Journal of Computational Methods in Molecular Design, 2013, 3 (2):10-25

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Discovery of novel EGFR inhibitors: In silico study and 3D-pharmacophore model generation.

Ola K. Sakka [a]*, Daisy H. Fleita[a], and Rafat M. Mohareb[b]

[a] *Department of Chemistry, American University in Cairo, P.O. Box 74, New Cairo, 11835, Egypt.*
[b]*Department of Chemistry, Faculty of Science, Cairo University, Giza, A.R., Egypt.*

ABSTRACT

In order to elucidate the essential structural features for Epidermal Growth Factor Receptor (EGFR) inhibitors, a ligand-based pharmacophore hypothesis was built on the basis of a set of twelve known EGFR inhibitors belonging to three different classes using Molecular Operating Environment (MOE) software. In a first step, three alignments, one for each group of compounds were generated. All of them were then submitted to MOE pharmacophore search in order to obtain a final pharmacophore model representative of the whole dataset. A pharmacophore model including three features was developed comprising one hydrogen-donor (F1) and two aromatic/ hydrophobic/ acceptor features (F2 and F3). The developed model was used to predict the activities of test set compounds by applying linear regression variable selection analysis. The model exhibited excellent linearity with correlation coefficient (r) value, i.e., 0.943, and squared predictive correlation coefficient (r^2) of 0.889 between experimental and predicted activity values of test set compounds. Our model demonstrated good performance in a separate test set of 25 compounds: it accurately identified 67.7% of the compounds of medium and high inhibitory activities and misclassified only 28.5% of the compounds with low inhibitory activities. The results proved our pharmacophore model to be a filter of great sensitivity and specificity.

Keywords: EGFR antagonists, pharmacophore, docking, MOE.

INTRODUCTION

Cancer is continuing to be a major health problem in developing as well as developed countries.[1,2] Surpassing heart diseases, it is taking the position number one killer due to various worldwide factors. Although major advances have been made in the chemotherapeutic management of some patients, the continued commitment to the laborious task of discovering new anticancer agents remains critically important. Protein tyrosine kinases (PTKs) are known for their role in cancer. The epidermal growth factor receptor (EGFR) belongs to the ErbB family, a subfamily of four closely related receptor tyrosine kinases: EGFR (ErbB1), HER2/neu (ErbB2), HER3 (ErbB3), and HER4 (ErbB4).[3] These receptors regulate intracellular signaling pathways mediating cell proliferation, differentiation, migration, survival, and adhesion.[4] ErbB family members, including the EGFR (ErbB1), are activated upon dimerization induced by binding their ligands, which are EGF, transforming growth factor-α (TGF-α) and Neu differentiation factor (NDF).[5] Over expression of EGFR leading to uncontrolled cell proliferation has been proven to be a definite cause of a significant number of human tumors (e.g. breast, ovarian, colon, and prostate) [6,7] and has been shown to affect proliferation, angiogenesis and cancer metastasis. Moreover, patients expressing high levels of EGFR usually

