ISSN 0975-6299 Vol.1/Issue-3/Jul-Sep.2010

International Journal of Pharma and Bio Sciences

MODELING OF MULTIDRUG RESISTANCE PROTEIN OF *LEISHMANIA DONOVANI* AND INSILICO PREDICTION OF ACTIVE DRUGS

SACHIN KUMAR[1]*, SATYANAND TYAGI[1], MOHIT SINGLA[1] AND AMIT KUMAR[2].

[1.] **K.N.G.D Modi Institute of Pharmaceutical Education & Research, Modinagar, Uttar Pradesh, India-201204**
[2.] **M.S. Ramaiah College of Pharmacy, Bengaluru, Karnataka, India**

Corresponding author sonuniper@gmail.com

ABSTRACT

Leishmaniasis is a protozoal disease which afflicts the world's poorest population with endemic distribution in 88 countries around the world and has been considered as a major tropical disease by the world health organization (WHO). P-glycoprotein (P-gp) is an ATP-dependent pump that exports a wide range of drugs from the cell, decreasing their intracellular concentration and preventing their cytotoxic activity. P-gp belongs to the ABC super family of transporters. A three dimensional structure of P-glycoprotein of *Leishmania donovani* was constructed through homology modeling using X-ray crystal structure of multi drug transporter protein sav1866 (pdb id: 2HYD) complex as template using Modeller 9v5. Virtual high throughput screening of one hundred fifty compounds (anthracyclines, alkaloids, xanthons, peptides and natural phenolic compounds) with the modeled P-glycoprotein (P-gp) for the potential P-glycoprotein inhibitors, using AutoDock4.0, GOLDv2.1 and Ligand Fit of discovery studio. Among all inhibitors, two compounds conivapton and nizatidine have shown highest ligand fit score while imatinib, telmisarton, dasatinib have shown better fitness score and formation of hydrogen bonds. These compounds may act as potential drug candidates against P-glycoprotein of *Leishmania donovani*.

KEYWORDS

P-glycoprotein (P-gp), Basic Local Alignment Search Tool (BLAST), genetic optimization of ligand docking (GOLD), World Health Organization (WHO), Visceral Leishmaniasis (VL), Nucleotide-Binding Domain (NBD),Multi Drug Resistance (MDR) , ATP-binding cassette (ABC) , hydrophobic vacuum cleaner (HVC), Transmembrane Domain (TMD). National centre for biotechnology information (NCBI), Protein Data Bank (PDB), Protein Information Resource **(**PIR) and Discovery Studio (DS).

INTRODUCTON

From World Health Organization (WHO) estimated the number of persons at risk to be around 350 million and the number of new cases to be 2,357,000 per year. *Leishmaniasis*, a vector-borne disease that is caused by obligate intra-macrophage protozoa (*Leishmania*) is endemic in large areas of the tropics, subtropics and the Mediterranean basin. This disease is characterized by both diversity and complexity. It is caused by more than twenty leishmanial species and is transmitted to humans by ~30 different species of *phlebotomine sandflies*[1].

Leishmaniasis consists of four main clinical syndromes: cutaneous leishmaniasis; muco-cutaneous leishmaniasis; visceral leishmaniasis (VL; also known as kala-azar); and post-kala-azar dermal leishmaniasis. In cutaneous leishmaniasis, the patient generally presents with one or several ulcer(s) or nodule(s) in the skin. Different species of *Leishmania* can infect the macrophages in the dermis, with variable clinical presentations and prognoses[2, 3].

VL is a systemic disease that is fatal if left untreated and is caused by the *Leishmania donovani* in East Africa and the Indian subcontinent and *Leishmania infantum* in Europe, North Africa and Latin America[4, 5]. There are two types of VL, which differ in their transmission characteristics: zoonotic VL is transmitted from animal to vector to human and anthroponotic VL is transmitted from human to vector to human.

In zoonotic VL, humans are occasional hosts and animals, mainly dogs, are the reservoir of the parasite. Zoonotic VL is found in areas of *L.infantum* transmission whereas anthroponotic VL is found in areas of *L.donovani* transmission [6].

PKDL is characterized by a macular, maculo-papular or nodular rash and is a complication of VL that is frequently observed after treatment in Sudan and more rarely in other East African countries and in the Indian subcontinent. It can also occur in immunosuppressed individuals in *L. infantum*-endemic areas. The interval between treated VL and PKDL is 0–6 months in Sudan and 6 months to 3 years in India. PKDL cases are highly infectious because the nodular lesions contain many parasites, and such cases are the putative reservoir for anthroponotic VL between epidemic cycles[7].

P-glycoprotein (P-gp) is a transmembrane protein, playing significant roles in the process of drug discovery and development and in pest resistance to pesticides[8, 9]. A subfamily of transmembrane proteins from the superfamily of ATP-binding cassette transporters that are closely related in sequence to P-glycoprotein when over expressed, they function as ATP-dependent efflux pumps able to extrude
Lipophilic drugs, especially antineoplastic agents, from cells causing multidrug resistance[10, 11]. P-glycoprotein is present in many normal cell types including secretary cells, and might protect them from naturally occurring xenobiotics[12, 13].
Recently, it has been described that nucleotide-binding domain (NBD) can be the target for inhibitors of P-gp-like transporters. Flavonoids constitute a well known class of natural inhibitors of different ATP-binding proteins[14]. P-gp is an ATP-dependent pump that exports a wide range of drugs from the cell, decreasing their intracellular concentration and preventing their cytotoxic activity.

P-gp belongs to the ABC superfamily of transporters. It consists of two homologous halves, each comprising a transmembrane domain involved in drug efflux, and a cytosolic (NBD) responsible for ATP binding and hydrolysis. The cytosolic domains are therefore attractive targets for the rational design of inhibitors against P-glycoprotein-like transporters[15].

PHYSIOLOGICAL EVENTS ASSOCIATED WITH DRUG RESISTANCE IN *LEISHMANIA:*

The physiological events associated with drug resistance in *Leishmania* include changes in P-gp expression, parasite infectivity (lipophosphoglycan, acid phosphatase and meta- 1 expression), incorporation of metabolites fundamental for the parasite survival (folates and nucleosides), xenobiotics conjugation and extrusion (trypanothione and Cyb expression), intracellular metabolism (dihydrofolate reductase-thymidylate synthetase, N acetylglucosamine-1-transferase and pterin transferase), host-parasite interaction (membrane fluidity) and parasite cell shape and promastigote-amastigote differentiation (tubulin phosphorylation). Black arrows indicate whether a given parameter is increased or decreased in the resistant parasite[16]. (Fig. 1)

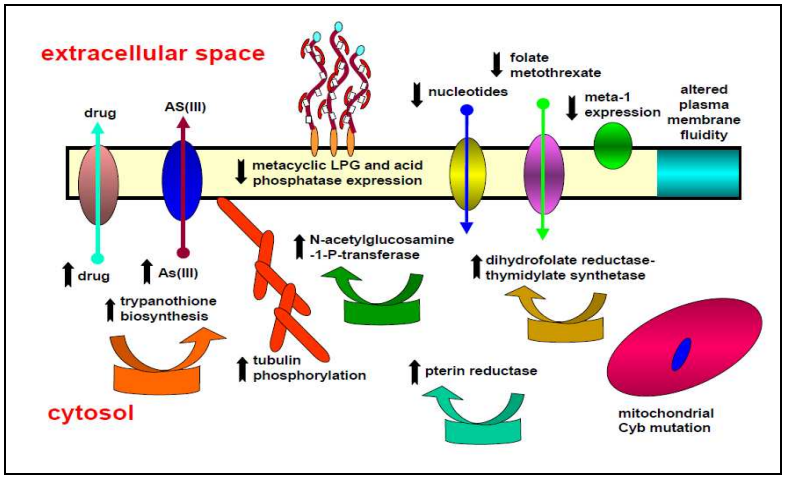

Fig 1: Physiological events associated with drug-resistance in *Leishmania*.

