Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 1778 – 1782

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Modeling Optimization and Computing (ICMOC-2012)

Homology Modeling of *P-glycoprotein* for Detecting Remote Protein Homologies

Anurag Pal*, Debahuti Mishra, Shruti Mishra and Sandeep Kumar Satapathy

Institute of Technical Education and Research, Siksha O Anusandhan Deemed to be University, Bhubaneswar, Odisha, India

Abstract

Proteins are polymers of amino acids and an important class of biological macromolecules present in all organisms. P-glycoprotein (P-gp) is one of the xenobiotic transport proteins implicated in multidrug resistance in neoplastic tissues. It is a cell membrane-associated protein that transports a variety of drug substrates. It is present in organ systems that influence drug absorption (intestine), distribution to site of action (central nervous system and leukocytes), and elimination (liver and kidney), as well as several other 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. P-gp (ABCB1) appears to have developed as a mechanism to protect the body from harmful substances. Drug resistance is the major constraint for chemotherapeutic agents used for the treatment of neoplastic diseases. The prediction of proper protein sequence and structure of protein help in many ways to medical science and in the field of bio-computing. The modelling technique is used for detecting remote protein homologies. In this paper, P-gp has been taken as the target sequence. The protein has been processed under molecular modelling. The creation of mathematical models of molecular properties and behaviour is modelling. To know the proper molecular model of the protein, the target sequence was matched with protein structure database to find the templates. The model has been designed using the modeller which is the homology modelling process. The target sequence has been matched with protein structure database with the help of BLAST. The maximum identity accession has been found and the structure was analyzed. The target sequence acted as a query and aligned with template structure. This protein modelling and structure designs are important for structure drug design, to minimize the time complexity and also to make the clinical trial process easier.

Keywords: P-glycoprotein (P-gp); Chemotherapeutic Agents; Remote Protein Homologies; Modeller; Structure Templates; Sequence Alignment; Structure Drug Design

1. 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. P-gps are part of a larger super-family of efflux transporters found in the gut, gonads, kidneys, biliary system, brain and other organs named the ATP-binding cassette family (ABCs). 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

* Tel.: +91-9437668393; fax: 91-674-2351880.
E-mail address: anuragom2000@yahoo.com

1877-7058 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.06.217

(BBB) [1]. 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. During evolution, sequence changes much faster than structure. It is possible to identify the 3D-structure by looking at a molecule with some sequence identity. Modeler is a computer program that models three-dimensional structures of proteins and their assemblies by satisfaction of spatial restraints. For this protein modeling purpose, different types of modeler are used. Modeler 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 Modeler will automatically calculate a model with all non-hydrogen atoms [2]. Modeler can also perform multiple comparisons of protein sequences and/or structures, clustering of proteins, and searching of sequence databases. The layout of the paper is as follows: section 2 deals with some of the related work with P-gps and its structure prediction; section 3 deals with our proposed workflow model; section 4 gives our experimental analysis and section 4 provides us with our conclusion and future work.

2. Related Work

Ramachandran *et al*.[3] proposed the molecular modelling technique and also the homology modelling of protein which applied in computational chemistry and computational biology. The different molecular modelling principles also used for p-gp modelling. Arias *et al*.[4] proposed intracellular trafficking of P-gp, modelling 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*.[5] proposed MDR-1-overexpression in HT 29 colon cancer cells grown in SCID mice after modelling the P-gp structure.

3. Proposed Model

Fig.1 depicts the proposed model for the selection of the P-gp dataset. Then search the protein structure database, use that target sequence as a query, save that sequence in .ali or .pir format or in fasta format for modeling, match the pattern and align by BLAST, find out the proper template, after alignment find the 3D-structure of max.identity for further study.