have poor prognosis. [8] Therefore, inhibitors of EGFR kinase activity may prove useful for therapeutic intervention in cancer as well as other proliferative disease. [9] Success in small molecule drug discovery against EGFR as an anticancer target has come from selective inhibition of its kinase activity using compounds that compete against ATP binding at the catalytic site such as the anilinoquinazoline derivatives gefitinib, Iressa and erlotinib that have recently been approved for the treatment of patients with advanced non-small cell lung carcinoma and are being clinically evaluated in patients with various types of cancers. [9] In addition, a great number of different structural classes of tyrosine kinase inhibitors has been reported and reviewed including pyrimido[4,5-*b*]-1,4-benzoxazepines [10], thiazolo[4,5-*d*]pyrimidines [11] ,diphenylamine 2,4'-dicarboxamides [12] and 1,4-dioxino quinazolines[13]. Thus, there exists a keen interest in understanding the structural determinants for substrates and inhibitors of EGFR. In this era, ligand-based pharmacophore modeling is playing an important role for the identification of ligand features for particular targets. The technique rigidly models the interaction between a ligand and its binding site in a specific binding situation. The result is a three-dimensional (3D) spatial arrangement of chemical features, which are derived using algorithms that take rules derived from chemical knowledge into account. A pharmacophore can be derived either in a structure-based manner by determining complementarities between a ligand and its binding site, or in a ligand-based manner, by flexibly overlaying a set of active molecules and determining those conformations that are able to be overlaid in such a way that a maximum number of important chemical features geometrically overlap .[14] A 3D pharmacophore model on compounds with observable structure diversity, if possible, will definitely lead to more universal and robust pharmacophore models for designing novel EGFR inhibitors. Numerous pharmacophore models for protein kinase inhibitors have been reported. [15, 16] Traxler et al. emphasized the importance of the presence of an acceptor–donor system in the ligand to mimick the anchoring of ATP to the active site of the enzyme and the presence of an aromatic feature to replace the ribose moiety of ATP conferring potency as well as selectivity for the EGFR- PTK ("sugar pocket") . [16] On the other hand, the pharmacophore model proposed by McGregor et al. indicated that the ligand-hydrogen bond acceptor interaction with the hinge region in the receptor is the most common interaction among kinase inhibitors and is made by N1 atom in ATP while a ligand hydrogen bond donor interaction on either side of the hinge region is present and is equivalent to the N6 atom of ATP. On the other hand, aromatic groups correspond to the 6- membered aromatic groups in ATP. In addition, hydrophobic features are found throughout the binding site but are most common in the hydrophobic inner region of the cleft and also in the parts occupied by the adenine and sugar moieties of ATP .[15]

However a limited number of models for EGFR inhibitors have been reported using a structurally diverse data set and to our knowledge, the correlation between pharmacophoric distances and predicted activities of possible EGFR inhibitors have never been attempted. Thus, in this study, pharmacophore modeling was brought into use by applying linear regression variable selection analysis to develop a novel 3D pharmacophore model that has not been reported earlier, with the sole purpose to assist the discovery of most potent EGFR inhibitors. The resulting validated pharmacophore model was then used to screen Maybridge database to identify structurally diverse EGFR inhibitors.

MATERIALS AND METHODS

Training Set

All the compounds under consideration were divided into training set and test set. The selection of a suitable training set is critical for the quality of automatically generated pharmacophore models. Our training set was composed of 12 compounds belonging to three different classes: pyrimido[4,5-*b*]-1,4-benzoxazepines[10], thiazolo [4,5-*d*]pyrimidines [11] and 1,4-dioxino quinazolines [13] which were demonstrated to be ATP competitive inhibitors of EGFR (Fig 1, Table 1) and were used to construct the pharmacophore model. Thirteen compounds (compounds **13-25**) were used as test set to evaluate the prediction capabilities of the generated model (Table 1) [17, 18]. Furthermore, to facilitate the modeling, the compounds were divided into three groups according to their activity data (expressed as IC_{50} values): highly active (0.09-3.40µM, +++), moderately active (3.43–11.29µM, ++), and least active (>11.29µM, +).This classification is highly beneficial when training the pharmacophore model with a broad range of activities, and also to access the estimation accuracy of pharmacophore quickly.

Energy Minimization and conformation generation

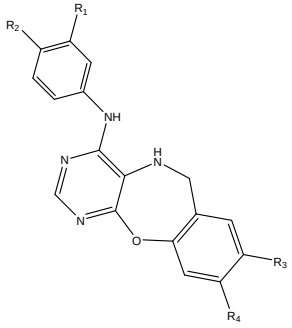
Prior to screening, all structures were built using 2D/3D editor sketcher in ChemDraw Ultra 8.0. [19] This allowed us to take into consideration molecular flexibility, thereby ensuring that fast-fitting would not be limited to rigid molecules with conformations already aligned to the pharmacophore. The molecular structures of **1**–**25** were energy

minimized within MOE (Molecular Operating Environment software, MOE 2008.10, Chemical Computing Group) using MMFF94 force field. This energy minimization methodology is capable of calculating constrained geometries through the use of chiral, distance, angle and dihedral restraints. [20] Hydrogen atoms and lone pairs were added to each molecule. Energy minimization was terminated when the root mean square gradient fell below 0.05. Force field partial charges were calculated prior to energy minimization. Conformational models were calculated using a 15 kcal energy cut off .The number of conformers generated for each substrate was limited to a maximum of 250.

