Research Article

Design of Novel Drug Molecules for the Treatment of Breast Cancer

Vijith V S[1], Harikrishnan K V[2], Krishnapriya S[4], Veena Varma[5], Asha Jose[1], Prasanna Ramani[3], Kumar E P[1], Krishnan Namboori P K[4*]

[1]Department of Pharmacology, Karpagam College of Pharmacy, Othakkalmandapam, Coimbatore, India.
[2]Department of Pharmaceutical Chemistry, Amrita School of Pharmacy, AIMS – Ponekkara P. O., Kochi, Kerala, India.
[3]Departments of Science, Amrita Vishwa Vidyapeetham University, Coimbatore, Tamilnadu, India.
[4]Computational Chemistry Group, Computational Engineering and Networking, Amrita Vishwa Vidyapeetham University, Coimbatore, Tamilnadu, India.
[5]AMRITA School of Biotechnology, AMRITA Vishwa Vidyapeetham, Amritapuri, Kollam, India.

***Corresponding author's E-mail:** n_krishnan@cb.amrita.edu

Accepted on: 21-03-2013; **Finalized on:** 31-05-2013.

ABSTRACT

In breast cancer, BRCA tumor suppressor gene repairs the damaged DNA by substituting with perfectly matching proteins. The required proteins are produced by the BRCA1 and BRCA2 genes. Hereditary mutation of such genes leads to produce undesirable proteins, which may not bind to DNA double strand for the repairing function. Mutated BRCA tumor suppressor gene extends the risk not only for causing breast cancer but also for developing other types of cancer. The technique involves a logic program for designing biologically active molecules, which overcome the limitations of traditional drug development. The proposed work suggests some new anti breast cancer drugs, designed through in-silico method. Fragment based drug designing technique (FBDD) has been adopted in this work. The designed ligands have been further subjected to screening through toxicity, pharmacokinetic and pharmacodynamic studies. The molecule EVO1, with maximum 'ligand enrichment factor', is found to be the most suitable anticancer ligand subject to further experimental evaluation.

Keywords: Breast Cancer, BRCA gene, Protein domains, Fragment based drug designing, I*n-silico* method.

INTRODUCTION

Breast cancer is a dreadful and common type of cancer occurring in both sexes, though found to be more common in women. Globally, every year, about 22.9% of cancer deaths in women are found to be due to breast cancer[1]. The cancer cells generally originate in the milk duct (*ductal carcinoma*) or the lobules (*lobular carcinoma*) of the breast, slowly covering the neighboring tissues and changing into malignant tumor.

BRCA1 and BRCA2 (Breast cancer susceptibility gene 1 and breast cancer susceptibility gene 2) are known as tumor suppressors and normal expression of these genes play a major role in repairing of damaged DNA. Over expression of BRCA1 and BRCA2 genes results in cancer risk[2]. Hence, BRCA proteins can be taken as the target for controlling the over expression of the genes.

The computer aided drug design (CADD) or computational drug design (CDD) has been accepted and appreciated as the designing phase of drug development[3]. Among various techniques used in CADD, a three dimensional structure based 'de nova drug design' strategy has come up as a major leading procedure. In the target-based de nova method, design of ligand is made according to the active site of the target. The method involves identification of active binding sites such as hydrophobic site, hydrogen bonding donor atoms, hydrogen bonding acceptor atoms, polar atoms etc. of the target molecules. Fragments or building blocks of the ligand would be made according to the nature of the binding sites of the target. All the designed fragments will be connected together to get a ligand template, which would be allowed to evolve in the binding pocket of the target. These ligands would be collected to make the combinatorial library. By proper scoring and filtering using interaction energy, pharmaco dynamic and pharmacokinetic attributes most suitable ligand could be identified. The technique is generally referred as 'fragment based drug design' (FBDD)[4]. In the present work, FBDD has been followed to identify the suitable ligands.

MATERIALS AND METHODS

The proteins corresponding to BRCA1 and BRCA2 genes have been collected from RSCB Protein Data Bank[5], the major protein repository. They are subjected to sequence analysis to generate theoretical pI, half-life, instability index, aliphatic index, and 'Grand Average Value of Hydropathicity (GRAVY)' using Protparam tool[6]. Secondary structure prediction has been carried out using SOPMA[7] and sub cellular locations have been identified using UniprotKB[8]. The proteins are scanned and their binding cavities have been analyzed through CASTp[9].

