Docking studies of Hemochromatosis protein with various compounds of the medicinal plants

O.S. Deepa[1], and Ani. R[2]

[1]Department of Mathematics, Amrita School of Engineering, Coimbatore

[2]Department of Computer Science and applications, Amrita School of Engineering, Amritapuri

Amrita Vishwa Vidyapeetham, Amrita University, India

***Corresponding author: E-Mail: os_deepa@cb.amrita.edu**

ABSTRACT

Hemochromatosis is a genetic disorder which leads to the accretion of iron in parenchymal organs leading to organ toxicity. Normal absorption of iron from daily food is 10% whereas people with hemochromatosis diseases can absorb iron four times more than the normal absorption. There is no proper medication and clinically proved medicines have side effects. Hence an alternative methods is the extraction of bioactive compounds from the medicinal plants to recognize the novel target. 17 compounds from 13 medicinal plants with 81 properties were collected from database. 81 properties with certain specific conditions were checked for accuracy using various machine learning techniques. The compounds that satisfies the pharmacological properties are docked with the mutated protein of hemochromatosis. The free energy and the interaction analysis were also discussed.

KEY WORDS: Hemochromatosis, bioactive compounds, medicinal plants, docking, Epigallocatechin -3 gallate, Aloin.

1. INTRODUCTION

Docking plays an important task in drug designing. Due to the presence of large amount of ligands in the databank, wet lab cannot be preferred directly. The role of bioinformatics helps to investigate the diseases using computational techniques at the molecular level. A genetic algorithm based docking analysis of breast cancer susceptibility protein with curcumin was carried by Permalatha (2013). Pharmacophore elucidation and docking studies on anti-inflammatory compounds of medical plants for ulcerative colitis was studied by Hamsa (2013). The amount of iron in the body is detected by HFE gene by interacting with other proteins on the cell surface. Hepcidin hormone which is regulated by HFE protein is produced by the liver and determines the percentage of iron that should be absorbed from the diet and the percentage of iron that should be released from storage sites in the body. Two major mutations are responsible for this disorder. More than 20 mutations are documented by the researchers in the HFE gene which forms basis of hereditary diseases called hemochromatosis - type 1. The replacement of the amino acid cysteine with the amino acid tyrosine at position 282 in the protein chain is the main cause of mutation. Also the replacement of the amino acid Glycine with the amino acid Histidine at position 105 is another cause of mutation.

Table.1. Amino acid change for the HFE protein

Entry name-HFE	Amino acid change	Condon change
1A6Z	Gly-Asp	G-D
1A6Z	Ser-Cys	S-C
1A6Z	Giy-Arg	G-R
1A6Z	Gln-His	Q-H
1A6Z	Ala-Val	A-V
1A6Z	Arg-Gly	R-G
1A6Z	Cys-Tyr	C-Y
1A6Z	Gln-Pro	Q-P
1DE4	Ary-Cys	R-C
1DE4	IIE-Thr	I-T
1DE4	Val-Ala	V-A

The protein sequence of HFE protein is obtained from the protein sequence database. It is found from PDB database that the protein 1A6Z has no ligand and there are 7 ligands in protein 1DE4 PDB entry. Hence the 3D structure of 1DE4 is considered for interaction and is viewed in RASMOL. The 3D structure enables the analysis of its interaction with suitable inhibitors.

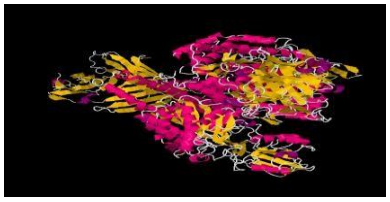

Figure.1. 3D structure of 1DE4 protein

It is considered to be complex, time-consuming, and expensive process for the development of new drug. Computer - Aided Drug Design (CADD) generally focuses on the regulation that depends on the computational methods to simulate drug - receptor interactions. CADD methods greatly rely on bioinformatics tools, applications and databases. In current scenario the use of bioinformatics tools leads to identifying therapeutic targets for meticulous diseases by using various software tools and is also used to recognize the drugs from the bioactive compounds of medicinal plants by doing virtual screening. These approaches have showed a momentous role in computer aided drug design. The field of drug design and discovery from medicinal plants are quicker, efficient and competent. Studies based on docking of Emblica officinalis compounds with DNA polymerase was done by Ramakanth (2012). Prashant Anthwal (2015) has discussed on the phytochemicals like azadirachtin, cardiofolisode and kutkin for the modulation of scretase enzymes in the treatment of Alzheimer's. An identification of variety of leaf diseases using data mining techniques was introduced by Sasirekha (2015).

