www.sphinxsai.com

International Journal of ChemTech Research
CODEN (USA): IJCRGG ISSN : 0974-4290
Vol.6, No.7, pp 3864-3882, Sept-Oct 2014

Homology Modeling and Docking Studies of MPT 51 Protein in *Mycobacterium leprae*

K.Praveen Kumar[1], U.Janardhan Reddy[1], A.Ranganadha Reddy[2*], N.ShyamalaDevi[3]

[1]School of Information Technology, Vignan University, Guntur, Andhrapradesh 522 213, India.
[2]School of Biotechnology, Vignan University, Guntur, Andhrapradesh 522 213, India.

[13]Departmentt. of Biotechnology, Sree Vidyanikethan Engg. College, Tirupathi, Andhrapradesh 517 102, India.

***Corres.author: rangaaluri@gmail.com**
Mobile no: 08019963972

Abstract: Mycobacterium leprae, the causative agent of leprosy, is known to secrete a number of highly immunogenic proteins that are thought to confer pathogencity, by mediating binding to host tissues. Among these, the secreted proteins known as the trimeric antigen 85 (Ag85) complexes and the related MPT51 protein, also termed as FbpC1. MPT51 protein binds to the fibronectin of the extracellular matrix. It may also have a role in host tissue attachment and virulence. The MPT51 proteins play a vital role in acyltransferase activity. These acyltransferases mainly involved in lipid biosynthesis. Lipids are not only the important components for energy production but are useful in building blocks for cell wall composition as well.

Considering the importance and lack of specific structure, of MPT-51 we have predicted the structure by using homology modeling method. Further they have performed docking studies with drug molecules like Dapsone, Naproxen, Tiaprofenic acid and Carprofen, Among these Carprofen was found to be the best drug molecule with least binding energy, Further we have selected Analogues of Carprofen and continued the next stage of study with docking studies to find a more better drug molecule than standard carprofen.

Keywords: Mycobacterium leprae, MPT51 protein, acyltransferase, Homology modeling, Docking.

INTRODUCTION

Vaccines are proven to be the greatest tools in the prevention of infectious diseases. The mycobacterial diseases leprosy and tuberculosis remain serious public health problems even today. Leprosy affecting 12 million people worldwide[1] and tuberculosis projected to result in 1 to 2 million deaths per annum. *Mycobacterium leprae*, also known as Hansen's bacillus, is an intracellular, pleomorphic, acid fast bacterium[2] that causes leprosy. Mycobacterium leprae is a Gram-positive, aerobic rod-shaped (bacillus) surrounded by the characteristic waxy coating unique to mycobacteria. In size and shape, it closely resembles Mycobacterium tuberculosis. Mycobacterium leprae stains with a carbol fuscin rather than with the traditional Gram stain it is all because of the existed thick waxy coating. It was the first bacterium identified that causes disease in humans.[3, 4]

Current available serological tests show a relatively low sensitivity, corresponds to the low prevalence of leprosy, limiting the application to early case finding and detection of infection.[5] Current treatments share a common basis made up of several combined drugs, particularly Rifampicin, Dapsone, and Clofazimine.[6] Antimicrobials most recently added to the therapy of leprosy include fluoroquinolones, tetracyclines, and

macrolides.[7] Supervised Multi-Drug Therapy (MDT) for fixed duration is highly effective for all forms of the disease.[8] The widespread implementation of MDT has been associated with a fall in the prevalence of leprosy but as of now there is no reduction in the case-detection rate globally.[9] Early diagnosis, adequate treatment, and health education are still the most important control measures.[10]

Attempts to identify strain variation in different isolates of *M*ycobacterium *leprae* have met with little success.[11] The Multi-Drug Therapy (MDT) has been very successful in reducing the prevalence of the leprosy, but still it is a major health problem globally[12]. MPT51 of *Mycobacterium tuberculosis* is an immune dominant recognized by serum antibodies from ~80% of human immunodeficiency virus-negative,smear-positive tuberculosis and leprosy patients. The protein MPT51 is produced by *M*ycobacterium *leprae*.[13] MPT51 is a non-catalytic α/β hydrolase that may be involved in *M*ycobacterium *leprae* adhesion mechanisms. The MPT51 expressed in the early stages of infection and is recognized by both healthy infected individuals and patients with active leprosy.[14]

In our study, we have modeled Three dimensitional structure of MPT-51 protein was, using Mol soft and MOE software. The final refined model was further assessed by RAPPER server (ramachandran plot analysis), and the results show that this model is reliable.

COMPUTATIONAL TOOLS

All calculations were carried out in Maestro v9.2 installed in Cadd-WS3 machine under 64-bit centos operating system placed in CADD department, Institute of Life Sciences. The machine was built up with:

A) 4 cores and 8 processers with Intel Xenon CPU E5620 @ 2.40GHZ
B) 16 GB RAM
C) NVidia Qudvo FX3800 Graphical Process Unit (GPU)
D) The PROCHECK analysis provides an idea of the stereo chemical quality of all protein chains in a given PDB structure. They highlight regions of the proteins which appear to have unusual geometry and provide an overall assessment of the structure as a whole.
E) Other Servers
   1) Primary sequence of the MPT 51 Protein was retrieved from Swiss Prot (accession number **Q05861**) from the ExPASy (Expert Protein Analysis System) proteomics serves of the Swiss Institute of Bioinformatics.
   2) Homology search for MPT 51 protein was carried out using BLAST software.
F) The crystal structure for MPT 51 protein (PDB ID: 1SFR) was obtained from PDB database RCSB.

DATABASES

Database: A database is a collection of information that is organized so that it can easily be accessed, managed, and updated.

Data mining: Data mining or knowledge discovery is the computer-oriented process of digging and analyzing large volumes of data and finally extracting the meaning of the data. Data mining can be performed on data represented in quantitative, textual, or multimedia forms.[15] Applications of data mining to bioinformatics include gene finding, protein function domain detection, function motif detection, protein function inference, disease diagnosis, disease prognosis, disease treatment optimization, protein and gene interaction network reconstruction, data cleansing, and protein sub-cellular location prediction.

Swiss-Prot: Swiss-Prot was was created in 1986 by Amos Bairoch and developed by the Swiss Institute of Bioinformatics and subsequently developed by Rolf Apweiler at the European Bioinformatics Institute.[16,17,18] Swiss-Prot aimed to provide reliable protein sequences associated with a high level of annotation a minimal level of redundancy and high level of integration with other databases.

Protein Data Bank: PDB consists of 3D (Three Dimensional) data of experimentally determined structures of proteins and nucleic acids[19] established at Brookhaven National Laboratory.[20] The archive is managed by the Worldwide Protein Data Bank organization (wwPDB), whose mission is to ensure that a single, global PDB data archive is and will remain freely and publicly available.[21]

PubChem: PubChem[22] is an open repository for experimental data identifying the biological activities of small molecules.[23].PubChem is a part of the Molecular Libraries and Imaging (MLI) component of the National Institutes of Health (NIH) Roadmap for Medical Research initiative.[24]

Drug Bank: Drug Bank is a unique bioinformatics/cheminformatics resource that combines detailed drug (i.e. chemical) data with comprehensive drug target (i.e. protein) information.[25] The database contains nearly 4800 drug entries including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs. Additionally, more than 2,500 non-redundant protein (i.e. drug target) sequences are linked to these FDA approved drug entries.[26, 27]

SOFTWARES USED:

Molegro Virtual Docker: Docking was performed by using Molegro Virtual Docker. It is an integrated platform for predicting protein-ligand interactions.[28,29] It handles all aspects of docking process from preparation of the molecules to determination of the potential binding sites of the target protein, and prediction of the binding modes of the ligand. It has been shown to yield higher docking accuracy than other state-of-the-art docking products (MVD: 87%, Glide: 82%, Surflex: 75%, FlexX: 58%).[30].Molegro Virtual Docker can be downloaded from the following server http://www.molegro.com.

Molecular Operating Environment: It is not a software package in the usual sense, but an integrated Methodology Development Platform; that is, a tool for chemical computing software development and deployment. Molecular Operating Environment integrates visualization, simulation and application development in one package. Custom methodology modules can be developed with the built-in high-performance data-parallel programming language SVL, the Scientific Vector Language.

BIOEDIT: BioEdit is a user friendly biological sequence alignment editor and sequence analysis program, for windows 95/95/NT systems [31].In-color alignment and editing with separate nucleic acid and amino acid color tables and full control over background colors. BioEdit can be downloaded from the following Servers for sequence analysis. .http://www.mbio.ncsu.edu/BioEdit/bioedit.html

MOLSOFT: ICM MOLSOFT algorithm was adopted for comparative modeling which provides an accurate and efficient module to build loops and side chains found non-identical in sequence. ICM molsoft algorithm contains robust modeling tools and high levels of accuracy with fast model building. [32, 33, 34,35,36,37]

DISCOVERY STUDIO: Molecular docking is a key tool in structural molecular biology and computer-assisted drug design. The goal of ligand—protein docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure.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 molecule. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions.

ARGUS LAB: Argus Lab is a molecular modeling program that runs on Windows 98, NT, and 2000. Argus Lab consists of a user interface that supports OpenGL graphics display of molecule structures and runs quantum mechanical calculations using the Argus compute server. The Argus compute server is constructed using the Microsoft Component Object Model

GENETIC OPTIMIZATION FOR LIGAND DOCKING: It is a program for calculating the docking modes of small molecules in protein binding sites and is provided as part of the GOLD Suite, a package of programs for structure visualization and manipulation (Hermes), for protein-ligand docking (GOLD) and for post-processing (Goldmine) and visualization of docking results.

TOOLS

Basic Local Alignment Search Tool (BLAST): Nowadays Similarity searching, including sequence comparison, is one of the principal techniques used by computational biologists and has found widespread use among biologists in general. The most popular tool for this purpose is BLAST[38,39] (**Basic Local Alignment Search Tool)** which performs comparisons between pairs of sequences, searching for regions of local similarity. NCBI BLAST is available from the NCBI[40] **(National Center for Biotechnology Information**).

Self-Optimized Prediction Method with Alignment (SOPMA): It is a self optimizing prediction method of alignment and is used for prediction of secondary structure of proteins. This method calculates the content of alpha helix, beta sheets, turns, random coils and extended strands.SOPMA[41] method predicts 69.5% of amino acids. The prediction of protein secondary structure is improved by 9% to 66%.SOPMA is neural network based methods; global sequence prediction may be done by this sequence method. SOPMA is available online on http://www.expasy.org .[42]

WEB BASED SERVERS

Real Space Automated Conformer Generation (RAPPER):RAPPER[43] is an ab initio conformational search algorithm for restraint-based protein modeling. It has been used for all-atom loop modeling, whole protein modeling under limited restraints comparative modeling, ab initio structure prediction, structure validation and experimental structure determination with X-ray and nuclear magnetic resonance spectroscopy. Web interfaces are available on this website for Ramachandran plot analysis.

CASTP (computed atlas of surface topography of proteins): Computed Atlas of Surface Topography of proteins (CASTp) locates and measures concave surface regions on 3D protein structures. This tool can be used to study surface features, binding sites, and functional regions of proteins.CASTp[44] is updated daily and can be accessed at http://cast.engr.uic.edu.

METHODOLOGY

To build a homology model of MPT 51 protein and later on docking studies were carried out using Argus Lab with the stable structure and finally performed flexible docking studies in GOLD.