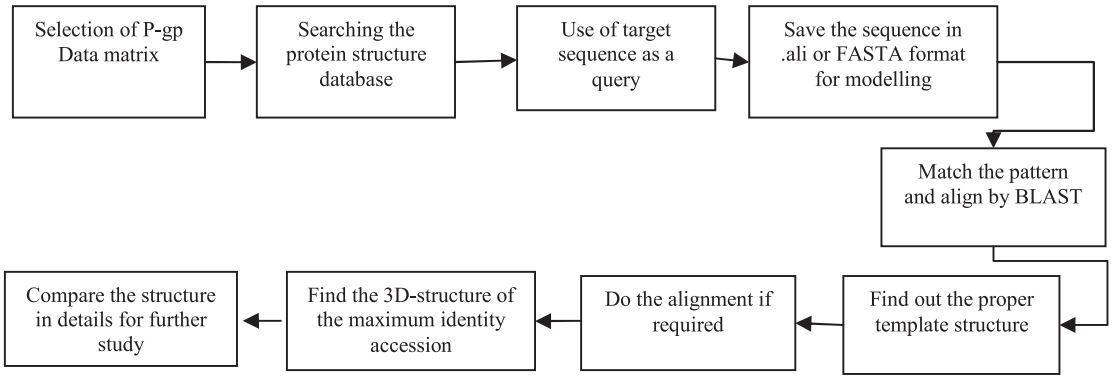


Fig. 1. Schematic representation of proposed model

4. Experimental Analysis and Evaluation

Step I: Initially the P-glycoprotein data set or sequence of the P-gp has been selected from [6]. This is the MDR1 or ABCB1 gene of human. P-gp has *n* unique id P08183. The sequence length of P-gp is1280 AA is shown in fig.2.

```
        10         20         30         40         50         60
MDLEGDRNGG AKKKNFFKLN NKSEKDKKEK KPTVSVFSMF RYSNWLDKLY MVVGTLAAII
        70         80         90        100        110        120
HGAGLPLMML VFGEMTDIFA NAGNLEDLMS NITNRSDIND TGFFMNLEED MTRYAYYYSG
       130        140        150        160        170        180
IGAGVLVAAY IQVSFWCLAA GRQIHKIRKQ FFHAIMRQEI GWFDVHDVGE LNTRLTDDVS
       190        200        210        220        230        240
KINEGIGDKI GMFFQSMATF FTGFIVGFTR GWKLTLVILA ISPVLGLSAA VWAKILSSFT
       250        260        270        280        290        300
DKELLAYAKA GAVAEEVLAA IRTVIAFGGQ KKELERYNKN LEEAKRIGIK KAITANISIG
       310        320        330        340        350        360
AAFLLIYASY ALAFWYGTTL VLSGEYSIGQ VLTVFFSVLI GAFSVGQASP SIEAFANARG
       370        380        390        400        410        420
AAYEIFKIID NKPSIDSYSK SGHKPDNIKG NLEFRNVHFS YPSRKEVKIL KGLNLKVQSG
       430        440        450        460        470        480
QTVALVGNSG CGKSTTVQLM QRLYDPTEGM VSVDGQDIRT INVRFLREII GVVSQEPVLF
       490        500        510        520        530        540
ATTIAENIRY GRENVTMDEI EKAVKEANAY DFIMKLPHKF DTLVGERGAQ LSGGQKQRIA
       550        560        570        580        590        600
IARALVRNPK ILLLDEATSA LDTESEAVVQ VALDKARKGR TTIVIAHRLS TVRNADVIAG
       610        620        630        640        650        660
FDDGVIVEKG NHDELMKEKG IYFKLVTMQT AGNEVELENA ADESKSEIDA LEMSSNDSRS
       670        680        690        700        710        720
SLIRKRSTRR SVRGSQAQDR KLSTKEALDE SIPPVSFWRI MKLNLTEWPY FVVGVFCAII
       730        740        750        760        770        780
NGGLQPAFAI IFSKIIGVFT RIDDPETKRQ NSNLFSLLFL ALGIISFITF FLQGFTFGKA
       790        800        810        820        830        840
GEILTKRLRY MVFRSMLRQD VSWFDDPKNT TGALTTRLAN DAAQVKGAIG SRLAVITQNI
       850        860        870        880        890        900
ANLGTGIIIS FIYGWQLTLL LLAIVPIIAI AGVVEMKMLS GQALKDKKEL EGSGKIATEA
       910        920        930        940        950        960
IENFRTVVSL TQEQKFEHMY AQSLQVPYRN SLRKAHIFGI TFSFTQAMMY FSYAGCFRFG
       970        980        990       1000       1010       1020
AYLVAHKLMS FEDVLLVFSA VVFGAMAVGQ VSSFAPDYAK AKISAAHIIM IIEKTPLIDS
      1030       1040       1050       1060       1070       1080
YSTEGLMPNT LEGNVTFGEV VFNYPTRPDI PVLQGLSLEV KKGQTLALVG SSGCGKSTVV
      1090       1100       1110       1120       1130       1140
QLLERFYDPL AGKVLLDGKE IKRLNVQWLR AHLGIVSQEP ILFDCSIAEN IAYGDNSRVV
      1150       1160       1170       1180       1190       1200
SQEEIVRAAK EANIHAFIES LPNKYSTKVG DKGTQLSGGQ KQRIAIARAL VRQPHILLLD
      1210       1220       1230       1240       1250       1260
EATSALDTES EKVVQEALDK AREGRTCIVI AHRLSTIQNA DLIVVFQNGR VKEHGTHQQL
      1270       1280
LAQKGIYFSM VSVQAGTKRQ
```

