Virtual screening Approach of drug designing for Colon Cancer

L.Pradeepthi Pemmaraju[*], R.Yerakayya Dora[**], P.Vagdevi Sai[***]

[*] Assistant Professor, Dept. of Biotechnology, MRPG College
[**] Research Scholar, Dept. of Human Genetics, Andhra University
[***] Dept. of Microbiology, MRPG College

***Abstract*:** Colorectal cancer is a cancer from uncontrolled cell growth in the colon or rectum or in the appendix. Symptoms typically include rectal bleeding and anemia which are sometimes associated with weight loss and changes in bowel habits. The most common mutation in colon cancer is inactivation of APC. Adenomatous polyposis coli (APC) also known as deleted in polyposis 2.5 (DP2.5) is a protein that in humans is encoded by the APC gene. APC is classified as a tumor suppressor gene. Tumor suppressor genes prevent the uncontrolled growth of cells that may result in cancerous tumors. The protein made by the APC gene plays a critical role in several cellular processes that determine whether a cell may develop into a tumor. The APC protein is a "brake" on the accumulation of β-catenin protein; without APC, β-catenin accumulates to high levels and translocates (moves) into the nucleus, binds to DNA, and activates the transcription of genes that are normally important for stem cell renewal and differentiation but when inappropriately expressed at high levels can cause cancer. One of the most exciting developments in cancer research in recent years has been the clinical validation of molecularly targeted drugs that inhibit the action of APC gene. Efforts are now being directed at identifying the tumor subtypes and patients who will benefit the most from these drugs. Agents directed against new molecular targets are also being explored.

***Index Terms*-** Colorectal cancer, drug designing, Protein modeling, Docking Studies, Virtual Screening.

I. INTRODUCTION

A few years ago, the National Institutes of Health (NIH) created the Biomedical Information Science and Technology Initiative (BISTI) to examine the current state of bioinformatics in the United States. BISTI's working definition of bioinformatics included its use in biomedical research, in particular for drug discovery and development programs. Bioinformatics was seen as an emerging field with the potential to significantly improve how drugs are found, brought to clinical trials and eventually released to the marketplace.

Computer-Aided Drug Design (CADD) is a specialized discipline that uses computational methods to simulate drug-receptor interactions. CADD methods are heavily dependent on bioinformatics tools, applications and databases. As such, there is considerable overlap in CADD research and bioinformatics.

Adenomatous Polyposis Coli (APC)

The *APC* gene on chromosome 5q21 encodes a 2,843-amino acid protein that is important in cell adhesion and signal transduction; beta-catenin is its major downstream target. *APC* is a tumor suppressor gene, and the loss of *APC* is among the earliest events in the chromosomal instability colorectal tumor pathway. The important role of *APC* in predisposition to colorectal tumors is supported by the association of *APC* germline mutations with FAP and attenuated FAP (AFAP). Both conditions can be diagnosed genetically by testing for germline mutations in the *APC* gene in DNA from peripheral blood leukocytes

By using the computer aided drug design in the colon cancer treatment, we can identify the mutated pathways by simulating the drug receptor interactions.

II. MATERIALS & METHODS

While finding an effective drug against Colon Cancer, many tools and software's have been used. Genome sequence of APC gene, NM code of the gene NM_001127510.2was taken from NCBI (National Center for biological information), and the sequence of the gene was taken in genbank format. The nucleotide sequence was taken from NCBI in FASTA format and was analyzed using Bioedit software. The FASTA format was submitted for various NCBI tools like ORF, ePCR, map viewer, and Vec-screen. BLAST was done to know the sequences producing significant alignments and also the distance BLAST trees. Sequencing was done using the multiple sequence alignment tool Clustalw. The tool Phylodraw was used to get the evolutionary relationship of the nucleotide sequence with that of others in the form of phylogenetic trees. In proteomics part like the NM codethe NPcode of the protein was taken from the NCBI as NP_001120982.1.The nucleic acid composition and the amino acid composition can be known by BIO-EDIT Software. Using the NP_0000336.1 of protein prediction of Primary, Secondary, Post Translational Modifications, Topology structures of a protein was done. GOR(Garnier Osguthorpe and Robson),SOPMA(self optimized prediction method) were used for secondary structure prediction of protein and to predict the post translational modifications the parameters used were signal, Net C Glyc, Net oGlyc, Net Acet, Net phos, Sulfinator. Topology prediction was done by SOSUI tool to predict solubility of protein.