Protein Preparation (Prep Wiz)

A typical PDB structure consists of heavy atoms, waters, cofactors, metal ions and can be multimeric. The structure generally has no information on bond orders, topologies, or formal atomic charges. So, **1SFR** (from the PDB) must be prepared by using protein preparation wizard (PrepWiz) of Schrödinger software. Protein preparation ensures that the **1SFR** protein structure was properly assigned with bond orders and correct number of hydrogens to make the structure compatible with the OPLS (Optimized Potential for Liquid Simulations) forcefields.[45] and the process of building the homology model Accelrys Discovery studio is used.

Homology modeling

The steps to creating a homology model are as follows:

- Identify homologous proteins and determine the extent of their sequence similarity with one another and the unknown.
- Align the sequences.
- Identify structurally conserved and structurally variable regions s
- Generate coordinates for core (structurally conserved) residues of the unknown structure from those of the known structure(s).
- Generate conformations for the loops (structurally variable) in the unknown structure.
- Build the side-chain conformations.
- Refine and evaluate the unknown structure.

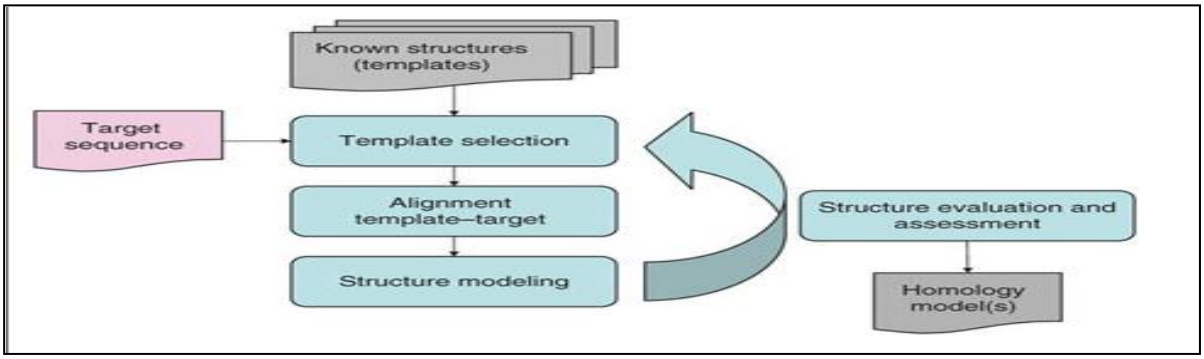


Figure-1: The four main steps of homology modeling of protein i.e. template selection, target–template alignment, model building and model quality evaluation. [46]

Docking: Docking studies of modeled stable MPT 51 protein with carprofen analogues (A) 2-(6-chloro-9H-carbazol-2-yl) propanoic acid Pubchem ID: 2581 , (B) 2-[5-chloro-2-(4-methylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 667685, (C) (2S)-2-(6-chloro-9H-carbazol-2-yl)propanoic acid) Pubchem ID: 3327305, (D) 2-[5-chloro-2-(4-ethylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 5023821, (E) 2-[5-chloro-2-(4-propan-2-

ylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 3950728 and (F) (2S)-2-(6-chloro-9H-carbazol-2-yl)propanoate Pubchem ID: 6919166 as shown in Figure-2(A)& 2(B) was carried out using GOLD software. These carprofen analogue molecules were retrieved from NCBI-PubChem Compound database.[47, 48]

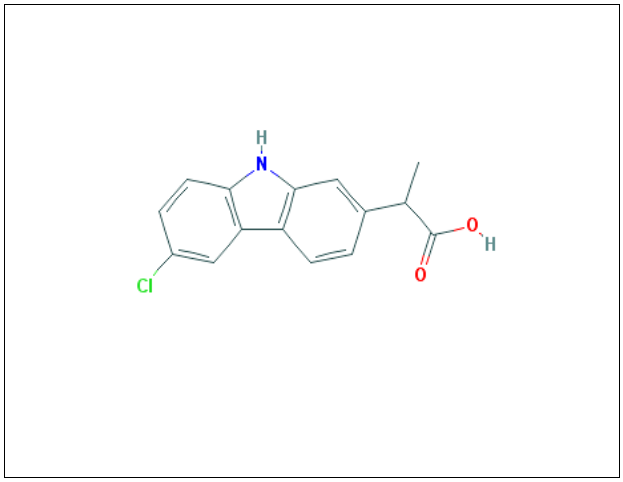
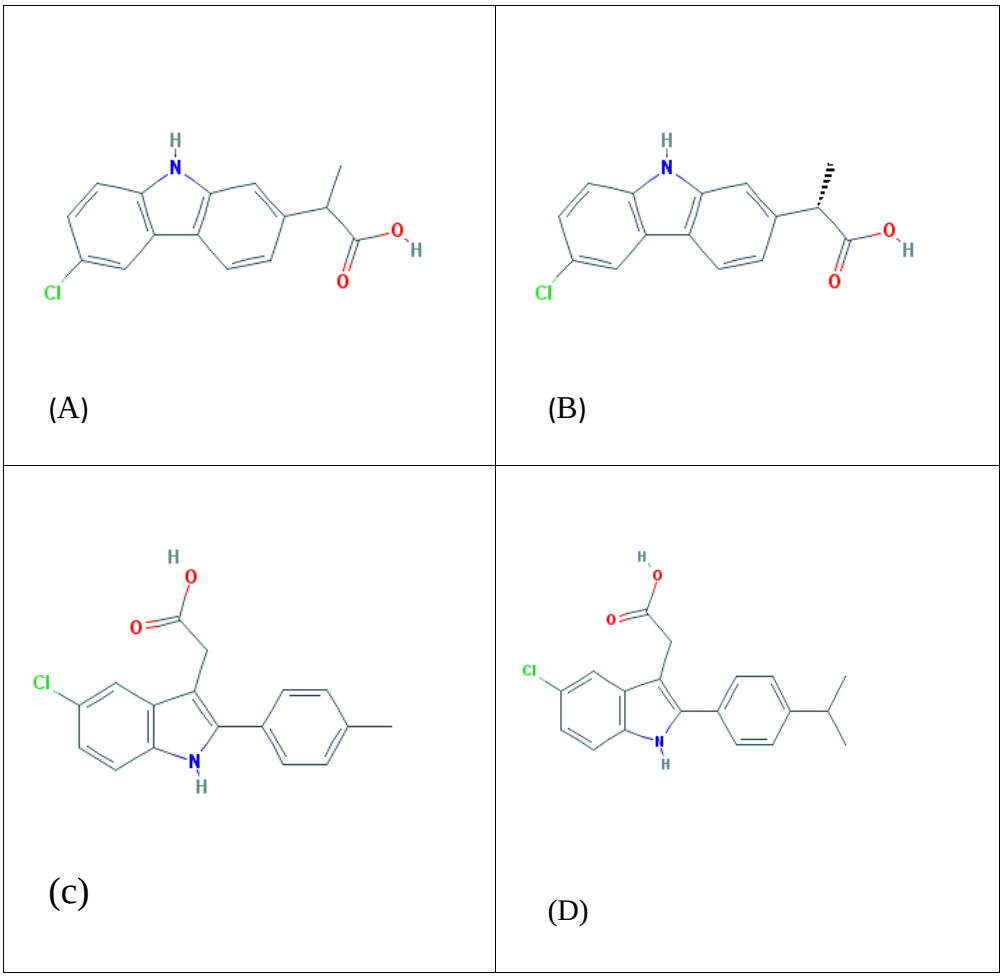


Figure-2(A): 2D Molecular structure of Carprofen molecule (RS-2-(6-chloro-9H-carbazol-2-yl) propanoic acid with Pubchem ID: 2581



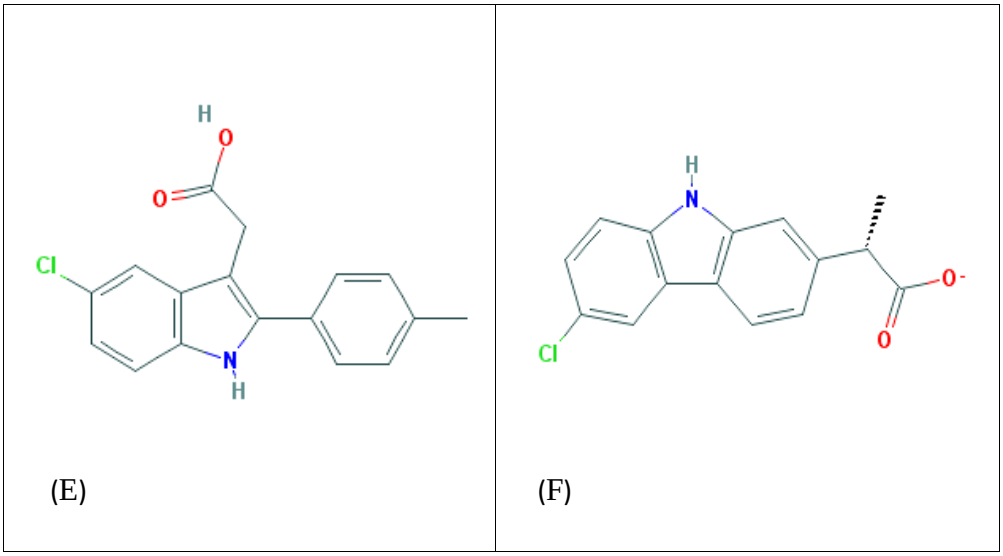


Figure-2(B): 2D Molecular structures of the ligands/inhibitors (A) 2-(6-chloro-9H-carbazol-2-yl)propanoic acid Pubchem ID: 2581 , (B) 2-[5-chloro-2-(4-methylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 667685, (C)(2S)-2-(6-chloro-9H-carbazol-2-yl)propanoic acid) Pubchem ID: 3327305, (D)2-[5-chloro-2-(4-ethylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 5023821, (E) 2-[5-chloro-2-(4-propan-2-ylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 3950728 and (F)(2S)-2-(6-chloro-9H-carbazol-2-yl)propanoate Pubchem ID: 6919166.

RESULTS AND DISCUSSION

Sequence Retrieval and Data Collection

The details of the protein MPT-51 like amino acid lengths of N-and C-terminus, helix lengths etc were obtained from Protein knowledgebase UniProtKB/Swiss-Prot database with accession number **Q05861.** The details of sequence annotation are shown below.

Query sequence of MPT-51 Protein

```
>sp|Q05861|MPT51_MYCLE MPT51 antigen OS=Mycobacterium leprae GN=mpt51 PE=3 SV=3
MKFVDRFRGAVAGMLRRLVVEAMGVALLSALIGVVGSAPAEAFSRPGLPVEYLQVPSPSMGRDIKVQFQNGGANSPALYLL
DGLRAQDDFSGWDINTTAFEWYYQSGISVVMPVGGQSSFYSDWYSPACGKAGCQTYKWETFLTSELPQYLQSNKQIKPTGS
AAVGLSMAGLSALTLAIYHPDQFIYVGSMSGLLDPSNAMGPSLIGLAMGDAGGYKAADMWGPSTDPAWKRNDPTVNVGTLI
ANNTRIWMYCGNGKPTELGGNNLPAKLLEGLVRTSNIKFQDGYNAGGGHNAVFNFPDSGTHSWEYWGEQLNDMKPDLQQYL
GATPGA
```

Sequence Analysis by Bioedit

Bioedit analysis of query protein MPT-51 shown in **Figure-3** and **Figure-4** indicates that the amino acid Glycine is present highest percentage compared to other residues and the molecular weight of the protein is 12.42.Similarly Helical wheel diagram shown in **Figure-5** obtained using Genetic Computer group shows the relative positions of amino acids, hydrophilic residues located in core region. Parker-HPLC program shown in **Figure-6** is used to identify hydrophilic regions, the plot reliably predict 4 hydrophilic regions above the sliding window, which are located in the three dimensional structure at amino acid positions 39-50,80-100,142-170 and 250-270 residues are hydrophilic regions in MPT-51.