Fig.2.Complete P-glycoprotein data sequence

Step II: After getting the P-gp data sequence, the structure data base has been selected [6]. The structure data base is generally a composition of huge protein structure data base. From Protein Data Bank (PDB) [8], the structure database was prepared.

```
>sp|P08183|MDR1_HUMAN Multidrug resistance protein 1 OS=Homo sapiens GN=ABCB1 PE=1 SV=3
MDLEGDRNGGAKKKNFFKLNNKSEKDKKEKKPTVSVFSMFRYSNWLDKLYMVVGTLAAII
HGAGLPLMMLVFGEMTDIFANAGNLEDLMSNITNRSDINDTGFFMNLEEDMTRYAYYYSG
IGAGVLVAAYIQVSFWCLAAGRQIHKIRKQFFHAIMRQEIGWFDVHDVGELNTRLTDDVS
KINEGIGDKIGMFFQSMATFFTGFIVGFTRGWKLTLVILAISPVLGLSAAVWAKILSSFT
DKELLAYAKAGAVAEEVLAAIRTVIAFGGQKKELERYNKNLEEAKRIGIKKAITANISIG
AAFLLIYASYALAFWYGTTLVLSGEYSIGQVLTVFFSVLIGAFSVGQASPSIEAFANARG
AAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLEFRNVHFSYPSRKEVKILKGLNLKVQSG
QTVALVGNSGCGKSTTVQLMQRLYDPTEGMVSVDGQDIRTINVRFLREIIGVVSQEPVLF
ATTIAENIRYGRENVTMDEIEKAVKEANAYDFIMKLPHKFDTLVGERGAQLSGGQKQRIA
```

```
IARALVRNPKILLLDEATSALDTESEAVVQVALDKARKGRTTIVIAHRLSTVRNADVIAG
FDDGVIVEKGNHDELMKEKGIYFKLVTMQTAGNEVELENAADESKSEIDALEMSSNDSRS
SLIRKRSTRRSVRGSQAQDRKLSTKEALDESIPPVSFWRIMKLNLTEWPYFVVGVFCAII
NGGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSLLFLALGIISFITFFLQGFTFGKA
GEILTKRLRYMVFRSMLRQDVSWFDDPKNTTGALTTRLANDAAQVKGAIGSRLAVITQNI
ANLGTGIIISFIYGWQLTLLLLAIVPIIAIAGVVEMKMLSGQALKDKKELEGSGKIATEA
IENFRTVVSLTQEQKFEHMYAQSLQVPYRNSLRKAHIFGITFSFTQAMMYFSYAGCFRFG
AYLVAHKLMSFEDVLLVFSAVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLIDS
YSTEGLMPNTLEGNVTFGEVVFNYPTRPDIPVLQGLSLEVKKGQTLALVGSSGCGKSTVV
QLLERFYDPLAGKVLLDGKEIKRLNVQWLRAHLGIVSQEPILFDCSIAENIAYGDNSRVV
SQEEIVRAAKEANIHAFIESLPNKYSTKVGDKGTQLSGGQKQRIAIARALVRQPHILLLD
EATSALDTESEKVVQEALDKAREGRTCIVIAHRLSTIQNADLIVVFQNGRVKEHGTHQQL
LAQKGIYFSMVSVQAGTKRQ
```

Fig.3.The target sequence in FASTA format

Step IV: The modeling was done using the modeler. For matching the patterns and alignment, find the sequence for which we have used BLAST [8]. In this tool or technique, the FASTA sequence was matched with PDB [7] by using BLOSUM62 [9] matrix. The result is shown below in fig.4.

Query ID -lcl|11549
Description -sp|P08183|MDR1_HUMAN Multidrug resistance protein 1 OS=Homo sapiens GN=ABCB1 PE=1 SV=3
Molecule type -amino acid
Query Length-1280
Database Name-pdb
Description-PDB protein database
Program-BLASTP 2.2.26

