

Drug-Target Interaction Prediction: Approaches for predicting interactions specific to patient treatment and drug design

Submitted in partial fulfilment of the required for the award of
Master of Science Degree in Bioinformatics and Data Science

By

J.JENNIFER(43739006)



**DEPARTMENT OF BIOINFORMATICS AND DATA
SCIENCE**

SCHOOL OF BIO AND CHEMICAL ENGINEERING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

**Accredited with Grade “A++” by NAAC JEPPIAAR
NAGAR, RAJIV GANDHISALAI, CHENNAI – 600119**

APRIL 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with “A++” grade by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Jennifer J(43739006) who carried out the project entitled “Drug-Target Interaction Prediction: Approaches for predicting interactions specific to patient treatment and drug design” under my / our supervision from January 2024 to April 2024

Internal Guide

External Guide (if Applicable)

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **JENNIFER J (43739006)**, hereby declare that the Project Report entitled “Drug-Target Interaction Prediction: Approaches for predicting interactions specific to patient treatment and drug design” done by me under the guidance of **Dr. JEMMY CHRISTY** is submitted in partial fulfilment of the requirements for the award of Master of Science degree in Bioinformatics and Data Science

DATE:

JENNIFER J

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

III ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. REKHA CHAKRAVARTHI M.E., Ph.D. Dean of Arts and Science** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my project guide **Dr. JEMMY CHRISTY** Department of Bioinformatics and Data Science for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Bioinformatics and Data Science** who were helpful in many ways for the completion of the project.

(JENNIFER J)

ABSTRACT

In drug discovery, predicting interactions between drugs and their target proteins is crucial. This study combines Swiss Target Prediction, a tool for target prediction, with Discovery Studio, a platform for molecular docking, to predict drug-target interactions. Initially, Swiss Target Prediction identifies potential protein targets for a given drug molecule based on structural similarity. Subsequently, Discovery Studio performs molecular docking simulations to predict the binding affinity and mode of interaction between the drug and its targets. The combined approach enhances the accuracy of drug-target interaction predictions, aiding in the design of novel therapeutic agents and understanding drug mechanisms.

TABLE OF CONTENTS

SERIAL NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF FIGURES	
1	INTRODUCTION	1
2	OBJECTIVE	2
3	MATERIALS AND METHODS	2
4	RESULT AND DISCUSSION	3
5	REVIEW ARTICLE	9
6	MINION NANOPORE SEQUENCING	11
7	REFERENCE	10
7	CERTIFICATES	12

	List of figures	
Sn.no	Figure name	Page no
1.	3D structure of CELECOXIB AND MAPK1	4
2.	2D structure of celecoxib and mapk1.	5
3.	Swiss Target Prediction Probability Score Table	5
4.	non-bond table	6
5.	molecule result table	6

INTRODUCTION:

Drug targets are specific molecules or biological structures involved in disease pathways that can be modulated by drugs to produce a therapeutic effect. Selecting the right drug target is crucial for the development of effective and safe medications. Public databases like PDB, PUBCHEM, THEMARKER AND SWISS TARGET were used to predict the target for the drug selected and to retrieve the structural information of the drug and the target. Researchers use various methods to identify drugs with repurposing potential, such as computational modeling, high-throughput screening, and data mining of electronic health records and scientific literature.

OBJECTIVE:

The main aim is to repurpose celecoxib which is used in the treatment of Rheumatoid arthritis (RA) as an (NASID) and it is metabolized by cox-2 enzyme which is considered as a remarkable biomarker for RA. Cyclooxygenase-2 (COX-2) is an enzyme involved in the synthesis of prostaglandins, which are lipid compounds that play a role in inflammation and pain signaling. COX-2 has been a target of interest for drug development due to its involvement in the inflammatory processes underlying the disease.