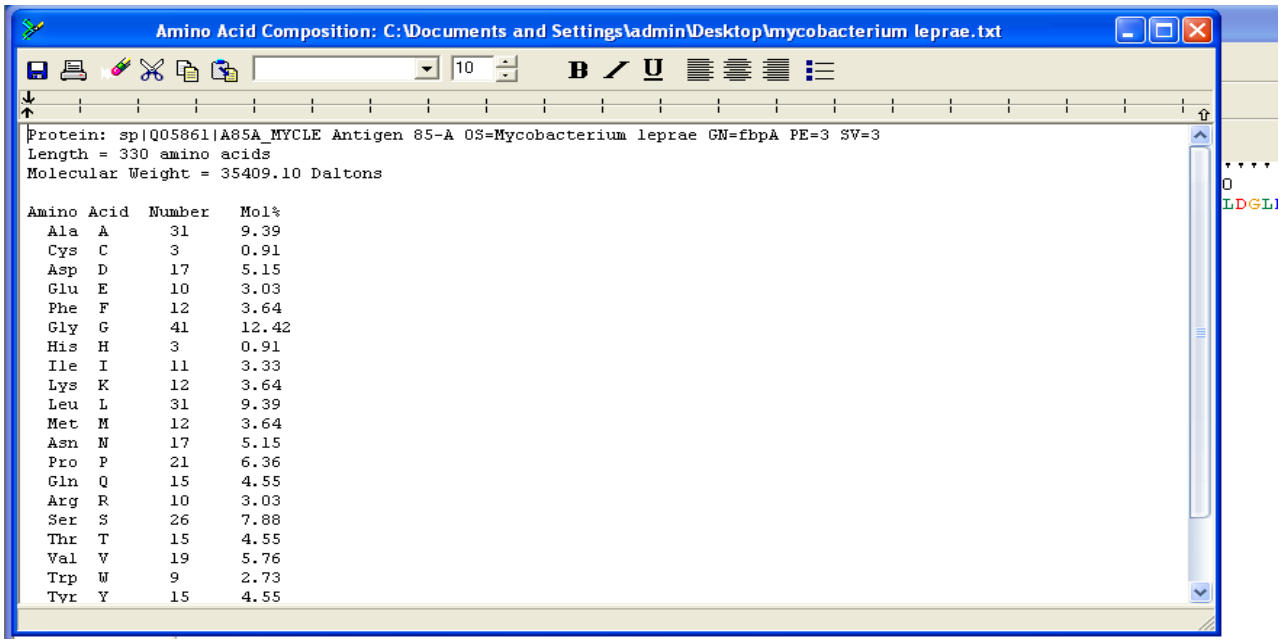


Figure-3: The above graph is showing amino acid composition of Ras-specific guanine nucleotide releasing factor 2 amino acid GLY is present highest percentage compared to other residues and the molecular weight of the protein is 12.42

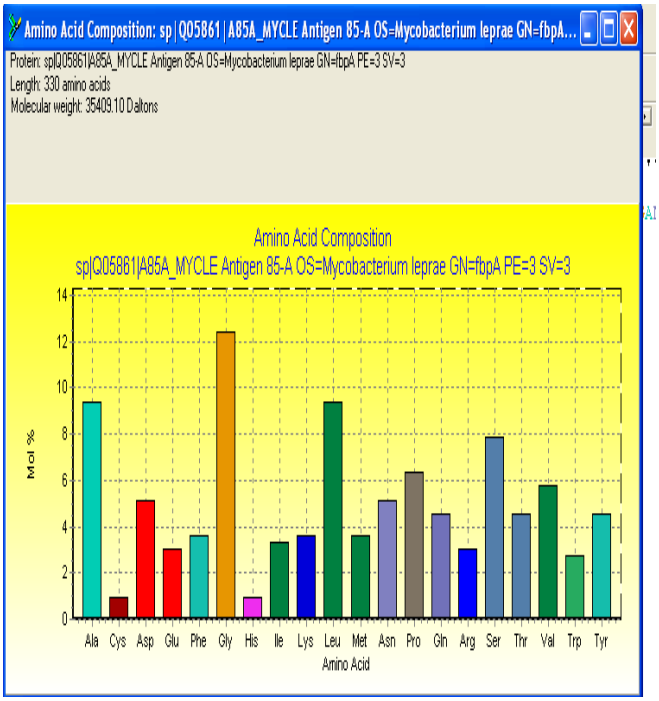


Figure-4: The above graph is showing amino acid composition of Query sequence, the amino acid Gly is present highest percentage compared to other residues.

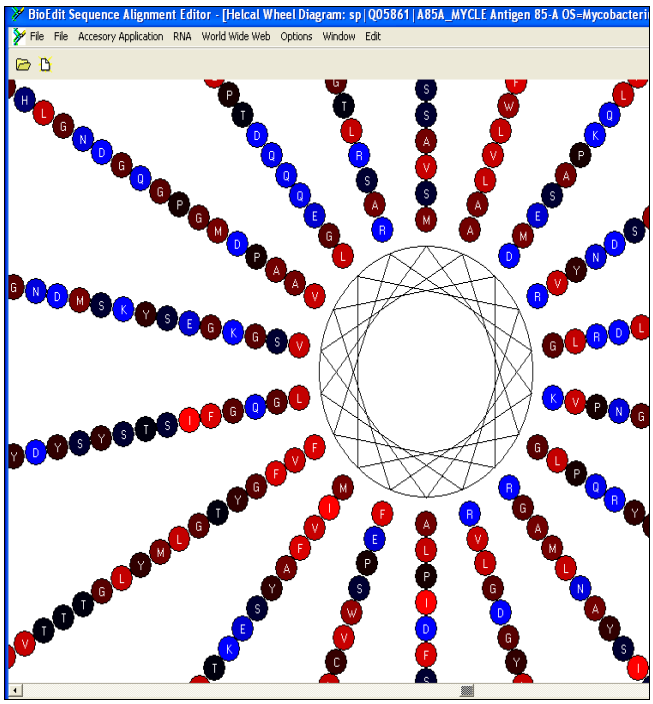


Figure-5: Wheel plot of the MPT-51-protein .the plot Shown obtained using Genetic Computer group HELICAL WHEEL PROGRAM, the diagram shows the relative positions of amino acids, hydrophobic residues located in core region

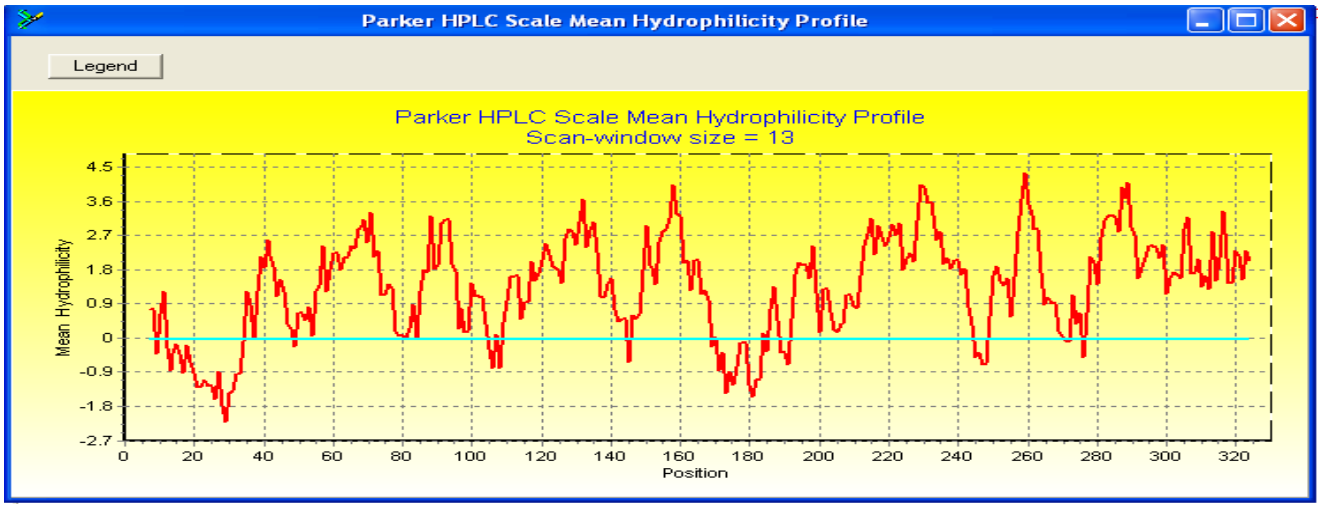


Figure-6: Hydrophilicity profile of above the slide window is indicating 39-50,80-100,142-170 and 250-270 residues are hydrophilic regions in MPT-51

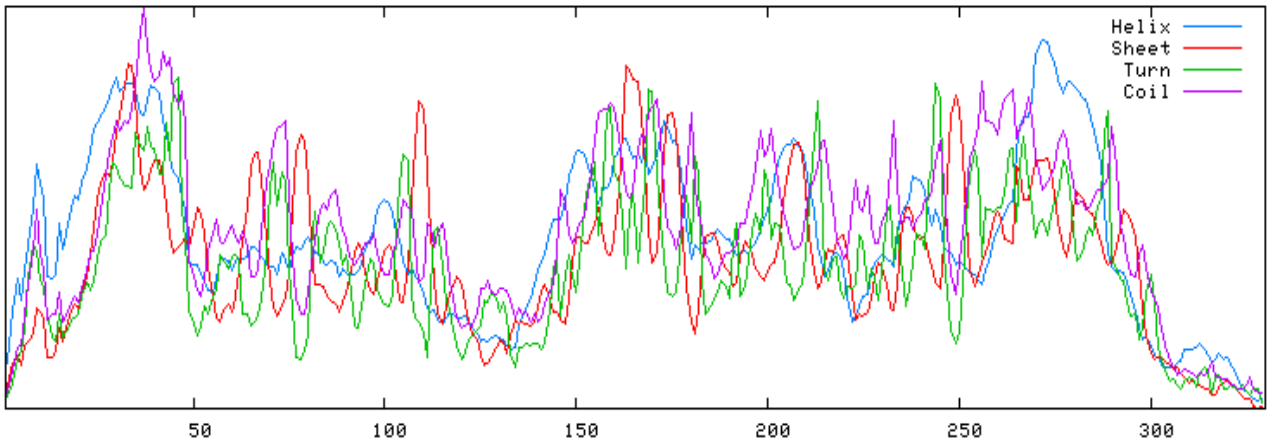
Secondary structure prediction by SOPMA

The secondary structures of proteins are the regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the alpha-helix and beta-sheet.Secondary structure analysis of MPT 51 protein was carried out using SOPMA as shown in **Figure-7**.SOPMA view predicted the following secondary structures.**Table-1** shows percentage of Alpha helix (Hh) is 36.67%%, percentage of 3_{10} helix (Gg) is 0.00% , percentage of Pi helix (Ii) is 0.00%, percentage of Beta bridge (Bb) is 0.00%, percentage of Extended strand (Ee) is 16.36% , percentage of Beta turn (Tt) is 6.97%, percentage of Bend region (Ss) is 0.00%, percentage of Random coil (Cc) is 40.00% percentage of Ambigous states is 0.00% and Other states is 0.00%.

```
GGANSPALYLLDGLRAQDDFSGWDINTTAFEWYYQSGISVVMPVGGQSSFYSDWYSPACGKAGCQTYKWE
cccccceeeeeetccccccccceeechhhhhhhhttteeeeecttccceeeeecccccccccccccchhh
TFLTSELPQYLQSNKQIKPTGSAAVGLSMAGLSALTLAIYHPDQFIYVGSMSGLLDPSNAMGPSLIGLAM
hhhhhhhhhhhhhhccccccccceeeeeeecttcchheeehhccttceeehhhhttcccccccccchhhhhhh
GDAGGYKAADMWGPSTDPAWKRNDPTVNVGTLIANNTRIWMYCGNGKPTELGGNNLPAKLLEGLVRTSNI
hhttccccceecccccccccccccchhhhhhhttttceeeeecccccccccccccchhhhhhhhhhhhhh
KFQDGYNAGGGHNAVFNFPDSGTHSWEYWGEQLNDMKPDLQQYLGATPGA
hhhhhhhhttccceeeecccccccchhhhhhhhhhhhhhhhhhtccccc
Sequence length :   330
```

