



Study on Molecular Modeling of P-Glycoprotein (P-gp) Structure Using Computational Approach

Anurag Pal*

Dept. of CSE, ITER

Siksha O Anusandhan University Bhubaneswar,
Odisha, INDIA

anuragom2000@yahoo.com

Abhisek Pal

Dept. of Pharmacology, SPS

Siksha O Anusandhan University Bhubaneswar,
Odisha, INDIA

abhipharma_2000@yahoo.co.in

Abstract: Proteins are sometimes described as long necklaces with differently shaped beads. Each bead is a small amino acid. These amino acids can join together to make thousands of different proteins. P-glycoprotein (P-gp) is one of the xenobiotic transport proteins implicated in multi-drug resistance in neoplastic tissues. It is a cell membrane-associated protein that transports a variety of drug substrates. P-gp (ABCB1) appears to have developed as a mechanism to protect the body from harmful substances. P-gp functions as a drug export pump that decreases intracellular concentrations of numerous chemotherapeutic agents. In this paper, P-gp has been taken as the target sequence and the protein is processed under molecular modeling. To know the proper molecular model of the protein, the target sequence is matched with the protein structure database with the help of protein data bank to find the templates. The target sequence has been matched with protein structure database with the help of bioinformatics tool and also with some of the other programming techniques. The maximum identity accession along with the P-gp structure is identified; its structure is analyzed and compared with the existing P-gp structure.

Keywords: P-glycoprotein (Pgp); Chemotherapeutic Agents; Remote Protein Homologies; Modeler; Structure Templates; Structure Drug Design; Drug Export Pump; molecular modeling

I. INTRODUCTION

P-glycoprotein (*P-gp*) is a cell membrane-associated protein that transports a variety of drug substrates. *P-gp* is one of the xenobiotic transport proteins implicated in multidrug resistance in neoplastic tissues. In cancer tissue with high expression of this protein, *P-gp* functions as a drug export pump that decreases intracellular concentrations of numerous chemotherapeutic agents [1]. *P-gp* (ABCB1) has been implicated as a primary cause of multidrug-resistance in tumors. The responsible gene- has been found to be MDR1. Many oncological drugs are ABCB1 (*P-gp*) substrates and are excluded from the brain at the blood-brain barrier (BBB) [2]. To predict or design the *P-gp* sequence and structure, the molecular modeling techniques are used and also in the fields of computational chemistry, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies.

Homology modeling is used to predict the 3D-structure of an unknown protein based on the known structure of a similar protein. Modeler is a computer program that models three-dimensional structures of proteins. It is most frequently used for homology or comparative protein structure modeling where the user provides an alignment of a sequence to be modeled with known related structures and it will automatically calculate a model with all non-hydrogen atoms [3]. Modeler can also perform multiple comparisons of protein sequences and/or structures, clustering of proteins and searching of sequence databases.

The rest of the paper has organized as follows: section 2 gives the idea about related work, section 3 represents preliminary concepts related to this paper; section 4 represents the proposed model; section 5 deals with experimental evaluation and result analysis and section 6 gives the conclusion.

II. RELATED WORK

Ramachandran *et al.* [4] proposed the molecular modeling technique and also the homology modeling of protein which is applied in computational chemistry and computational biology. James *et al.* [5] proposed the sequence alignment of the protein using the perl script with comparison to other protein sequence. Fu *et al.* [6] proposed intracellular trafficking of *P-gp*, modeling the *P-gp* and intracellular trafficking pathways for *P-gp* and participation of different. *Rab proteins* depend on cellular polarization and choice of primary culture, cell line or neoplasm. Schumacher *et al.* [7] proposed MDR-1-overexpression in HT 29 colon cancer cells grown in SCID mice after modeling the *P-gp* structure. Alexandridou *et al.* [8] proposed the process of pattern matching of protein sequence with help of Perl script or *Bio-Perl*, also known as Perl pattern matching.

