Available Online at www.jddt.in

ISSN: 2320 - 4230

Journal of Drug Discovery and Therapeutics 2 (24) 2014, 09-17

RESEARCH ARTICLE

Emerging drug discovery paradigm in non small cell lung cancer: pharmacophore modeling, atom-based 3D- QSAR and virtual screening of novel EGFR inhibitors

Joseph Gerad Rakesh[1], Bappunhi Shaheera[2], Palani Aarumugam[3], Ramamurthi Vaiyshnavi[4], Ramamoorthy Thulasibabu[*]

[1]Department of Microbiology, Sri Venkateshwaara Medical College Hospital and Research Centre, Ariyur, Puducherry – 605102.

[2]SNMV College of Arts and Science, Malumichampatti, Coimbatore–641 021, India.

[3] Vector Control Research Centre (VCRC), Medical Complex, Indira Nagar, Puducherry–605006, India.

[4] Sri Venkateswara College of Engineering, Sriperumbudur, Chennai-602 117, India.

Received 25 November 2014; Accepted 4 December 2014

ABSTRACT

EGFR is a significant drug target approached in the domain of the wide prevailing cancer type, non- small cell lung cancer and search for its potent inhibitor is an emerging opportunity in cancer drug discovery. Computational framework of our study outsmarts the existing strategies and new insight into EGFR inhibitors. To figure out the higher biological behavior of the N-benzyl-N-(X-2-hydroxybenzyl)-N0-phenylureas and thioureas derivatives, functional spots requirement were substantially discovered using pharmacophore modeling and atom- based 3D-QSAR. Having the eye on five crucial pharmacophore features AADHR: namely two acceptors (A), one hydrogen bond donor (D), one hydrophobic group (H), one aromatic ring (R), then unique pharmacophore was identified. As the succession of this pharmacophore concept, statistical parameters for better construction of best model, correlation coefficient R^2 =0.9803 for training set compounds and Q^2=0.7099 for test set compounds were outcome of the atom based 3D-QSAR analysis. Then we engaged with virtual screening method to screen existing drug molecules of zinc database to identify exceedingly capable inhibitor for EGFR. Ultimately, scaffold with distinct qualities for drug-like species, top four of them were reported.

KEYWORDS: EGFR inhibitors, pharmacophore modeling, atom-based 3D-QSAR, PHASE, glide.

INTRODUCTION:

Cancer is a major public health issue in the United States and other regions of the world. One in four deaths in the United States is due to lethal cancer types observed in human population. In this study our immediate focus is to develop a clinically valid potent inhibitor for non small cell lung cancer, the most dominant lung cancer type effecting 85-90% of peoples with lung cancer. The Epidermal Growth Factor Receptor (erbB1) one among the receptor tyrosine kinase family [EGFR (ErbB-1/HER1), ErbB-2 (HER2), ErbB-3 (HER3) and ErbB-4 (HER4)] is involved in various fundamental cellular functions like cell proliferation, differentiation, metastasis and survival[1]. EGFR is expressed in tumors like prostrate, breast, lung, head and neck cancer[2]. EGFR is expressed in all types of cell, activation in non-cancer cell types that are involved in cell growth and proliferation might lead to tumor progression[3]. The tyrosine kinase family consisting of four receptors structurally possesses an extra cellular ligand-binding domain, a transmembrane domain and an intracellular domain with tyrosine kinase activity[4]. The binding of epidermal growth factor (EGF) or transforming growth factor α (TGFα) to EGFR leads to receptor-linked tyrosine kinase activation causing various effects like cell differentiation, apoptosis inhibition, maturation, metastasis, migration and angiogenesis[5,6]. The compounds inhibiting the tyrosine kinase activity of EGFR after binding to EGF and hence regulating the cellular signal transduction has evolved as anticancer therapeutic agents. Among the four major classes of anti-EFGR agents, tyrosine kinase inhibitors (TKIs) gefitinib, erlotinib and lapatinib which are 4-anilinoquinazoline derivatives have been clinically approved[7]. In the year of 2008-2009, compounds such as thienopyrimidine, thiazolo [4,5-d] pyrimidine, arylaminopyrimidine, pyrazolo [3, 4-d] pyrimidine, and pyrrolotriazine derivatives were studied as anti-EFGR agents [8,9,10,11,12,13]. Even though these molecules served the purpose they also had adverse side effects with majority of people. Designing novel and potent EGFR tyrosine kinase inhibitors is believed to be a

*Corresponding author: Thulasibabu Ramamoorthy || E-mail: thulasibabu.icmr@gmail.com

promising approach for anticancer therapy and it is the heart of this study.