The binding sites or 'hot spots' of the target have been located from the naturally occurring complexes of the protein molecules. The protein molecules are subjected to molecular dynamic (MD) simulation[10] to identify the conformational involvement in biological activity. The identified conformers are further subjected to docking with the natural ligands, providing maximum freedom to the ligand molecules using the CDOCKER tool of Accelerys Discovery Studio 2.1[11] with CHARMm force field. Interactional variation on different conformers is the result of conformational involvement in biological activity. If the interactions are identical for all conformers,

then the molecule will not be treated as conformationally biased. The highly interacting binding sites have been chosen for generating ligands. The fragments of ligands corresponding to these interacting hotspots are identified using the 'de novo receptor protocol' of Discovery studio. Based on these fragments, series of new ligand molecules have been evolved through 'de novo evolution protocol' and these molecules constitute the combinatorial library for new anti cancer drug molecules.

The ligands in the combinatorial library have been screened based upon the interaction score and pharmacokinetic properties like absorption, distribution, metabolism and excretion (ADME) of the molecules, which determine the potency and pharmacological behavior. Further screening has been carried out based upon properties like aqueous solubility, partition coefficient (logP), blood brain barrier (BBB) penetration and cytochrome P450 enzyme (CYP 450) inhibition. Toxicity parameters like hepatotoxicity, mutagenicity (using Ames test), carcinogenicity, and rat acute toxicity have been considered for the final screening of ligands using admetSAR[13]. Pharmacophoric points of the ligands have been identified using Ligand Scout 2.1[14].

Evaluation of the Ligands

The new ligands identified have been subjected to evaluations studies such as ligand efficiency, lipophilic efficiency, fit quality, IC50 value and ligand efficiency scale for identifying their efficiency as anticancer agents. These properties are compared with that of the known drug molecules to compute the *enrichment factors*. Four known drug molecules, *anastrozole, capecitabine, toremifene, vedafaxine* have been selected as standard for efficiency analysis and their properties have been evaluated with the new drug molecules.

a) Ligand efficiency

Ligand efficiency is a measurement of the binding energy per atom of a ligand to its binding partner, such as a receptor or enzyme. Ligand Efficiency (LE) can be also defined as the ratio of Gibbs free energy (ΔG) to the number of non-hydrogen atoms of the compound (Equation1).

$$LE = (\Delta G) / N \quad (1)$$

where, ΔG, the binding energy of the complex is given by equation 2:

$$\Delta G = -RT \ln K \quad (2)$$

N is the number of non-hydrogen atoms. The units of ligand efficiency are kcal/mol per non-hydrogen atom.

Ligand efficiency is a simple metric for assessing whether a ligand derives its potency from optimal fit with the target protein or simply by virtue of making many contacts. It shows generally a dependency on ligand size. Ligand efficiency drops dramatically when the size of the ligand increases. Ligand efficiency is used in drug discovery programs to assist in narrowing focus to lead compounds with optimal combinations of physicochemical properties and pharmacological properties. It is frequently used to evaluate fragment compounds in fragment-based drug discovery[15].

b) Lipophilic efficiency (LiPE)

It is also known as ligand lipophilic efficiency. It is a parameter used to evaluate the quality of compounds, linking potency and lipophilicity in an attempt to estimate drug likeness. LiPE is defined as the pIC50 (or pEC50) of the compound of interest minus the LogP of the compound (Equation 3).

$$LiPE = pIC_{50} - \log P \quad (3)$$

Lipophilic efficiency is used to compare compounds of different potencies and lipophilicities. It is a measure of efficiency of a ligand exploits its lipophilicity to bind to a given target. It has been reported that a lipophilic efficiency greater than 5 combined with $c \log P$ values between 2 and 3 is considered to be optimal for a promising drug candidate.

c) Ligand Efficiency Scale

It is a size-independent ligand efficiency value. This was obtained by fitting the top ligand efficiency versus heavy atom count to a simple exponential function (Equation 4).

$$LFscale = \left[0.104 + 0.65e^{-(0.037*HA)}\right] \quad (4)$$

d) Fit quality

Fit quality is the ratio between the ligand efficiency and the normalized ligand efficiency scale (Equation 5).

$$FQ = LE / LEscale \quad (5)$$

Fit quality scores close to 1.0 or above indicate near optimal ligand binding, while low fit quality scores are indicative of suboptimal binding.