2. METHODOLOGY

Identification of potential ligands: Many research papers have reported that the compounds extracted from herbal have high medicinal values against many diseases. From extensive literature study, a preliminary survey of 17 compounds were selected from 11 different medicinal plants.

Table.2. Compounds and source of the medicinal plants

Compounds	Source	Compounds	Source
Curcumin	*Curcuma domestica*	Epigallacatechin gallate	*Camellia sinesis*
Demethoxy Curcumin	*Curcuma domestica*	Aloin	*Aloe vera*
Tumerone	*Curcuma domestica*	Emodin	*Aloe vera*
Allicin	*Allium sativum*	Rutin	*Calendula officinalis*
Ursolic acid	*Tulsi*	Lawsone	*Lawsonia inermis*
Rosmarinic acid	*Tulsi*	Gingerol	*Zingiber*
Luteolin	*Tulsi*	Epigallocatechin	*Camellia sinesis*
Nimbin	*Azadirachta indica*	Cyanidin	*Hibiscus sabdariffa*
Quercetin	*Azadirachta indica*		

From the drug data base the 3D structures of ligands from various plants can be found.

ADMET studies: The ligands can be considered for structure based drug discovery if the pharmacological properties of the ligands are satisfied. The computational biology tools are used for estimating the essential components like absorption, distribution, metabolism, excretion and toxicity (ADMET) for a lead molecule. Hence, all 17 ligands identified from preliminary studies were tested for their drug-likeliness, ADME profile and toxicity analysis using PreADMET tools. The ADME properties consist of the rate of absorption, distribution, metabolism and excretion. Absorption determines the movement of the substance in to the blood circulation and distribution elucidates how substances are propagating throughout the tissues and fluids of the body agents. Topological analysis and other basic properties were also calculated from various online web servers. The decrease in intricacy of drug discovery process have overcome by screening original therapeutic targets by computer aided methods. The screening of lead molecule with excellent pharmacological property and drug likeness is a monotonous assignment in drug development procedure. The investigation of biologically active compounds with constructive ADMET and drug-likeness properties are done by many computer aided techniques.

Screening of ligands: 81 properties were collected from various online web-servers like ADMET, drug-likeness, Molecular weight, Molecular surface area, alogP, tPSA, Rotable hydrogen bond, Plat index, Randii index, Balaban index, Wiener index, bioavailablility, etc. A study on the classification of chemical compounds extracted from medicinal plants decides the acceptability of the compound in drug discovery. A machine learning approach may be useful to classify the compounds based on the attribute values. The attribute values are predicted based on the ranges from research literatures. Highly hydrophilic and highly hydrophobic molecules have been eluded. Extremely high percentage of aromaticity is not usually chosen as it may be carcinogenic. High value of molar refractive index is preferred. Molecules with high polar surface area are considered for ligands. The expected logP value should be more than 2 to lie in the acceptable region of drug designing. Solubility is the major criteria to be considered for a molecule to be a ligand. $CaCO_2$ cell predicts the level of permeability of all compounds and MDCK predicts the level of absorption. Skin permeability and Human Intestinal Absorbtion (HIA) are also important for screening the compounds as ligands. The AMES test is most commonly used as an initial potential for drugs. Based on these conditions the ranges for 81 properties were fixed and then normalized. Classification algorithms using Machine learning techniques such as Decision Tree, Naive Bayes and Random Forest are used to analyse the set of attributes of the chemical compound found in medicinal plants. There are two class labels considered in the training data sample. The class labels used in the data set are Medium acceptable and highly acceptable compounds for drug design. Random forest, Rotation forest, J48 and Navie bayes gave 50%, 80% and 100% accuracy for different ranges

of 81 properties. Random forest ensemble classifier to predict the coronary heart disease using risk factors was studied by Ani (2015). The 3D structure of the ligands are given in figure.

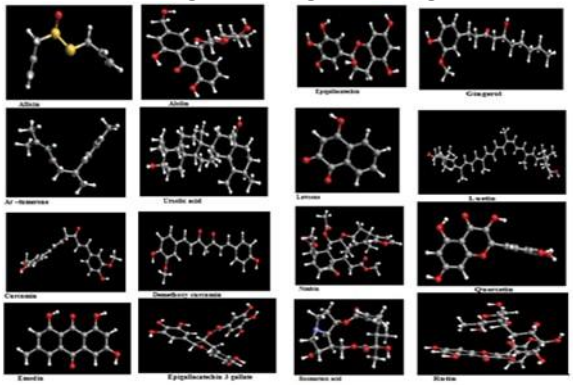

Figure.2. 3D structure of ligands from medicinal plants

Docking studies: The most significant process in drug discovery is the constructive means to develop an perceptive of protein–ligand interactions by computational docking or computer-aided docking. Computational docking is the procedure of computationally foretelling the position and binding affinity of the ligand in the binding pocket of the protein. Docking methods depend on search algorithm which calculates the position of the ligand in the binding pocket and a scoring function which approximates the binding affinity, by predicting how strong the ligand interacts with the protein.

