INTERNSHIP REPORT

Under the Supervision of Dr. Sukanta Mondal

BITS Pilani, Goa Campus

(01.12.2022 – 31.12.2022)

Computational Aspects of the Structural Biology of Transmembrane Protein

Submitted By:

Sanjay S. – 21P050020

M.Sc. Marine Biotechnology (2021-2023)

School of Biological Sciences and Biotechnology,

Goa University, Goa.

Birla Institute of Technology & Science, Pilani
K K Birla Goa Campus

Date: 04-Mar-2023

INTERNSHIP CERTIFICATE

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. Sanjay S., student of M.Sc. Marine Biotechnology, School of Biological Science and Biotechnology, Goa University, Goa, has successfully completed one month internship in the field of structural bioinformatics from 01-Dec-2022 to 31-Dec-2022 under the guidance of Sukanta Mondal, Ph.D., Associate Professor, Biological Sciences, BITS Pilani, K K Birla Goa campus, India.

As a student, Sanjay, is polite and soft-spoken. His approach is enthusiastic, sincere and perseverant. I wish his success in his life and career.

Sincerely,

Sukanta Mondal

Sukanta Mondal, Ph.D.
Associate Professor
Dept. of Biological Sciences
BITS Pilani – K.K. Birla Goa Campus

Birla Institute of Technology & Science, Pilani
K K Birla Goa Campus , Near NH 17B, Bypass Road
Zuarinagar 403726, Goa, India

Tel: +91 0832 2580149
E-mail: suku@goa.bits-pilani.ac.in
http://www.bits-pilani.ac.in/Goa/

Preface and Acknowledgement:

This internship report is a document summarizing my experience and learning during my internship at BITS Pilani Goa Campus, conducted from 01.12.2022 to 31.12.2022. During my internship, I was exposed to a wide range of activities and gained valuable insight into the field, computational structural biology.

BITS Pilani Goa Campus is a renowned institution for higher education and technologies. Located on the lush green hills of South Goa, the campus is well known for its focus on interdisciplinary research and education in the fields of Science, Technology, Engineering and Management. The research conducted by the students at BITS Pilani Goa Campus has resulted in a number of publications in world-renowned journals. It offers a wide range of courses and research opportunities, enabling students to gain the necessary knowledge and skills to pursue a career in this field.

I am grateful for the opportunity to learn and work in BITS Pilani Goa Campus under Dr. Sukanta Mondal on Computational Structural Biology, which allowed me to apply the knowledge and skills I have acquired through my academic studies.

This report aims to provide a comprehensive overview of my internship experience. It includes my internship objectives, the activities I undertook, and a description of the outcomes I achieved.

I would like to thank Department of Biotechnology (DBT), India for giving me this opportunity.

I would like to express my gratitude towards Dr. Sukanta Mondal of BITS Pilani Goa Campus for his immense support, guidance and for providing me with knowledge & motivation throughout the internship period. I would also like to thank PhD Scholar, Miss. Mohita Mahajan for her support on the same.

I am highly indebted to Dean Prof. Savita S. Kerkar and Vice-Dean Prof. Sanjeev C. Ghadi for the facilities provided to accomplish this internship. I would like to thank Dr. Meghanath Prabhu and Dr. Samantha Fernandes for their support and advices to get and complete internship in above said organization. I am extremely great full to my department staff members and friends who helped me in successful completion of this internship.

Sanjay S. - 21P050020

Table of Contents:

Introduction:

This internship report provides information about the work done by me as an intern in order to gain knowledge during the 4-week internship period. My internship work in computational structural biology was primarily concerned with comprehending and applying computational biology to the understanding of transmembrane proteins.

Computational structural biology is a field of study that utilizes software to computationally analyze the structure of biological molecules such as proteins, nucleic acids, and carbohydrates. Computational approaches can be used to predict the structure of a molecule from its sequence or to compare the structures of different molecules.

Computational structural biology also encompasses methods for predicting the binding sites of proteins, the energetics of protein-protein interactions, and the dynamics of proteins and nucleic acids.

Computational structural biology has become increasingly important in understanding transmembrane proteins. It can be used to identify potential drug targets, predict protein structures, and study protein-protein interactions. Additionally, computational approaches can be used to determine the three-dimensional structures of these proteins and how they interact with their environment.

By understanding the structure of transmembrane proteins, researchers can better understand their function and gain insight into how they affect the biological processes they are involved in.

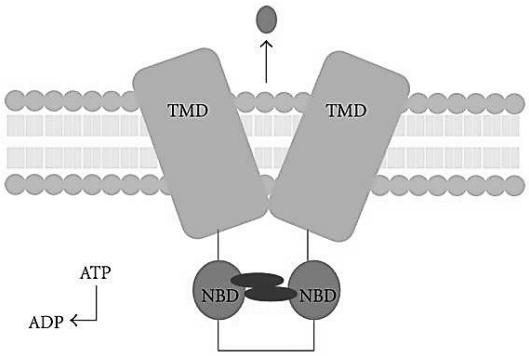
Computational methods are also used to analyze the dynamics of transmembrane proteins and identify possible conformational changes that may be involved in the function of the proteins. Finally, computational approaches can be used to evaluate the thermodynamic and kinetic properties of transmembrane proteins and predict their stability and activity.

Instead of taking a linear approach to the learning process, my internship mentor Dr. Sukanta Mondal wanted me to learn and work in a comprehensive way. He wanted me to learn and implement various computational techniques with a single subject protein in order to build a comprehensive understanding of transmembrane protein rather than just learning about different computational techniques with random examples.

Objective 1:

My first objective is to thoroughly educate myself on transmembrane proteins, their biological functions, and their structural importance. Human MRP1, an ABC type transport membrane protein, was the focus of my study. Therefore, my first task is to review them. The very next task is to do a literature study of the different computational approaches and the related software and databases that have been presented for use in research on transmembrane proteins like P-glycoprotein and other MRP proteins like MRP2, MRP4, etc.

ABC (ATP-Binding Cassette) type transport systems are membrane-associated proteins found in the cells of all living organisms. These proteins are involved in the movement of a variety of molecules across the cell membrane. The ABC transport system is a group of proteins that are responsible for the active transport of molecules, such as ions, sugars, and amino acids, from one side of the cell membrane to the other.