III. PRELIMINARIES

A. *P-glycoprotein and Multi-Drug Resistance (MDR):*

In recent years there has been increasing focus on the role of the drug transporter, *P-gp* with regard to drug penetration into the brain. Studies using mice devoid of functional *P-gp* have revealed that *P-gp* at the blood-brain barrier (BBB) can exert a profound effect on the ability of some drugs to enter the brain, e.g. cardiovascular drugs (digoxin, quinidine), opioids (morphine, loperamide, methadone), HIV protease inhibitors, the new generation of antihistamines, and some antidepressants and antipsychotics [9].

On the other hand, multidrug resistance, the principal mechanism by which many cancers develop resistance to

chemotherapy drugs is a major factor in the failure of many forms of chemotherapy. It affects patients with a variety of blood cancers and solid tumors, including breast, ovarian, lung, and lower gastrointestinal tract cancers [10].

B. Molecular Modeling and Homology Modeling using Modeler:

Molecular modeling is a collection of (computer based) techniques for deriving, representing and manipulating the structures and reactions of molecules, and those properties that are dependent on these three dimensional structures [11]. *Remote Protein Homology* is one of the most intensively researched problems in bioinformatics. Researchers are increasingly depending on computational techniques to classify proteins into functional or structural classes by means of homologies [12]. Homology Modeling is the 3D structure of a target protein and that has also not been solved empirically by X-ray crystallography or NMR. If an empirically determined 3D structure is available for a sufficiently similar protein (50% or better sequence identity would be good) then we can use the software that arranges the backbone of the sequence identical to the template. This is homology modeling [13]. Modeler is used for homology or comparative modeling of protein three-dimensional structures. The user provides an alignment of a sequence to be modeled with known related structures and Modeler automatically calculates a model containing all non-hydrogen atoms [14].

C. Perl and Bio-Perl:

Perl is a high-level, general-purpose, interpreted dynamic programming language. Though *Perl* is not officially an acronym, there are various acronyms in usage, such as: Practical Extraction and Reporting Language [15]. On the other hand, *Bio-Perl* is a collection of *Perl* modules that facilitate the development of *Perl* scripts for bioinformatics applications. It has played an integral role in the Human Genome Project [16].

IV. PROPOSED MODEL

Figure 1 depicts the proposed model to the selection of the dataset of *P-gp*. The protein structure database is searched with the help of protein structure data bank. The target sequence is used as a query for sequence alignment, pattern matching. Also for many purposes, the sequence in .ali or .pir or in fasta format is saved for modeling. The pattern is matched and aligned by BLAST; the sequence is aligned using *Bio-Perl* and the pattern is matched using *Bio-Perl*. The methods are compared in detail for a study to find out the proper template. After alignment the 3D-structure of maximum identity is found for further study of the structural analysis of various model and structures.



Figure 1. Proposed work flow model

V. EXPERIMENTAL EVALUATION

The total procedure is explained in the following steps.

Step I: Initially, the *P-gp* data set or sequence has been selected [17]. This is the MDR1 or ABCB1 gene of human. *P-gp* has a unique id P08183. The main function of *P-gp* is energy-dependent efflux pump responsible for decreased drug accumulation in multidrug-resistant cells. The sequence length of *P-gp* is 1280 AA. A sample sequence of *P-gp* is shown in figure 2.

```

10  20  30  40
MDLEGDRNGG AKKKNFFKLN NKSEKDKKEK KPTVSFMSF
50  60  70  80
RYSNWLKLY MVVGTAAII HGAGLPLMML VFGEMTDIFA
.....
1230 1240 1250 1260
AREGRTCIVI AHRSTIQNA DLIVVFQNGR VKEHGTHQQL
1270 1280
LAQKGIFYFSM VSVQAGTKRQ
  
```

Figure 2. A sample *P-gp* data sequence

Step II: After getting the *P-gp* data sequence, the structure data base was selected. The structure data base is generally a composition of huge protein structure data base. From PDB [18], the structure database was prepared.