MATERIALS AND METHODS:

- PROTEIN DATA BANK
- SWISSTARGETPREDICTION
- PUBCHEM
- THERAPEUTIC MARKER

Are the databases used to retrieve structural and function information of the drug and the target. TheMarker is a database which contains the information about therapeutic markers based on safety biomarker, pharmacodynamic biomarker, monitoring biomarker, predictive biomarker and surrogate biomarker are the 5 types. Swiss target prediction is an online database which predicts the possible molecular targets (genes, proteins) based on similarity principle which generally states that two similar molecules are prone to have similar properties. The predicted targets are those having the actives displaying the highest similarity with the query molecule. The drug associated with rheumatoid arthritis was chosen from themarker and the smiles notation was taken from Pubchem. The smiles was given as an input on swisstargetprediction and the top 100 predicted targets were predicted. Targets with high probabiltity scores were taken into consideration. Hubbed genes were found to predict promising molecular target. The structure of the drug and the target were retrieved and uploaded on Discovery Studion for docking. Molecular docking is done to see how the drug interacts with the targets and find the active sites of interaction

RESULTS AND DISSCUSSION:

Chosen drug: celecoxib

Chosen target: cox-2

Predicted target: mapk1 [Mitogen activated protein kinase – 1]

Organism: Rattus norvegicus [358 residues] *similar to Homo sapiens with 360 residues.

Domain: 23-311

Structure ID: 3QYZ [from PDB]

Celecoxib exerts its therapeutic effects by selectively inhibiting COX-2, an enzyme involved in the synthesis of prostaglandins, which are mediators of inflammation, pain, and fever. By blocking COX-2, celecoxib reduces the production of prostaglandins responsible for inflammatory responses, leading to decreased pain and inflammation. MAPK1, also known as ERK2 (Extracellular Signal-Regulated Kinase 2), is a member of the mitogen-activated protein kinase (MAPK) family. It plays a crucial role in cellular signaling pathways involved in cell growth, differentiation, survival, and inflammation. MAPK1 is part of the MAPK/ERK pathway, which is activated by various stimuli, including pro-inflammatory cytokines like tumor necrosis factor-alpha (TNF-alpha) and interleukin-1 (IL-1). Activation of MAPK1 leads to the phosphorylation of downstream targets involved in inflammatory signaling cascades. The predicted target 3QYZ was docked with celecoxib and the following results were obtained. The binding takes place on site 1 of the active sites of the protein