Pharmacophore Model Generation

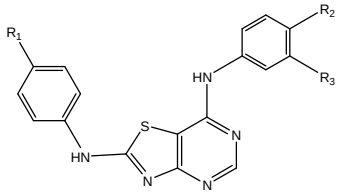
All molecules with their associated conformations were regrouped including their biological data.

The developed model can be based on one conformation of the most active molecule. Alternatively, it can be based on an alignment of several active molecules, a method that is used here. Such an alignment can be obtained by using MOE's flexible alignment, and all conformations of the molecules were considered for the alignment. After assigning MMFF94 charges to all molecules, flexible alignment was employed to scan and rank overlays of each class of compounds based on steric, electrostatic field, hydrophobic areas overlap, hydrogen bond donors and acceptors overlap. Each alignment is given a score that quantifies the quality of the alignment in terms of both internal strain and overlap of molecular features. Methodologies based upon 3D alignment for finding biologically active ligands generally make use of the qualitative assumption that if two ligands align well, they will possess similar biological activity. [21] Using the MOE pharmacophore consensus search module, setting tolerance to 1.40 and threshold to 100% a pharmacophoric model was generated using a scheme which was comprised of four different annotation points (H-bond donor, H-bond acceptor, hydrophobic and aromatic features). Starting from the best geometries obtained by conformational analysis, three alignments were derived, one for each class of compounds characterized by the highest accuracy and overlap of the most active compounds in each set. Using pharmacophore consensus as implemented in MOE, three different pharmacophore ligand-based models were developed in order to highlight the most important key features shown by each group of compounds belonging to the dataset (Fig 2-4). In addition, the availability of an x-ray structure of Erlotinib bound to the EGFR binding site (PDB ID: 1M17) was used as a template to create a fourth pharmacophore (Fig 5). Results indicated that all of them share the following requirements: two hydrophobic/aromatic features, one hydrogen bond donor function and two hydrogen acceptor functions (Table 2). To further validate this hypothesis, all the compounds in the training set were subjected to flexible alignment (Fig 6). Pharmacophore consensus was used again and indicated that three features contribute to the pharmacophore of EGFR which are spatially oriented so as to form a triangle: two hydrophobic/aromatic/acceptor moieties (F2 and F3) which are 5.48 A° apart from each other, in addition to a donor (F1) feature situated at 2.94 A° and 3.14 A° from both F2 and F3 features respectively (Fig 6). The default radii for F1, F2 and F3 features were set to 0.5A°, 0.9A° and 0.8A° respectively. Due to the wide structural diversity of our data set, the acceptor features failed to align together so they were omitted from the pharmacophore.



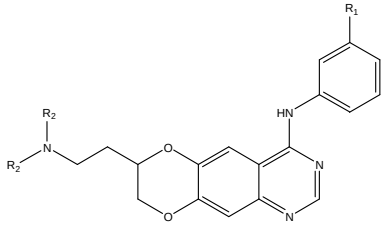
1-4

Pyrimido[4,5-*b*]-1,4-benzoxazepines



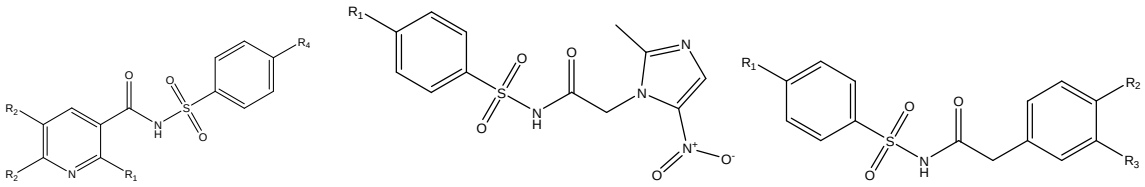
5-8

Thiazolo[4,5-*d*]pyrimidines



9-12

[1,4]-dioxino quinazolines



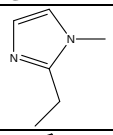
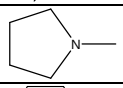
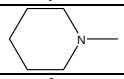
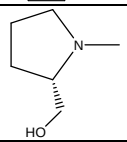
13-18 N-phenylsulfonylnicotinamides bezenesulfonamides