P-GLYCOPROTEIN:

P-glycoprotein (abbreviated as P-gp) is a well-characterized human ABC-transporter of the MDR/TAP subfamily. It is extensively distributed and expressed in normal cells such as those lining the intestine, liver cells, renal proximal tubular cells, and capillary endothelial cells comprising the blood-brain barrier. P-gp is also called ABCB1, ATP-binding cassette sub-family B member 1, MDR1, and PGY1[17].P-gp plays a large role in the distribution and elimination of many clinically important therapeutic substances. Prescription and OTC drugs, foods and substances made by the body may be inhibitors and/or inducers of these transporters. Some drugs like cyclosporin are both substrates and inhibitors of P-gp, other drugs like nifedipine are inhibitors only and some drugs like digoxin are only substrates[18]. Another example is loperamide and quinidine. Loperamide is an opiate antidiarrheal that is normally kept out of the brain by the blood brain barrier due to transport away from the brain by P-gp. When given with quinidine which inhibits P-gp, more loperamide can enter the brain and cause respiratory depression[19].P-glycoprotein is a 170 kDa membrane-bound protein which has been implicated as a primary cause of multidrug-resistance in tumours. The responsible gene- MDR1 has recently been reclassified as part of the ABCB subfamily and rename PGY-1, and a website of ATP-Binding Cassette Transporters has been established at http://nutrigene.4t.com/humanabc.htm . An understanding of the physiological regulation of these transporters is key to designing strategies for the improvement of therapeutic efficacy of drugs that are their substrates for P-gpactivity[20].

The ATP-binding cassette (ABC) proteins are ubiquitous and most of these proteins mediate

transport across biological membranes. The ATP-binding domains of the ABC proteins include the Walker A and B motifs and the “signature” or “C” motif just upstream of the Walker B site which distinguishes members of the ABC super family from other ATP-binding proteins. The sequence conservation of the ABC domains has allowed the isolation of new ABC genes by hybridization, degenerated PCR, and by inspection of DNA sequence databases. The latter strategy is now the most efficient one and inventories and classification of ABC proteins have been made for several genomes. The family of ABC transporters is one of the largest families of proteins. The human ABC transporters have been divided in at least seven subclasses, while several more subclasses have been described for bacteria. Not all ABC transporters have transmembrane domains and not all appear to be involved in drug resistance. Several mechanisms could account for increased activities of ABC transporters leading to drug resistance. In addition to ABC transporters, other membrane proteins of parasites are involved in drug resistance. For example, down regulation of transport systems can lead to arsenical resistance in trypanosomes, a 170-kDa transmembrane glycoprotein from the superfamily of ATP-binding cassette transporters. It serves as an ATP-dependent efflux pump for a variety of chemicals, including many antineoplastic agents. Over expression of this glycoprotein is associated with multidrug resistance[21].

ROLE OF P-GLYCOPROTEIN IN MDR:

P-glycoprotein (P-gp) is a transmembrane protein, playing significant roles in the process of drug discovery and development and in pest resistance to pesticides. When over expressed, they function as ATP-dependent efflux pumps able to extrude lipophilic drugs, especially antineoplastic agents, from cells causing multidrug Resistance (drug resistance, multiple). Although P-glycoprotein share functional similarities to Multidrug Resistance-Associated Proteins they are two distinct subclasses of ATP-binding cassette transporters, and have little sequence homology. Protozoan parasites are important causative agents of morbidity and mortality throughout the world a problem further complicated by the emergence of drug resistance in these parasites. Mechanisms of drug resistance include the following: decreased uptake of the drug into the cell, loss of drug activation, alterations in the drug target, and over-expression of a well-known multiple drug transporter protein[22].

P-glycoprotein is responsible for multidrug resistance in tumour cell lines and are thought to have a physiologic role in exporting cellular metabolites. There has been report that a P-gp gene in the H region of the trypanosomatid protozoan *Leishmania* confers resistance to heavy metals when present in multiple copies. One of the major known mechanisms of resistance to antileishmanial drugs is the form of multidrug resistance (*MDR*) caused by the multidrug transporter P-gp, which are members of the ABC transporter family[23]. P-gp is first reported by Juliano and Ling in 1976. P-gp functions as an ATP-dependent membrane efflux pump that maintains the intracellular drug concentration below cytotoxic levels. Human P-gp is a phosphorylated and glycosylated protein and consists of two halves that share a high degree of similarity. It has 1280 amino acids, 12 membrane-spanning domains, and two nucleotide binding sites. The polypeptide component of the P-gp has a molecular weight of 120 to 140 KD. There is a single glycosylation locus in the extracellular region of the N-terminal half of the molecule. The extent of glycosylation can be variable, producing P-gp with molecular weights ranging from 135 to 180 kDa. The P-gp activity is regulated to some extent by phosphorylation. Monoclonal antibodies to P-gp inhibit growth of MDR cells and result in increased accumulation of drug. P-gp expression is regulated by various factors, including xenobiotics and hormones. Heat shock proteins, oncogene transfection, cell differentiation or proliferation could modulate its expression[24].

Mechanism of drug efflux:

Various models were proposed to explain the mechanism of xenobiotic extrusion by P-gp; however, the exact site of substrate interaction

with the protein is not well resolved. The three prevalent models pore model, flippase model and hydrophobic vacuum cleaner (HVC) model, explains the efflux mechanism to certain extent. Among these HVC model has gained wide acceptance in which P-gp recognizes substrates embedded in the inner leaflet of plasma membrane and transported through a protein channel[25]. Recently, Rosenberg et al[26] reported that three-dimensional conformation of P-gp changes upon binding of nucleotide to the intracellular nucleotide-binding domain. ATP binding and hydrolysis was found to be essential for functioning of P-gp, where one molecule of drug is effluxed at the expense of two molecules of ATP elucidated the catalytic cycle of P-gp, which expands the opportunity for the development of P-gp inhibitors, comprises of two cycles where drug and nucleotide binding sites coordinately function to efflux out the substrates by an ATP driven energy-dependent process. The drug and ATP initially binds to the protein at their own binding sites, where nucleotide hydrolyses to ADP yields energy for the extrusion of drug. Subsequent release of ADP completes another catalytic cycle, bringing P-gp molecule back to the original state, where it again binds to both substrate and nucleotide to initiate the next cycle[27]. (Fig.2)

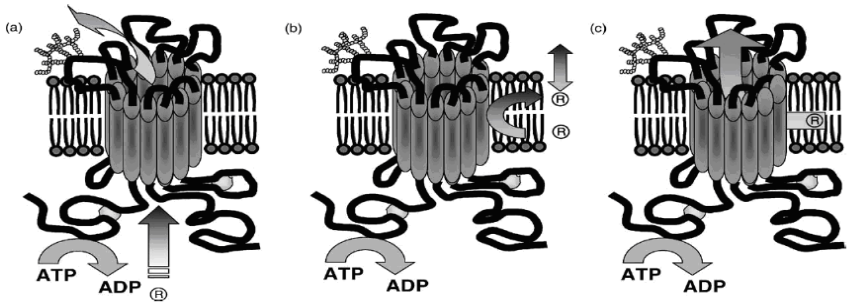

Fig 2: Mechanism of drug efflux by P-gp.
(a) Pore model, (b) flippase model and (c) hydrophobic vacuum cleaner model.

