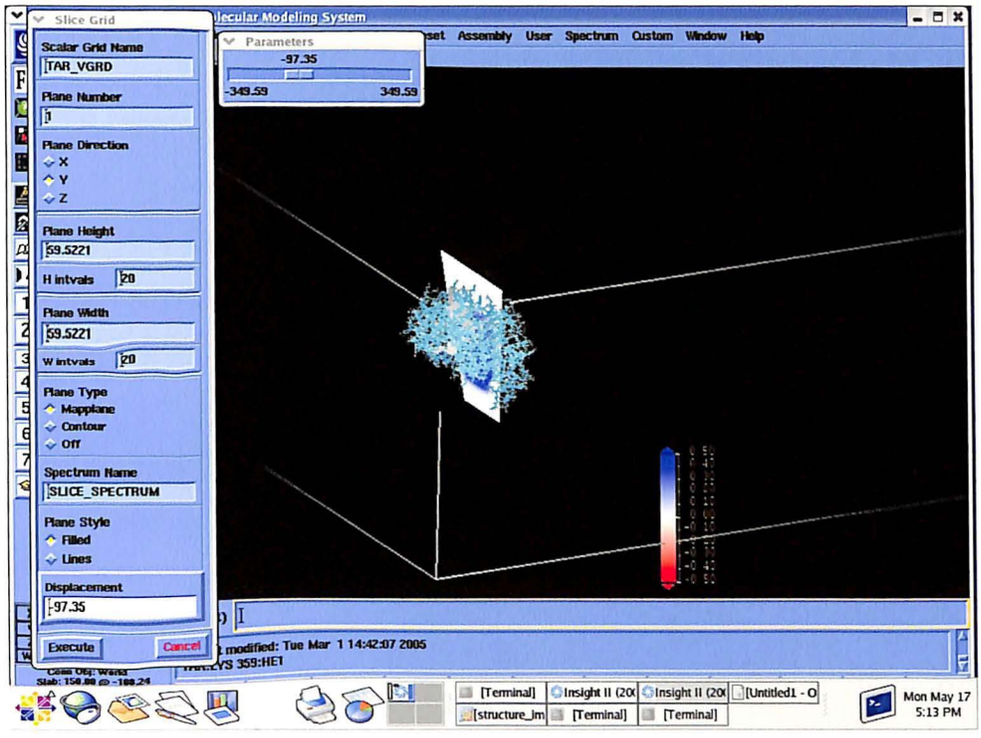


Fig 21: Result of Grid Docking showing Grid Slice

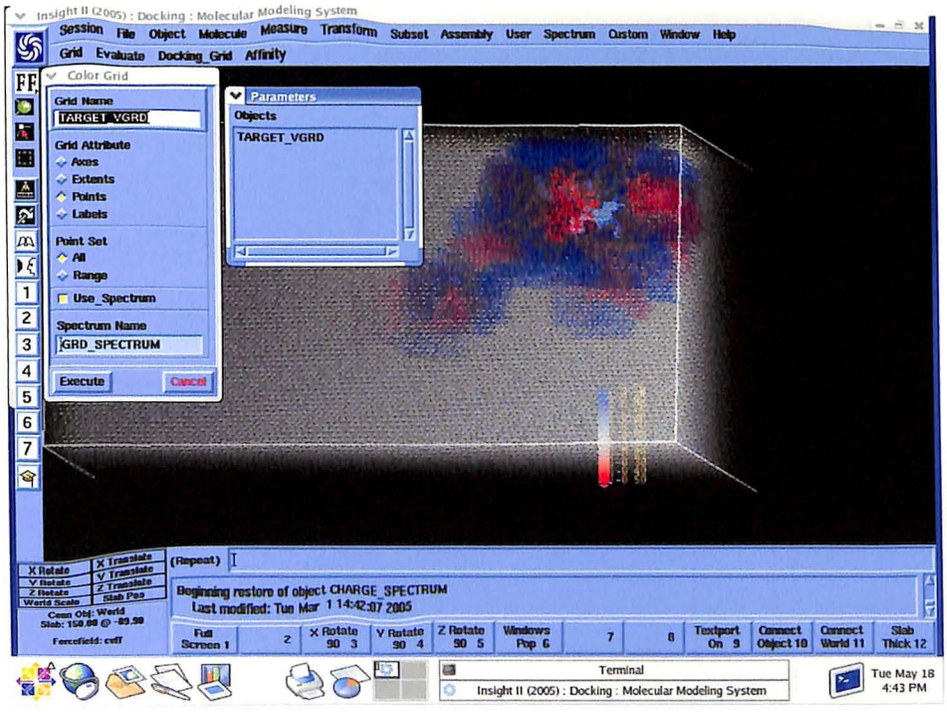


Fig 22: Result of the Grid Docking showing Color of the Grid & CHARGE_SPECTRUM

TOXICITY ANALYSIS RESULTS

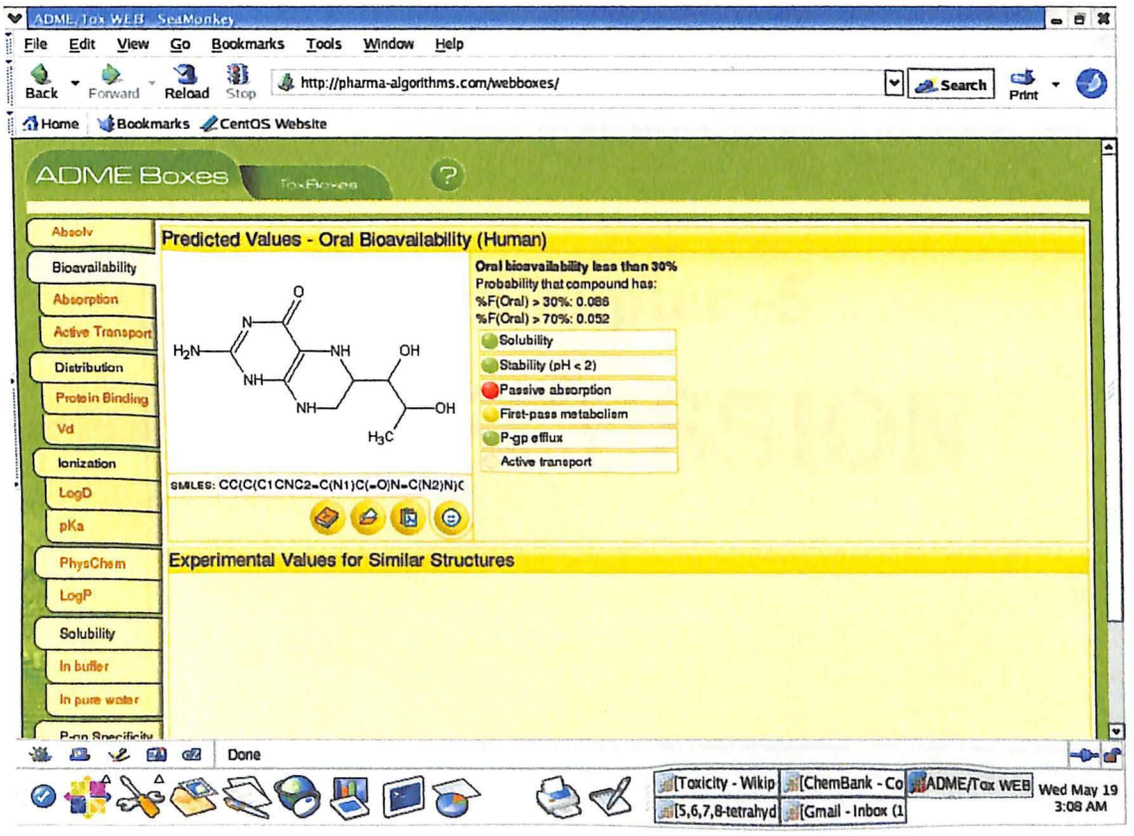


Fig 23: Oral Bioavailability Values.

Chapter -5

DISCUSSION

DISCUSSION

We had Built 3D Model of the Autism Disease Related Protein L1 CAM 5 and Dock with Tetrahydrobioptrin ligand which is a anti-autism agent .

Homology Modeling using InsightII need a template sequence using which the model of the target sequence predicted ,We had Built ten sample model and take the Highest verified 3D value and did its energy minimization.

To find the template sequence we had performed a BLAST experiment and the Result shown in Figure 1 .We observed the parameter of score , E-value , Gap and Identity and based on these we had choose the best suitable template having pdb id 1U13 of identity 35% and Gap 0.

Figure 2 shown the Accession number of target protein , sequence length 475, gene name CAPS2 of Homo sapiens & PDB page shown that no exact matches only fuzzy structure found.