Step III: The target sequence or the *P-gp* data sequence used as a query for structure data base. This query sequence was saved in .ali or fasta format. The sample fasta format was shown below in figure 3 and the sample .ali format is shown below in figure 4.

```

>sp|P08183|MDR1_HUMAN Multidrug resistance protein
1 OS=Homo sapiens GN=ABCB1 PE=1 SV=3
MDLEGDRNGGAKKKNFFKLNKSEKDKKEKKPTVSFMSFRYSN
WLDKLYMVVGTAAIIHGAGLPLMMLVFGEMTDIFANAGNLEDL
MSNITNRSNDINDL.....LAQKGIFYFSMVSVQAGTKRQ
.....
  
```

Figure 3. The target sequence in .fasta format

The sample .ali format is shown below in figure 4.

```

>P1; P-gp
Sequence: P-gp::: 0.00:0.00
MDLEGDRNGGAKKKNFFKLNKSEKDKKEKKPTVSFMSFRYSN
WLDKLYMVVGTAAIIHGAGLPLMMLVFGEMTDIFANAGNLEDL
MSNITNRSNDINDT.....LLAQKGIFYFSMVSVQAGTKRQ*
.....
  
```

Figure 4. The target sequence in .ali format

Step IV: The modeling started using the modeler. For matching the patterns, alignment and finding the sequence BLAST (Basic Local Algorithm Search Tool) [19] was used. In this tool, the fasta format of target sequence was matched with PDB (Protein Data Bank) [19] by using

BLOSUM62 matrix [20]. The result was shown below in figure 5 and figure 6.

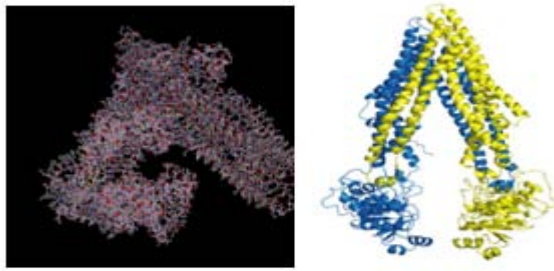


Figure 5. Structure of P-glycoprotein

P-gp

Query ID-|cl|77207

Description-Pgp

Molecule type-amino acid

Query Length-1280

Database Name-pdb

Description-PDB protein database

Program-BLASTP 2.2.26

Figure 6. BLAST aligns and matches the target sequence (or) query with PDB

Step V: Among the entire structure database the proper template structure was found. The alignment was done if at all required.

Step VI: After the template was found, select the appropriate protein or accession which was identified in between 40% -90%, for detailed study. Figure 7 shows the accession number 3NH6 protein was the required identity. Also, the 3D –structure of the protein was found. The 3NH6 protein which has 51% of identity was also taken as the template structure. Sample sequence of 3NH6 in .fasta format and the structure is as shown in figure 8.

Molecular Description of 3NH6:

Classification: Transport Protein

Structure Weight: 33644.26

Molecule: ATP-binding cassette sub-family B member 6, mitochondrial

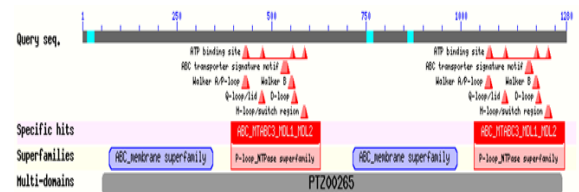
Polymer: 1 Type: protein Length: 306

Chains: A

Fragment: Nucleotide Binding Domain (UNP residues 558-842)