19-21 Metronidazole–sulfonamides

22-25 Phenylacetyl

Fig. 1 General structures of data set compounds 1-25

Table 1 Structures of compounds in data set

	No.	**R_1**	**R_2**	**R_3**	**R_4**	**IC $_{50}$(µM)**
Pyrimido[4,5-*b*]-1,4-benzoxazepines	1	Br	H	OCH_3	OCH_3	0.34
	2	Cl	F	OCH_3	H	0.9
	3	Cl	F	H	OCH_3	1.0
	4	Cl	F	F	H	1.0
Thiazolo[4,5-d] pyrimidines	5		F	Cl		0.006
	6		F	Cl		0.004
	7		F	Cl		0.010
	8		F	Cl		0.011
[1,4]-dioxino quinazolines	9	Br	N-methyl piperazine			0.036
	10	Cl	Morpholine			0.022
	11	Br	morpholine			0.042
	12	Br	piperidine			0.052
N-phenylsulfonyl nicotinamides	13	H	Br	H	Cl	0.09
	14	H	Br	H	Me	12.11
	15	H	H	Cl	Br	29.07
	16	H	H	H	Me	24.83
	17	Cl	H	H	Me	27.05
	18	H	H	Cl	Cl	18.15
Metronidazole–sulfonamides	19	H	-	-	-	2.94
	20	Me	-	-	-	3.43
	21	Br	-	-	-	2.17
Phenylacetyl bezenesulfonamides	22	H	F	H	-	6.74
	23	Me	Cl	H	-	13.73
	24	Br	Cl	H	-	11.29
	25	F	OMe	H	-	21.84
Erlotinib	26	-	-	-	-	0.03

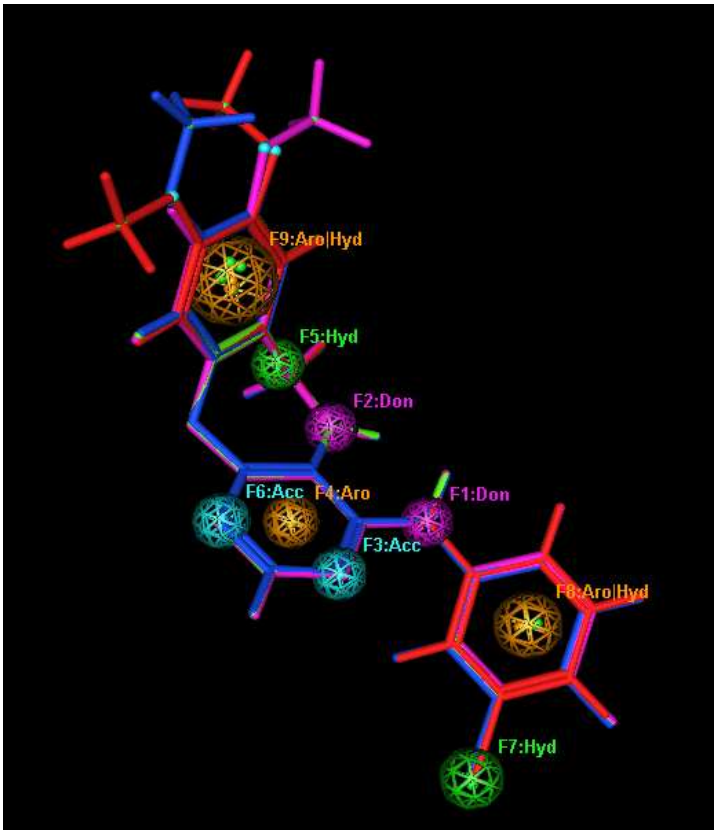


Fig. 2 Flexible alignment of azepines 1-4 mapped to the pharmacophoric model for EGFR activity. Pharmacophore features shared by compounds 1-4 are color coded: F1 and F2: purple for a hydrogen bond donor, F3 and F6: cyan for a hydrogen bond acceptor, F4, F8 and F9: orange for hydrophobic/ aromatics, F5 and F7: green for hydrophobics