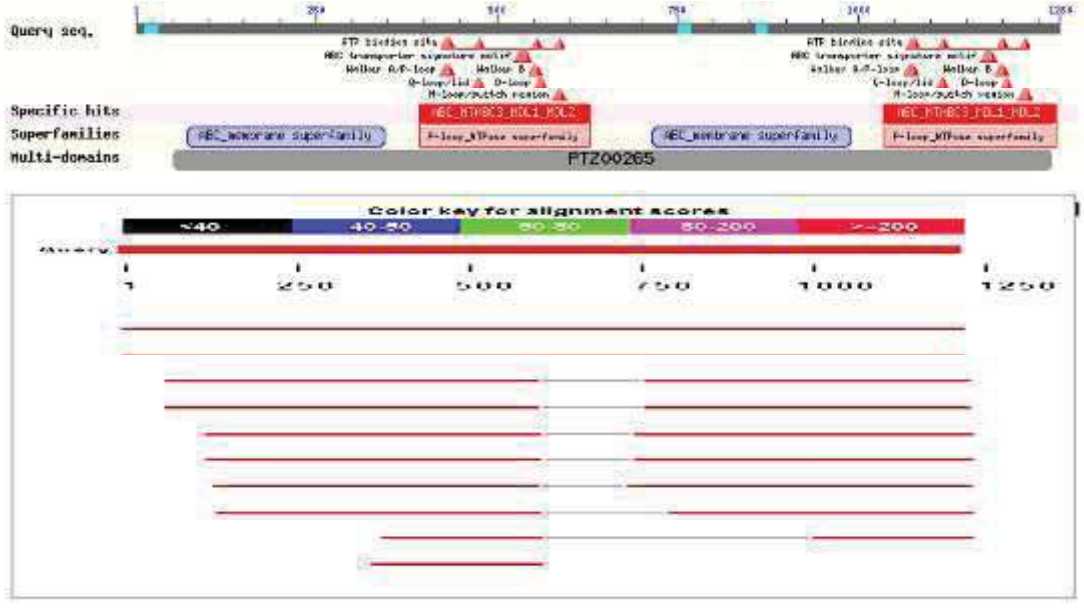


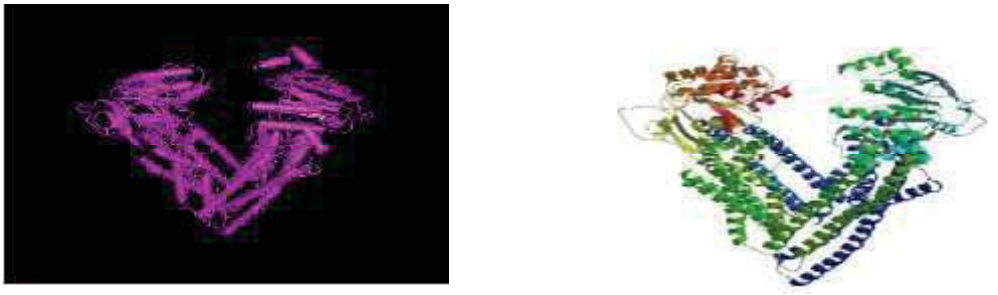
Fig.4. Using the BLAST to target sequence (or) query and PDB

Step V: Among all the structure data bases the proper template structure was found. If required alignment was done.

Step VI: After the template was found, then the appropriate protein or accession which was maximum identity has been selected for detailed study. Here, the accession no.3G60_A protein was the max. identity. Also the 3D –structure was found.

Step VII: The target protein structure was found with the help of modeller and the model has been designed. Both the structures were compared.

Step VIII: The template structure and target protein structure is shown in fig.5 (a) and fig. 5(b).



(a) (b)

Fig.5. (a) Structure of 3G60_A and (b) Approximate Structure of P-gp

5. Conclusion

With the help of the modeler the modeling of P-glycoprotein is done. Modeler is needed for molecular modeling of P-gp. Before predicting or designing any structure of protein, the homology modeling is to be done. Here, P-gp acts as a target and also represent as a query. The similar type of protein was found using BLAST. All the alignment and matching sequencing are done with the help of the modeler and BLAST technique. The required template also discovered. After the comparative study of the target one and template one; the modified model or design has been prepared. In further study, the structure of P-gp and P-gp like protein can be compared and modified. To get the better result and advance research work, the changes will also occur and design or model the 3D structure as per our requirement for better and effective solution.

References

[1] Christopher JM, Matthew WL, Kim LRB, Gary MP. Pharmacokinetic and Pharmacodynamic Implications of P-glycoprotein Modulation. *Pharmacotherapy Publications*; 2010; 21: 7:1-6.

[2] Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Martí-Renom MA, Madhusudhan MS, Yerkovich B and Šali A. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Research*; 2008; 31:3375-3380.

[3] Ramachandran KI, Deepa G, Namboori K. Computational Chemistry and Molecular Modeling Principles and Applications. *Springer*; 2008; ISBN 978-3-540-77302-3.

[4] Fu D, Arias IM. Intracellular trafficking of P-glycoprotein. *International Journal of Biochemistry and Cell Biology*; 2011; 42: 11: 1864-1875.

[5] Schumacher U, Nehmann N, Adam E, Mukthar D, Slotki IN, Horny HP, Flens MJ, Schlegelberger B, Steinemann D. MDR-1-overexpression in HT 29 colon cancer cells grown in SCID mice. *Acta Histochem;* 2011.

[6] *http://www.uniprot.org/uniprot/P08183.*

[7] *http://www.rcsb.org/pdb/home/home.do*

[8] *http://blast.ncbi.nlm.nih.gov/Blast.cgi*

[9] Styczynski MP, Jensen KL, Rigoutsos I, Stephanopoulos G. "BLOSUM62 miscalculations improve search performance. *Nat. Biotech.;* 2008; 26(3): 274-275