MATERIALS AND METHODS:

Our approach basically dwells with Contemporary drug discovery approach, ligand based design of the drug molecules are theoretically experimented and validated using schemes namely pharmacophore modeling, 3D-QSAR, virtual screening by PHASE. PHASE is a versatile solution for pharmacophore perception, structural alignment, activity prediction and three dimensional high throughput screening. The data set of 36 N-benzyl-N-(X-2-hydroxybenzyl)-N0-phenylureas and thioureas derivatives were obtained from literature[14]. The 2D chemical structures of N-benzyl-N-(X-2-hydroxybenzyl)-N0-phenylureas and thioureas derivatives were sketched using Chem Draw[15]. The biological activity of all of the compounds was converted into pIC_{50} by taking negative logarithm of IC_{50}. The ligands were prepared using Ligprep application of Schrodinger suite where in the ligands geometry were converted to 3D structure, optimized and conformation states were generated by incorporated Epik module. The optimization parameters were set as; OPLS-2005 force field for the torsional search method sampling, distance-dependent dielectric solvent model, pre-minimization for maximum of 50 steps and post-minimization for maximum of 100 steps. Conformers with a maximum energy difference of 10 kcal/mol relative to the global energy minimum conformer were obtained[16,17,18]. Activity threshold valve was set forth to an extent, 6.500 are of most active agents, and 5.000 are of inactive in nature. The set of compounds remaining are of moderately active were listed in the table 1 and it paves way for the generation common pharmacophore hypothesis. Further the Pharmacophore sites generated based on PHASE, in-built six features, namely hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P) and aromatic ring (R) were defined by a set of chemical structure patterns with using SMARTS queries. The common pharmacophore hypothesis (CPH) was predicted from a set of variants using a tree-based partition algorithm with a maximum tree depth of 5. The final size of the pharmacophore box that governs the tolerance on matching was 1Å. After creating the sites, a set of 2 active compounds was used for scoring the CPHs. The threshold for the root mean square deviation (RMSD) of 2 Å was applied for the inter-site distances. The atom based 3D QSAR models were generated by applying the maximum of five partial least square (PLS) factors. A rectangular grid with spacing of 1.0 Å was defined to encompass the space occupied by the aligned training set molecules[19, 20]. Each of these models was validated using a test set of eight molecules that were not considered during model generation. Pharmacophore from all conformations of the two most active compounds were examined, and those that contained identical sets of features with similar spatial arrangements were grouped together. Based on the pharmacophoric sites, the number of sites can be set to any value between 4 and 7 survival score and molecule having best survival score is chosen.

To most comprehensive docking method, Glide is adopted to find the binding mode prediction between the crystal structure, retrieved from protein data bank (PDB ID: 1M17)[21] and ligand molecule examined for drug-like properties that are clinically acceptable. Pre-process before docking essentially includes adding hydrogen atom, deleting unwanted water molecule, and removing the co-crystal ligand and correspondingly energy minimization of the protein structure was done till it hits 0.30Å cut off value, by the dint of protein preparation panel of Schrodinger suite[22]. Gathering of active residues found in the protein were done with highlighting information acquired through previous literature. The residue involved in hydrogen bond interaction is MET 769, apart from that MET 769 and LEU 768 are involved in pi-pi interaction and pi – cation interaction respectively. Hydrophobic interactions were observed in the residues namely LEU 820, GLY 772, and LEU 694. It is often challenging for a theoretical biologist to end up with accurate force field patterns for docking the two molecules in a prescribed grid space. OPLS 2005 was used for this computational job, and the grid enclosing box was structured theoretically in a way it is totally centered towards the co-crystal ligand. To achieve the accuracy in the molecular docking, Extra Precision (XP) is chosen and results were highly promising to pick the novel small molecule with competitive biological properties with varied scaffold steric spots[23]. Just like a knock out, the pharmacophore with fewer features is pushed from large set of molecules 2, 64,000 curated small molecule databases ZINC based on fitness score[24]. The compounds with ideal properties to come under the roof of Lipinski Rule of Five were further screened by docking method [25].

Table1: Data set used for 3D-QSAR analysis with actual and predicted activities.