e) IC_{50} Value

Half maximal inhibitory concentration (IC50) is a measure of the effectiveness of a compound in inhibiting biological or biochemical function. This quantitative measure indicates the efficiency of a particular drug or other substance (inhibitor) to inhibit a given biological process (or component of a process). In other words, it is the half maximal (50%) inhibitory concentration (IC) of a substance (50% IC, or IC50). It is commonly used as a measure of antagonist drug potency in pharmacological research. IC50 values are used to calculate the lipophilic efficiency of known ligands. A regression equation is generated for identifying the unknown IC50 value by using the IC 50 values and other parameters such as $A \log P$, pKa, pKb, number of hydrogen bonding acceptors (HBA), number of hydrogen bonding donors (HBD), number of rotatable bonds etc. of the known drug

molecules using the online tool 'XURU'. Regression equation for Anticancer Drug molecules (equation 6):

$$IC50(y) = -91.20171929x_1 - 87.65810566x_2 - 229.9661444x_3 - 67.0139673x_4 + 6.648309562x_5 - 11.58145736x_6 + 341.5006512 \quad (6)$$

Where, $x_1 = c\log P$, $x_2 = \log S$, $x_3 = HBD$, $x_4 = HBA$, $x_5 = PSA$ and $x_6 = \text{Rotable bonds}$.

Enrichment factor of the new ligands with respect to ligand efficiency, fit quality, IC50 value and lipophilic efficiency have been computed (Equation 7) with respect to the common anticancer drugs, anastrozole, capecitabine, toremifene and vedafaxine.

$$\text{Enrichment factor } (\phi) = \frac{\Delta(property)}{|\text{Reference property}|} \quad (7)$$

RESULTS AND DISCUSSION

On protein characterization, excepting *2JIT* and *2R4B*, instability index of all other targets has been found to be less than 40, supporting high structural stability. The aliphatic index varies from 70 to 95 corresponding to high thermodynamic stability of these protein molecules. Normally cancer target protein molecules are expected to be thermodynamically stable to withstand the higher rate of metabolism going on in cancer cells. Negative value of 'grand average hydropathy value (GRAVY) supports the protein molecules to be hydrophilic. The pI value of protein molecules ranges from 4.5 to 6.5. The lower pI probably helps the protein molecules to exist in the Zwitter ionic form even in the slightly acidic cancer tissues. The sub cellular location of the target molecules is found to be nucleus, cytoplasm and melanoma. Other than *1LOB*, all other protein molecules are having a high half life period. This clearly highlights the kinetic stability of the targets and mutational stability of the corresponding genes. The above properties are enlisted in Table 1.

In the secondary structure prediction of proteins, it has been found that all these molecules are structurally stabilized by efficient folding technique (Table 2). The low percentage of β- turn and the high percentage of α – Helix support the above factor.

The protein surface scanning using CASTp[9, 16] reveals a number of possible pockets with available volume varying from 200 (Å) to 8112.8 (Å).

During docking with the conformers obtained by Molecular Dynamic (MD) simulation, it has been found that different conformers interact with the ligands using different binding sites clearly supporting the 'conformationally biased nature' of the target molecules. However during ligand-protein interaction, ligands interact with the target in the most possible conformation facilitating maximum stability to the complex formed.

Table 1: Protein characterization for target proteins

PDB I.D	GRAVY	Half-life (Hrs)	Sub cellular location	Instability index	Aliphatic index	pI
1LOB	-0.324	7.2	Cytoplasm, melanoma, nucleus.	29.72 (stable)	70.65	4.59
1T15	-0.288	100	Nucleus	34.16 (stable)	78.6	5.50
2JIT	-0.213	30	Cytoplasm	44.24 (unstable)	95.69	5.59
2J6M	-0.220	30	Cytoplasm	44.04 (unstable)	95.69	5.59
1T29	-0.327	100	Nucleus	37 (stable)	77.70	5.78
2R4B	-0.317	30	Cytoplasm	40.08 (unstable)	90.11	6.61

Table 2: Secondary structure analysis results

Sl. No	PDB I.D	α – Helix (%)	Random coil (%)	Extended strand (%)	β turn (%)
1	1LOB	12.88	47.21	31.76	8.15
2	1T15	39.64	31.53	21.62	7.21
3	2JIT	46.79	31.80	16.21	5.20
4	2J6M	45.87	32.42	15.6	6.12
5	1T29	36.84	35.96	20.16	6.58
6	2R4B	38.32	37.38	17.76	6.54

Table 3: Results of ADME analysis of the generated ligands

Sl.no.	Ligands	Absorption	Solubility	BBB penetration	CYP 450 inhibition	logP
1	EVO 1	good	good	low	non- inhibitor	0.855
2	EVO2	good	good	low	non -inhibitor	0.551