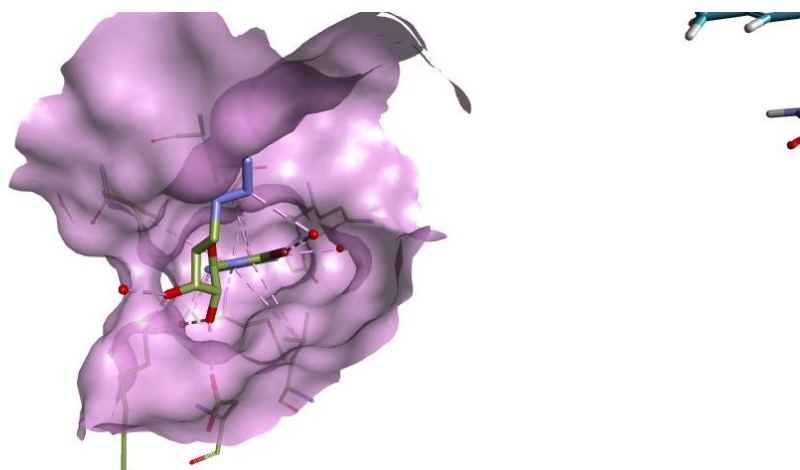


Figure 1 3D structure of CELECOXIB AND MAPK1

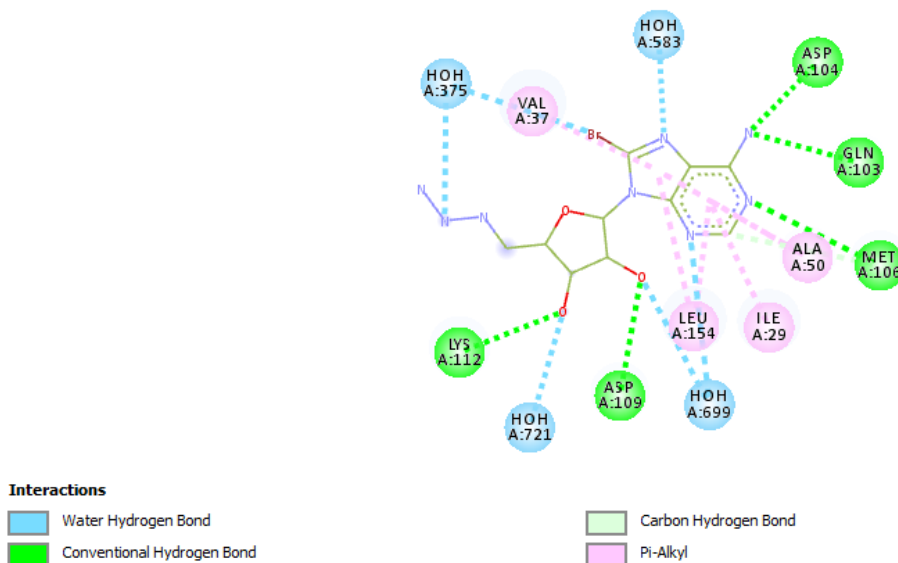


Figure 2: 2D structure of celecoxib and mapk1.

The amino acids interacting with celecoxib are LYS112, ASP109, LEU154, ILE29, MET106, GLN103, ALA50, ASP104 in the A chain of the protein. These residues interact by conventional hydrogen bonds and hydrophobic bonds. These residues are part of the structure of MAPK1 and are involved in its enzymatic activity and interactions with other molecules, including potential drug compounds like celecoxib.

FUNCTIONS OF CONVENTIONAL HYDROGEN BONDS:

VAL37: part of overall structural framework of mapk1.

MET106: initiation of protein synthesis, hydrophobic core formation.

LYS112: interacts with substrates cofactors and other ligands.

GLN103: signalling pathways and regulatory factor.

ASP109: acts as catalytic base assisting the transfer of phosphate group from ATP to substrate.