ATP-binding domain of ABC transporters:

ATP-binding domain of ABC transporters is a water-soluble domain of transmembrane ABC transporters. ABC transporters belong to the ATP-Binding Cassette superfamily, which uses the hydrolysis of ATP to translocate a variety of compounds across biological membranes. ABC transporters are minimally constituted of two conserved regions: a highly conserved ATP binding cassette (ABC) and a less conserved transmembrane domain (TMD). These regions can be found on the same protein or on two different ones. Most ABC transporters function as a dimer and therefore are constituted of four domains, two ABC modules and two TMDs[28]. (Fig.3)

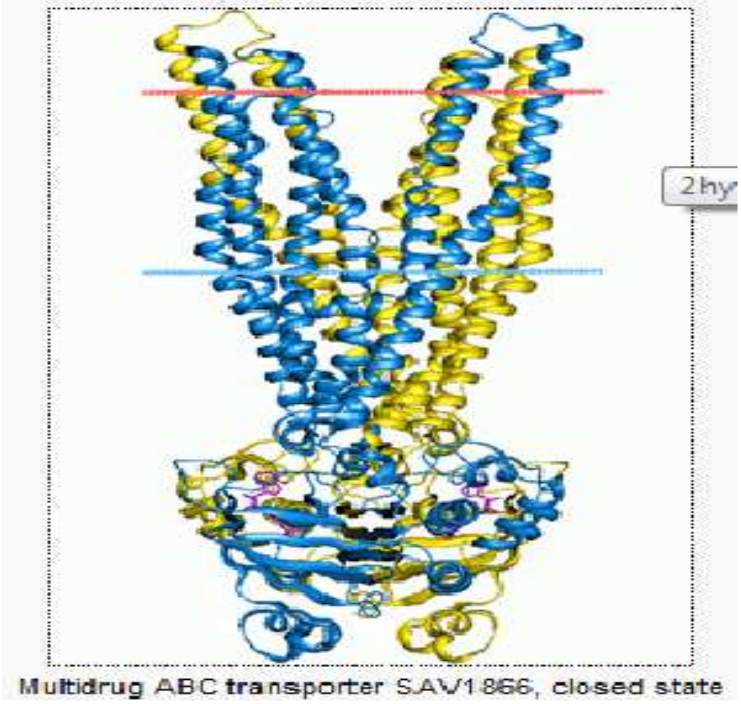

Fig 3: Structure of Multi Drug ABC Transporter SAV1866.

Structure:

P-gp is a 170 kDa transmembrane glycoprotein, which includes 10-15 kDa of N-terminal glycosylation. The N-term half of the molecule contains 6 transmembrane domains, followed by a large cytoplasmic domain with an ATP-binding site, and then a second section with 6 transmembrane domains and an ATP-binding site that shows over 65% of amino acid similarity with the first half of the polypeptide[29].

Biological function:

ABC transporters are involved in the export or import of a wide variety of substrates ranging from small ions to macromolecules. The major function of ABC import systems is to provide essential nutrients to bacteria [30].

MATERIALS AND METHODS:

Protein Structure Prediction:

Knowledge of the three-dimensional structure is a prerequisite for rational drug design. X-ray crystallography, NMR spectroscopy and electron microscopy are most important experimental methods to obtain detailed structural information. Protein and nucleic acid sequence methods are now well advanced and available in many laboratories. As a result sequence databases such as Swiss-Prot, TrEMBL (http://www.expasy.ch/), and the protein information resource (PIR), (http://www-nbrf.georgetown.edu/), have been growing rapidly in recent years. In contrast the determination of protein structure by NMR or X-ray crystallography has tended to proceed much more slowly. Hence there are many important proteins where the sequence is available but the three-dimensional structure is not yet known. The gap between the number of known sequences and the number of known structures is widening rapidly and the most successful theoretical approach to bridging this gap is homology modeling. The objective of homology modeling is to build a 3D model for a protein of which the structure is unknown (the target) on the basis of sequence similarity to proteins of known structure (the templates). The technique uses experimentally determined protein

structures to predict the conformation of another protein that has a similar amino acid sequence. The method relies on the observation that in nature the structural conformation of a protein is more highly conserved than its amino acid sequence and that small or medium changes in sequence typically result in only small changes in the 3D structure. Homology modeling or comparative modeling methods are able to predict the 3D structure of a protein sequence by using information derived from a homologous protein of known structure.

Outline of the Modeling Schemes:

Generally, the process of homology modeling involves four major steps – Fold Assignment, Sequence Alignment, Model Building and Model Refinement. The fold assignment process identifies proteins of known 3D structure (template structures) that are related to the polypeptide sequence of unknown structure (the target sequence). Next, a sequence database of proteins with known structures (e.g. the PDB-sequence database) is searched with the target sequence using sequence similarity search algorithms (e.g. BLAST). Following identification of a distinct correlation between the target protein and a protein of known 3D structure, the two protein sequences are aligned to identify the optimum correlation between the residues in the template and target sequences.

Retrieval of Protein Sequence of P-glycoprotein from NCBI:

Protein sequence of P-glycoprotein has been retrieved from protein database cited at National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov/). Details of the protein sequences shown below: Accession No. AAA02976

Modeling of P-glycoprotein Involves Use of Various Softwares and Tools:

- **For protein analysis**
  - ❖ http://www.expasy.org/
- **For retrieval of sequences and template selection and comparison of sequence**
  - ❖ NCBI (www.ncbi.nlm.nih.gov)
  - ❖ CLUSTAL W (www.ebi.ac.uk/clustalw)
  - ❖ BLAST (www.ncbi.nlm.nih.gov/blast)
- **For protein modelling :**
  - ❖ MODELLER 9V5
  - ❖ SWISS MODEL (http:/swismodel.expasy.org)
  - ❖ WHAT IF (www.cmbi.kun.nl/whatif)
  - ❖ VERIFY -3D (Discovery Studio)
  - ❖ DISCOVERY STUDIO (Accelrys)
  - ❖ NIH PROTEOMIC SERVER
- **For active site prediction**
  - ❖ Q Site Finder
- **For designing of library**
  - ❖ ChemSketch
  - ❖ Hyper chem.
- **For Docking**
  - ❖ GOLD 2.1
  - ❖ AutoDock 4.0
  - ❖ Ligand fit (Accelrys DS 2.0)
- **For ADMET prediction**
  - ❖ Discovery studio (Accelrys DS 2.0)
- **Visualization tool**
  - ❖ Rasmol
  - ❖ Pymol
  - ❖ Discovery Studio visualizer.

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 likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply

significant structural similarity. Homology modeling predicts structure based on sequence homology with known structure. It is also known a comparative modeling. The principle behind it is that two proteins share a high enough sequence similarity they are likely to have very similar three dimensional structures. If one of the protein sequences has a known structure, then the structure can be copied to the unknown protein with a high degree of confidence. Homology modeling produces an all atom model based on alignment with templates protein.

The overall procedure consists of five steps.