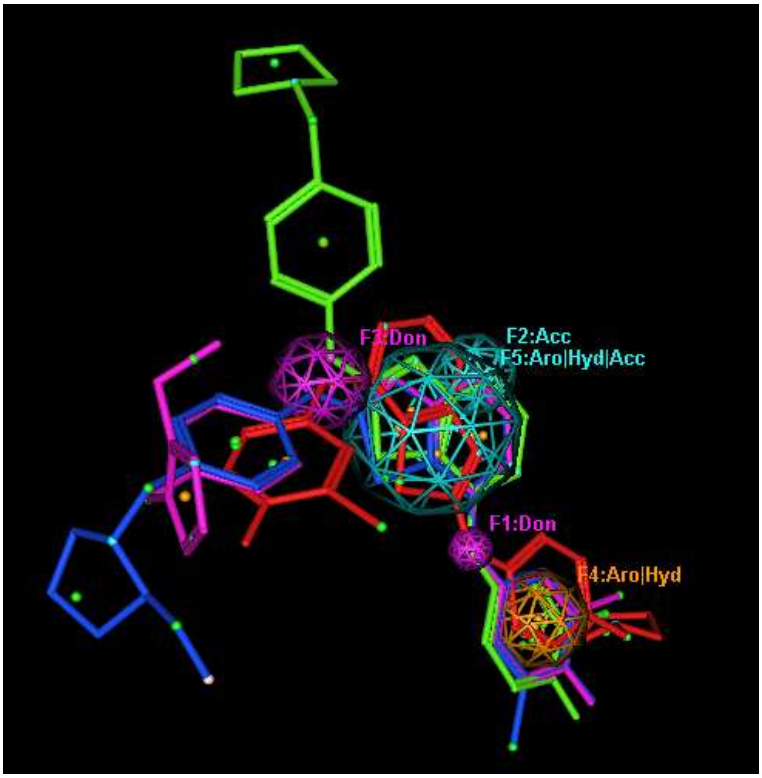


Fig. 3 Flexible alignment of thiazolo [4,5-*d*] pyrimidines 5-8 mapped to the pharmacophoric model for EGFR activity. Pharmacophore features shared by compounds 5-8 are color coded: F1 and F3: purple for a hydrogen bond donor, F2 and F5: cyan for a hydrogen bond acceptor or aromatic/ hydrophobic/acceptor feature, F4: orange for hydrophobic aromatics

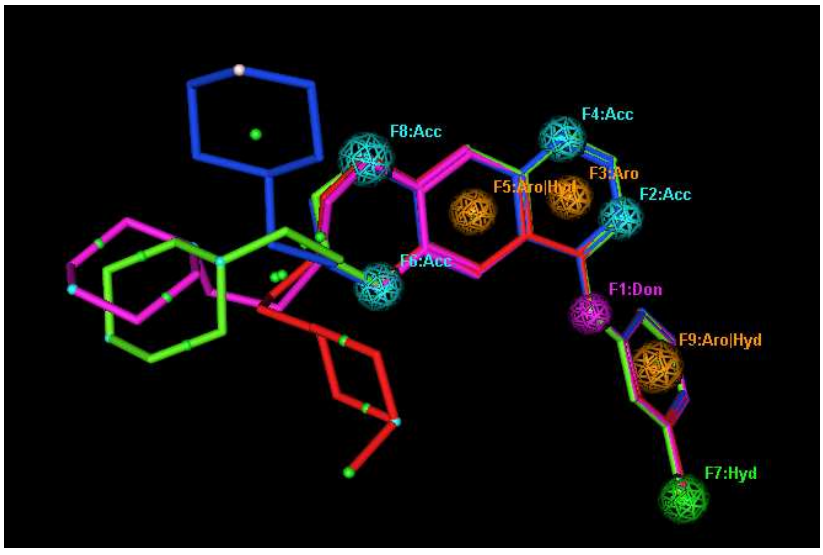


Fig. 4 Flexible alignment of [1,4]-dioxino quinazolines 9-12 mapped to the pharmacophoric model for EGFR activity. Pharmacophore features shared by compounds 9-12 are color coded: F1: purple for a hydrogen bond donor, F2, F4, F6 and F8: cyan for a hydrogen bond acceptor, F3, F5 and F9: orange for hydrophobic/ aromatics, F7: green for hydrophobics