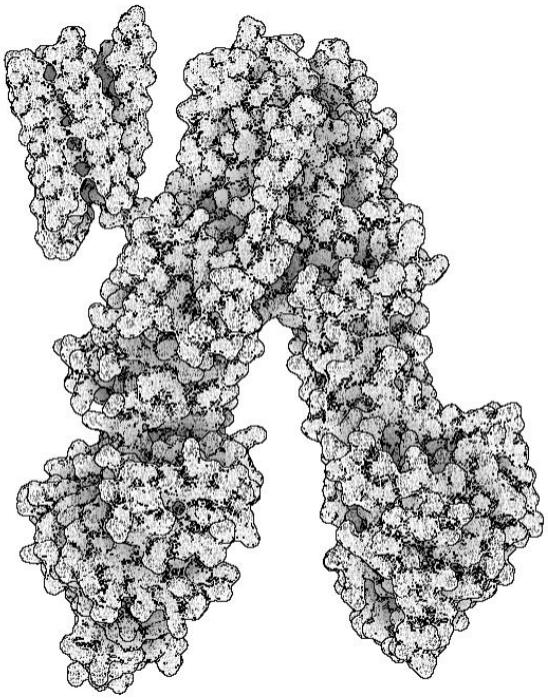
The ABC type transport proteins are composed of two main domains, namely the Nucleotide-binding protein and the Transmembrane domains. The ATP binds to the Nucleotide-binding domain and responsible for providing the energy required for the transport to occur, while the transmembrane domains form the "gate" through which the molecules pass.

The ABC transport protein system is essential for the efficient functioning of a cell, as it helps regulate the concentration of various molecules in the cell. Without the ABC transport system, the concentration of molecules inside and outside the cell would be out of balance, leading to a variety of cellular malfunctions. The ABC type transport protein system is also involved in the movement of toxins and other harmful substances out of the cell and into the extracellular environment. This is essential for the maintenance of homeostasis and for the prevention of cell death due to toxic overload.

In addition, the ABC type transport protein system plays a role in the regulation of gene expression. By controlling the movement of molecules into and out of the cell, the ABC transport system can affect the expression of certain genes, leading to either increased or decreased production of proteins.

MRP1 (Multidrug Resistance-associated Protein 1) is a transmembrane protein that is involved in the transport of a wide variety of molecules, including drugs and their metabolites, across the plasma membrane of cells. MRP1 is found in the brush border membrane of the kidney, liver, and intestine, where it mediates the active transport of drugs and other molecules, such as glutathione conjugates and certain metabolites, across the membrane.

MRP1 belongs to the ATP-binding cassette (ABC) transporter superfamily, which consists of proteins that transport molecules across membranes by means of ATP binding and hydrolysis. The MRP1 protein contains an N-terminal (Transmembrane domain), a highly conserved ABC transporter domain, and a C-terminal (Nucleotide Binding domain). The N-terminal Transmembrane domain is composed of 12 transmembrane helices, which form a ring-like structure in the membrane. The highly conserved ABC transporter domain contains two ATP-binding sites, an ATP-hydrolyzing site, and a regulatory domain. The C-terminal hydrophilic domain is responsible for binding the substrate molecule and facilitating its transport across the membrane.



MRP1 plays an important role in drug resistance. It has been shown to be involved in the active efflux of a variety of drugs, including some chemotherapy drugs, antiviral drugs, and antibiotics, from cells, thus conferring resistance to those drugs. In addition, MRP1 has been implicated in the development of multidrug resistance in cancer cells.

The expression of MRP1 is regulated by multiple mechanisms, including transcriptional and post-transcriptional regulation. Transcriptional regulation of MRP1 is mediated by several transcription factors, including STAT3, NF-kB, and HIF-1α.

Post-transcriptional regulation of MRP1 is mediated by a variety of factors, including miRNAs and small molecules, such as glucocorticoids. In conclusion, MRP1 is an important transmembrane protein that is involved in the transport of a wide variety of molecules, including drugs and their metabolites, across the plasma membrane of cells. The expression of MRP1 is regulated by multiple mechanisms, including transcriptional and post-transcriptional regulation.

The theoretical computation and computer simulation of the characteristics and behavior of molecules is known as molecular modelling. The modelled system may merely have one ligand or protein, but it may also contain protein-protein complexes, protein-ligand complexes, or a sizable complex with membrane structures that contain proteins. Homology Modeling of protein, Molecular Docking studies, Molecular Dynamic studies, Quantum Mechanics, QSAR and Pharmacophore Modeling are the part of Molecular Modeling.

In order to create 3D structures of receptors and to better understand drug-receptor interactions, homology modelling research and molecular docking simulations are essential tools. Molecular docking is a computational method used to predict the preferred orientation of one molecule to a second when bound to each other to form a stable complex. It is used to study protein-ligand binding and protein-protein interactions, and can help researchers on studying the molecular interactions. Molecular docking involves using a computer algorithm to calculate the relative energies of different orientations of the two molecules, and then selecting the configuration that has the lowest free energy. The method can also be used to predict the binding affinity of a ligand for a target protein, as well as its binding mode.

In computational structural biology, studies of quantum mechanics are widely employed to prepare docking structures with geometric optimization. To determine non-covalent interactions in ligand-receptor interactions and to do free energy estimates in bigger systems, new techniques have been devised.

The energy calculations such as Homo-Lomo calculations, partial charge calculations for the structure optimization are done through different methods, namely:

- DFT – Density Functional Theory.
- HT – Hartree Fock Theory.
- Semiempirical quantum model PM6.
- The GFN2-xTB (G: Geometries; F: Frequencies; N: Noncovalent interactions) tight-binding based electronic structure method.

The GFN2-xTB method was more accurate to experimental methods compared to MD simulations and MMPB(GB)SA calculations. The QM study is confined to 1000 atoms which is its major drawback when comes to transmembrane proteins.

The following is a list of various computational methods and related software that were studied as part of the first objective.