1. **Template selection**
2. **Sequence alignment**
3. **Loop refinement**
4. **Side chain refinement**
5. **Model evaluation**

The 'information content' of sequence and structure can be used to gain an insight into the molecular function of proteins. The sequence-to-function is the most commonly used approach for prediction and annotation of protein function. Sequence-based function prediction methods rely on sequence similarity and evolutionary relationship between the proteins of known and unknown functions. However, sequence-based methods are limited by their ability to detect similarities between distantly related proteins. The 3D structure of proteins is more conserved than the sequence. Therefore, structure-based function prediction approaches are considered to be more reliable, as function can be assigned to proteins irrespective of studying their evolutionary relationship. However, structure-based approaches are limited by the availability of protein structures. Hence, a unified approach involving an array of methods based on sequence and structural similarities should be used to assign/predict the molecular function of proteins.

Finding active site using Q-Site Finder

Identification of active site and binding sites in the Model:

For the purpose of this study amino acid sequence of P-glycoprotein was obtained from the NCBI protein database (Accession No. AAA02976). The active site residues of P-glycoprotein are (**His251, Glu252 and His255**) of *Leishmania donovani*. The active site will help in designing of selective P-glycoprotein inhibitors. This selectivity is very important and useful for selective inhibition of *Leishmania* parasite. (Fig.4)

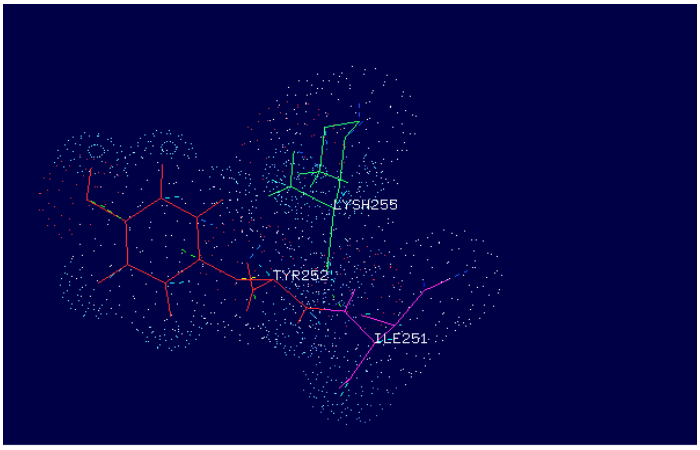

Fig 4: Active site residues of P-glycoprotein of *L.donovani*

Q-Site Finder

It uses the interaction energy between the protein and a simple Vander Waals probe to locate energetically favourable binding sites. Energetically favourable probe sites are clustered according to their spatial proximity and clusters are then ranked according to the sum of interaction energies for sites within each cluster. There is at least one successful prediction in the top three predicted sites in 90% of proteins tested

when using Q-Site Finder. This success rate is higher than that of a commonly used pocket detection algorithm (Pocket-Finder) which uses geometric criteria. Additionally, Q-Site Finder is twice as effective as Pocket-Finder in generating predicted sites that map accurately onto ligand coordinates. It also generates predicted sites with the lowest average volumes of the methods examined in this study.

Designing of chemical library:

ACD/ChemSketch: is the chemical structure drawing program most frequently used by organic chemists. Most of the chemical structures that you see in texts, journal articles, and on the web were drawn in ChemSketch. Besides drawing structures, ChemSketch is also useful for:

- Naming a compound.
- Determining the structure of a named compound.
- Predicting the NMR spectrum of a compound.
- Finding physical data.
- Finding the molecular weight of fragments for mass spectroscopy

ACD/ChemSketch is the powerful all-purpose chemical drawing and graphics package from ACD/Labs developed to help chemists quickly and easily draw molecular structures, reactions, and schematic diagrams, calculate chemical properties, and design professional reports and presentations.

ACD/ChemSketch includes:

• Structure mode for drawing chemical structures and calculating their properties

• Draw mode or text and graphics processing.

• Additional modules that extend the ChemSketch possibilities.

Protein-Ligand Interaction Study: For protein-ligand interaction study of the flexible ligands to the receptors, following softwares are used:

1. **GOLD 2.1**
2. **AUTODOCK4.0**
3. **Ligand Fit (Discovery Studio2.0)**

Protein Ligand Interaction:

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule (ligand or macromolecule) to a second (macromolecule) 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 scoring functions.

AutoDock4.0:

AutoDock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. AutoDock actually consists of two main programs: AutoDock performs the docking of the ligand to a set of grids describing the target protein; AutoGrid pre-calculates these grids.

AutoDock Tools, or ADT: is the free GUI for Auto Dock developed by the same laboratory that develops Auto Dock. It is used to set up, run and analyze Auto Dock dockings and is contour Auto Grid affinity maps, as well as compute molecular surfaces, display secondary structure ribbons, compute hydrogen-bonds, and do many more useful things. (Fig.5)

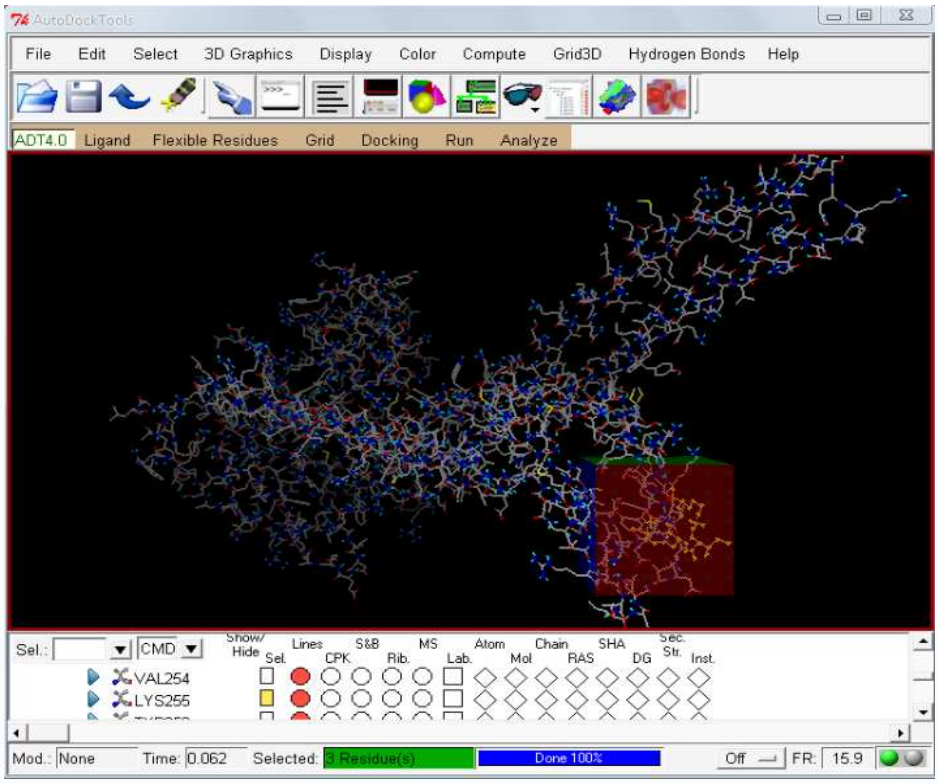

Fig 5: Showing AutoDock tool window.

RESULTS AND DISCUSSION

BLAST search was performed against protein data bank (PDB) with the default parameter to find suitable template for homology modeling. Sequence was aligned and the one that showed the maximum identity with high score and lower e- value and 42% sequence identity were used as a reference structure to build a 3D model for P-glycoprotein respectively. The P-glycoprotein was modeled by means of comparative modeling procedure using the PDB having ID “2hyd” as template for P-glycoprotein.