Organism Homo sapiens

UniProtKB: Q9NP58



Distribution of 17 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

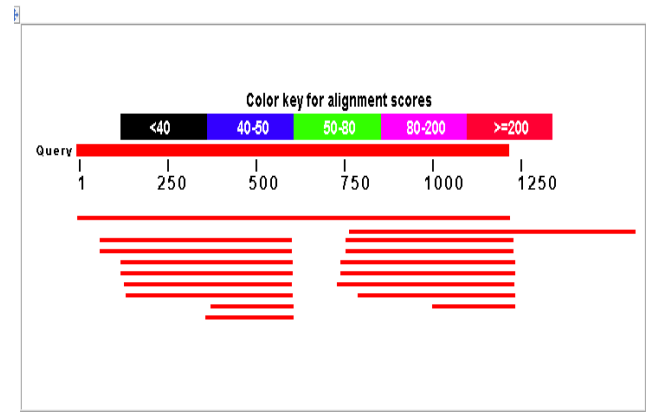


Figure 7. Structure of 3NH6Protein

```
>3NH6:A|PDBID|CHAIN|SEQUENCEMGSSHHHHHHSSGLVPRGSHM
FIDMENMFDLLKEETEVKDLPGAGPLRFQKGRIEFENVHFSYADGR
ETLQDVSTVQ
QLQQGQEETSEDTPQTMER
```

Figure 8. Fasta format of 3NH6 protein sequence

Step VII: The target protein sequence was aligned and matched with the protein structure database to find the proper identical protein sequence for designing the model and for a comparative study approach.

Step VIII: The *Bio-Perl* codes for the extraction of the protein sequence are stated below. Also, the alignment or sequencing the protein with the protein structure database and pattern matching of the target one with the template one is shown below. *Bio-Perl* algorithm for extracting the target protein sequence from the protein sequence databank: sequence and its regular expressions.

Input: protein file name, protein file, *P-gp* fragment, long protein.

Output: the sequence data of required protein: *P-gp*

Method: # Reading protein sequence data from a file.

The filename of the file containing the protein sequence data # \$proteinfilename =

Pgpfragment.pep;

PROTEINFILE for readability.

open(PROTEINFILE, \$proteinfilename);

Now we do the actual reading of the protein sequence from the

```
# file by using the angle brackets < and > to get
the input from the
# filehandle. We store the data into our variable
$protein.
#while ($protein = <PROTEINFILE>) {
# $longprotein. = $protein; }
# Now that we've got our data, we can close the
file.
#close PROTEINFILE;
# Print the protein onto the screen
#print "Here is the protein:\n\n";
#print $longprotein; #exit;
```

Bio-Perl algorithm for alignment and sequencing of *P-gp* with respect to protein structure database.

Input: sequence alignments, sequences, target query.

Output: sequence alignment of target one with the other protein structure database.

Method:

```
#a course on sequence alignments
#parameters: aln1 aln2
#matrix=idmat: 10 pour unmatch, -10 pour unmm
# read the two sequences from two fasta format
file:
#extract the names and the sequences
# get rid of the newlines, spaces and numbers
foreach $seq (@seq_list0)
foreach $seq (@seq_list1)
# get rid of the newlines, spaces and numbers
# split the sequences
for ($i=0; $i<=$#name_list0; $i++)
for ($i=0; $i<=$#name_list1; $i++)
if ($res0[0][$i-1] eq $res1[0][$j-1]){
$match;
}
else {
$mismatch ;
}
if ($sub>$del &&
$sub>$ins){
$mat[$i][$j]=$sub;$tb[$i][$j]=0;
}
elseif ($del>$ins){
$mat[$i][$j]=$del;$tb[$i][$j]=-1;
}
else {
$mat[$i][$j]=$ins;$tb[$i][$j]=1;
}
}
}
#print ">$name_list0 [0]\n";
for ($i=$aln_len-1; $i>=0; $i--){
print $aln0[$i];
}
#print "\n";
#print ">$name_list1 [0]\n";
for ($j=$aln_len-1; $j>=0; $j--){
print $aln1[$j];
}
#print "\n";
```

Bio-Perl algorithm for sequence pattern matching of *P-gp* and identical protein.