Compound	R^1	R^2	R^3	IC50 (µM)	Actual pIC_{50}	Predicted pIC_{50}
1a	H	Cl	F	1.98	5.703	5.51
2a	H	Br	F	2.24	5.649	5.63
3a	H	Me	F	21.71	4.663	4.64
4a	Cl	Cl	F	4.15	5.381	5.4
5a	Br	Br	F	4.68	5.329	5.34
6a	Me	Me	F	27.35	4.563	4.63
7a	H	Cl	OH	0.96	6.017	6
8a	H	Br	OH	1.07	5.97	5.94
9a	H	Me	OH	7.51	5.124	5.17
10a	Cl	Cl	OH	1.86	5.73	5.85
11a	Br	Br	OH	1.12	5.95	5.88
12a	Me	Me	OH	9.26	5.033	4.87
13a	H	Cl	-	32.12	4.493	4.63
14a	H	Br	-	36.88	4.433	4.64
15a	H	Me	-	<50	4.301	4.32
16a	Cl	Cl	-	<50	4.301	4.4
17a	Br	Br	-	<50	4.301	4.41
18a	Me	Me	-	<50	4.301	4.05
1b	H	Cl	F	0.67	6.173	6.07
2b	H	Br	F	0.98	6.008	6.05
3b	H	Me	F	8.54	5.068	5.01
4b	Cl	Cl	F	1.34	5.872	5.5
5b	Br	Br	F	1.76	5.754	5.71
6b	Me	Me	F	19.41	4.711	4.76
7b	H	Cl	OH	0.08	7.096	6.97
8b	H	Br	OH	0.12	6.92	5.94
9b	H	Me	OH	3.46	5.46	5.51
10b	Cl	Cl	OH	0.56	6.251	6.24
11b	Br	Br	OH	0.87	6.06	6.26
12b	Me	Me	OH	5.83	5.234	6.11
13b	H	Cl	-	18.12	4.741	4.74
14b	H	Br	-	22.04	4.656	6.74
15b	H	Me	-	43.21	4.364	4.34
16b	Cl	Cl	-	24.65	4.608	4.56
17b	Br	Br	-	31.94	4.495	4.56
18b	Me	Me	-	<50	4.301	4.11

RESULTS:

On the whole, 36 N-benzyl-N-(X-2-hydroxybenzyl)-N0-phenylureas and thioureas derivatives served as the starter steric conformers for the pharmacophore generation. The elucidation of the common pharmacophore feature is indispensible for this approach. For that, molecules were classified with reference to pIC_{50} value as more active (>6.500), less

active (< 5.000) and intermediates are coined as moderately approachable for high throughput screening. Out of 208 five point pharmacophore hypothesis, the best hypothesis was picked up for the 3D QSAR study based on the survival score.

The distinct and unique pharmacophore hypotheses AADHR.94 consists of two hydrogen bond acceptor (A), one hydrogen bond donor(D), one hydrophobic group (H), one aromatic ring (R) is identified, alignment of the most active compounds with best pharmacophore hypothesis were depicted in figure 1.

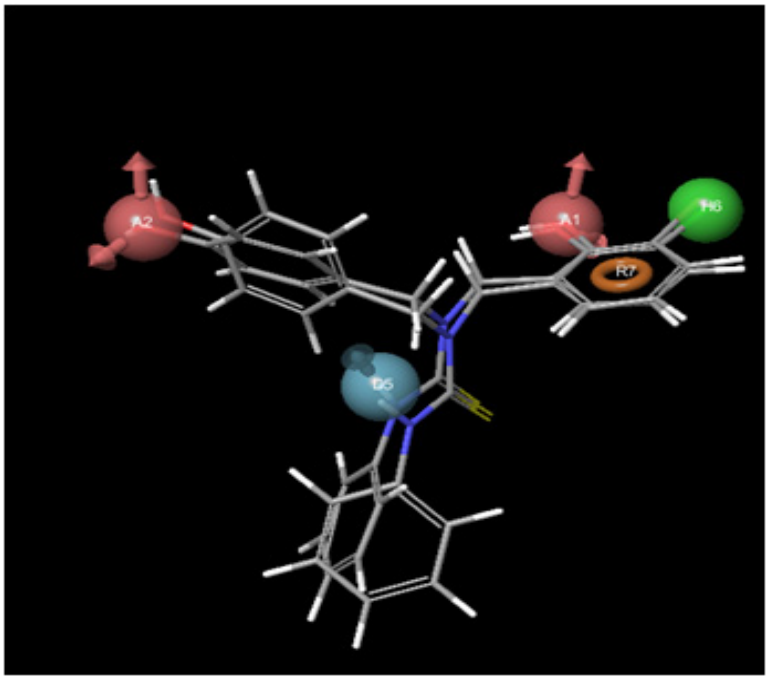

Figure 1: Alignment of most active compounds with best common pharmacophore hypotheses.

Characteristic angles and distance between the pharmacophore are featured in the figure 2 and 3. In the atom based 3D-QSAR model, generation of the dataset of 28 training set compounds and 8 test compounds are chosen randomly to calculate partial least square values.