Database:

Databases are collections of data related to biological processes and molecules, such as genes, proteins, DNA sequences, and drug information. They are used to store, organize, and analyze large amounts of biological data. Examples of bioinformatics databases include the National Center for Biotechnology Information (NCBI) GenBank, UniProt, the Protein Data Bank (PDB), and the Human Genome Project.

Here I have listed major databases focused on structural biology.

Sl. No	**DATABASE TITLE**	**REMARKS**
01.	RCSB	Protein structural data bank.
02.	Binding DB	Contains 8,623 protein targets, and 1,015,369 small molecules.
03.	Drug Bank	about drug and drug products.
04.	Protein Data Bank (PDB)	Contains models of protein structures generated from X-ray crystallography, NMR spectroscopy, and other methods.
05.	Macromolecular Structure Database (MSD)	Contains structures of large complex macromolecules.
06.	Structural Genomics Database (SGD)	Aims to provide information about the structure and function of gene products.
07.	Structural Classification of Proteins (SCOP)	A hierarchical classification of protein structures, based on their evolutionary and structural relationships.
08.	DUD-E	about Docking Decoys.
09.	VARIDT 1.0	Protein database especially for ABC transport protein.
10.	CHARMM-GUI	for membrane embedded dynamic studies.

Protein Modeling and Visualization:

Protein modeling is the process of using computer software to create a three-dimensional representation of a protein molecule. The aim is to replicate the physical structure of the protein as closely as possible, to create an accurate model of the protein. Protein modeling can be used for a variety of purposes, including designing drugs that target the protein, predicting the structure of proteins based on their amino acid sequence, and studying the interactions between proteins and other molecules. Protein visualization tools are computer programs used to create an image of a protein's structure. They enable scientists to analyze a protein's shape and size, as well as its 3D conformation. By visually inspecting protein structures, researchers can gain insights into how the protein might interact with other molecules, its role in a biological system, and its susceptibility to disease-causing mutations.

Here I have listed major computational tools available for Protein Modeling and Visualization.

Sl. No	**TITLE**	**REMARKS**
01.	UCSF Chimera	A widely used software for protein visualization, with support for volumetric data and molecular surface representation.
02.	MolMol	An open-source molecular visualization program, with support for 3D structure viewing and surface representation.
03.	PyMOL	An open-source molecular graphics system, with support for 3D structure viewing and ray-tracing.
04.	VMD	A molecular visualization program, with support for 3D structure viewing, molecular dynamics, and surface representation.
05.	RasMol	An open-source molecular visualization program, with support for 3D structure viewing and surface representations.
06.	Jmol	An open-source molecular visualization program, with support for 3D structure viewing and ray-tracing.
07.	YASARA	An open-source molecular visualization program, with support for 3D structure viewing and molecular dynamics.

Homology Modeling:

Homology modeling is a powerful tool for predicting a protein's 3-dimensional structure based on the amino acid sequence. It is an important tool in modern biology and is used in a variety of applications such as drug design, functional annotation, and biophysical studies. In homology modeling, a 3-dimensional structure of a protein is predicted by comparing its amino acid sequence to that of a known protein structure. The known protein's structure is used as a reference and its amino acid sequence is aligned to that of the target protein. This alignment is then used to generate a 3-dimensional model of the target protein.

Homology modeling is important because it can be used to predict the structure of a protein of unknown structure. This is important for understanding the function of the protein, as the structure often dictates its function. It can also be used to design drugs that target specific proteins, as well as to predict the effects of mutations on a protein's structure and function. Homology modeling is not perfect, though. It relies on the assumption that the known protein structure is a good reference for the target protein, and this is not always the case. Additionally, the accuracy of the model decreases as the level of sequence similarity between the target and reference proteins decreases. Nevertheless, it is an invaluable tool in modern biology and has enabled us to gain a better understanding of the structure and function of proteins.

Here I have listed major computational tools available for Homology Modeling.

Sl. No	**TITLE**	**REMARKS**
01.	Maestro 11	Software Package.
02.	Discovery Studio	Software Package.
03.	Alpha fold	AI based.
04.	Swiss Model	Webserver based.
05.	Itasser	Webserver based.

Molecular Docking:

Molecular docking is a computational method used to predict the preferred orientation of one molecule to another when bound to each other to form a stable complex. It is used to predict the binding affinity of a molecule to a receptor and can be used to determine which molecules may bind to a target receptor and how they may interact with it. Molecular docking is an important tool in drug design as it can be used to identify potential drug candidates and to determine their binding affinities to a target receptor. It can also be used to study the protein-ligand interactions and to gain insights into the mechanism of action of a drug. Additionally, molecular docking can be used to study the biological activities of a given ligand and to identify novel drug targets.

Here I have listed major computational tools available for Molecular Docking.

Sl. No	**TITLE**	**REMARKS**
01.	Autodock	Accurate, fast, and easy to use; can be used to dock ligands against a wide range of protein structures.
02.	MolDock	It was designed with the goal of providing an efficient and accurate method for predicting the binding of small molecules to macromolecules.
03.	Molegro Virtual Docker	offers high-quality docking based on a novel optimization technique combined with a user interface experience focusing on usability and productivity.
04.	Surflex dock module	historically focused on re-docking the cognate ligand of a well-determined protein-ligand complex to measure.
05.	PatchDock	Geometry-based molecular docking algorithm. It is aimed at finding docking transformations that yield good molecular shape complementarity.
06.	HadDock	High Ambiguity Driven protein-protein docking is an information-driven flexible docking approach
07.	SwissDock	SwissDock is based on the protein-ligand docking program EADock DSS and has a simple and integrated interface.

Molecular Dynamic Simulations:

Protein molecular dynamic simulations are a powerful tool in understanding and predicting the behavior of proteins. These simulations can be used to study a variety of biological processes and interactions, including protein folding, protein-protein interactions, and protein-ligand interactions. The use of molecular dynamic simulations has become increasingly important in the study of proteins, as they provide a detailed look at a protein's behavior in a given environment. By understanding the behavior of a protein, researchers can better predict how it will interact with other molecules and how its activity may be altered. This knowledge can then be used to design more effective drugs and therapies.

Here I have listed major computational tools available for Molecular Dynamics.