We had done the secondary structure prediction for the target protein and analysis its Alpha Helix, Beta Sheet and Random Coil Region using the Online Tools SOPMA and GOR4 Result page are showing Figure 3 & 4 the Alpha helix, 3_{10} helix, Pi helix ,Beta bridge, Extended strand, Beta turn, Bend region, Random coil, Ambigous states and Others.

Figure 5 shown the InsightII Homepage within this it indicate the MSI logo , Frequently used buttons, Viewer toolbar & module toolbar.

We Are Upload the target and Template in InsightII to Build Comparative model. Figure 6 shows the target and template sequence and the 3D Structure of the template. The template sequence appear in Capital Letter Because it have already 3D Structure , and the target sequence appeared in small letter because it has no 3D structure and we

have to predict the structure. The ten predicted structure are formed showing in the table 3.

We had gone for online 3D Structure validation tools like Rampage and Varify_3d to choose the best model out of these ten model.

Figure 7 shown the Homepage of Verify 3D. In which I puted highest value for show the graph.

Figure 8 shown the Highest Graph value of the Varify_3D result is 0.57 of the Model protein . TAR.B99990007.pdb.

Table 4 shown the Evaluation residues in which indicated favoured region 429(90.7), in allowed region 36 (7.6%), and in outlier region 8(1.7%) .

Figure 9 shown Rampage of the target protein having Proline favoured , Proline allowed region & Glycine favoured, Glycine allowed region. In the Proline favoured region the dark black spot shown most favourable region.

Figure 10 & 11 shown the Ramachandran plot in InsightII for the model protein & in PROCHECK.

Figure 12 & 13 shown the energy minimization step 3 & step 296. Maximum derivative, potential (kcal/mol) ,Electrostatic , vdw repulsive , vdw dispersive , total nonbond , angle, bond in step 3 to step 296 gradually decreases & vwd dispersive,torsion & cut-of-plane gradually increases.

Figure 14 shown the ligand structure.Figure 15 shown the ligand drawn in the InsightII page. Figure 16 shown the optimization of ligand.

Figure 17 shown the ligand binding to the active site of the receptor. Blue color shows the Ligand, Red color shows the Active site & Sky color shows the Receptor molecule.

Figure 18 PatchDock Homepage & Figure 20 shows the Rasmol structure of the target after PatchDock result.

Figure 21 shows the Grid Slice_spectrum & the parameter intersects the target protein.Figure 22 also shows the color Grid spectrum & the parameter of charge spectrum.

Figure 23 shows the ADME Boxes of Toxicity results only showing the Bioviailability values.

In Docking Process we docked the modeled with Tetrahydrobiopterin which is best fit the model protein.Tetrahydrobiopterin is the Anti Autism Drug .

Chapter-6

SUMMARY

SUMMARY

Study of "Homology Modeling" & "Docking" of ligand gives the knowledge about the structural conformation of protein & how the bonding conformation & interaction energy between small organic molecules & biological receptors. These computational methods can be used to determine the affinity with which a ligand will bind to target receptor. We have outline the various steps involved in the process of homology modeling & methods of protein modeling. Homology modeling has been applied to L1 CAM 5 protein & ten targets were generated in both modeler & InsightII. Out of these ten target (models) the best one was chosen according to the highest score obtained by verify 3D with that target energy minimization was done docking also performed with the same target & four different ligands. From a drug design stand point it was seen that the target or the receptor binds in the suitable way as docking energy for that ligand is the least one. So it can be used as a drug for "Autism" disease. There is no known cure for Autism disease .

However, many different types of treatment can be used to alleviate symptoms & to modify the disease process. The goal of treatment in this disease must be two folds to alleviate the symptoms of the patient here & now to prevent the future destruction of the brain if the disease is left unchecked. These two goals may not always coincide: while pain relievers may achieve the first goal, they do not have any impact on the long-term consequences. For these reasons, most authorities believes that Autism should be treated in the vast majority of patients, by at least one specific anti Autism medication named Tetrahydrobiopterin to which other medications & non-medical interventions can be added as needed.

In our studies we have tried to show the accuracy of ligand & its action as a inhibitor. Our ligand Tetrahydrobiopterin can be used for the further study as it is an anti-inflammatory analgesic & antipyretic drug.It is used in the treatment of Autism more work is needed to determine its usefulness as drug.