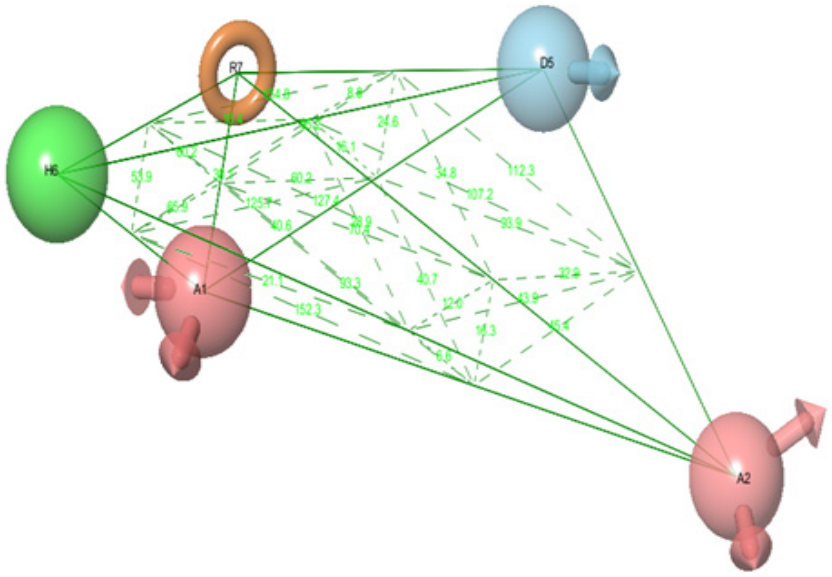

Figure 2: Pharmacophore hypothesis and angles between pharmacophoric sites.

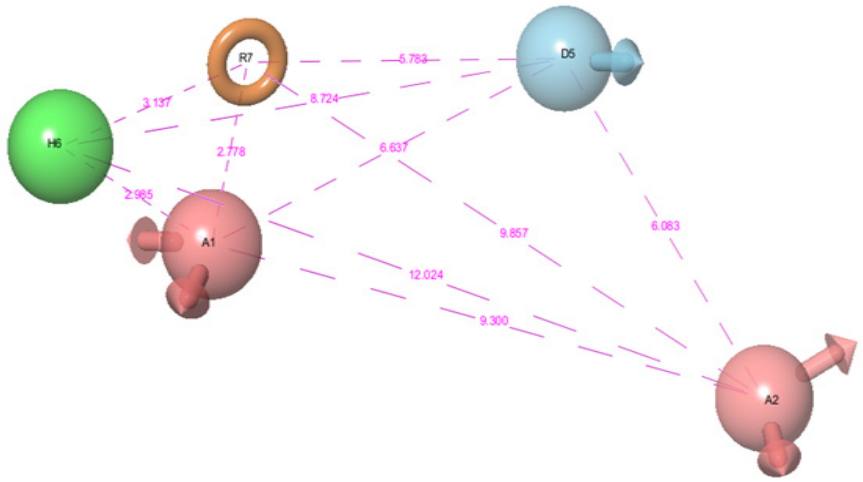

Figure 3: Pharmacophore hypothesis and distance between pharmacophoric sites.

The hypotheses yielded significant 3D-QSAR model with good PLS statistics. The statistical results of atom-based 3D-QSAR were reported in table 2. The training set correlation were characterized by PLS factor (R^2= 0.9803, SD= 0.1154, F= 346, P= 3.876e-020). The test set correlation were characterized by PLS factor (Q^2= 0.7099, RMSE= 0.4915, Pearson-R= 0.8459). The observed and predicted activity of the training and test set are shown in figure 4.

Table 2: Atom-based 3D-QSAR results for best four common pharmacophore hypotheses by PLS.

Hypotheses	SD	R^2	F	P	RMSE	Q^2	Pearson - R	Stability
ADHRR	0.0921	0.9874	346	3.876e-020	0.5649	0.6168	0.8075	0.4853
ADRRR	0.1224	0.9778	193.9	2.009e-017	0.60770	0.5565	0.825	0.7532
DDHRR	0.1145	0.9806	222.1	4.67e-018	0.5881	0.5846	0.7892	0.5938
AADHR	0.1154	0.9803	218.6	5.533e-018	0.4915	0.7099	0.8459	0.5978

SD= standard deviation of the regression, R^2= correlation coefficient
P= significant level of variance ratio, f= variance ratio
Q^2=for the predictive activities, RMSE=root-mean-square error
Pearson-R=correlation between the predicted and observed activity for the test set.