Sl. No	**TITLE**	**REMARKS**
01.	CHARMM-GUI	Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations.
02.	MDWeb	personal workspace providing standard protocols to prepare structures, run standard molecular dynamics simulations and to analyse trajectories.
03.	NAMD	It is noted for its parallel efficiency and is often used to simulate large systems.
04.	AMBER	Assisted Model Building with Energy Refinement (AMBER) is a family of force fields for molecular dynamics of biomolecules
05.	GROMACS	One of the most widely used open-source and free software codes in chemistry, used primarily for dynamical simulations of biomolecules. It provides a rich set of calculation types, preparation and analysis tools. Several advanced techniques for free-energy calculations are supported.

Objective: 2

My second objective was to build a homology model of the human MRP1 transmembrane protein. The absence of the experimental structure of the human MRP1 necessitates the development of a homology model for the respective. Transmembrane proteins are embedded in the cell membrane and can be difficult to study experimentally because they are insoluble and difficult to extract in their native environment. Additionally, the techniques used to determine protein structure, such as X-ray crystallography and nuclear magnetic resonance spectroscopy, require large, well-ordered protein crystals, which are difficult to achieve with transmembrane proteins due to their highly hydrophobic nature.

The experimental structures of bovine MRP1 are available in the database, Protein Data Bank (PDB IDs: 5UJ9, 5UJA, 6UYO, and 6BHU). As a result, the plan is to construct a homology model of Human MRP1 using available experimental structures. As previously stated, a good predicted model structure can resemble the template structure to some extent and can be used for studies.

Transmembrane proteins have a challenging topology: they are embedded in the membrane, frequently have an extended conformation, and can have large extracellular and intracellular domains. This makes homology model generation for transmembrane proteins more difficult and time-consuming than for normal proteins or enzymes. Furthermore, the stability of the protein is greatly influenced by the membrane environment, and it is challenging to anticipate how this environment will affect the protein. Finally, it can be challenging to find appropriate template proteins because the amino acid sequence of transmembrane proteins is frequently more variable than that of other proteins.

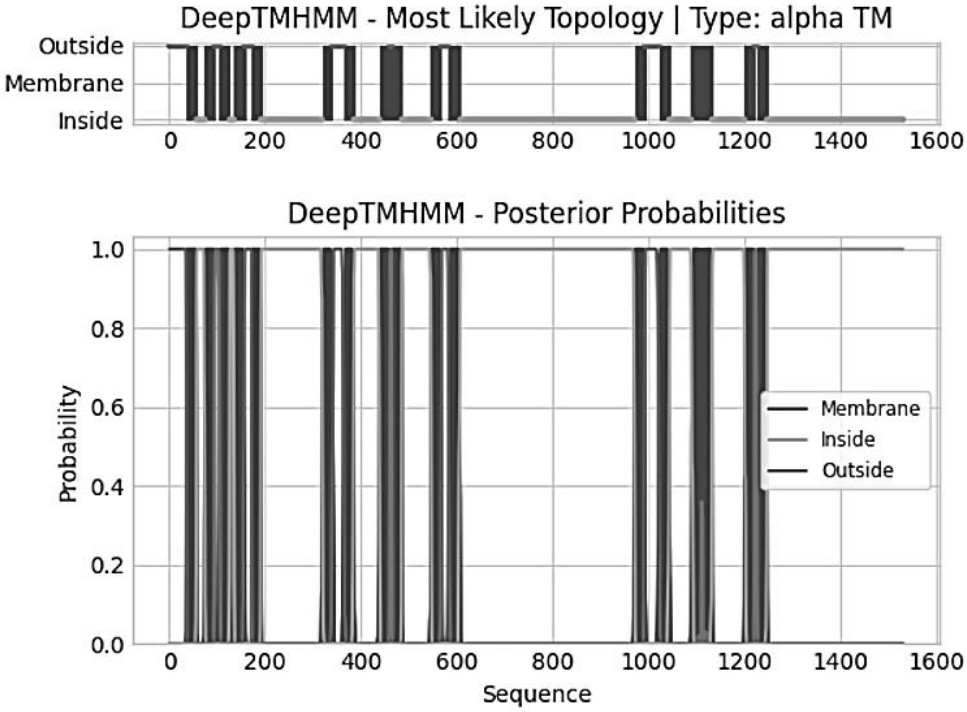
Step 1: Retrieval of target protein sequence for homology model generation.