Table-1: Secondary structures predicted by SOPMA

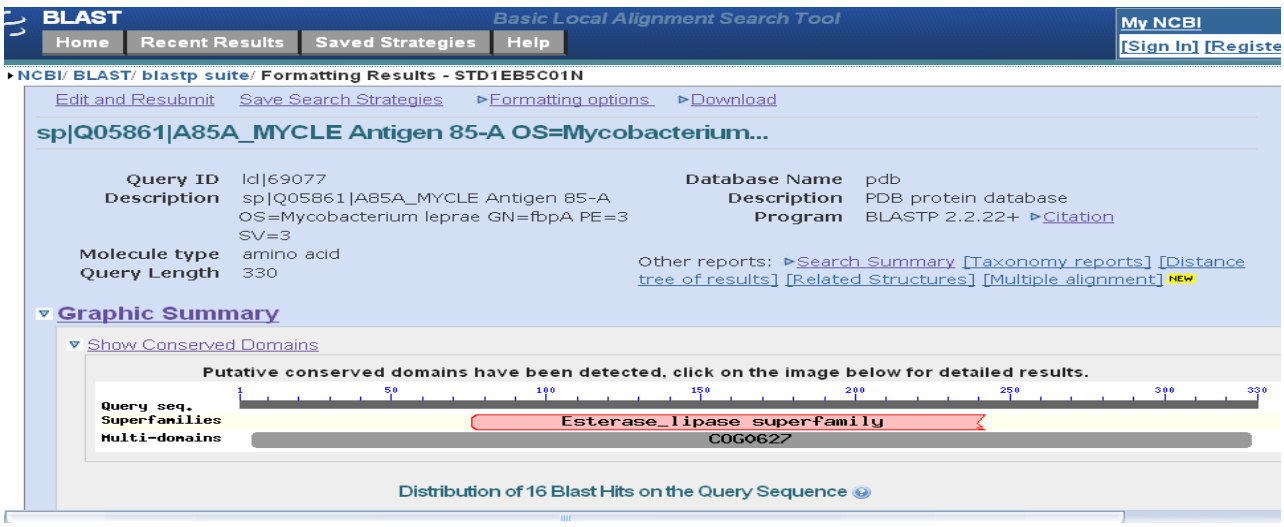
Protein structure, Unit	No. of amino acids	Percentage of structural, Unit
Alpha helix((Hh)	121	36.67%
3_{10} helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand(Ee)	54	16.36%
Beta turn (Tt)	23	6.97%
Bend region (Ss)	0	0.00%
Random coil (Cc)	132	40.00%
Ambigous states	0	0.00%
Other states	0	0.00%




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

Figure-7: Secondary structure analysis of MPT-51 Protein from SOPMA server

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between protein or nucleotide sequences. The program compares nucleotide or protein sequences to sequence in a database and calculates the statistical significance of the matches. The target sequence i.e., MPT-51 (UniProt ID: **Q05861**) was searched against the protein database by using BLAST tool. From the BLAST results as shown in **Figure-8** we observed that four proteins (PDB IDs: 1SFR-A, 1FON-A, 1DQY-A, 3HRH-A) are showing the maximum identity with the target sequence. Among the four proteins obtained, **1SFR-A** selected for further proceedings.



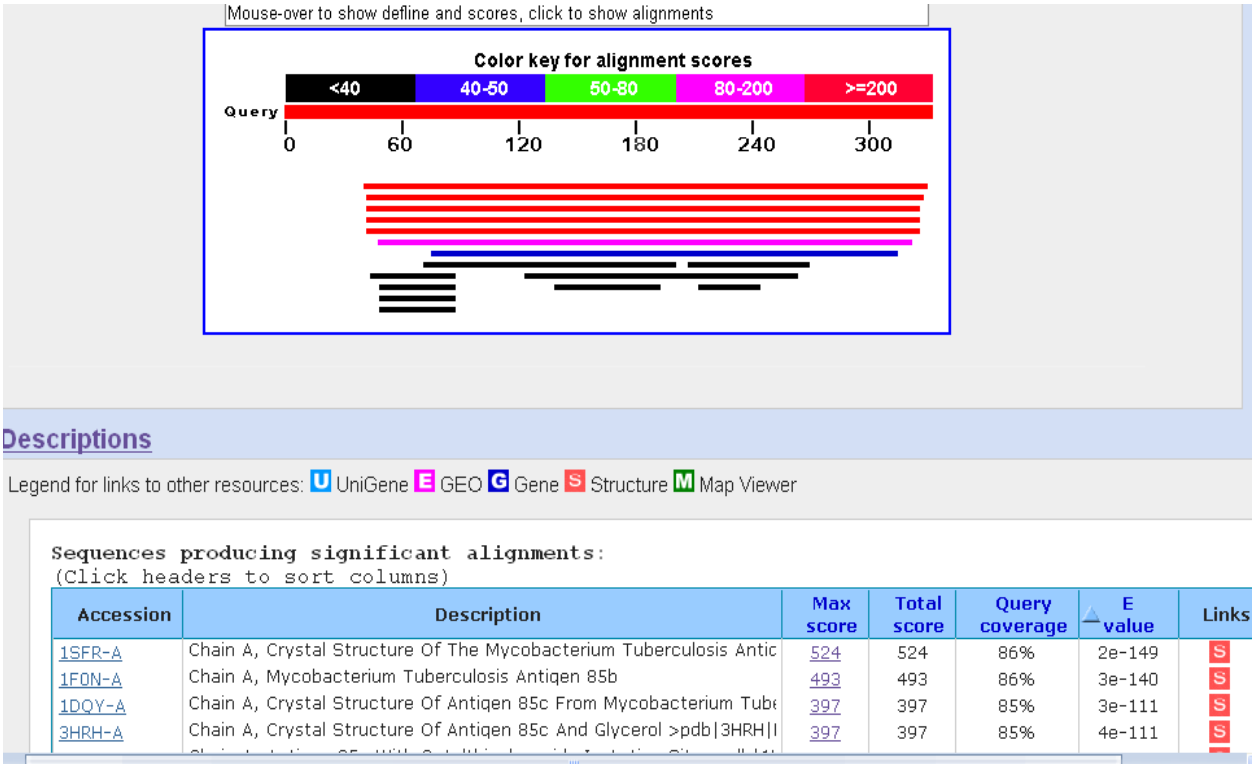


Figure-8: BLAST result for MPT-51 protein

Three-Dimensional Structure Prediction of MPT51 Protein by MOLSOFT and MOE Software

Three dimensitional structure of MPT-51 Protein (Q05861) as shown in **Figure-9** ,**Figure-10** were predicted by using the tool MOLSOFT ICM and MOE by taking 1SFR-A as template which was obtained through BLAST results by taking Q05861 as query sequence and performing Protein BLAST against the protein sequence database. Energy minimization of the modeled three dimensitional structure of MPT 51 protein is carried by using MOE as shown in **Figure-11.**

Homology modeling by MOLSOFT

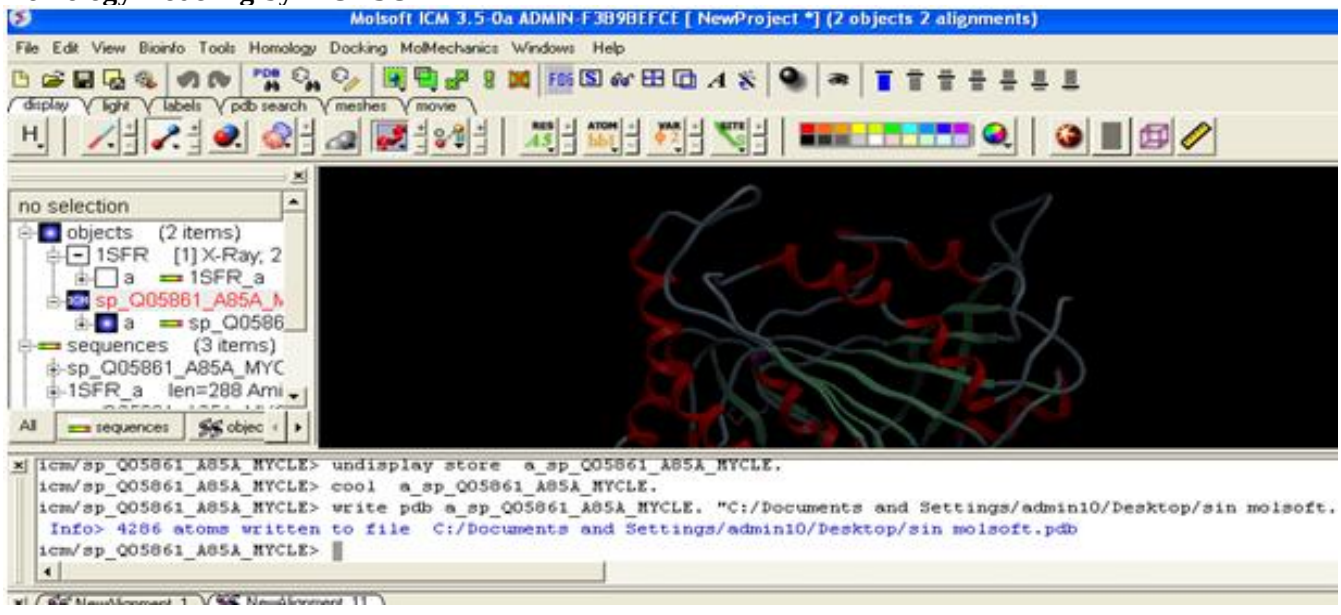


Figure-9: MPT-51 protein structure was predicted by Mol soft ICM, by taking template 1SFR

Modeling of MPT51 protein In MOE:

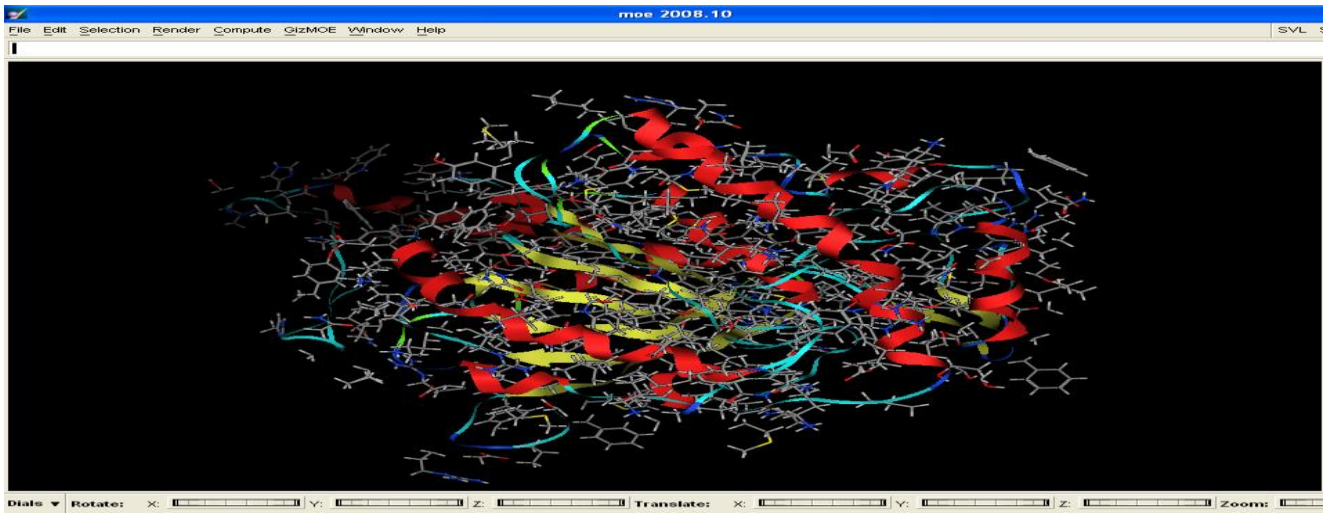


Figure-10: Modeled structure of MPT-51 protein using MOE

Energy Minimization in MOE:

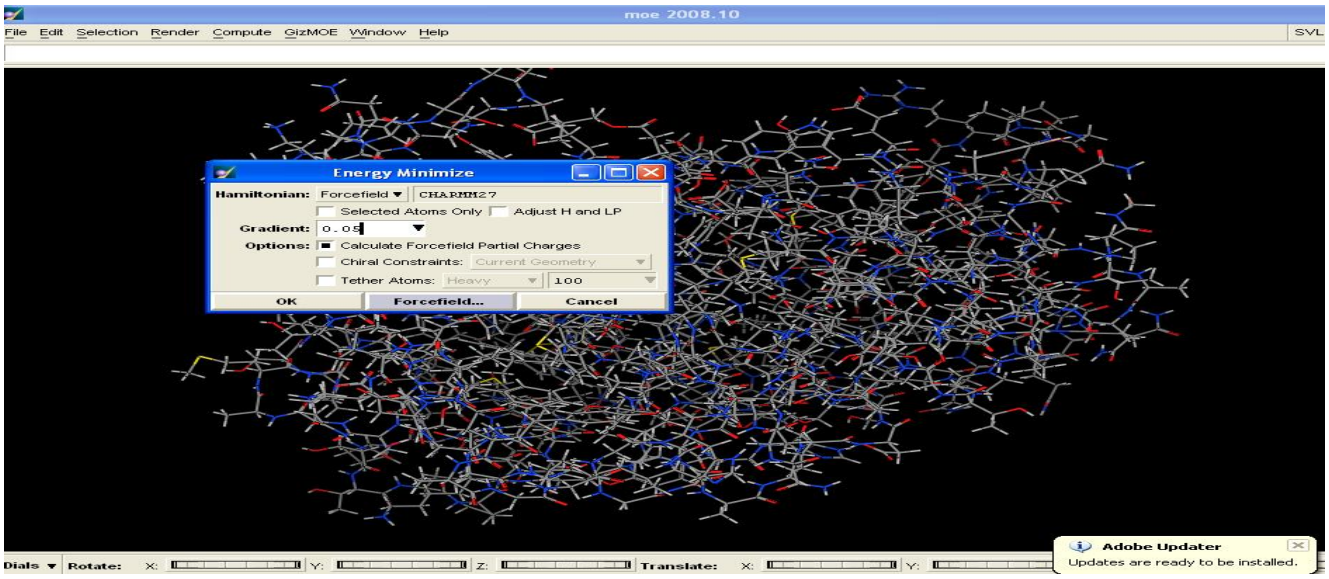


Figure-11: Modeled MPT-51 protein structure after energy minimization using MOE

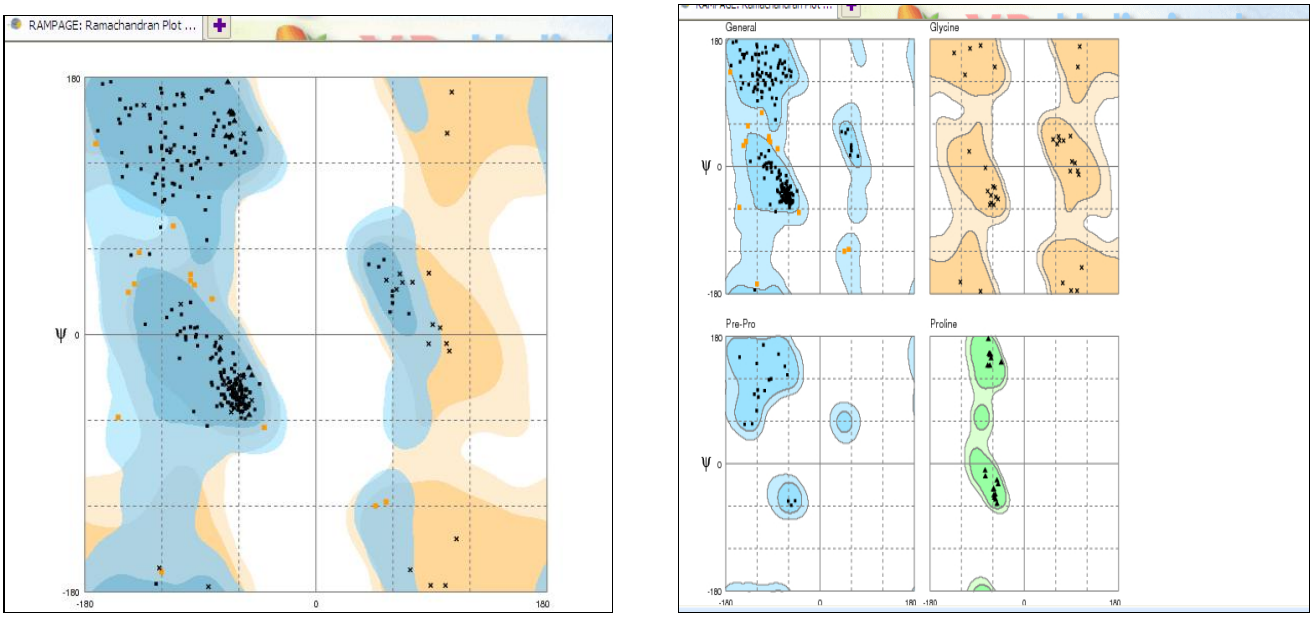
Assessment of the Homology Model of MPT 51 Protein

The validation of the final model was carried out using Ramachandran plot computed with PROCHECK, program by checking the detailed residue-by-residue stereo-chemical quality of a protein structure. The PROCHECK is used for stereo chemical assessment of the model. The criteria for analysis of stereochemistry of the model includes,

1) Main chain conformation in acceptable regions of the Ramachandran plot.
2) Planar peptide bonds.
3) Side chain conformations that correspond to those in rotamer library.
4) Hydrogen bonding of polar atoms if they are buried.
5) No bad atom-atom contacts.
6) No holes inside the structure.

Ramachandran Plot analysis by RAPPER

A Ramachandran plot (also known as a Ramachandran map or a Ramachandran diagram or a [Φ,Ψ] plot), developed by Gopalasamudram Narayana Ramachandran and Viswanathan Sasisekharan is a way to visualize dihedral angles Ψ and Φ of amino acid residues in protein structure. It shows the possible conformations of Φ and Ψ angles for a polypeptide. Hence, Ramachandran plot is a useful way of assessing the stereo chemical quality of a protein structure. The final refined model was further assessed by RAPPER server (ramachandran plot analysis), and the results show that MPT-51 protein had 273 (95.1%) residues in favoured region against (~98.0% expected), 12 (4.9%) residues in allowed region against (~2.0% expected) and 0 (0.0% residues in outlier region as shown in **Figure-12** which shows that the final model is reliable. .



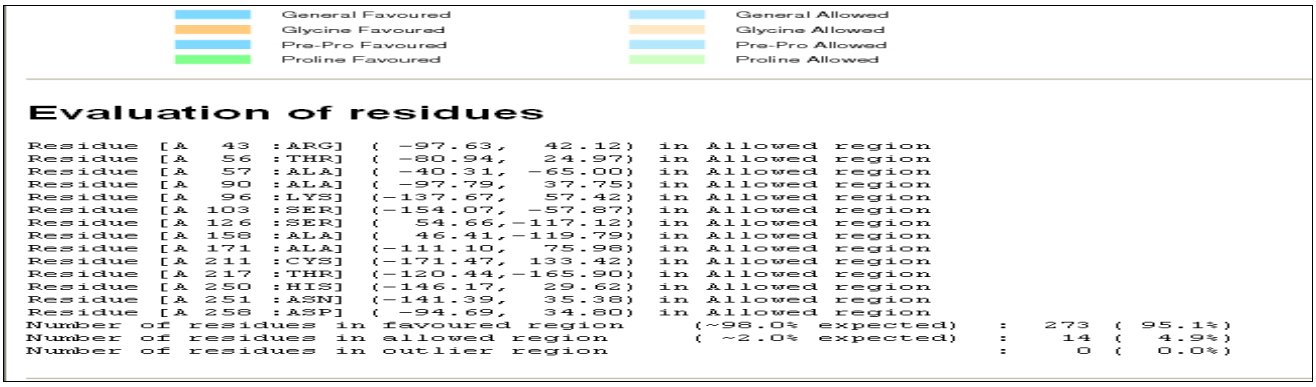


Figure-12: Protein validation study by RAPPER Server

Docking studies in ARGUS Lab:

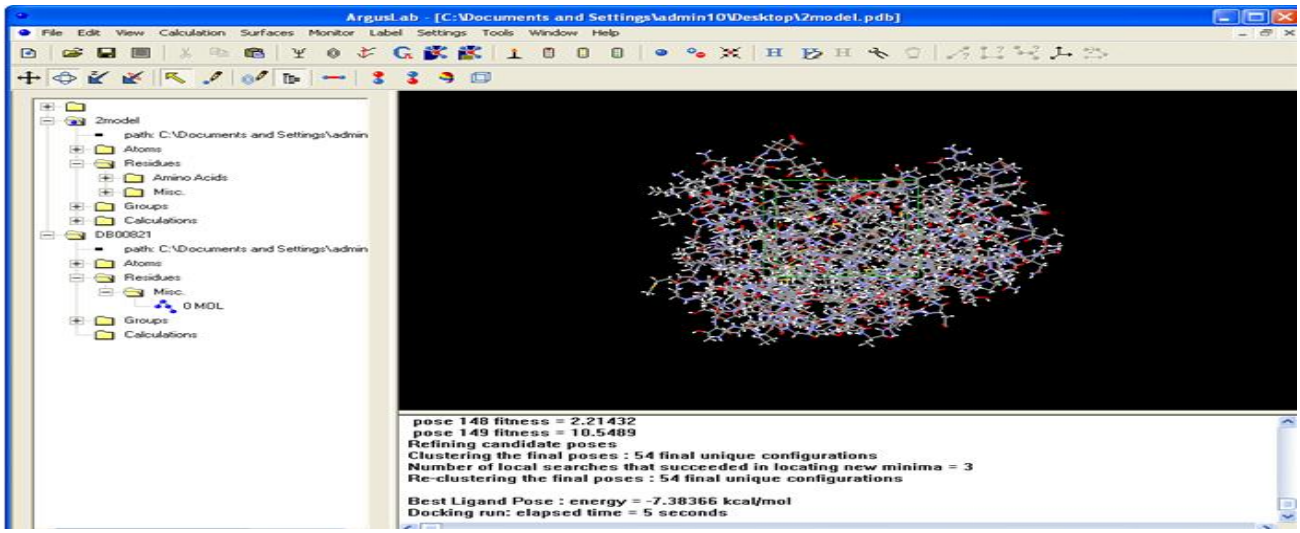


Figure-13: Docking studies of MPT 51 protein with Dapsone, caprofen, Tiaprofenic acid and Naproxen in ARGUS Lab