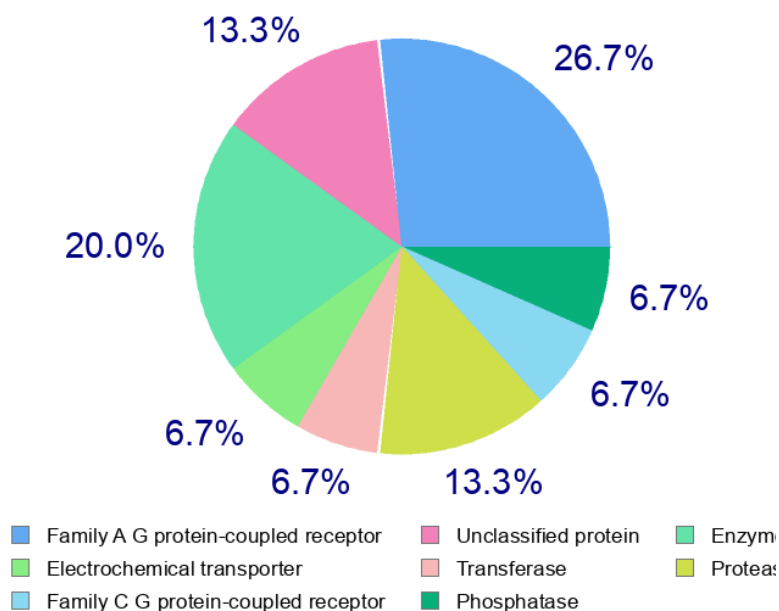


Figure 3 piechart from swisstargetprediction

Target	Common name	Uniprot ID	ChEMBL ID	Target Class	Probability*	Known actives (3D/2D)
Cyclooxygenase-2	Ptgs2	P35355	CHEMBL2977	Family A G protein-coupled receptor		312 / 232 ↓
Alpha-2b adrenergic receptor (by homology)	Adra2b	P19328	CHEMBL266	Unclassified protein		3 / 1 ↓
Serotonin transporter (by homology)	Slc6a4	P31652	CHEMBL313	Enzyme		38 / 1 ↓
Dopamine transporter (by homology)	Slc6a3	P23977	CHEMBL338	Electrochemical transporter		25 / 1 ↓
Carbonic anhydrase II (by homology)	Ca2	P27139	CHEMBL4706	Transferase		19 / 96 ↓
Serine/threonine-protein kinase mTOR (by homology)	Mtor	P42346	CHEMBL1075134	Protease		11 / 0 ↓
Thymidylate synthase (by homology)	Tyms	P45352	CHEMBL4341	Family A G protein-coupled receptor		10 / 0 ↓
Sphingosine 1-phosphate receptor 1 (by homology)	S1pr1	P48303	CHEMBL1914263	Family C G protein-coupled receptor		22 / 0 ↓
Cytochrome P450 17A1 (by homology)	Cyp17a1	P11715	CHEMBL4430	Enzyme		20 / 0 ↓
Hexokinase type IV	Gck	P17712	CHEMBL3882	Family A G protein-coupled receptor		1 / 0 ↓
Serotonin 1b (5-HT1b) receptor	Htr1b	P28564	CHEMBL3459	Family A G protein-coupled receptor		4 / 0 ↓
Dihydrofolate reductase (by homology)	Dhfr	Q920D2	CHEMBL2363	Enzyme		24 / 0 ↓
Arachidonate 5-lipoxygenase-activating protein (by homology)	Alox5ap	P20291	CHEMBL1921661	Phosphatase		47 / 0 ↓

Figure 4 probability table

NAME	DISTANCE	CATEGORY	TYPES	FROM	TO
A:DQ2402	3.41503	Hydrogen Bond	Water Hydrogen Bond;Carbon Hydrogen Bond	A:DQ2402:C2	A:HOH750:O
A:DQ2402	3.66133	Hydrogen Bond	Water Hydrogen Bond;Carbon Hydrogen Bond	A:DQ2402:C13	A:HOH750:O
A:HOH525	2.60587	Hydrogen Bond	Water Hydrogen Bond;Conventional Hydrogen Bon	A:HOH525:O	A:DQ2402:O1
A:HOH561	2.9601	Hydrogen Bond	Water Hydrogen Bond;Conventional Hydrogen Bon	A:HOH561:O	A:DQ2402:N3
A:MET106	2.92306	Hydrogen Bond	Conventional Hydrogen Bond	A:MET106:N	A:DQ2402:N5
A:LYS112:I	2.71838	Hydrogen Bond	Conventional Hydrogen Bond	A:LYS112:NZ	A:DQ2402:O3
A:LYS112:I	3.31706	Hydrogen Bond	Conventional Hydrogen Bond	A:LYS112:NZ	A:DQ2402:O4
A:DQ2402	3.10272	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N1	A:ASN152:OD1
A:DQ2402	2.73913	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N1	A:ASP165:OD2
A:DQ2402	3.01239	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N1	A:ASP165:OD2:B
A:DQ2402	3.36065	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N10	A:DQ2402:O1
A:DQ2402	3.38829	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N2	A:SER151:O
A:DQ2402	2.98918	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N2	A:ASN152:OD1
A:DQ2402	3.01442	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:N4	A:ASP104:O
A:DQ2402	2.9171	Hydrogen Bond	Conventional Hydrogen Bond	A:DQ2402:O4	A:ASP109:OD2
A:DQ2402	3.01306	Hydrogen Bond	Carbon Hydrogen Bond	A:DQ2402:C9	A:MET106:O

Figure 5: non-bond table

NAME	PLP1	PLP2	JAIN	PMF	DOCKSCORE
6FLE					
2662	41.1	38.32	-1.#J	18.46	43.074
2662	46.66	39.01	-1.#J	5.82	36.013
2662	58.15	52.25	-1.#J	18.08	34.27
2662	35.69	45.81	-1.#J	21.02	34.173
2662	50.15	41.69	-1.#J	4	33.615
2662	47.55	49.51	-1.#J	20.2	32.792
2662	58.74	45.39	-1.#J	7.61	31.442
2662	16.92	20.49	-1.#J	25.3	28.412
2662	23.96	30.47	-1.#J	18.24	27.319
2662	57.77	50.79	-1.#J	12.09	26.867

Figure 6: molecule result table

Applications and Learnings from Network Analysis of Networks in Action

In many different fields, network analysis has become a potent technique that provides an insightful understanding of complex systems and their dynamics. To shed light on network analysis's uses and the insights it can provide, we explore four major themes in this article.