The sequence of the query protein (hMRP1) was retrieved from the UniProt database with ID P33527, having 1531 amino acids.

```
>sp|P33527|MRP1_HUMAN Multidrug resistance-associated protein 1
MALRGFCSADGSDPLWDWNVTWNTSNPDFTKCFQNTVLVWVPCFYLWACFPFYFLYLSRH
DRGYIQMTPLNKTKTALGFLLWIVCWADLFYSFWERSRGIFLAPVFLVSPTLLGITMLLA
TFLIQLERRKGVQSSGIMLTFWLVALVCALAILRSKIMTALKEDAQVDLFRDITFYVYFS
LLLIQLVLSCFSDRSPLFSETIHDPNPCPESSASFLSRITFWWITGLIVRGYRQPLEGSD
LWSLNKEDTSEQVVPVLVKNWKKECAKTRKQPVKVVYSSKDPAQPKESSKVDANEEVEAL
IVKSPQKEWNPSLFKVLYKTFGPYFLMSFFFKAIHDLMMFSGPQILKLLIKFVNDTKAPD
WQGYFYTVLLFVTACLQTLVLHQYFHICFVSGMRIKTAVIGAVYRKALVITNSARKSSTV
GEIVNLMSVDAQRFMDLATYINMIWSAPLQVILALYLLWLNLGPSVLAGVAVMVLMVPVN
AVMAMKTKTYQVAHMKSKDNRIKLMNEILNGIKVLKLYAWELAFKDKVLAIRQEELKVLK
KSAYLSAVGTFTWVCTPFLVALCTFAVYVTIDENNILDAQTAFVSLALFNILRFPLNILP
MVISSIVQASVSLKRLRIFLSHEELEPDSIERRPVKDGGGTNSITVRNATFTWARSDPPT
LNGITFSIPEGALVAVVGQVGCGKSSLLSALLAEMDKVEGHVAIKGSVAYVPQQAWIQND
SLRENILFGCQLEEPYYRSVIQACALLPDLEILPSGDRTEIGEKGVNLSGGQKQRVSLAR
AVYSNADIYLFDDPLSAVDAHVGKHIFENVIGPKGMLKNKTRILVTHSMSYLPQVDVIIV
MSGGKISEMGSYQELLARDGAFAEFLRTYASTEQEQDAEENGVTGVSGPGKEAKQMENGM
LVTDSAGKQLQRQLSSSSSYSGDISRHHNSTAELQKAEAKKEETWKLMEADKAQTGQVKL
SVYWDYMKAIGLFISFLSIFLFMCNHVSALASNYWLSLWTDDPIVNGTQEHTKVRLSVYG
ALGISQGIAVFGYSMAVSIGGILASRCLHVDLLHSILRSPMSFFERTPSGNLVNRFSKEL
DTVDSMIPEVIKMFMGSLFNVIGACIVILLATPIAAIIIPPLGLIYFFVQRFYVASSRQL
KRLESVSRSPVYSHFNETLLGVSVIRAFEEQERFIHQSDLKVDENQKAYYPSIVANRWLA
VRLECVGNCIVLFAALFAVISRHSLSAGLVGLSVSYSLQVTTYLNWLVRMSSEMETNIVA
VERLKEYSETEKEAPWQIQETAPPSSWPQVGRVEFRNYCLRYREDLDFVLRHINVTINGG
EKVGIVGRTGAGKSSLTLGLFRINESAEGEIIIDGINIAKIGLHDLRFKITIIPQDPVLF
SGSLRMNLDPFSQYSDEEVWTSLELAHLKDFVSALPDKLDHECAEGGENLSVGQRQLVCL
ARALLRKTKILVLDEATAAVDLETDDLIQSTIRTQFEDCTVLTIAHRLNTIMDYTRVIVL
DKGEIQEYGAPSDLLQQRGLFYSMAKDAGLV
```

Step 2: Transmembrane region/topology prediction.

The transmembrane topology of the protein was predicted using the webserver "DeepTMHMM: A Deep Learning Model for Transmembrane Topology Prediction and Classification" for the query protein using the retrieved protein sequence.



The number of transmembrane regions predicted is 17, and the regions were also depicted for the reference. which can be used to compare and correlate once a homology model is generated.

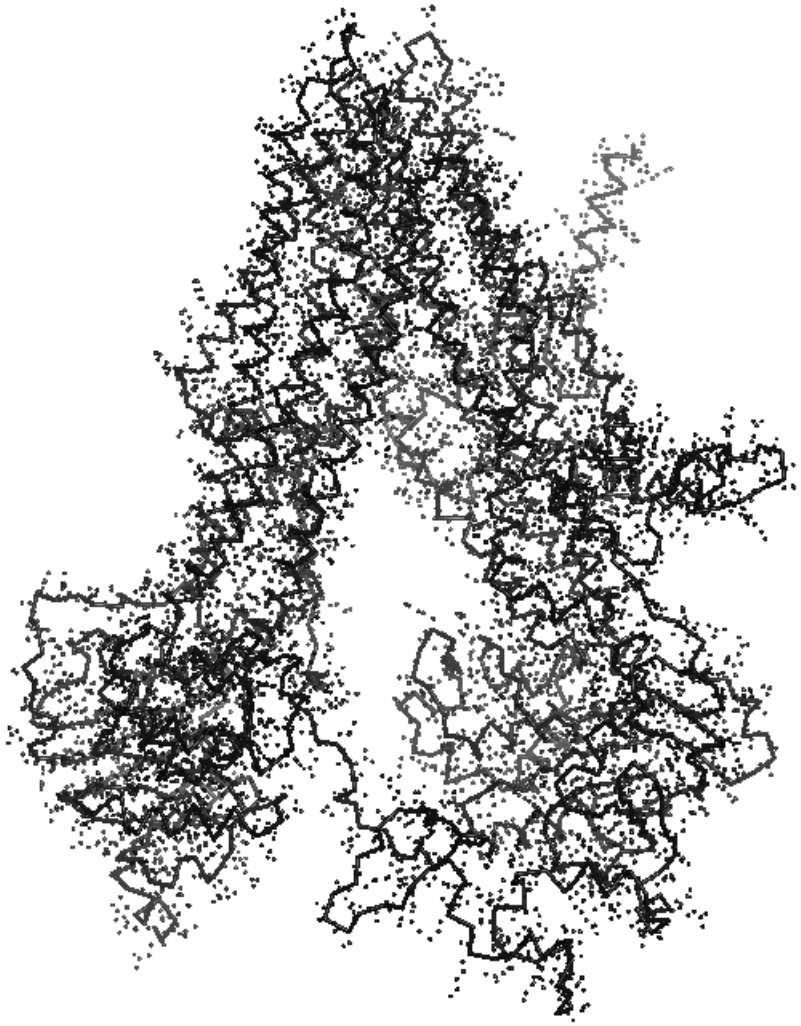
Step 3: Template selection (Experimental Structure) for Model Generation.

The existence of the experimental structures in the RCSB was determined through manual search and BLAST (Protein Data Bank). Four experimental structures were found to be predominant within MRP1 itself, namely 5UJ9, 5UJA, 6BHU, and 6UYO. Only the 5UJ9 Apo-form existed among the four structures mentioned above; the others were bound to ATP (6BHU,6UYO) and substrate (5UJA). Because the idea was primarily to generate the Apo-form, the experimental structure that already existed in apo-form, 5UJ9, was chosen as the template for homology model generation.

Step 4: Homology Model Generation.

The homology model for the query protein was generated using the Swiss Model web server, as it is one of the most commonly used and easily available tools for building homology models. The retrieved sequence is loaded either by manual copy-paste or as a FASTA file format in the input. Then it displays the available templates, and by selecting a template, the models are generated and can be viewed in the server.

The above-mentioned steps were carried out for the query sequence using the 5UJ9 experimental structure available as a template, and the model was generated and saved for further assessment and study.