Input: sequence line, pattern, match count, filename.

Output: shows the matches.

Method:

```
#enter the search pattern
#$filename = $argv[0];
#$linenum = $matchcount = 0;
#print ("matches found:\n");
#while ($line = <>) {
if ($line =~ /$pattern/) {
filename and line number
$linenum = 0;
$filename = $argv[0];
```

```
if ($matchcount == 0) {
print ("no matches found.\n"); }
else {
total no. of matches }
```

VI. CONCLUSION

With the help of the modeler, the modeling of *P-glycoprotein* takes place. Here, *P-gp* which acts as a target sequence also represents as a query. The similar type of protein is also found using BLAST. All the alignment, matching and sequencing occurred with the help of the modeler and BLAST technique. All the processes had taken place by *Bio-Perl*. In our work, for a systematic comparative study, both the bioinformatics tool and *Bio-Perl* were used to predict the sequence and structure of the *P-gp*.

VII. REFERENCES

- [1] Christopher J. M., Matthew W., Lamb K., Brouwer L.R. and Pollack G.M. 2010. Pharmacokinetic and Pharmacodynamic Implications of *P-glycoprotein* Modulation. *Pharmacotherapy Publications*, 21, 7, 1-6.
- [2] Eswar N., John B., Mirkovic N., Fiser A., Ilyin V.A., Pieper U., Stuart A.C., Martí-Renom M.A., Madhusudhan M.S., Yerkovich B. and Šali A. 2008. Tools for Comparative Protein Structure Modeling and Analysis. *Nucleic Acids Research*, 31, 3375-3380.
- [3] Šali A. 2011. Modeler: A program for Protein Structure Modeling Release. *Modeler Manual*.
- [4] Ramachandran K.I., Deepa G. and Namboori K. 2008. Computational Chemistry and Molecular Modeling Principles and Applications. *Springer*, 21, 397, 111-121.
- [5] James. T. Mastering Perl for Bioinformatics. 2003. O'Reilly Pub.
- [6] Fu D., Arias I.M. 2011. Intracellular trafficking of *P-glycoprotein*. *Int. Journal of Biochem. and Cell Biology*, 44, 3, 461-464.
- [7] Schumacher U., Nehmann N., Adam E., Mukthar D., Slotki I.N., Horny H.P., Flens M.J., Schlegelberger B. and Steinemann D. 2011. MDR-1-overexpression in HT 29 colon cancer cells grown in SCID mice. *Acta Histochem.*
- [8] Alexandridou A., Dovrolis N., Tsangaris G., Nikita K. and Spyrou G. 2011. PepServe: A Web Server for peptide analysis, clustering and visualization. *Nucleic Acids Res.*, 39, 381-384.
- [9] <http://en.wikipedia.org/wiki/P-glycoprotein>.
- [10] Bartsevich V.V. and Juliano R.L. 2000. Molecular Pharmacology: Cancer multidrug resistance. *Nature Biotechnology*, 18, 18 – 20.
- [11] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Longman.
- [12] Liu B., Wang X., Lin L., Dong Q. and Wang X. 2008. A discriminative method for protein remote homology detection and fold recognition combining Top-n-grams and latent semantic analysis. *BMC Bioinformatics*, 9, 510, 1-16.

- [13] Martz E. 2001. Homology Modeling for Beginners. Molecular Visualization Workshop
- [14] http://en.wikipedia.org/wiki/Molecular_modeling
- [15] www.perl.org
- [16] <http://en.wikipedia.org/wiki/BioPerl>
- [17] <http://www.uniprot.org/uniprot/P08183>.
- [18] <http://www.rcsb.org/pdb/home/home.do>
- [19] <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- [20] <http://en.wikipedia.org/wiki/BLOSUM>