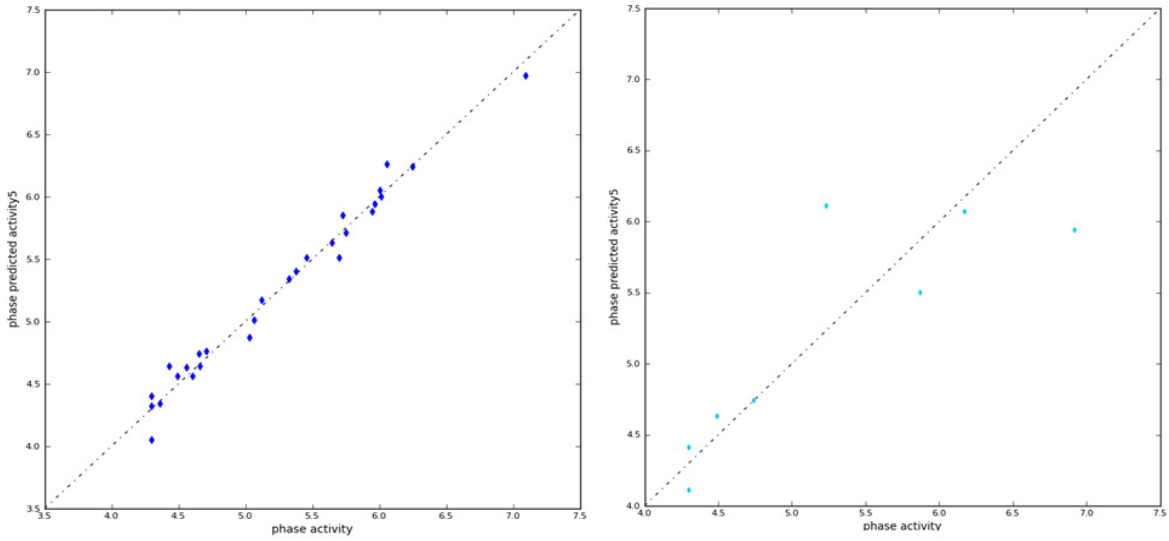

Figure 4: Graph of observed versus predicated biological activity of training set (a) and test set compounds (b).

Analysis of QSAR model: The contour maps were derived from the atom-based 3D QSAR and the map reveals the required positions of the highly must for the enhanced biological activity. The favorable and unfavorable region of the contour maps for most active compound 7b was represented in figure 5. In the representation cubes, the blue color region indicate the favorable activity the region enhance the activity of the compound. While red color region indicates unfavorable region of the compound these region decrease the activity of the compound

H-bond donor: presence of secondary pyrrole /indole nitrogen capable of acting as hydrogen bond donor seemed to have a favorable effect on activity. However, hydrogen bond donor in the side chain affected activity adversely (Figure. 5A); Hydrophobicity: hydrophobic group (methyl) at N-position of pyrrole/indole ring negatively correlated with activity, whereas as its presence on side-chain nitrogen had a positive influence on activity (Figure. 5B); Electron-withdrawing: presence of electron-withdrawing carbonyl group exercised a favorable effect on the activity (Figure. 5C).

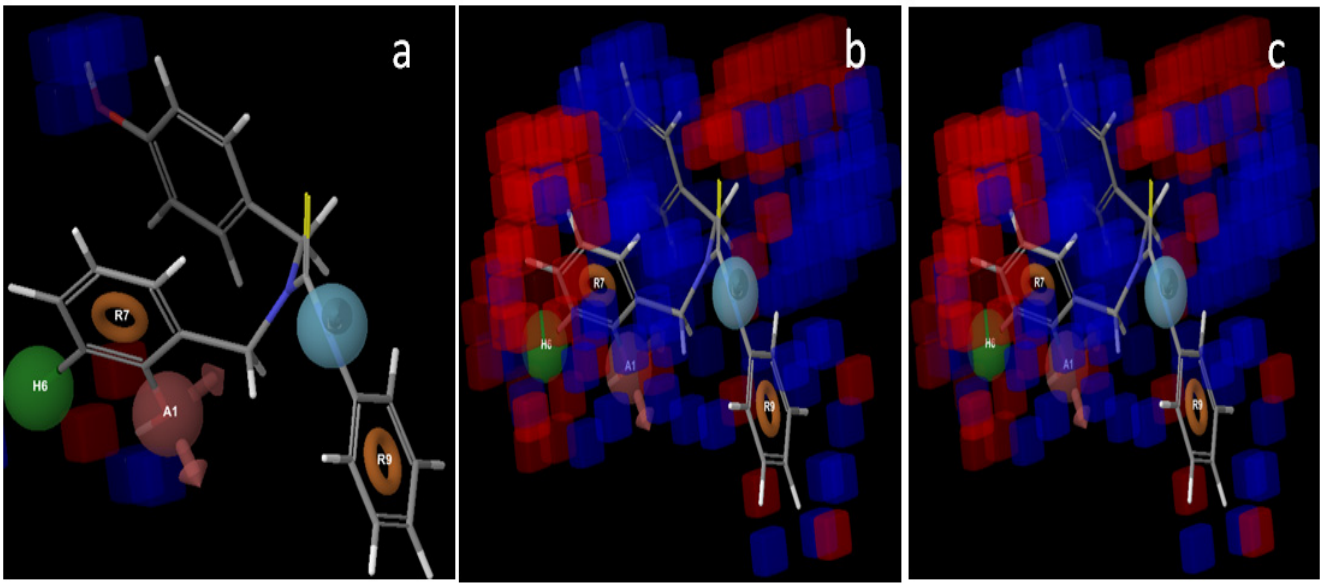

Figure 5: Pictorial representation of the contours generated using the QSAR model blue cubes indicates favorable region, red cubes indicates unfavorable region for the activity. QSAR model visualized in the context of favorable and unfavorable hydrogen bond donor with most active compound 7b (a), QSAR model visualized in the context of favorable and unfavorable hydrogen hydrophobic group with most active compound (b), QSAR model visualized in the context of favorable and unfavorable electron withdrawing features with most active compound (c).