3. RESULTS AND DISCUSSION

In our proposed work, the compound cyanidine is removed before docking process as it does not satisfy the most of the pharmacological properties. The software used for docking is iGEMDOCK. A docking software iGEMDOCK constructs protein- ligand interaction profile based on three interactions namely electrostatic, hydrogen bonding and Vanderwaal's interactions. It also tabulates the energy based scoring table for the active components from medicinal plants. The interaction analysis of 1DE4 with the components of the medicinal plants is given below:

Table.2. The interaction analysis of 1DE4 with the components of the medicinal plants

Compounds	**Energy** (Kcal/mol)
Aloin	-57.88
Tumerone	-41.38
Urosolic acid	-51.06
Curcumin	-39.32
Embodin	-53.75
Epigallocatechin- 3 gallate	-64.22
Epigallocatechin	-50.92
Gingerol	-48.9
Lawsone	-40.17
Lutein	-41.39
Nimbin	-32.52
Quercetin	-45.09
Rosmarinine	-49.69
Rutin	-54.94

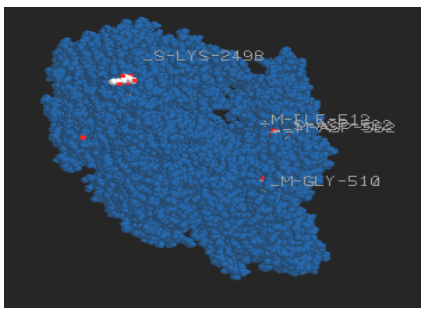

Figure.3. Interaction analysis of 1DE4- Epigallocatechin -3 gallate from igemdock

4. CONCLUSION

Epigallocatechin -3 gallate has the minimum energy of -64.22 (Kcal/mol). It is also found that the compound Epigallocatechin -3 gallate has a molar refractive index less, MlogP is more than 2, soluable in water, skin premeability is high, satisfies Lipinskin's rule, medium level of premeability by Caco-2 cell, high level of absorption by MDCK and human intestinal absorption is 83%. Hence Epigallocatechin -3 gallate from green tea medicinal plant can be considered as one of the potential drug for hematochromotosis.

REFERENCES

Ani R, Aneesh Augustine, Akhil N.C, Deepa O.S, Random forest ensemble classifier to predict the coronary heart disease using risk factors,proceeding of the International conference on soft computing systems, Springer, 397, 2015.

Hamsa N.S, Pharmacophore elucidation and docking studies on anti-inflammatory compounds of medicinal plants for ulcerative colitis, Asian Journal of pharamaceutical and clinical research, 6 (3), 2013.

Lee S.K, Lee I.H, Kim H.J, Chang G.S, The PreADMET: PC-Based program for batch prediction of ADME properties, Euro QSAR, 2004, 9.5-10.

Pradeepa K, Antibacterial acitivity of leaf extract of Delonix elata and molecular docking studies of luteolin, Journal Biochem Tech, 3 (5), 2012, S193-S197.

Prashant Anthwal, Prabhakar Semwal, Taranjeet Kapoor, Madhu thapliyal, Ashish Thapliyal, Azadirachtin, cardiofolisode and kutkin can be vital phytochemicals for the modulation of scretase enzymes for the treatment of Alzheimer's: An Insilco analysis, Asian Journal of pharmaceutical and clinical research, 8 (4), 2015, 108-112.

Premalatha V, A Genetic algorithm based approach for comparative docking analysis of breast cancer suspectibility protein interaction with curcumin, International Journal of Pharma and Biosciences, 4 (1B), 2013, 325-338.

Ramakath K, Mohammad Munwar, Sreenivasulu M, Javasimha Ravalu D, Docking studies of *Emblica officinals* compounds with DNA polymerase, International Journal of Applied Biology and Pharmaceutical Technology, 3 (4), 2012, 406-413.

Sasirekha N, An identification of variety of leaf diseases using various Data mining techniques, International Journal of advanced research in computer and communication engineering, 4 (10), 2015.

Yang J.M, and Chen C.C, GEMDOCK: A generic evolutionary method for molecular docking, Proteins: Structure, Function and Bioinformatics, 55, 2004, 288-304.