Table-2: Docking of MPT 51 molecule with Dapsone

S. No	Pocket ID	Drug name	Amino acid	Best Ligand Pose Energy Kcal/mol
	31	Dapsone(DB00250)	40 Asp	-8.34932
2	31	Dapsone(DB00250)	42Leu	-10.3365
	31	Dapsone(DB00250)	43Arg	-8.1756
4	31	Dapsone(DB00250)	44Ala	-8.39502
5	31	Dapsone(DB00250)	45Gln	-8.3712
6	31	Dapsone(DB00250)	48Phe	-8.87328
	Total Ligand Pose Energy			= -55.10508

Table-3: Docking of MPT 51 molecule with Tiaprofenic acid

S.No	Pocket ID	Drug name	Amino acid	Best Ligand Pose Energy Kcal/mol
1	31	Tiaprofenic cid(DB01600)	40 Asp	-7.70375
2	31	Tiaprofenic cid(DB01600)	42Leu	-13.2092
3	31	Tiaprofenic cid(DB01600)	43Arg	-8.79831
4	31	Tiaprofenic cid(DB01600)	44Ala	-8.87241
5	31	Tiaprofenic cid(DB01600)	45Gln	-10.0244
6	31	Tiaprofenic cid(DB01600)	48Phe	-10.8092
	Total Ligand Pose Energy			= -57.55750

Table-4: Docking of MPT 51 molecule with Caprofen

S. No	Pocket ID	Drug name	Amino acid	Best Ligand Pose Energy Kcal/mol
1	31	Carprofen(DB00821)	40 Asp	-7.38366
2	31	Carprofen(DB00821)	42Leu	-12.7922
3	31	Carprofen(DB00821)	43Arg	-10.0551
4	31	Carprofen(DB00821)	44Ala	-8.86619
5	31	Carprofen(DB00821)	45Gln	-9.83644
6	31	Carprofen(DB00821)	48Phe	-10.5393
	Total Ligand Pose Energy			= -58.84990

Table-5: Docking of MPT 51 molecule with Naproxen

S.No	Pocket ID	Drug name	Amino acid	Best Ligand Pose Energy Kcal/mol
1	31	Naproxen-DB00788)	40 Asp	-6.62535
2	31	Naproxen-DB00788)	42Leu	-11.473
3	31	Naproxen-DB00788)	43Arg	-7.52224
4	31	Naproxen-DB00788)	44Ala	-6.98671
5	31	Naproxen-DB00788)	45Gln	-8.73265
6	31	Naproxen-DB00788)	48Phe	-9.20775
	Total Ligand Pose Energy			= -50.54770

Virtual Screening of Compounds by MOLGROW VIRTUAL DOCKER:

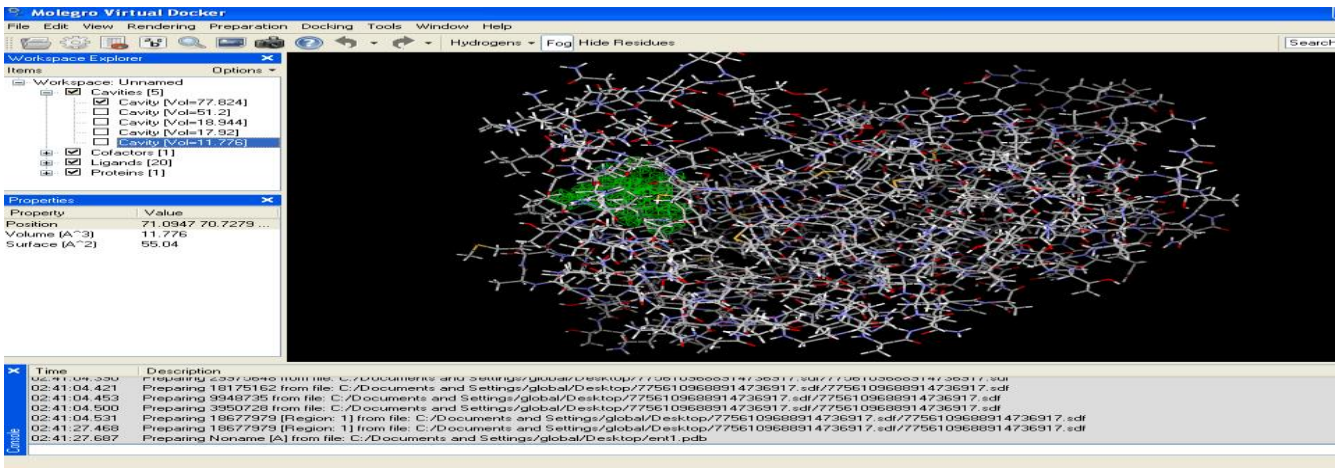


Figure-14: shows active site (Cavity) pocket of our modeled protein ,here we have got 4 different active site pockets based on highest area and volume we have selected the pocket 1 is the best one.

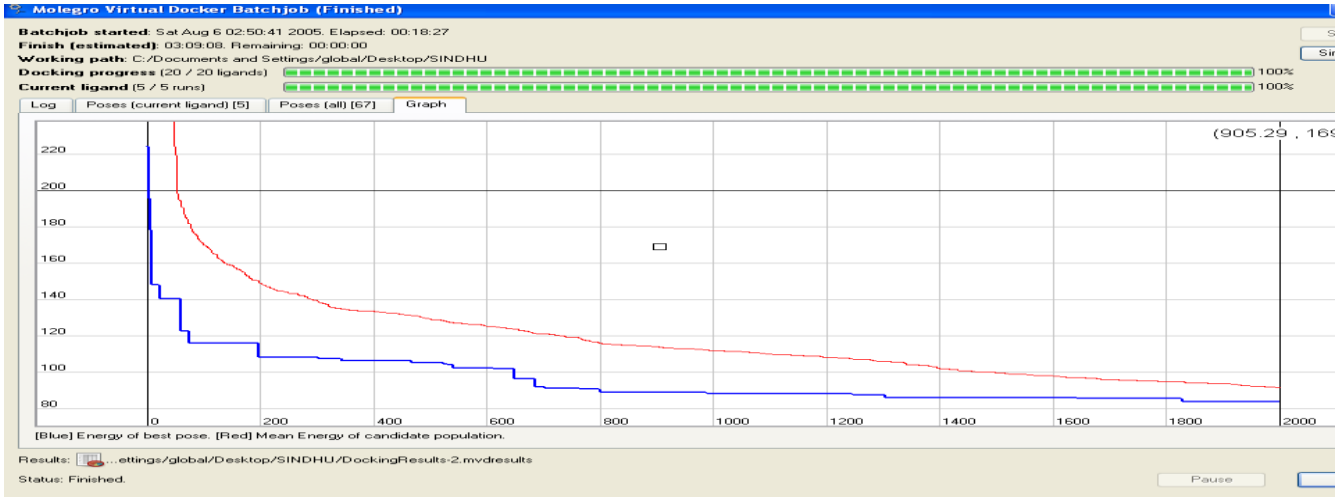


Figure-15: Graph docking curve of carprofen with MPT 51 using MVD software Blue –shows energy of best pose, red – indicate energy of candidate populations

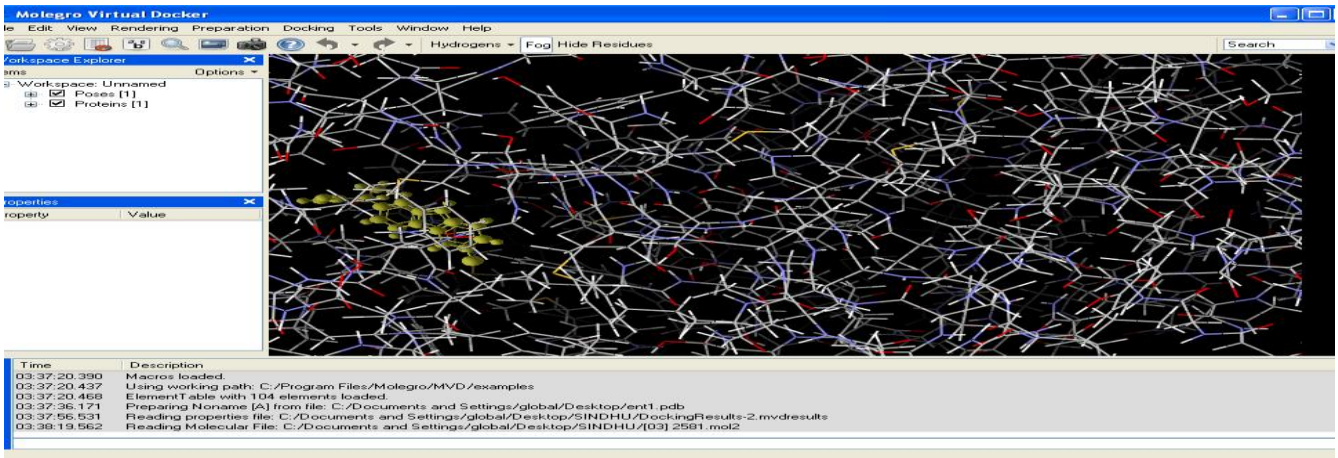


Figure-16: represents ligand moelcule bounded in the active site pocket of modeled protein MPT51

The stable structure is used for docking studies with Dapsone ,Carprofen, Tiaprofenic acid and Naproxen in ARGUS LAB and Molgrow Virtual Docker as shown in Figure-13,Figure-14,Figure-15and Figure-16 respectively .The interactions between the modeled stable MPT 51 protein structure and ligand

molecules and their energy values in stable conformations are shown in Table-2,Table-3,Table-4 and Table-5 respectively. Among all of them carprofen got least energy (-58.84990Kcal/mol) and high affinity.

Identificatin of active sites of proteins and DNAs are often associated with structural pockets and cavities were carried out by using CASTp. CASTp is updated daily and can be accessed at http://cast.engr.uic.edu. CASTp predicted different active site Pockets based on area and volume we have selected the best pocket as 17 is shown in Figure-17.