The best hypotheses were utilized to screen Zinc database. Totally 1,000 were obtained by above mentioned pharmacophore hypotheses. Molecular docking was carried out for 1000 compounds and 36 ureas and thioureas derivatives using glide program. Glide, a docking panel comes with Maestro, well known for its accuracy and potential in finding the best conformer for the lead molecule assessment. Here, co-crystal ligand erlotinib are removed from crystal structure of EGFR (PDB ID: 1M17) and re-docking was performed with Crystal structure of EGFR and surprisingly RMSD valve was found to be 0.115Å. This, in turn efficacy and dynamicity of Glide is clearly showed**.** The docking simulation of Zinc24830081 into EGFR formed single hydrogen bonds with bond distance of 2.01Å. The side chain hydrogen atom of MET769 act as hydrogen bond donor to interact with oxygen atom of the lead molecule Zinc24830081. The glide score were calculated -8.43 Kcal/mol. The docked result of Zinco1312245 into EGFR contains two hydrogen bond interactions. The docking score was observed -6.83.0 Kcal/mol. The side chain hydrogen atom of the MET 769 and LYS721 were well interacted with oxygen atom of the lead molecule with bond length about 2.06Å and 2.27Å. The oxygen atom of the lead molecule act as hydrogen bond acceptor, hydrogen atom of the protein molecule act as hydrogen bond donor. The docking simulation of Zinc01838681 and Zinc01262238 into EGFR formed single hydrogen bond interaction with Met 769 with bond length of 1.91Å and 2.44Å respectively. The binding mode of hit molecules with target protein structure was depicted in figure 6 and 2d structure of them was represented in figure 7.