The sequence alignment was carried out using the ClustalW and MODELLER9V5 program. The alignment was manually refined at some loop region of the template. The academic version of MODELLER 9V5 was used for the model building. Backbones of core region of the protein were transferred directly from the corresponding co-ordinates of templates. Side chains conformations for backbone was generated automatically by homology. Out of 5 models generated by MODELLER, the one with the best Dope score and with the VERIFY 3D profile was subjected to energy minimization. The distance –dependent dielectric constant e=1.0 and non bonding cut-off 14Å CHARM force field and CHARM all atoms charges were used for energy minimization . Initially 800 step steepest descent algorithms was used to remove close van der Waals contacts, followed by 1000 iteration conjugates gradient minimization until the maximum derivatives is less than 20.0 kcal.mol.nm. All hydrogen atoms were included during the calculation. The above energy minimization was started with the core main chain, then all core side chain, all calculation was perform using Accelrys DS modeling 2.0 software suite. During this step the quality of initial model was improved VERIFY 3D to check the residue profile of the three dimensional models .In order to assess the stereo chemical qualities of three dimensional model PROCHECK analysis was performed and Ramachandran plot was drawn.

Choice of template:

Comparative modeling usually starts with searching the PDB of known protein structures using the target sequence as the query. This search is generally done by comparing the target sequence with the sequence of each of the structures in the database. The most popular programs in the class include FASTA and BLAST.

Once a list of potential templates is obtained using searching methods, it is necessary to select one or more templates that are appropriate for the particular modeling problem. The quality of a template increases with its overall sequence similarity to the target and decreases with the number and length of gaps in the alignment. The simplest template selection rule is to select the structure with the highest sequence similarity to the modeled sequence. (Table 1)

BLAST result of P-glycoprotein:

Table 1

BLAST result of P-glycoprotein showing % identity with template

PDB ID	Name of template	identity
2hyd	Multi drug transporter protein sav1866 complex	42%

Sequence alignment:

Alignment of sequences with their templates structure was done using the ClustalW. The software also takes into account structural information from the template when constructing an alignment. The MODELLER script was used for aligning all target sequences in the Ali file with their corresponding template structures in the PDB files. Finally, the alignment was written out in two formats, PIR and PAP. The PIR format is used by MODELLER in the subsequent model building stage, while the PAP alignment format is easier to inspect visually. In the PAP format, all identical positions are marked with a '*'. The details of modeling and sequence alignment scripts are submitted as supplementary material. (Fig.6)

Homology Modeling:

A 3D model of the target sequence was constructed with the auto model class of MODELLER9V5 to generate five similar iterative models of the target sequence based on its template structure and the alignment input file 'filename.ali' (PIR format). The 'best' model was selected by picking the model with the lowest MODELLER objective function value, which is reported in the second line of the model PDB file. (Fig.7)

```
CLUSTAL 2.0.10 multiple sequence alignment


2hyd            MIKRYLQFVKPYKYRIFATIIVGIIKFGIPMLIPLLIKYAIDGVINNHALTTDEKVHHLT 60
target          ---------APSATAFTESRAAAYEIFKAIDRVPPVDIDAGGVPVTGFKQSIEFRNVRFS 51
                         *    :  :  ..   *     :* :   * .  :...  : : :  :::

2hyd            IAIGIALFIFVIVRPPIEFIRQYLAQWTSNKILYDIRKKLYNHLQALSARFYANNQVGQV 120
target          YPTRPDMILFRDLNLTIKCGQKVAFSGSSGCGKSSMIGLIQRFYDPVGGAVLCD---GVD 108
                 .    :::*  :. .*:  ::   . :*.    .:   : .. :.:.. . .:   *

2hyd            ISRVINDVEQTKDFILTGLMNIWLDCITIIIALSIMFFLDVKLTLAALFIFPFYILTVYV 180
target          MRELCLHDWRNQIGIVSQEPNLFAGTMMENVRVGKPNATEEEVIEACRQAN--------- 159
                : .:  .  :.:  *::   *:: . :   : :.     : ::  *.

2hyd            FFGRLRKLTRERSQALAEVQGFLHERVQGISVVKSFAIEDNEAKNFDKKNTNFLTRALKH 240
target          VHDTIMSLPDQYNTPVGAVGSQLSGGQKQRIAIARALVKRPPVLLLDEATSALDRKSEME 219
                ... : .*. : . .:. * . *    :   .:    ::   .  :*: .: :  ::  .

2hyd            TRWNAYSFAAINTVTDIGPIIVIGVGAYLAISGSITVGTLAAFVGYLELLFGPLRRLVAS 300
target          VQR------SLDQLMQRGGMTVIVIAHRLATIRNVDCIYYVKYDGAEGSKITESG-TFDE 272
                .:       ::: : : * : ** :.  **   .:     . : *     :      . .

2hyd            FTTLTQSFASMDRVFQLIDEDYDIKNGVGAQPIEIKQGRIDIDHVSFQYNDNEAPILKDI 360
target          LMELGGEFAAMARIQGVSTGDARSGARGQDGGKANDFLNAILDEAALAQLDEEAPRTARQ 332
                :  *  .**:* *:  :   *                .  .  :*..::   *:***

2hyd            NLSIEKGETVAFVGMSGGGKSTLINLIPRFYDVTSGQILIDGHNIKDFLTGSLRNQIGLV 420
target          KVPIEELAKWEVNGVSVGFRRLMGMNKDKSWAVALG---ILGSAVGGAARPTSTILMGYM 389
                ::.**:  .  . *:* * :  :     : : *: *   * *  : .    :    :* :

2hyd            QQDNILFSDTVKENILLGRP-TATDEEVVEAAKMANAHDFIMNLPQGYDTEVGERGVKLS 479
target          LR--VLGEYSVNKNVEQLRSGTNLYAPLFIVFAVANFLGWILHGFYGYAGEHLT------ 441
                 :  :* . :*:::*:   *. *     :. .  :**  .:*::   **  *

2hyd            GGQKQRLSIARIFLNNPPILILDEATSALDLESESIIQEALDVLSKDRTTLIVAHRLSTI 539
target          --TKIRVLLFRQIMRQDMNFFDIPGRDAGTLAG---------MLSGDCEAVHQLWGPSIG 490
                   * *: : * ::.:   ::   . .*  * .         :** *  ::      *

2hyd            THADKIVVIENGHIVETGTHRELIAKQGAYEHLYSIQNL 578
target          LKVQTVCIVASGIVVSF-----------IYQWKLALVAL 518
                 :.:.: :: .* :*.             *:   ::  *
```

Fig 6: Clustal W showing multiple sequence alignment of amino acid sequence of p-glycoprotein of *L.donovani* & PDB template"2hyd". The alignment was carried out with Clustal W.