Table 4: Results of Toxicity Analysis

Ligands	Mutagenicty	Hepatotoxicity	Carcinogenicity	Acute toxicity in rat (LD50, mol/kg)
EVO1	Non-Mutagenic	Non-Toxic	Non-Carcinogenic	**2.4800**
EVO2	Non-Mutagenic	Non-Toxic	Non-Carcinogenic	**2.4009**

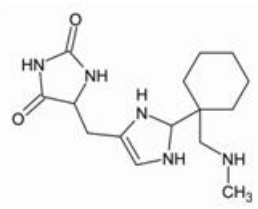
Table 5: Lipophilic Efficiency, IC50, ligand efficiency and Fit quality of Drug Molecules

Sl. No.	Ligands	Lipophilic Efficiency	IC_{50} value	Ligand efficiency value	Fit Quality value
1	ANASTROZOLE	2.7438	5.5398	-0.0002916	-0.000785984
2	CAPECITABINE	3.0182	3.4922	0.00003125	0.000103477
3	TOREMIFENE	-0.3024	5.9126	-0.0000322	-0.000103871
4	VEDAFAXINE	1.9076	4.9436	-0.0000333	-0.000104717
5	EVO1	4.1129	4.3689	0.000235	0.000652778
6	EVO2	5.0241	6.7561	0.0033	0.008894879

Table 6: Enrichment factors of the new ligands

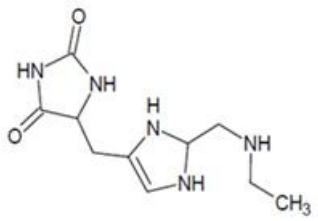
Property Ligands		Anastrozole	Capecitabine	Toremifene	Vedafaxine
Ligand efficiency	Evo1	1.8059	6.5200	8.2981	8.0571
	Evo2	12.3169	104.6000	103.4845	100.0990
Fit quality	Evo1	1.8305	5.3084	7.2845	7.2337
	Evo2	12.3169	84.9600	86.6339	85.9421
IC50	Evo1	0.2114	0.2510	0.2610	0.1162
	Evo2	0.2196	0.9346	0.1427	0.3666
Lipophilic efficiency	Evo1	0.4990	0.3627	14.6009	1.1561
	Evo2	0.8311	0.6646	17.6141	1.6337

The fragments of ligands from each breast cancer susceptible protein have been generated. All together, 551 new ligand molecules have been evolved and all of them have been put up in the combinatorial library. A preliminary screening has been done using the CDOCKER score, which gives out 59 highly interacting ligands. Pharmacokinetic properties of ligand molecules have been taken as the second screening criterion, in which only 7 of them have been predicted as having optimum conditions. Further screening using toxicity conditions, solubility and log P suggests (Tables 3 and 4) two ligands to be most suitable for anticancer properties. Thus, *5-[(2-{1-[(methylamino) methyl] cyclohexyl}-2, 3 -dihydro-1H-imidazol-4-yl) methyl] imidazolidine-2, 4-dione (EVO1) and 5-({2-[(ethylamino)methyl]-2, 3-dihydro-1H-imidazol-4-yl} methyl) imidazolidine-2, 4-dione (EVO2)* are the two ligand molecules found to be non mutagenic, non toxic, non carcinogenic, non hepatotoxic and with maximum interaction possibility with the target molecules (Figure 1 and Figure 2).



5-[(2-{1-[(methylamino) methyl] cyclohexyl}-2, 3 -dihydro-1H-imidazol-4-yl) methyl] imidazolidine-2, 4-dione

Figure 1: Structure and IUPAC name of EVO1



5-({2-[(ethylamino)methyl]-2, 3-dihydro-1H-imidazol-4-yl} methyl) imidazolidine-2, 4-dione

Figure 2: Structure and IUPAC name of EVO2

Furthermore, 3D pharmacophoric features of the ligands like hydrogen bond donar (HBD), hydrogen bond acceptor (HBA), hydrophobicity, positive ionizable (P.I) centers and total polar surface area (TPSA) have been studied.