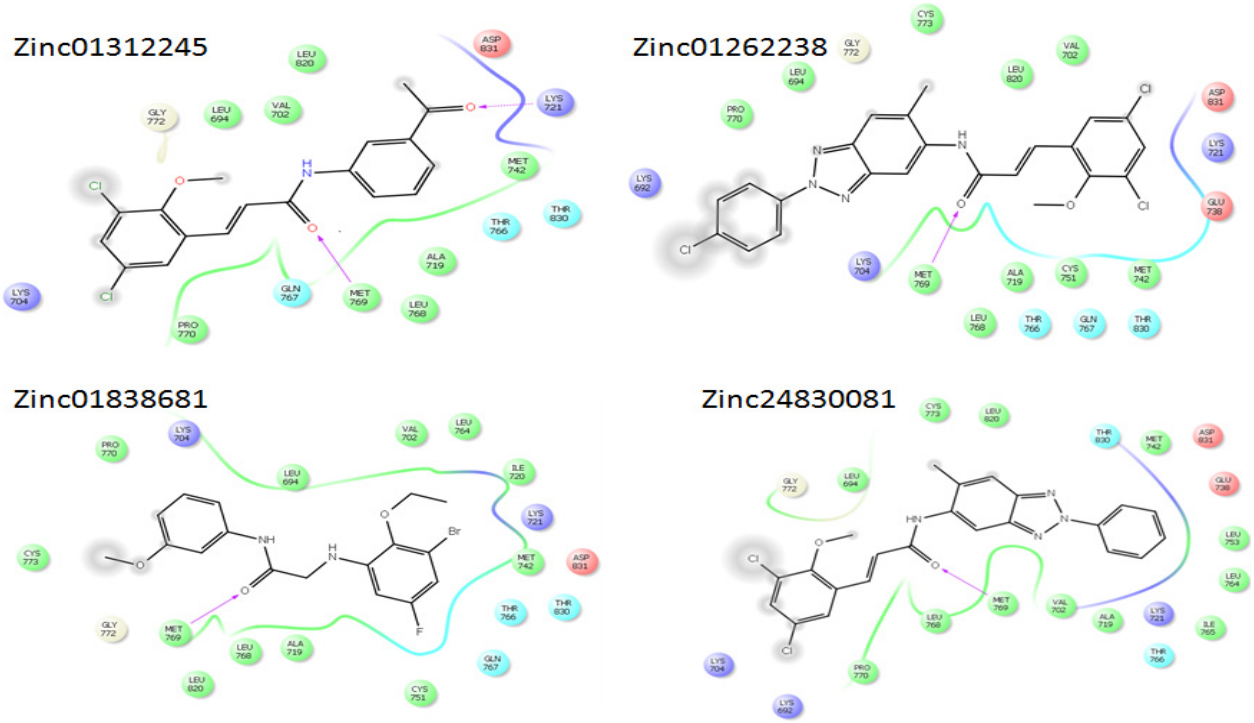

Figure 6: The binding mode of the four hit compounds

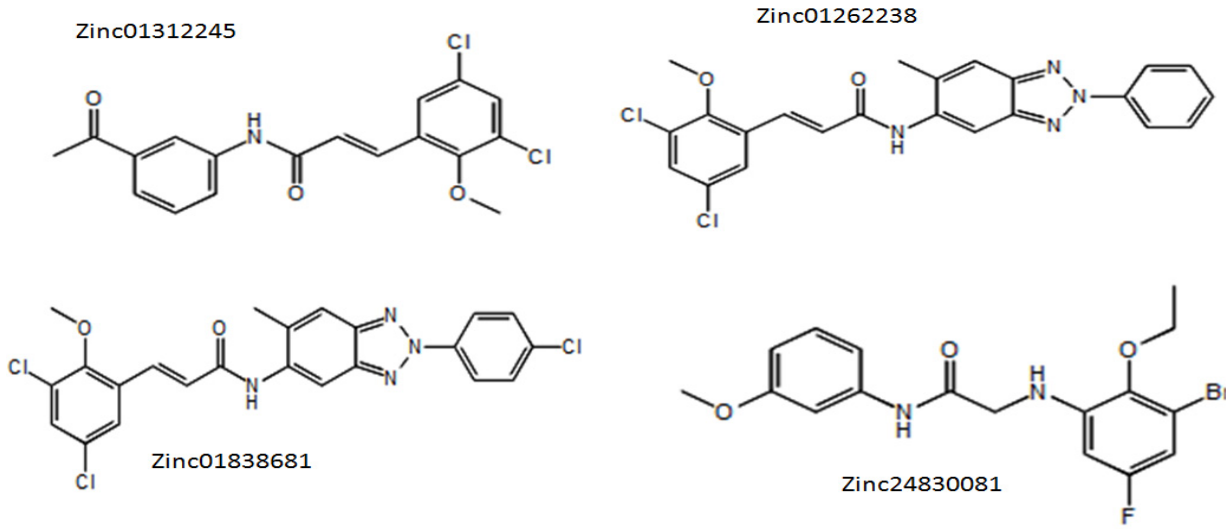

Figure 7: 2D Representation of hit molecules - Zinc01312245- (E)-N-(3-acetylphenyl)-3-(3,5-dichloro-2-methoxy-phenyl)acrylamide. Zinc01262238-(E)-3-(3,5-dichloro-2-methoxy-phenyl)-N-(6-methyl-2-phenyl-benzotriazol-5-yl)acrylamide.Zinc01838681-(E)-N-[2-(4-chlorophenyl) -6-methyl-benzotriazol-5-yl]-3-(3,5-dichloro-2-methoxy-phenyl)acrylamide. Zinc24830081- 2-[(3-bromo-2-ethoxy-5-fluoro-phenyl) amino]-N-(3-methoxyphenyl) acetamide

DISCUSSION:

The essence of recent advances in computational drug discovery helps in finding lead like molecules in short span of time. In this study our effort to elect novel inhibitors using ligand based designing approaches involved pharmacophore modeling and 3D QSAR. The pharmacophore hypothesis and statistical values for generated model is very robust. Observing all other toxicity profiles of the hit molecules, we strongly believe four categorized molecules would be a potent candidate in cancer drug discovery. And also need for its further *in vitro* studies is very clear.

CONCLUSION:

In this study, pharmacophore modeling, atom-based 3D-QSAR was developed for urea and thioureas derivatives of EGFR inhibitors using PHASE. A five point

pharmacophore with two hydrogen bond acceptor (A), one hydrogen bond donor (D), one hydrophobic group (H), one aromatic ring (R) as pharmacophore features were assessed by 3D-QSAR model. The hypotheses (AADHR) having good statistical results with R^2= 0.9803, Q^2=0.7099. Further, visualization of the 3D-QSAR model in the context of the molecules under study provides details of the relationship between structure and activity, and thus provides explicit indications for design better analogues. The results exposed that the substitution of OH. H, Cl enhances the activity of the compounds. Then the best hypothesis (AADHR) was utilized as 3D query to screen against Zinc database. More than 1000 compound were obtained from pre screening, finally four potent lead molecules identified by molecular docking studies. Four new lead compounds obtained can be targeted for further experimentation leading to clinical trials.

ACKNOWLEDGEMENT:

We thank whole heartedly Dr.K. Kovendan, young scientist, Bharathiar University for critical review of the manuscript. Authors thank Dr.P.Shanmugavel, Assistant Professor, Centre for bioinformatics for providing computational facilities to carry out the work.