Homology modeling was done using SPDBV software where the FASTA format and two other templates collected from the swiss model were taken to model the loop build protein. The active site analysis was done using Q-site

finder to know the amino acids at the active sites. Using this side chain protein drug designing was done. In drug designing the drugs that act on the protein were taken from the drug bank and the process of INDIVIDUAL DOCKING was performed to the drugs with the active site amino acids of the side chain protein using software Argus Lab. The similar molecules for the above drug molecules were taken using chem. Officeand their IUPAC names were converted to structures from Chem. Draw ultra software.

Finally database creation was done using VEGA ZZ software and database docking was done to find the effective drug which docks with the maximum amino acids in the active site with the least energy value. QSAR (quantitative structure activity relationship) properties of the drug were evaluated by using hyper chem. software. Computer aided chemistry was done for the drug using Cache software.

METHODOLOGY FLOWCHART

- Target and Template Identification
- Find nucleotide and amino acid composition by using bioedit software
- Protein modeling using SPDBV software
- Loop building procedure
- Individual docking using Argus lab
- Database creation using VEGAZZ software
- Database docking procedure using Argus lab
- Find UV visible and IR transition using CAChe (computer aided chemistry) software
- Find QSAR properties using HYPERCHEM software

III. RESULTS & DISCUSSIONS

1. Bioedit result: DNA molecule:
gi|307133685|ref|NM_001127510.2| Homo sapiens adenomatous polyposis coli (APC), transcript variant 2, mRNA
Length = 10848 base pairs
Molecular Weight = 3297190.00 Daltons, single stranded
Molecular Weight = 6575615.00 Daltons, double stranded
G+C content = 39.74%
A+T content = 60.26%
Nucleotide Number Mol%

A	3675	33.88
C	2136	19.69
G	2175	20.05
T	2862	26.38

Protein: gi|4557761|ref|NP_000242.1| DNA mismatch repair protein Msh2 isoform 1 [Homo sapiens]
Length = 934 amino acids
Molecular Weight = 104737.73 Daltons
Individual docking results using ARGUS LAB:
Standard market drugs like capacitabine, Flurouracil, Irinotecan, Leucovorin, Levamisol, thotraxate structures are taken from the drug bank and indivual docking of these drugs with different amino acids were analysed using Argus Lab software to identify the best ligand pose energy released during the docking.
Similar molecules for these standard drugs were modeled using Argus Lab.

A database is created with all these molecules in Vega ZZ. Virtual Screening of these drugs are done through protein Vega ZZ database docking method. The results obtained show that cinnamyl isovalerate is interacting at the lowest energy level with all the amino acids in the potential active site, and through Cache analysis the lead Cinnamyl isovalerate is potential to UV and IR transitions. The QSAR Properties are also give the immense potential that the Selected lead is the Best Among the molecules. So, we consider that ligand as the potential lead molecule to target the protein APC in treating the disease.

IV. CONCLUSION

The current research is intended to perform comparative homology modeling and virtual screening of the gene APC. The sequence was retrieved from NCBI database having accession number |NM_001127510.2 and verified in Swiss port database having accession number NP_000242.1.Specific protein causing the disease is identified as APC INHIBITOR and homology modeling of the protein is done using SPDBV (Swiss PDB Viewer). Active site analysis is done through Q site finder method and the active site amino acids are noted.

REFERENCES

1. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005).
2. *"General Information About Colon Cancer". NCI. May 12, 2014. Archivedfrom the original on July 4, 2014. Retrieved June 29, 2014.*
3. *"SEER Stat Fact Sheets: Colon and Rectum Cancer". NCI. Archived from the original on June 24, 2014. Retrieved June 18, 2014.*
4. Middleton SB, Frayling IM, Phillips RK. Desmoids in familial adenomatous polyposis are monoclonal proliferations. Br J Cancer. 2000 Feb;82(4):827-32.Citation on PubMed.
5. https://www.ncbi.nlm.nih.gov/
6. https://www.drugbank.ca/
7. https://blast.ncbi.nlm.nih.gov/Blast.cgi

AUTHORS

First Author – L.Pradeepthi Pemmaraju, MSc,assistant professor, Dept. of Biotechnology,MRPG College
pemmarajudeepthi@gmail.com
Second Author – R.Yerakayya Dora, Research Scholar, Dept. of Human Genetics, Andhra University
Dora.rayapureddy@gmail.com
Third Author – P.Vagdevi Sai, Dept. of Microbiology, MRPG college
Vagdevi7702@gmail.com
Correspondence Author – L.Pradeepthi Pemmaraju, pemmarajudeepthi@gmail.com

8121998340