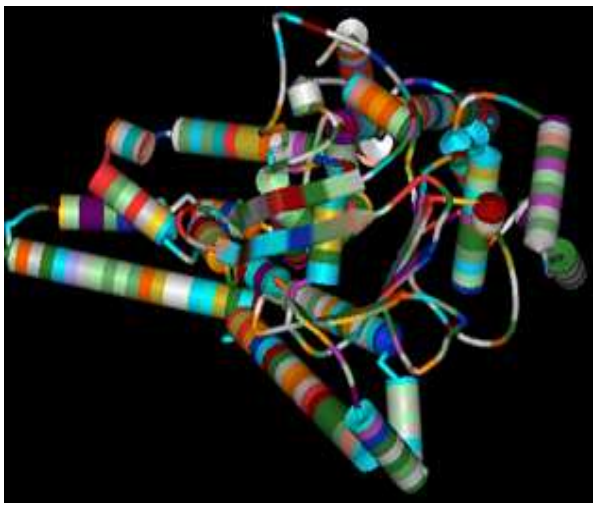

Fig 7: Representation of modeled P-glycoprotein image of *L.donovani.*

Refinements:

Loop Refinement:

In the sequence alignment for modeling, there are often region caused by insertion an deletion producing gaps in sequence alignment. The gaps cannot be directly modeled, creating "loops" in the model. Closing the gaps requires for modeling, which is very difficult problem in homology modeling and is also a major source of error. There are two technique are used to approach the problem: database searching method and ab-initio method. The rough model has various loops in the structure which are called as invalid region.

The rough model has been subjected to loop refining using the Discovery Studio (DS). The obtained model was subjected to again loop refinement. There were some invalid regions in the model of P-glycoprotein that were subjected to loop refinement. (Table 2)

Table 2

VERIFY 3D SCORE: Model of P-glycoprotein before loop refinement

Expected high score	236.743
Expected low score	106.534
P-glycoprotein model (best) score	169.23*

The model was subjected to loop refinement by Discovery Studio the first invalid region was refined and there were increment in verify 3D score. These invalid regions were again loop refined. There was increment in the VERIFY3D score near to the high value but no further improvement in verify 3D score. (Table 3)

Table 3

VERIFY 3D SCORE: Final Model of P-glycoprotein after loop refinement

Expected high score	236.743
Expected low score	106.534
P-glycoprotein model (best) score	173.832*

Side Chain Refinement:

Side chain refinement was performed by the discovery studio of model of p-glycoprotein. But there was no improvement in the model of P-glycoprotein but the VERIFY3D score was increased and assumed that the model is good model. The final model has following VERIFY3D score after side chain refinement. (Table 4, 5)

Table 4

VERIFY 3D SCORE: Model of P-glycoprotein before side chain refinement**.**

Expected high score	236.743
Expected low score	101.534
P-glycoprotein model (best) score	173.832*

Table 5

VERIFY 3D SCORE: Model of P-glycoprotein after side chain refinement.

Expected high score	236.743
Expected low score	106.534
P-glycoprotein model (best) score	176.897*

Model Validation:

Validation of the model was carried out after the refinement process using Ramachandran map calculation computed with the PROCHECK program. The phi and psi distribution of Ramachandran map of non-glycine, non-proline residue are summarize in Fig 8 and Table 6. The model has 89% of residue in the most favoured region.

Table 6

Ramachandran plot calculation for 3D model of P-glycoprotein with the PROCHECK program

Residues	No. of Amino acid	Percentage
Residue in most favoured region	396	89%
Residue in the additionally allowed zones	39	8.8%
Residue in the generously region	3	0.7%
Residue in disallowed region	7	1.6%

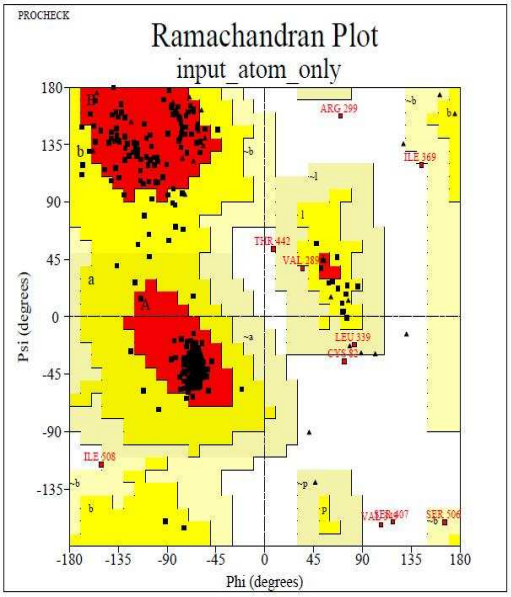

Fig 8: Ramachandran plot: The residues marked red shows the disallowed and generously allowed residues.

Retrieval of P-glycoprotein inhibitors:

Potent inhibitors of P-glycoprotein are retrieved from various sources mentioned below:

P-glycoprotein Inhibitors:

Table 7
GOLD Fitness score of 21 ligand molecules of P-glycoprotein inhibitors

Drugs	Fitness	S(hb_ext)	S(vdw_int)	S(hb_int)	S(vdw_ext)
IMATINIB	75.76	0	58.56	0	-5.76
TELMISARTON	77.68	0	58.99	0	-5.42
PANTAPRAZOLE	59.71	0	40.68	0	-2.72
NIZATIDINE	56.44	0	44.94	0	-5.35
NELFINAVIR	53.25	0	46.17	0	-10.23
MOEXIPRIL	59.65	0	46.29	0	-4
INDINAVIR	59.79	0	54.84	0	-4.54
FOSINOPRIL	65.98	0	51.54	0	-4.9
ERLOTINIB	56.92	0	43.39	0	-6.42
DARUNAVIR	59.61	0	48.02	0	-2.74
CILOSTAZOL	57.46	0	42.29	0	-0.69
ATROVASTATIN	57.48	0	51.78	0	-13.71
AMIODARONE	61.14	0	49.04	0	-6.3
AZLOCILLIN	57.98	0	26.64	0	-2.35
AZTREONAM	57.55	0	41.08	0	-5.38
CANDOXATRIL	65.08	0	58.34	0	-15.13
CONIVAPTON	67.48	0	50.05	0	-1.35
DARIFENACIN	64.12	0	47.94	0	-1.8
DASATINIB	67.41	0	50.32	0	-1.79
DELAVIRDINE	65.45	0	45.5	0	-2.89
NEFAZODONE	62.35	0	46.67	0	-1.82

Table 7 shows the GOLD fitness score of 21 compounds of P-glycoprotein inhibitors. Imatinib, Telmisartan inhibitor shows the highest GOLD fitness score. Nelfinavir shows the lowest GOLD fitness score. Pantaprazole, dasatinib, amiodarone and conivapton also show good fitness score.

TOP TEN DRUGS SELECTED AND COMPARISION OF RESULT OF AUTODOCK, GOLD, DISCOVERY STUDIO

Table 8
Showing Discovery studio score of Top ten compounds from different class of inhibitors.

Antileishmanial Drug & P-glycoprotein inhibitor	Best pose	Ligand-Fit Score	No. of H-bonds	Amino Acid Residue &Distance
KETOCONAZOLE	10	45.304	5	ARG156:HH21-KETOCONAZOLE:04 ARG156:HH21-KETOCONAZOLE:03 ARG118:HE-KETOCONAZOLE:CL2 ASN159:HD21-KETOCONAZOLE:03 LYS187:HZ3-KETOCONAZOLE:05
IMATINIB	10	47.24	4	LYS187:H21-IMATINIB:05 VAL349:H21-IMATINIB:01 ARG352:H21-IMATINIB:01 ARG353:H21-IMATINIB:05
TELMISARTON	10	45.305	6	ARG352:HH21-TELMISARTON:02 ARG353:HH12-TELMISARTON:01 ARG156:HH22-TELMISARTON:N6 ARG391:HH12-TELMISARTON:02 GLN483:HE22-TELMISARTON:02 ARG118:HE-TELMISARTON:01
CONIVAPTON	10	55.769	2	ARG118:HE-CONIVAPTON:02 MET473118:H55-CONIVAPTON:05
AMIODARONE	10	35.62	4	ARG352:HH21-AMIODARONE:03 ARG352:HH22-AMIODARONE:03 ARG353:HH21-AMIODARONE:05 ARG353:HH22-AMIODARONE:03
DASATINIB	10	44.23	3	ASN423:HD21-DASATINIB:N7 MET473:H48-DASATINIB:O5 LEU470:H48-DASATINIB:O5
DIAZEPAM	10	44.738	4	GLN221:HE22-DIAZEPAM:O2 ARG352:HH12-DIAZEPAM:CL1 ARG353:HH22-DIAZEPAM:N4 ARG391:HH12-DIAZEPAM:O2
NELFINAVIR	10	33.137	3	ARG118:HE-NELFINAVIR:O4 GLU275:H85-NELFINAVIR:OE1 GLU219:H85-NELFINAVIR:OE2
PANTAPRAZOLE	10	42.758	6	ARG352:HH21-PANTAPRAZOLE:N9 LYS187:H21-PANTAPRAZOLE:O6 VAL349:HN-PANTAPRAZOLE:F2 ARG353:HE-PANTAPRAZOLE:F2 ARG391:HH12-PANTAPRAZOLE:O6