Step 5: Homology Model Validation.

The purpose of homology model validation is to get an idea of the degree of correctness of the model generated. Usually, the primitive model generated has to undergo model refinement to get a better model. By comparing the validation parameters of the primitive and refinement models, the degree of refinement can be determined. But the energy refinement techniques were not done due to the reason that the implicit and explicit solvent parameters of the software algorithm are not reasonable to be used for transmembrane proteins. They need specific dynamic simulation parameters to resemble the protein in approximation. Therefore, the validation was carried out with the primitive model itself.

Necessary validation parameters such as ERRAT SCORE, VERIFY 3D SCORE, PROVE, PROCHECK, MOLPROBITY and QMEAN were analyzed.

- ERRAT - ERRAT (Error Rate Assessment Tool) is a software tool used to assess the accuracy of a homology model of a protein. It uses statistical methods to determine the reliability of the model, and compares it to a set of known structures.

  ERRAT can be used to evaluate the quality of a homology model, and to identify regions of the protein that may need additional refinement or modification.
  - The ERRAT score calculated for the model generated was 96.724, which passes the parameter criteria (Greater than 95% are considered good).

- VERIFY 3D - Verify3D is a method used in homology model validation that compares the predicted 3D structure of a protein to its known structure. It assesses the similarity between the predicted and known structures at both the local and global levels.

  It uses a Z-score to compare the predicted structure to a database of structurally similar proteins and then calculates an overall score to assess the accuracy of the model. This score can then be used to assess the quality of the homology model.
  - The VERIFY3D score calculated for the model generated had less than 80% (70.1) of its residue with <=0.1, and the model failed the parameter criteria.

- PROVE - PROVE stands for Protein Validation by Evolutionary Relationship. PROVE evaluates the similarity of the amino acid sequences of two protein structures, and compares the evolutionary relationships of the sequences to the predicted structure of the target protein. If the sequences are highly similar, and the structure of the target protein is similar to that of the homology model, then the model is deemed to be valid.
  - The PROVE score calculated for the model generated was 0.82, which passes the parameter criteria (Should be less than 1).

- PROCHECK - PROCHECK is a program used to validate homology models. It evaluates the stereochemical quality of protein structures by calculating a number of parameters and plotting them in a Ramachandran plot. The program also calculates a number of other parameters such as the overall energy of the structure, its secondary structure composition, and the root-mean-square deviation from the native structure. PROCHECK also checks for any possible clashes between atoms, as well as any possible gaps or clashes between the main-chain and side-chain atoms.

  - The PROCHECK results were not that great for the model. Only 89.1% of the residues were in the favored regions in the Ramachandran Plot.

```
 +----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
  |                                                                            |
  | /var/www/SAVES/Jobs/1269261/saves.pdb   1.5                 1328 residues |
  |                                                                            |
*| Ramachandran plot:   89.1% core    9.9% allow    0.8% gener    0.3% disall |
  |                                                                            |
*| All Ramachandrans:   41 labelled residues (out of1326)                     |
+| Chi1-chi2 plots:      4 labelled residues (out of 790)                     |
  | Side-chain params:    5 better     0 inside      0 worse                   |
  |                                                                            |
*| Residue properties: Max.deviation:     4.7             Bad contacts:    0 |
*|                     Bond len/angle:    8.2    Morris et al class:  1  1  2 |
  |                                                                            |
  | G-factors           Dihedrals:  -0.09  Covalent:  -0.03    Overall:  -0.05 |
  |                                                                            |
+| Planar groups:    95.3% within limits   4.7% highlighted       7 off graph |
  |                                                                            |
 +----------------------------------------------------------------------------+
```

- MOLPROBITY - MolProbity is a web-based program used for validating and improving homology models. It utilizes a combination of biochemical and physical methods to evaluate the quality of a protein structure. It provides an assessment of the stereochemical quality of the model, including bond lengths and angles, and clashes between atoms, as well as the overall energy of the structure. It also identifies potential problems such as poorly defined regions, poorly packed sidechains and incorrect protonation states. MolProbity can help to identify errors in the model and suggest ways to improve it, such as through manual improvements or running additional simulations.
  - The MOLPROBITY results was not that great depicting model has to be refined to increase the degree of correctness.

Protein Geometry	Poor rotamers	2	0.17%	Goal: <0.3%
	Favored rotamers	1126	97.49%	Goal: >98%
	Ramachandran outliers	15	1.13%	Goal: <0.05%
	Ramachandran favored	1237	93.29%	Goal: >98%
	Rama distribution Z-score	-0.47 ± 0.21		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	19	1.52%	Goal: 0
	Bad bonds:	3 / 10641	0.03%	Goal: 0%
	Bad angles:	95 / 14434	0.66%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 51	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	50	3.8%	Goal: <1.0%
	CA Geometry outliers	13	0.98%	Goal: <0.5%
Additional validations	Tetrahedral geometry outliers	2		

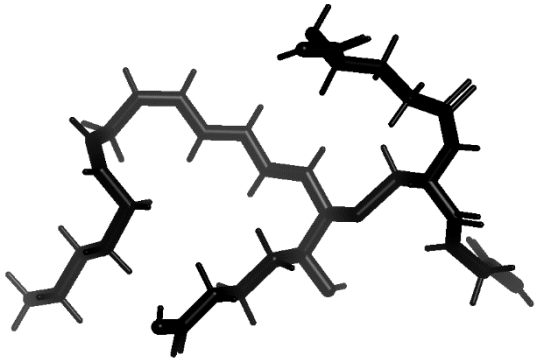
- The RMSD (Root Mean Square Deviation) was calculated by aligning both the homology model generated and experimental structure (5UJ9) to know about degree of structural difference.