Network Analysis in Online Communities and Social Media

Social media platforms now play a crucial role in our everyday lives by influencing conversations, viewpoints, and fashions. With the use of network analysis tools, we may examine the composition of social networks, pinpoint key nodes, locate communities, and comprehend the dynamics of information flow. For example, network analysis has been used by researchers to examine patterns of information dissemination on social media sites like Twitter, pinpointing important influencers and the channels via which viral content spreads.

Network Pharmacology: Utilising Network Analysis to Identify Drug Targets

Network pharmacology is a field in medicine that uses network analysis to find intricate relationships between pharmaceuticals, targets, and biological pathways. Researchers can find possible drug targets, forecast medication side effects, and optimise drug combinations for better therapeutic outcomes by building and analysing drug-target networks. This strategy has the potential to speed up the medication development and discovery process.

Network Sturdiness and Resilience: A Comprehensive Overview of System Dynamics

Networks encounter difficulties including outages, malfunctions, and intrusions in a variety of systems, including social networks, power grids, and transportation networks. Through scenario modelling, the identification of crucial nodes or linkages, and the development of techniques to improve system stability and recovery, network analysis aids in the evaluation of these systems' robustness and resilience. This information is essential for creating efficient and resilient infrastructures.

Ethical Considerations in Network Analysis and Data Privacy

As we harness the power of network analysis, ethical considerations and data privacy become paramount. Analyzing network data raises questions about consent, anonymity, data ownership, and potential biases. Researchers and practitioners must navigate these ethical challenges responsibly, ensuring transparency, informed consent, data anonymization, and compliance with regulations like GDPR and HIPAA.

ORIGINALITY REPORT

5

%
SIMILARITY
INDEX

5%

INTERNET
SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

fastercapital.com

Internet Source

3%

2

arxiv.org

Internet Source

2%

Exclude quotes Off

Exclude bibliography On

Exclude matches Off

GENERAL COMMENTS

REVIEW ARTICLE

Applications and Learnings from Network Analysis of Networks in Action

In many different fields, network analysis has become a potent technique that provides an insightful understanding of complex systems and their dynamics. To shed light on network analysis's uses and the insights it can provide, we explore four major themes in this article.

Network Analysis in Online Communities and Social Media

Social media platforms now play a crucial role in our everyday lives by influencing conversations, viewpoints, and fashions. With the use of network analysis tools, we may examine the composition of social networks, pinpoint key nodes, locate communities, and comprehend the dynamics of information flow. For example, network analysis has been used by researchers to examine patterns of information dissemination on social media sites like Twitter, pinpointing important influencers and the channels via which viral content spreads.

Network Pharmacology: Utilising Network Analysis to Identify Drug Targets

Network pharmacology is a field in medicine that uses network analysis to find intricate relationships between pharmaceuticals, targets, and biological pathways. Researchers can find possible drug targets, forecast medication side effects, and optimise drug combinations for better therapeutic outcomes by building and analysing drug-target networks. This strategy has the potential to speed up the medication development and discovery process.

Network Sturdiness and Resilience: A Comprehensive Overview of System Dynamics.

Networks encounter difficulties including outages, malfunctions, and intrusions in a variety of systems, including social networks, power grids, and transportation networks. Through scenario modelling, the identification of crucial nodes or linkages, and the development of techniques to improve system stability and recovery, network analysis aids in the evaluation of these systems' robustness and resilience. This information is essential for creating efficient and resilient infrastructures.

Ethical Considerations in Network Analysis and Data Privacy.

As we harness the power of network analysis, ethical considerations and data privacy become paramount. Analyzing network data raises questions about consent, anonymity, data ownership, and potential biases. Researchers and practitioners must navigate these ethical challenges responsibly, ensuring transparency, informed consent, data anonymization, and compliance with regulations like GDPR and HIPAA.

References:

Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.

Börner, K., & Scharnhorst, A. (2009). *Visual Concept Exploration in Digital Libraries*. Springer.

Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology*, 4(11), 682–690.

Watts, D. J. (2004). *Six Degrees: The Science of a Connected Age*. W. W. Norton & Company.

MinION NANOPORE SEQUENCING

INTRODUCTION

MinION (Mk1C- model) is a compact, portable device, combining the hardware for running nanopore

sequencing experiments with fully integrated compute used for base calling and onward analysis. The data generated by the device can be accessed by copying it in a pendrive and visualising on a computer.