				GLN483:HE21-PANTAPRAZOLE:N9
NIZATIDINE	10	48.986	3	VAL160:HN-NIZATIDINE:S2 GLN221:HE22-NIZATIDINE:N5 SER223:HG-NIZATIDINE:O4

Table 9

Showing GOLD Fitness, Discovery & AutoDock score of Top ten compounds from different class of inhibitors.

Antileishmanial Drug & P-glycoprotein inhibitor	DISCOVERY STUDIO	AUTODOCK 4.0	GOLD
KETOCONAZOLE	45.304	-3.90	66.02
IMATINIB	47.24	-3.36	75.76
TELMISARTON	45.305	-2.35	77.68
CONIVAPTON	55.769	-5.45	67.48
AMIODARONE	35.62	-5.87	61.14
DASATINIB	44.23	-5.95	67.41
DIAZEPAM	44.738	-4.32	61.1
NELFINAVIR	33.137	-2.66	53.25
PANTAPRAZOLE	42.758	-4.92	59.71
NIZATIDINE	48.986	-5.75	56.44

From GOLD ligand protein interaction study the Table 8 shows the GOLD fitness score of top 10 compounds shown valid fitness score. Imatinib & Telmisarton inhibitor shows the highest GOLD fitness score. Conivaptan & Dasatinib also shows good fitness score as shown in the table. Table 9 also shows the comparision of the results of AUTODOCK, GOLD & DISCOVERY STUDIO between modeled P-glycoprotein and its inhibitors.

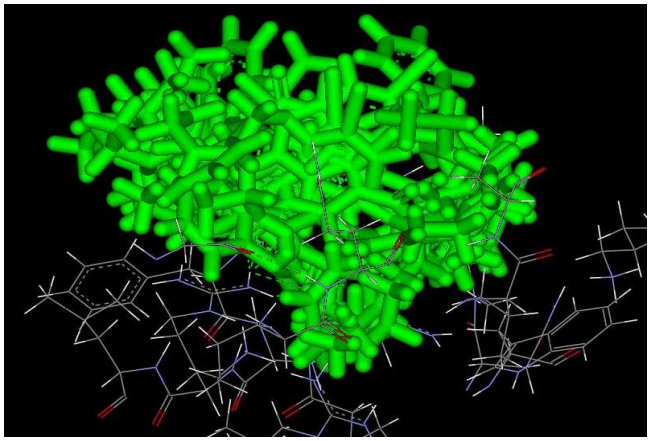

Fig 8: Protein ligand interaction of Amiodarone & P-glycoprotein active site residues showing H-bonds by Discovery Studio. H-bonds are shown by green dotted line.

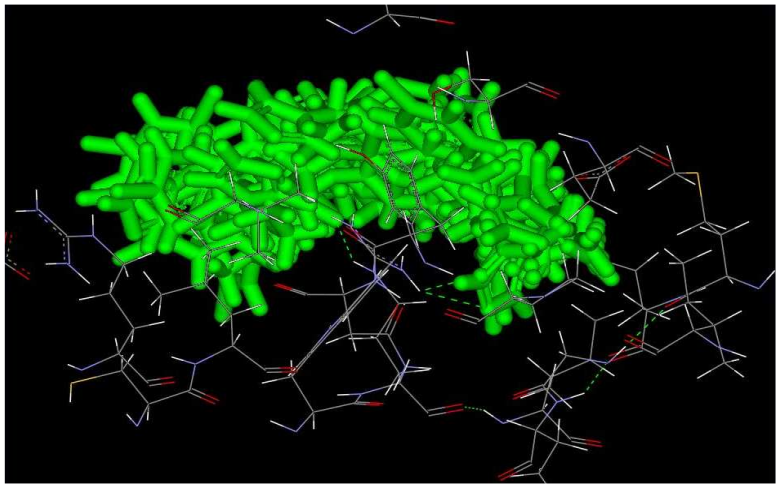

Fig 9: Protein ligand interaction of Ketoconazole & P-glycoprotein active site residues showing H-bonds by Discovery Studio. H-bonds are shown by green dotted lines.

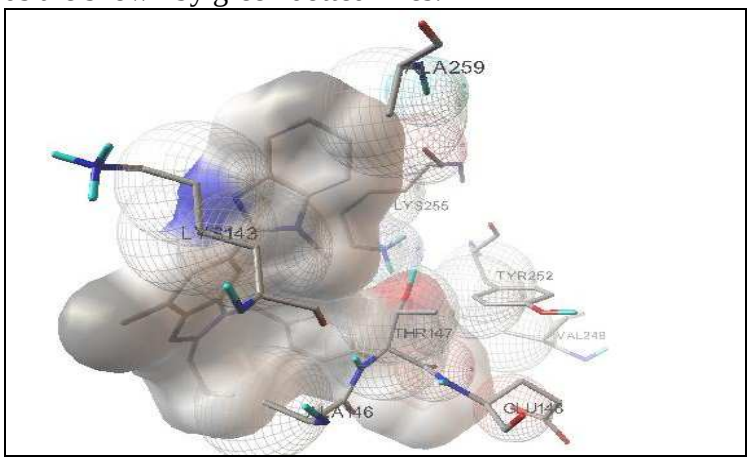

Fig 10 : Protein ligand interaction of Telmisarton & P-glycoprotein active site residues showing H-bonds and close contacts by AutoDock.

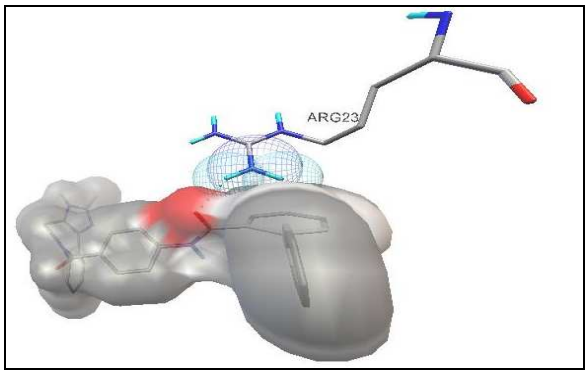

Fig 11: Protein ligand interaction of Conivapton & P-glycoprotein active site residues showing H-bonds and close contacts by AutoDock.