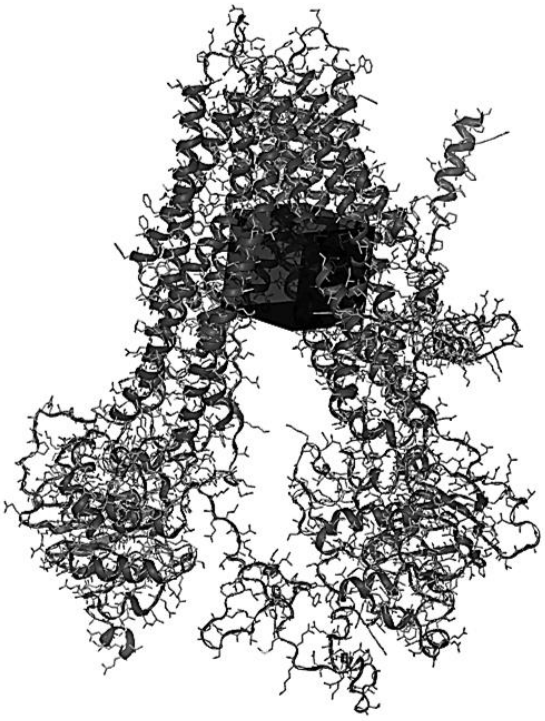
- The RMSD calculated was 0.058 which is closer to Zero therefore depicts the structural reliability.

- The structure is ‘Mesh’ form is the 5UJ9 and the one in ‘Ribbon’ form is the Homology Model generated.

Objective: 3

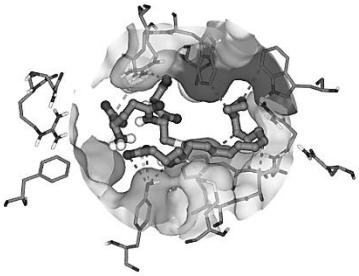
The third objective is to perform docking studies with the model protein. Leukotriene C4, which is one of the natural substrates for the transmembrane protein, was selected as the ligand for docking. The docking was carried out in the Patch Dock webserver, and the results were interpreted. The top 10 hits with the highest docking score were selected for analysis. Out of the top 10, the pocket resembling the pocket of the experimental structure was selected, and the active site prediction was carried out and compared.





- The above are the representation of 3d structure of Leukotriene C4 molecule followed by the pocket considered for docking of the ligand to protein molecule.

Docking was carried out and the active site residues involving with the ligand was estimated.



- The active residues estimated were = L-389, Y-440, W-553, F-594, M-601, R-1196, Y-1242, W-1245, R-1249.

Sl. No	5UJA	DOCKED MODEL
01.	H-335	-
02.	L-389	L-389
03.	Y-440	Y-440
04.	W-553	W-553
05.	F-594	F-594
06.	M-601	M-601
07.	R-1196	R-1196
08.	Y-1242	Y-1242
09.	N-1244	-
10.	W-1245	W-1245
11.	R-1249	R-1249

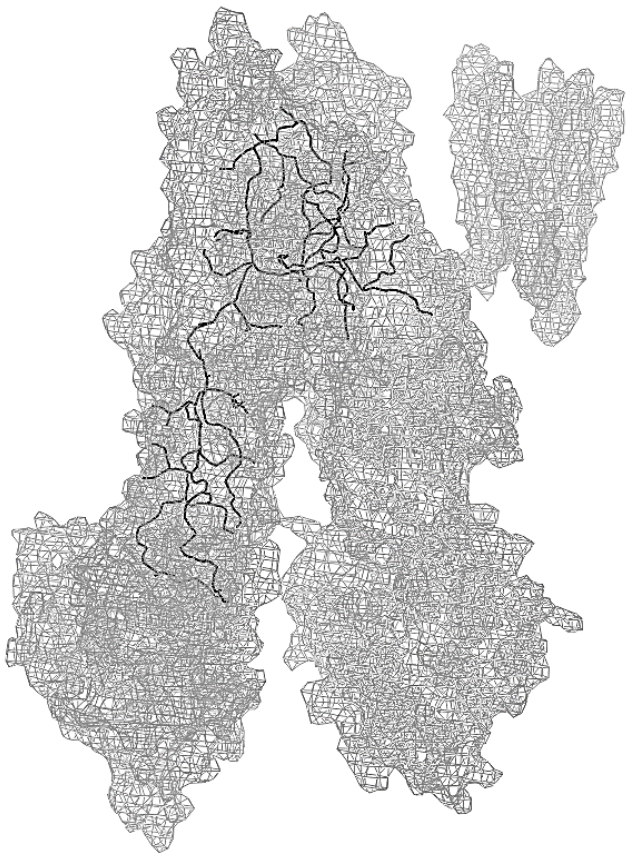
Almost all the residues that were predicted as active site residues for the docked structure were some of the predicted active site residues of the experimental structure 5UJA. This depicts the degree of correctness of the docking is relatively good and reliable.

Objective: 4

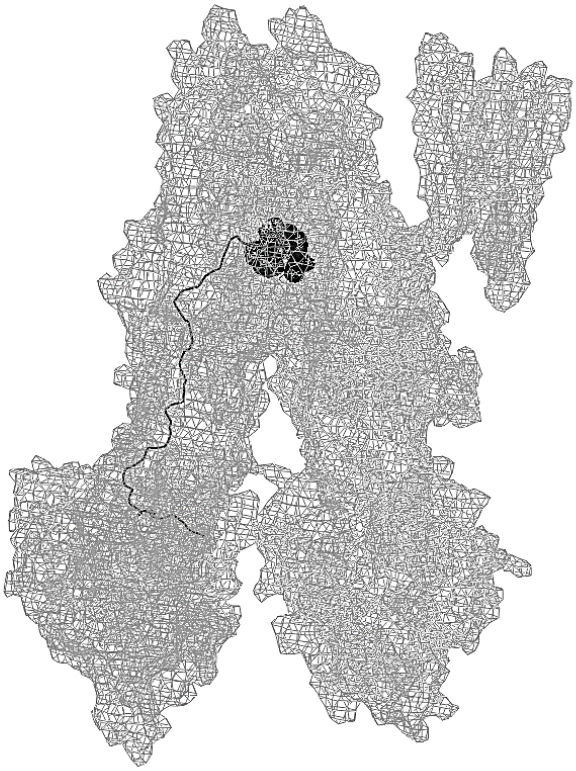
My last objective is to study and work on TUNNEL ANALYSIS by CAVERWEB online server. Tunnel analysis of ligand in protein is a method used to determine the location and size of binding sites in proteins. It involves the analysis of the molecular interactions between a ligand (small molecule) and a protein in order to understand the structural determinants of ligand binding and its intrinsic passage. Tunnel analysis helps to identify and characterize sites of interaction between ligands and proteins, and can be used to predict the affinity of protein-ligand interactions and how they intrinsically transport. It can also be used to evaluate the therapeutic potential of a given ligand. Tunnel analysis can also provide insight into the dynamics of ligand binding and can provide valuable information for the study of intrinsic transportation.