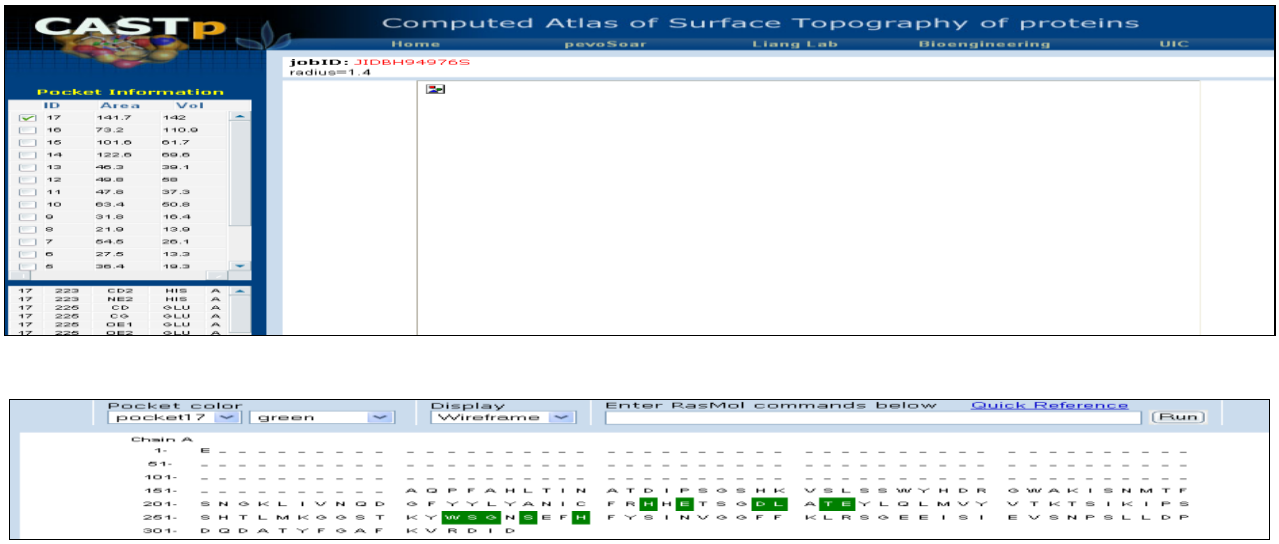


Figure-17: CASTp predicted different active site Pockets based on area and volume we have selected the best pocket as 17

DOCKING IN GOLD:

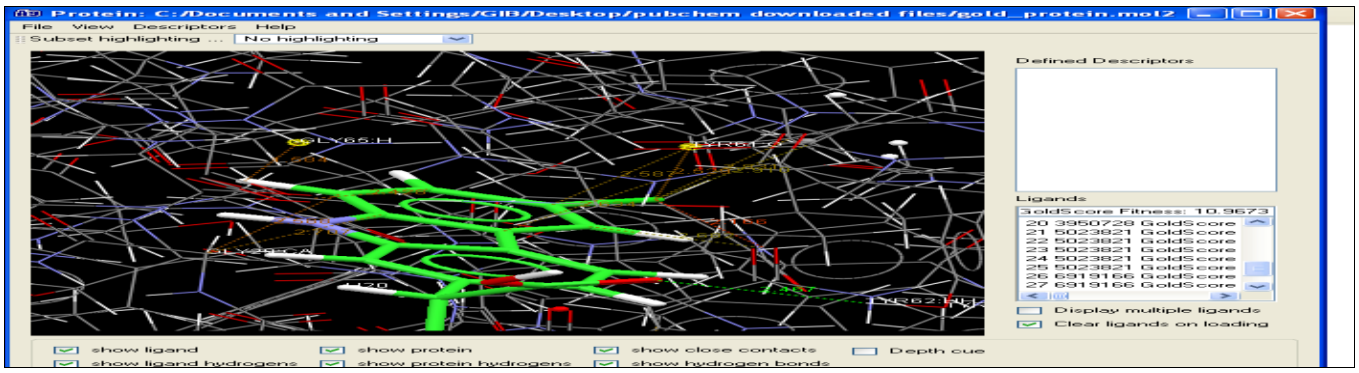


Figure -18: The analogues of carprofen 2-[5-chloro-2-(4-propan-2-ylphenyl)-1H-indol-3-yl] acetic acid molecule was interacting with active site residues (H bonds interaction formed between Tyr29, Gly62, Gly65, Tyr61)

GOLD Results: In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex[49].Further studies we have taken six analogs of carprofen that is (A) 2-(6-chloro-9H-carbazol-2-yl)propanoic acid with Pubchem ID: 2581 , (B) 2-[5-chloro-2-(4-methylphenyl)-1H-indol-3-yl]acetic Acid with Pubchem ID: 667685, (C) (2S)-2-(6-chloro-9H-carbazol-2-yl)propanoic acid with Pubchem ID: 3327305, (D) 2-[5-chloro-2-(4-ethylphenyl)-1H-indol-3-yl]acetic Acid with Pubchem ID: 5023821, (E) 2-[5-chloro-2-(4-propan-2-ylphenyl)-1H-indol-3-yl]acetic Acid with Pubchem ID: 3950728 and (F) (2S)-2-(6-chloro-9H-carbazol-2-yl)propionate Pubchem with ID: 6919166 and performed flexible docking studies in GOLD, to compare standard carprofen.we have found one of the carprofen analogue 2-[5-chloro-2-(4-propan-2-ylphenyl)-1H-indol-3-yl]acetic acid with Pubchem ID: 5023821 shown better interaction with Modeled MPT51 protein as shown in Figure-18. In this study it was found that Tyr29, Gly62, Gly65, Tyr61 are important for strong hydrogen bonding interaction with the ligands.

Calculation of ADMET by DISCOVERY STUDIO:

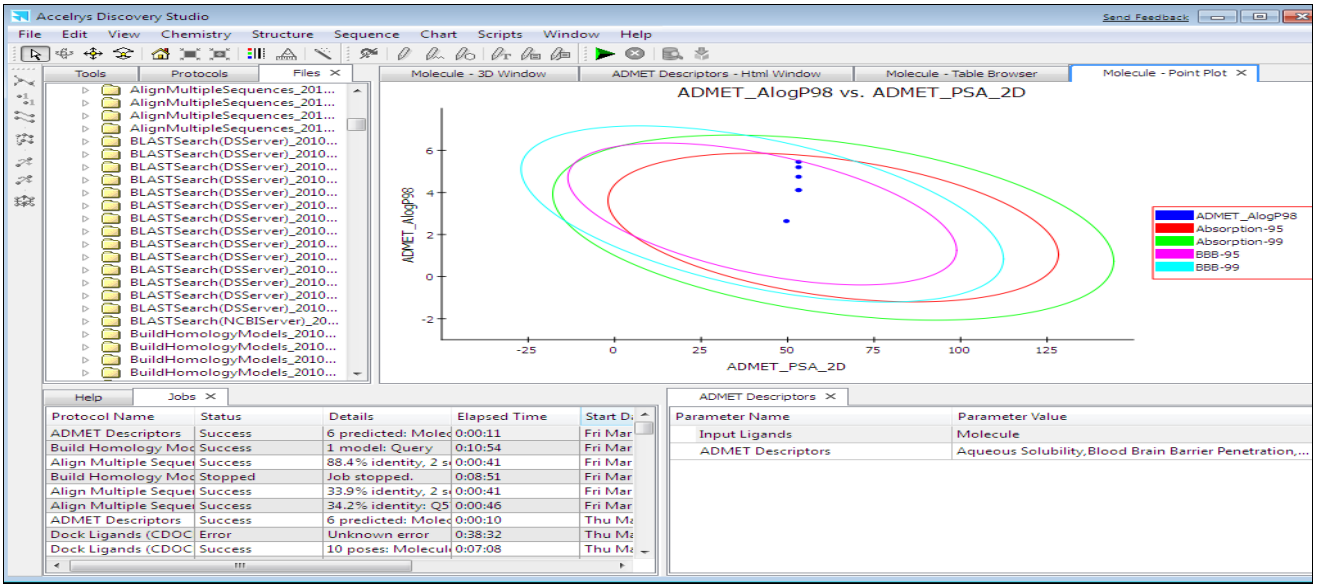
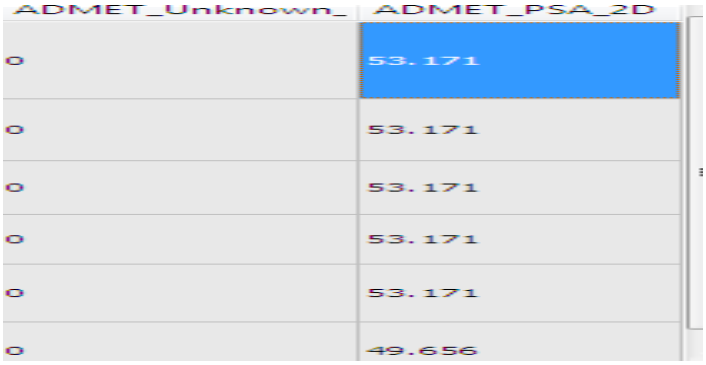


Figure -19: Plot of Polar Surface Area (PSA) vs.LogP for a standard and test set showing the confidence limit 95% and 99% confidence limit ellipses corresponding to the blood brain barrier and intestinal absorption models

	Structure	Name	Index	PUBCHEM_COM	PUBCHEM_C	PUBCHEM_	PUBCHEM_MN	PUBCHEM_	PUBCHEM_A	PUBCHEM_ATO
1		2581	1	2,581	0.6	2	-0.57 0.03 0	19	0	1
2		3327305	2	3,327,305	0.6	3	-0.15 -0.65 -0.57 0.03	21	0	0
3		3950728	3	3,950,728	0.6	4	-0.18 -0.65 -0.57 0.03	22	0	0
4		5023821	4	5,023,821	0.6	4	-0.18 -0.65 -0.57 0.03	23	0	0
5		667685	5	667,685	0.6	2	-0.18 -0.65 -0.57 0.03	19	1	0
6		6919166	6	6,919,166	0.6	2	-0.18 -0.9 -0.9 0.03	19	1	0

	PUBCHEM_SHAPI	PUBCHEM_SHAPE_S	PUBCHEM_SHAPE_	PUBCHEM_COORDI	ADMET_BBB	ADMET_BBB_Le	ADMET_Absorptio	ADMET_Solu
1	12 1.68 0.99 0.42 0.78 1.84	827.341	219	2 5 255	0.276	1	0	-5.365
2	10.66 2.78 1.04 1.22 0.94	912.212	242.9	2 5 255	0.472	1	0	-5.616
3	12.2 2.7 1.03 3.88	951.09	255.9	2 5 255	0.613	1	0	-5.938
4	13.43 2.72 1.04 4.75	992.61	268.7	2 5 255	0.691	1	0	-6.219
5	11.99 1.68 0.99 0.43	827.27	219.2	2 5 255	0.276	1	0	-5.365
6	11.73 1.68 1.03	829.129	219.2	2 5	-0.124	2	0	-4.64



ADMET_Unknown_	ADMET_PSA_2D
0	53.171
0	53.171
0	53.171
0	53.171
0	53.171
0	49.656

Figure-20: Caliculating ADMET properties of Carprofen analogues by using Descriptors like absorption, aqueous solubility, blood brain barrier plasma protein binding, CYP2D6 binding and hepatotoxicity,

IN SILICO ADMET ANALYSIS

In Insilco ADMET analysis both the standard drugs and test compounds were screened for ADMET properties using the ADMET Descriptors for in silico screening in DS 2.0 (Discovery Studio 2.0 is a product of Accelrys Inc, San Diego, CA.). The ADMET properties such as absorption, aqueous solubility, blood brain barrier plasma protein binding, CYP2D6 binding and hepatotoxicity, were evaluated for these molecules within human.