REFERENCES:

1. Riese DJ, Stern DF Specificity within the EGF family ErbB receptor family signaling network. Bioessays 1998; 20(1):41-8.
2. Salomon DS, Brandt R, Ciardiello F, Normanno N Epidermal growth factor-related peptides and their receptors in human malignancies Crit Rev Oncol Hematol. 1995; 19(3):183-232.
3. Normanno N, Bianco C, De Luca A, Maiello MR, Salomon DS. Target-based agents against ErbB receptors and their ligands: a novel approach to cancer treatment. Endocr Relat Cancer. 2003; 10(1):1-21.
4. Hubbard SR, Till JH. Protein tyrosine kinase structure and function Annu Rev Biochem. 2000; 69:373-98.
5. Woodburn JR. The epidermal growth factor receptor and its inhibition in cancer therapy Pharmacol Ther. 1999; 82(2-3):241-50.
6. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001; 2(2):127-37.
7. Ghosh S, Liu XP, Zheng Y, Uckun FM. Rational design of potent and selective EGFR tyrosine kinase inhibitors as anticancer agents. Curr Cancer Drug Targets. 2001; 1(2):129-40.
8. Waterson AG, Petrov KG, Hornberger KR, Hubbard RD, Sammond DM et al., Synthesis and evaluation of aniline headgroups for alkynyl thienopyrimidine dual EGFR/ErbB-2 kinase inhibitors. Bioorg Med Chem Lett. 2009; 19(5):1332-6.
9. Rheault TR, Caferro TR, Dickerson SH, Donaldson KH, Gaul MD et al., Thienopyrimidine-based dual EGFR/ErbB-2 inhibitors. Bioorg Med Chem Lett. 2009; 19(3):817-20.
10. Lin R, Johnson SG, Connolly PJ, Wetter SK, Binnun E, Hughes TV, Murray WV, Pandey NB, Moreno-Mazza SJ, Adams M, Fuentes-Pesquera AR, Middleton SA. Synthesis and evaluation of 2,7-diamino-thiazolo[4,5-d] pyrimidine analogues as anti-tumor epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors. Bioorg Med Chem Lett. 2009; 19(8):2333-7.
11. Xu G, Abad MC, Connolly PJ, Neeper MP, Struble GT et al ., 4-Amino-6-arylamino-pyrimidine-5-carbaldehyde hydrazones as potent ErbB-2/EGFR dual kinase inhibitors. Bioorg Med Chem Lett. 2008; 18(16):4615-9.
12. Ducray R, Ballard P, Barlaam BC, Hickinson MD, Kettle JG, Ogilvie DJ, Trigwell CB. Novel 3-alkoxy-1H-pyrazolo[3,4-d]pyrimidines as EGFR and erbB2 receptor tyrosine kinase inhibitors Bioorg Med Chem Lett. 2008;18(3):959-62.
13. Dai Y, Guo Y, Frey RR, Ji Z, Curtin ML, Ahmed AA, Albert DH et al Thienopyrimidine ureas as novel and potent multi targeted receptor tyrosine kinase inhibitors. J Med Chem. 2005;48(19):6066-83.
14. Li HQ, Yan T, Yang Y, Shi L, Zhou CF, Zhu HL. Synthesis and structure-activity relationships of N-benzyl-N-(X-2-hydroxybenzyl)-N'-phenylureas and thioureas as antitumor agents. Bioorg Med Chem. 2010; 18(1):305-13.
15. *htp://www.cambridgesoft.com.*
16. Ligprep, Version 2.7, Schrodinger, LLC, New York, NY.
17. Halgren TA MMFF VI: MMFF94s option for energy minimization studies. J Comp Chem 1999; 20 (7): 720-729.
18. Jorgensen WL, Mazxwell DS, Tirado- Rives J. Development and testing of the OPLS all –atom force field on conformational energetic and properties of organic liquids. Journal of the American Chemical Society 1996;118 (45): 111225-11236.
19. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screenibng: 1Methodology and preliminary results. J Comput Aided Mol Des 2006; 20(10-11):647-71.
20. Phase, version 3.6. Schrödinger, New York, NY, 2013.

21. Stamos J, Sliwkowski MX, Eigenbrot C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J Biol Chem. 2002; 277(48):46265-72.
22. Protein preparation Wizard Maestro (2013) New York: Schrodinger LLC.
23. Glide, Version 6.0 (2013) Schrodinger, LLC, New York, NY.
24. Irwin JJ, Shoichet BK. ZINC- a free database of commercially available compounds for virtual screening. J Chen Inf Model 2005; 45(1):177–182.
25. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development. Advanced Drug Delivery Reviews 1997; 1; 46(1-3):3-26.