The tunnel analysis was carried out for the experimental structure 5UJA (having leukotriene C4 as its ligand) in order to study how the substrate intrinsically binds to the transmembrane and gets transferred.

The PDB structure was loaded into the webserver and it calculates and produces details of the ligand, binding pocket and much. Later the actual process of tunnel analysis is carried out.



- 41 Tunnel possibilities were estimated for the intrinsic binding and transportation. The length of the tunnels and the bottleneck radius (the maximal probe size which can fit in the narrowest part of the tunnel) of the tunnels were calculated and listed out for the better understanding of the intrinsic transportation possible and the residues within.



- The tunnel 41 shows complete coverage with maximum bottleneck radius, therefore this tunnel has selected for the study of intrinsic ligand transport.

- The Ligand Transport study of the Ligand through all the 41 tunnels were analyzed and the 41st tunnel holds better energy parameters (Less Binding Energy) and therefore acts as good tunnel with maximum coverage length and bottle neck radius.

bottleneck radius [Å]:	0.9
length [Å]:	112.0
distance to surface [Å]:	58.7
curvature:	1.9
throughput:	0.01
number of residues:	103
number of bottlenecks:	1

Conclusion:

So, this was all about the work I did with the internship opportunity I got. The report highlighted the various activities conducted throughout the duration of the internship, from the research of the project, to the practical application of the knowledge gained.

The comprehensive study and work helped me to understand how I should keep my approach in research and how to create a balanced and better workflow. With this internship I also learnt a lot about structural biology of transmembrane proteins and about the transport mechanism. I got to know, study and work on many computational methods and software within.

Overall, it was found that the internship was an invaluable experience and it provided me an opportunity to gain knowledge and experience in the research and application of the subject, and to develop an understanding of the fundamentals of the discipline. The internship also provided a platform for the development of communication and interpersonal skills, which will be invaluable in future professional endeavors. Additionally, the internship enabled me to gain an insight into the complexities of the field, and to understand the importance of accurate and thorough research.

And I hope that I could reproduce the work to better extent and add some extra bytes the work. This also motivated me to new ideas which I would like to investigate further.

Bibliography:

1. Bank, R. C. S. B. P. D. (n.d.). 5UJ9: Cryo-EM structure of bovine multidrug resistance protein 1 (MRP1). RCSB PDB. Retrieved March 2, 2023, from https://www.rcsb.org/structure/5uj9
2. Bank, R. C. S. B. P. D. (n.d.). 5UJA: Cryo-EM structure of bovine multidrug resistance protein 1 (MRP1) bound to leukotriene C4. RCSB PDB. Retrieved March 2, 2023, from https://www.rcsb.org/structure/5uja
3. Bank, R. C. S. B. P. D. (n.d.). 6BHU: Cryo-EM structure of ATP-bound, outward-facing bovine multidrug resistance protein 1 (MRP1). RCSB PDB. Retrieved March 2, 2023, from https://www.rcsb.org/structure/6bhu
4. Bank, R. C. S. B. P. D. (n.d.). 6UY0: Cryo-EM structure of wild-type bovine multidrug resistance protein 1 (MRP1) under active turnover conditions. RCSB PDB. Retrieved March 2, 2023, from https://www.rcsb.org/structure/6uy0
5. BioLib. (n.d.). BioLib. BioLib. Retrieved March 2, 2023, from https://dtu.biolib.com/DeepTMHMM
6. Caver web - tool for the analysis of tunnels and channels in protein structures. (n.d.). Retrieved March 2, 2023, from https://loschmidt.chemi.muni.cz/caverweb/
7. Cole, S. P. C. (2014). Multidrug resistance protein 1 (MRP1, ABCC1), a “multitasking” ATP-binding cassette (ABC) transporter. Journal of Biological Chemistry, 289(45), 30880–30888. https://doi.org/10.1074/jbc.r114.609248
8. Efficient unbound docking of rigid molecules - bioinfo3d.cs.tau.ac.il. (n.d.). Retrieved March 2, 2023, from http://bioinfo3d.cs.tau.ac.il/PatchDock/wabi02.pdf
9. Errat. UCLA. (n.d.). Retrieved March 2, 2023, from https://www.doe-mbi.ucla.edu/errat/
10. Humana Press. (n.d.). Homology modeling. SpringerLink. Retrieved March 2, 2023, from https://link.springer.com/book/10.1007/978-1-61779-588-6
11. Leukotriene C4. Uses, Interactions, Mechanism of Action | Drug Bank Online. (n.d.). Retrieved March 2, 2023, from https://go.drugbank.com/drugs/DB08855

12. Model. SWISS. (n.d.). Retrieved March 2, 2023, from https://swissmodel.expasy.org/
13. MolProbity. (n.d.). Retrieved March 2, 2023, from http://molprobity.biochem.duke.edu/
14. Pontius, J., Richelle, J., & Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. Journal of Molecular Biology, 264(1), 121–136. https://doi.org/10.1006/jmbi.1996.0628
15. Team, E. B. I. W. (n.d.). Procheck and Procheck-NMR. PROCHECK home page. Retrieved March 2, 2023, from https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
16. UniProt website. UniProt. (n.d.). Retrieved March 2, 2023, from https://www.uniprot.org/uniprotkb/P33527/entry
17. Verify3D. UCLA. (n.d.). Retrieved March 2, 2023, from https://www.doe-mbi.ucla.edu/verify3d
18. Yalcin-Ozkat, G. (2021). Molecular modeling strategies of cancer multidrug resistance. Drug Resistance Updates, 59, 100789. https://doi.org/10.1016/j.drup.2021.100789