In silico ADMET properties such as ADMET BBB level, as absorption, aqueous solubility, blood brain barrier plasma protein binding, CYP2D6 binding and hepatotoxicity CYP2D6, AlogP98 and PSA were studied for the standard compounds from standard data set and test compounds from test data set. An ADMET model was generated that predicts the human intestinal absorption (HIA) after oral administration of the inhibitors tested. The intestinal absorption model includes 95% and 99% confidence ellipses in the ADMET_PSA_2D and ADMET_AlogP98 plane (Figure-19). There are four prediction levels for the absorption of compounds as good (0), moderate (1), poor (2) and very poor (3). These levels are defined by the 95% (red line) and 99% (green line) confidence ellipsoids (Figure-19). The upper limit of PS A_2D value for the 95% confidence ellipsoid is at 131.62, while the upper limit of PSA_2D value for the 99% confidence ellipsoid is at 148.12 (Figure-19). Based on the in silico ADMET analysis it was found that the test compound 2-[5-chloro-2-(4-propan-2-ylphenyl)-1H-indol-3-yl]acetic acid (Pubchem ID: 5023821) as shown in Figure-20 fulfilled the ADMET descriptors criteria at the optimal level among the test compounds tested whereas all the compounds from the standard dataset fulfilled the ADMET descriptors criteria.

CONCLUSION

Present study, we have constructed a 3D model of MPT-51 protein, using Molsoft and MOE software. Obtained a refined model after energy minimization. The final refined model was further assessed by RAPPER server (ramachandran plot analysis), and the results show that this model is reliable. The stable structure is further used for docking studies with Dapsone ,Carprofen, Tiaprofenic acid, Naproxen in ARGUS lab, among all of them carprofen shows least energy -58.84990Kcal/mole, further studies we have taken the analogues of Carprofen performed flexible docking studies in GOLD to compare standard carprofen.we have found one of the carprofen analogue 2-[5-chloro-2-(4-propan-2-ylphenyl)-1H-indol-3-yl]acetic acid Pubchem ID: 5023821 shows better interaction with Modeled MPT 51protein .The interaction between the protein and ligands proposed in this study are useful for understanding the potential mechanism of protein and ligand binding. As well known, hydrogen bonds play important role for the structure and function of biological molecules. In this study it was found that Tyr29, Gly62, Gly65, Tyr61 are important for strong hydrogen bonding interaction with the ligands.

REFERENCES

1. John D, Daniel E. Infectious keratitis in leprosy. British Journal of Ophthalmology, 1999, 83: 173–176.
2. McMurray DN Mycobacteria and Nocardia: .In Baron S. Et al., eds.Baron's Medical Microbiology (4th ed.).University of Texas Medical Branch. 1996 :ISBN 0-9631172-1-1.

3. Hansen GHA "Undersøgelser Angående Spedalskhedens Årsager (Investigations concerning the etiology of leprosy)". *Norsk Mag. Laegervidenskaben* (in Norwegian) 1874. 4:pp 1–88.
4. Irgens L "The discovery of the leprosy bacillus".*Tidsskr nor Laegeforen* 2002:Vol 122 No 7: pp708–9. PMID 11998735.
5. Fine PEM, Pounighaus JM, Burgess P, Clarkson JA, Draper CC. Seroepidemiological studies of leprosy in northern Malawi based on an enzyme-linked immunosorbent assay using synthetic glycoconjugate antigen. Int J Lepr Other Mycobact Dis 1998;Vol 66:pp243–254.
6. A. K. Shabaana, N. P. Shankernarayan and K. Dharma lingam *Mycobacterium leprae* 18-kDa heat shock protein gene is polymorphic Current Science, 2003: Vol 84, NO. 1, 64-70.
7. Luis Fernandez J, Rangel Mayoral JF, Liso Rubio FJ.A review on Hansen's disease. Farm Hosp 2004; 28:pp123-129.
8. Axel Kroger,V. Pannikar , M. T. Htoon , A. Jamesh , K. Katoch , P. Krishnamurthy , K. Ramalingam , Shen Jianping , Vitthal Jadhav, M. D. Gupte9 and P. Manickam International open trial of uniform multi-drug therapy regimen for 6 months for all types of leprosy patients: rationale, design and preliminary results, Tropical Medicine and International Health ; 2008:vol 13 no 5 :pp 594–602.
9. Viroj Wiwanitkit Analysis of Mycobacterium leprae Genome: In silico Searching for Drug Targets Southeast Asian J Trop Med Public Health 2005: Vol 36 No 4:pp225-227.
10. Leprosy: the disease, World Health Organization.
11. Yoo-chul shin, Hyejon Lee,Hyeyoung Lee, Gerald P.Walsh, Joo-Deuk Kim, and Sang-Nae Cho Variable Numbers of TTC repeats in Mycobacterium leprae DNA from leprosy patients and use in strain differentiatio The Journal of Clinical Microbiology , 2000; Vol 38 No 12 : pp4535–4538
12. Dr Maria Neira , Director Control Prevention and Eradication, Disease Elimination and Eradication
13. Rinke De Wit et al The Mycobacterium leprae Antigen 85 Complex Gene Family:Identification of the Genes for the 85A, 85C,and Related MPT51 Proteins Infection and Immunity, 1993, Vol. 61, No. 9: pp3642-3647.
14. Michelle Cristina Guerreiro dos Reis et al Health Care Workers Humoral Immune Response against GLcB, MPT51 and HSPX from Mycobacterium tuberculosis BJID. 2009, Vol 13: pp417-421.
15. Tobias Scheffer and Ulf Leser, Data Mining and Text Mining for Bioinformatics,Proceedings of the European Workshop on Data Mining and Text Mining for Bioinformatics, (2003)
16. Bairoch, A.; Apweiler, R. "The SWISS-PROT protein sequence data bank and its new supplement TREMBL".*NucleicAcidsResearch* 1996:24 (1)21–25. doi:10.1093/nar/24.1.21. PMC 145613.PMID 8594581.
17. Bairoch, A. "Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times!". *Bioinformatics* 2000:16 (1): 48–64. doi:10.1093/bioinformatics/16.1.48. PMID 10812477.
18. Séverine Altairac, "Naissance d'une banque de données: Interview du prof. Amos Bairoch". *Protéines à la Une*, August 2006. ISSN 1660-9824.
19. Berman, H. Acta Crystallogr A: Foundations of Crystallography 2008, Vol 64,:pp88–95.
20. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.;Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.;Tasumi, M. J Mol Biol 1977, 112,:pp535–542.
21. Berman, H. M.; Henrick, K.; Nakamura, H. Nat Struct Biol 2003, 10, 980.
22. http://pubchem.ncbi.nlm.nih.gov
23. Bolton E, Wang Y, Thiessen PA, Bryant SH. PubChem: Integrated Platform of Small Molecules and Biological Activities. Chapter 12 IN *Annual Reports in Computational Chemistry*, Elsevier: Oxford, UK; 2008, Vol 4 pp. 217-240.
24. http://nihroadmap.nih.gov/molecularlibraries/
25. David S. Wishart,* Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey Drug Bank: a comprehensive resource for insilico drug discovery and exploration Nucleic Acids Res. 2006; Vol 34:pp668-672.
26. Dannenberg AJ, Zakim D. Chemoprevention of Colorectal canser through inhibition of Cycloxygenase-2 Semin Oncol 1999; Vol 26:pp499-504.
27. Egil F. Biochemistry of Cycloxygenase(COX-2) inhibitors and molecular pathology of COX-2 in neoplasia.Crit Rev Clin Lab Sci 2000; 37:431-502.
28. Kawamori T, Rao CV, Seibert K, Reddy BS. Chemopreventive activity of celecoxib, a specific cyclooxygenase-2 inhibitor, against colon carcinogenesis.Cancer Res. 1998; Vol 58 No 3:pp409–412.
29. Reddy B.S, Rao C.V, Seibert K. Evaluation of cyclooxygenase-2 inhibitor for potential chemopreventive properties in colon carcinogenesis *Cancer Res* 1996; Vol 56 No 20: pp4566-69.

30. G. Ratnavali, N. Kanaka Durga Devi, K. Bhavya Sri1, J. Kalyan Raju1, B. Sirisha and R. Kavitha 2011 An attempt to screen top colorectal cancer drugs by using Molegro Virtual Docker Annals of Biological Research, 2011;Vol 2 No 1 : 114-126.
31. HallTA BioEdit:auser-friendly biologicalsequencealignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 1999. Vol 41:95-98.
32. Schapira M., Raaka B.M., Samuels H.H. and Abagyan R. In silico discovery of novel Retinoic Acid Receptor agonist structures BMC Structural Biology. 2001: 1:1.
33. Schapira M., Abagyan R. and Totrov M. Structural model of nicotinic acetylcholine receptor isotypes bound to acetylcholine and nicotine BMC Struct Biol .2002: Vol 2 No 1:1. doi:10.1186/1472-6807-2-1
34. Schapira, M., R. Abagyan, andM. Totrov.. Nuclear hormone receptor targeted virtual screening. J Med Chem 2003: Vol 46 No 14:pp3045-3059.
35. Schapira, M., B.M. Raaka, S. Das, L. Fan, M. Totrov, Z. Zhou, S.R. Wilson, R. Abagyan, andH.H. Samuels.. Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. Proc Natl Acad Sci U S A. 2003: Vol 100 No 12: pp7354-7359.
36. Schapira, M., B.M. Raaka, H.H. Samuels, andR. Abagyan.. Rational discovery of novel nuclear hormone receptor antagonists. Proc. Natl. Acad. Sci. U. S. A. 2000: Vol 97 No 3:1008-1013.
37. Abagyan, R., S. Batalov, T. Cardozo, M. Totrov, J. Webber, andY.Y. Zhou.. Homology modeling with internal coordinate mechanics: Deformation zone mapping and improvements of models via conformational search. 1997 Proteins: 29-37.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, 215:403-410.
39. NCBI BLASt [http://www.ncbi.nlm.nih.gov/BLAST/
40. Alexander Pertsemlidis* and John W Fondon Having a BLAST with bioinformatics (and avoiding BLASTphemy) *Genome Biology* 2001, 2:reviews2002-reviews2002.10 doi: 10.1186/gb-2001-2-10-reviews2002.
41. Prashant v. Thakare, Uddhav s.chaudhari Madura s.makhe Vishal p.Deshmukh Renuka R.Kurtkoti Secondary structure prediction and Phylogenetic analysis of salt tolerant proteins Global journal of molecular sciences 2010 Vol 5 No 1 :pp 30-36.
42. http://www.expasy.org
43. de Bakker, P.I., DePristo, M.A., Burke, D.F., and Blundell, T.L. Ab initio construction of polypeptide fragments: Accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the Generalized Born solvation model. Proteins 2003. Vol 51:pp21-40.
44. Dundas J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues, Nucl Acids Res, 2006:Vol 34,pp116-118.
45. Laurence J. Miller, Quan Chen, Polo C.-H. Lam, Delia I. Pinon, Patrick M. Sexton, Ruben Abagyan, and Maoqing Dong, J. Biol. Chem, 2011, Vol 286, pp15895– 15907.
46. L. Bordoli et al. Protein structure homology modelling using SWISS-MODEL workspace. Nature Protocols 4, 1-13 2009. doi:10.1038/nprot.2008.197.
47. NCBI-PubChem Compound database [http://pubchem.ncbi.nlm.nih.gov/.
48. Uthaman Gowthaman1, Mannu Jayakanthan1 and Durai Sundar Molecular docking studies of dithionitrobenzoic acid and its related compounds to protein disulfide isomerase: computational screening of inhibitors to HIV-1 entry BMC Bioinformatics 2008, Vol 9(Suppl 12):S14 doi:10.1186/1471-2105-9-S12-S14
49. Lengauer T, Rarey M. "Computational methods for biomolecular docking". *Curr. Opin. Struct. Biol.* 1996: Vol 6 No 3: pp402–406. doi:10.1016/S0959-440X(96)80061-3.PMID 8804827.

*****