Ligand evaluation

a) Ligand Efficiency

Ligand efficiency values of the new drug molecules have been computed and compared with that of the available drug molecules. The result implies that the new designed ligands possess good ligand efficiency while comparing with that of the available drug molecules.

b) Fit Quality

Fit quality of the ligands indicates the efficiency of the ligand molecule to bind with a particular target protein molecule. Fit quality of the designed ligand molecules have been identified and compared with that of the

available drug molecules .The result shows that the new designed molecules have high fit quality when compare with that of the available molecules.

c) IC_{50} Value

IC_{50} values of all the designed molecules as well as the available drug molecules have been identified. The results indicates that the designed molecules have promising inhibitory concentration when compared to that of the available molecules, so that they are capable for exerting good anticancer activity.

d) Lipophilic Efficiency

Lipophilic efficiencies of the ligand molecules have been identified from the IC_{50} values of the molecules. Lipophilic efficiencies of the designed molecules and the available drug molecules have been compared and the results indicate that the new drug molecules have a very high lipophilic efficiency than that of the available drug molecules, which concluded that the designed molecules are capable for existing as an anti breast cancer agent. (Table 5).

The enrichment values of the new ligands have been listed in Table 6. Ligand efficiency of Evo2 is found to be high. Similarly, fit quality with the respective targets, IC50 and Lipophilic efficiency values are favoring Evo2.

CONCLUSION

Computational drug design has been used for identifying new anticancer ligand molecules for breast cancer. Fragment based de novo technique has been employed in this work. All modern drug design strategies such as docking score, ADME score, toxicity, BBB, solubility and log P have been extended in this work. EVO1 and EVO2 are found to be most suitable for the specific targets. On fine tuning using hydrophobicity pharmacophoric point, EVO 2 is found to be more suitable towards breast cancer. The ligand efficiency studies indicate that the two new ligand molecules are highly active against breast cancer in which EVO 2 shows higher activity than EVO 1. The enrichment factors of the compound Evo2 are effectively high in choosing the compound for further in vitro and in vivo analysis. The margin of efficiency of Evo2 over Evo1 may be due to the existence asymmetric centers in EVO 1, which may make the molecule conformationally biased.

REFERENCES

1. Sariego J, Breast cancer in the young patient, The American surgeon, 76 (12), 2010, 1397–1400.
2. Saxena S., Contribution of germline BRCA1& BRCA2 sequence alterations to breast cancer in Northern India, *BMC Medical Genetics,* 7, 2007, 1-12.
3. Kapetanovic I M, Computer-Aided drug discovery and Development (CADDD): in-silico-chemico-biological approach, *Chemical & Biological Interaction,* 171 (2), 2008, 165–176.
4. Congreve M, Carr R, Murray C and Jhoti HA, A “rule of three” for fragment-based lead discovery, Drug Discovery Today, 8, 2003, 876-77.
5. Berman H M., The Protein Data Bank, *Nucleic Acid Research,* 28, 2000, 2207-2215.
6. Gasteiger E, Hoogland C, Gattiker A. et al., ExPASy: The proteonomics server for in-depth protein knowledge and analysis, *Nucleic Acid Research,* 31, 2003, 3784-3788.
7. Geourjon C. and Deleage G., SOPMA: A Self-optimized method for protein secondary structure prediction, *Protein Engineering,* 7 (2), 1994, 157-64.
8. Jain E., Infrastructure for the life sciences: Design and implementation of the uniprot website, *BMC Bioinformatics,* 10, 2009, 136.
9. Liang J, Edeksbrunner H, Woodward C, Anatomy of protein pockets and cavities: Measurement of binding site geometry and implication for ligand design, *Protein Science,* 1998, 71884-1897.
10. Duraant J D, McCammon J A, Molecular Dynamics and Drug Discovery, *BMC Biology,* 9, 2011, 417-421.
11. Diego S, Discovery Studio Modeling Environment Release 2.1, Accelerys Software Inc, 2007.
12. Wu G, Robertson D H, Vieth M, Detailed analysis of grid-based molecular docking: A case study of CDOCKER–A CHARMm-based Moleclar dynamics docking algorithm, *Journal of Computational Chemistry,* 24 (13), 2003, 1549-1562.
13. Shen J, Cheng F, Xu Y, Li W, Tang Y, ESTIMATION OF ADMET properties with substructure pattern recognition, *Journals of chemical information and modeling,* 50 (6), 2010, 1034-1041.
14. Wolber G, Langer T, Ligand scout: 3D pharmacophores derived from protein-Bound Ligands and Their use as Virtual screening Filters, *Journals of chemical information and modelling,* 45(1), 2005, 160-169.
15. Dundas J, Ouynag Z., Tseng J. CASTp: Computed Atlas of Surface Topography of Protein with structural and topographical mapping of functionally annotated residues, *Nucleic acid Research, 2006,* 34, 116-118.
16. Zapatero C A, Ligand efficiency indices for effective drug discovery, Expert Opin. Drug Discov., 2(4), 2007, 469-488.

Source of Support: Nil, **Conflict of Interest:** None.