CONCLUSION:

Leishmania P-glycoprotein is potential target for treating visceral leishmaniasis. In order to understand structural features, homology modeling of P-glycoprotein of *Leishmania donovani* was carried out taking P-glycoprotein of sav1866 (pdb id:2hyd) as template. The refined model of the protein was obtained by further refinement of raw model by loop and side chain refinements. The model showed good overall structural quality as validated by using different validation tools like PROCHECK, verify 3D score. In addition the model shows good native protein folding. The ligand binding site was determined by sequence alignment and Q site finder analysis of various species of *Leishmania*. These are the residues **His251, Glu252, His255** of P-glycoprotein of *Leishmania donovani*. Thus, developed P-glycoprotein model was taken for screening a chemical library of PubChem and Drug Bank for finding the best P-glycoprotein inhibitor of *L.*donovani.

Ligand protein interaction study of various categories of inhibitors i.e. anthracyclines, xanthons, natural phenolic compound, alkaloids and the modeled P-glycoprotein of *Leishmania donovani* reveals that the active site is inhibited as some of the compounds have shown good fitness score(GOLD).

This is the first report of ligand protein interaction about the currently used drugs for the treatment of leishmaniasis and P-glycoprotein of *L.donovani*. The study of inhibitory capacity of antileishmanial drugs is a prerequisite for design of novel drug candidates against *Leishmania* species. Homology modeling of P-glycoprotein of *Leishmania* shed new light on the ligand binding features. From ligand protein interaction study, it is learned that few drugs which have been implicated for leishmaniasis treatment are specific for particular strain (ketoconazole to *Leishmania donovani*). Different ligand binding conformations were observed in both Discovery studio 2.0 and GOLD 2.1.

It is hypothesized that the compounds conivapton and nizatidine which have shown the highest ligand fit score in discovery studio and inhibitors like imatinib, telmisarton, dasatinib which have shown the highest binding affinity in GOLD, may be the best inhibitors of P-glycoprotein of *L.donovani*.

REFERENCES:

1. Pearson R. D, Sousa A Q, Clinical spectrum of Leishmaniasis. Clin .Infect. Dis, 22, 1–13, (1996).
2. Arevalo J. et al., Influence of leishmania (viannia) species on the response to antimonial treatment in patients with American tegumentary leishmaniasis. J. Infect. Dis, 195, 1846–1851, (2007).
3. Dedet J P, Pratlong F, In Manson's Tropical Diseases (eds Cook, G. C. & Zumla, A. I.), 1339–1364, (2003).
4. Lukes J. et al., Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. Proc. Natl Acad. Sci. USA, 104, 9375–9380, (2007).
5. Mauricio I L, Stothard J R., Miles M A, The strange case of Leishmania chagasi. Parasitol. Today, 16, 188–189, (2000).
6. Alvar J, Canavate C, Molina R, Moreno J, Nieto J, Leishmaniasis. Adv. Parasitol, 57, 1–88, (2004).
7. Addy M, Nandy A, Ten years of kala-azar in west Bengal, Part I. Did post-kala-azar dermal Leishmaniasis initiate the outbreak in 24-Parganas Bull. World Health Organ, 70, 341–346, (1992).
8. Wrin J, Reeder J, Liu D, Ring D B, High-affinity monoclonal antibodies against P-glycoprotein. Clin Immunol Immunopathol, 76, 44-51, (1995).
9. Gupta S, Aggarwal S, P-glycoprotein in cells of the human immune system. The Immunologist, 43, 86-90, (1996).
10. Miles F, Role of Multidrug resistance and its pharmacological modulation in acute myeloid leukemia. Leukemia, 10, 937-942, (1996).
11. Filipits M, Suchomel R W, Dekan G et al., MRP and MDR1 gene expression in primary

breast carcinomas. Clin Cancer Res, 2, 1231-1237, (1996).
12. Drach D, Zhao S, Drach J et al, Subpopulations of normal peripheral blood and bone marrow cells express a functional multidrug resistant phenotype. Blood, 80, 2729-2734, (1992).
13. Klimecki W T, Gutscher B W, Grogan T M, Dalton W S, P-glycoprotein expression and function in circulating blood cells from normal volunteers. Blood, 83, 2451-2458, (1994).
14. Victoria J M, Chiquero M J, Conseil G, Dayan G, Pietro A, Barron D, Castanys S, Gamarro F, Correlation between the affinity of flavonoids binding to cytosolic site of Leishmania tropica multidrug transporter and their efficiency to revert parasite resistance to daunomycin. Biochemistry, 38, 1736–1743, (1999).
15. Middleton E, Kandaswami C, The impact of plant flavonoids on mammalian biology: implications for immunity, inflammation and cancer. Pharmacol Rev, 52(4), 673-751, (2000).
16. Dey S, Papadopolou B, Haimeur A, Roy G, Grondin K, Dou D, Roser B P, Ouellette M., High level arsenite resistance in Leishmania tarentolae is mediated by an active extrusion system. Mol Biochem Parasitol, 67, 49-57, (1994).
17. Hedman A, Meijer D K, The stereoisomers quinine and quinidine exhibit a marked stereo selectivity in the inhibition of hepatobiliary transport of cardiac glycosides. J Hepatol, 28(2), 240-9, (1998).
18. Koren G, Woodland C, Ito S, Toxic digoxin-drug interactions the major role of renal P-glycoprotein. Vet Hum Toxicol, 40(1), 45-6, (1998).
19. Sadeque A J, Wandel C, He H, Shah S, Wood A J, Increased drug delivery to the brain by P-glycoprotein inhibition. Clin. Pharmacol. Ther, Sep, 68(3), 231-7, (2000).
20. Sukhai M, Piquette-Miller M, Regulation of the multidrug resistance genes by stress signals. J Pharm. Pharm Sc, 3(2), 268-80, (2000).
21. Higgins C.F, ABC transporters from microorganisms to man. Ann Rev. Cell Biol, 8, 67-113, (1992).
22. Fardel O, Lecureur V, Guillouzo A, The P-glycoprotein multidrug transporter. Gen. Pharmacol, 27**,** 1283-1291, (1996).
23. Chen C J, Chin J E, Ueda K, Clark D P, Pastan I, Gottesman M M, Roninson I B, Molecular biology of drug resistance. Cell Biol , 47, 381-389, (1986).
24. Bradley G, Juranka P F, Ling V, Mechanism of multi drug resistance. Biochim. Biophys. Acta, 948, 87-128, (1988).
25. Higgings C F, Gottesman M M, Is the multidrug transporter a flippase Trends. Biochem Sci, 17**,** 18–21, (1992).
26. Rosenberg M F, Kamis A B, Collaghan R, Higgins C F, Ford R C, Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major conformational changes in the transmembrane domains upon nucleotide binding. J .Biol .Chem, 278(10)**,** 8294-9, (2003).
27. Gottesman M M, Pastan I, Biochemistry of multidrug resistance mediated by the multidrug transporter. Annu Rev Biochem, 62, 385–427, (1993).
28. Michael D, The Human ATP-Binding Cassette (ABC) Transporter Super family. National Library of Medicine (US), NCBI, (2002).
29. Stephen A, Jodie Yu, Ward A, Yue W, Chittaboina S, Zhuo R, Patina M H, Yen phuong T, Zhang T Q, Urbatsch L, Chang G, Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science, 323, 1718-1722, (2009).
30. Andrew W, Christopher L, Jodie Yu, Christopher B Roth, Geoffrey Chang , Flexibility in the ABC Transporter MsbA: Alternating Access with a Twist. Proc Natl Acad Sci, 104(48), 19005-19010, (2007)