Technical specifications

Component	Specification
Size and weight	H 33 mm x L 142 mm x W 118 mm; 455 g
Power	• Supplied with a 6.3-19.6 VDC power supply • Max rated current 10 A • Max rated power 60 W
Compute spec	1 TB SSD Storage, 8 GB RAM, GPU embedded analysis accelerator (ARM processor 6 cores, 256 core GPU)
Connectivity	• 1x USB 2.0 • 1x eSATA (currently non-functional) • 1x microSD card • 1x 1 GB Ethernet • Wi-Fi connectivity
Pre-loaded software	Linux OS, MinKNOW and Guppy
Security	Kensington Nano Security Slot
Environmental conditions	• Designed to sequence at +10°C to +30°C* • Do not cover vents on the top or sides of the device • Indoor use • Altitude up to 2,000 metres • Maximum relative humidity 80% for temperatures up to 31°C, decreasing linearly to 50% relative humidity at 40°C • Applicable pollution degree 2 in the intended environment

The flow cell is made up of an array of micro scaffolds with 512 channels, each containing 4 nanopores. (2048 nanopores in total). Since the nanopores are made up of proteins, they can be denatured due to various reasons. Therefore, no. of active pores should be checked before loading the sample.



Fig.1 sequencing machine in which the flow cell is placed



fig.2 Flow cell

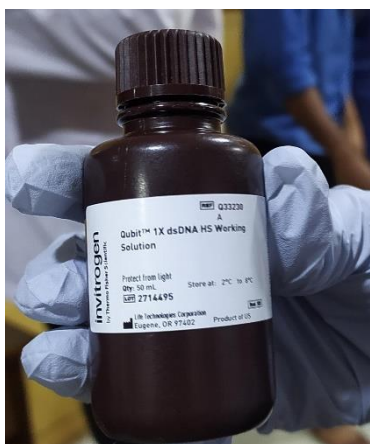
The motor proteins present above the nanopore is made up of helicase which helps to unwind the dsDNA and allows the ssDNA to pass through the nanopore to the +ve side of the membrane. Each base gives a characteristic reduction in the ionic current allowing the DNA to be sequenced.

DNA ISOLATION:

Phenol-chloroform method is used to isolate the genomic DNA. An aqueous layer will be formed in which the genomic DNA is present.

DNA QUANTIFICATION:

The concentration of DNA should be 1microgram per microlitre. Qubit fluorometer is used to check the concentration of the isolated DNA.



LIBRARY PREPARATION:

DNA library is prepared by PCR technique. DNA is separated from the buffer solution by using magnetic beads (AMPure XP beads). AMPure XP beads binds to the DNA and forms magnetic beads, which are separated by placing on a magnetic stand

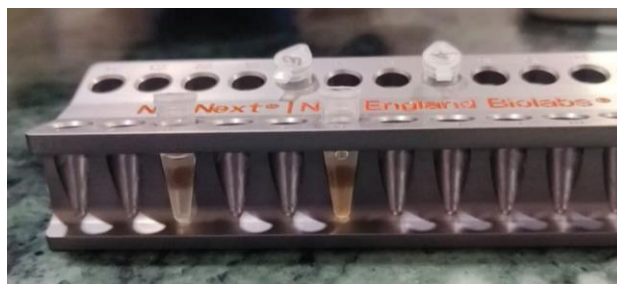


Fig.4 AMPure XP beads
after adding AMPure XP beads

Fig.5 DNA separated by placing on a magnetic stand

NATIVE BARCODE LIGATION:

Barcoding is important before loading the sequence. 24 barcodes(maximum) can be used with 24 different samples and can be sequenced at a time. Barcodes are provided by the manufacturers themselves.

ADAPTER LIGATION:

Adding adapters aids in the attachment of the sequence to the nanopore.

FLOW CELL LOADING:

The storage buffer is removed gently from the flow cell and then 75 microlitres of the sample is loaded.

The sequence was obtained after running the cell for approximately 16 to 17 hours. The sequence data is stored in the machine initially which can be copied to a pendrive to be accessed.