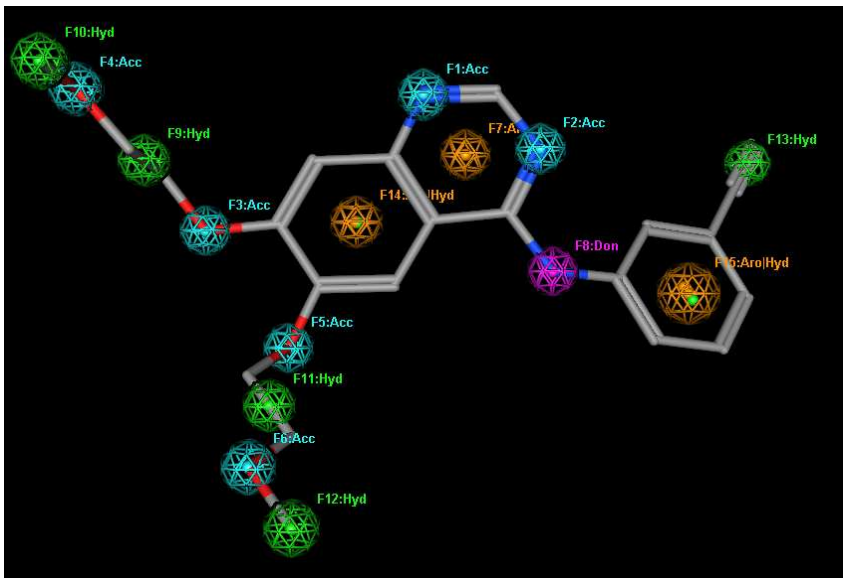


Fig. 5 Pharmacophore model for Erlotinib

Table 2 All Pharmacophoric features of each class

pounds	Don	Acc	Aro/hyd	Hyd
Azepines	two	two	three	two
Thiazolo[4,5-*d*]pyrimidines	two	two	two	-
1,4-dioxino quinazolines	one	four	three	one
Erlotinib	one	six	three	five

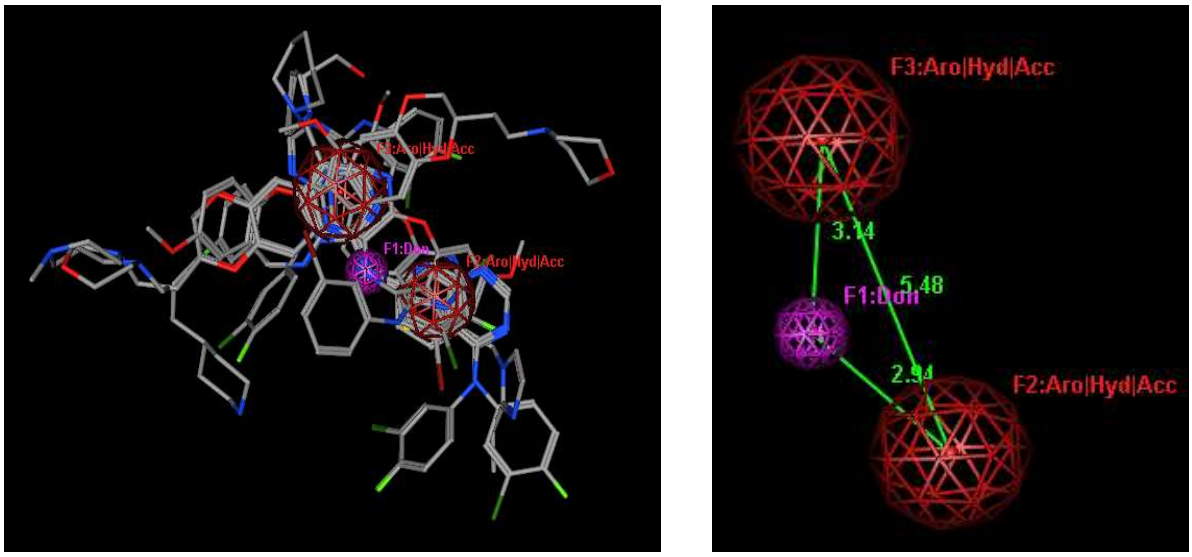


Fig. 6 Flexible alignment of training set compounds 1-12 and Erlotinib (left panel) and the best predicted pharmacophore features and geometries which are required for EGFR activity(right panel). Pharmacophore features are color coded: F1: purple for a hydrogen bond donor, F2 and F3: red for aromatic/hydrophobic /acceptor feature

Validation of pharmacophoric hypothesis

Test Set Prediction

The validity and predictive character of our model were further assessed by predicting the activity of test set molecules .This was performed by studying the significance of distances between different structural features and EGFR inhibitory activity of the test set. Accordingly, linear regression variable selection analysis was applied using SPSS software, version 20.0, SPSS Inc., USA. Regression analysis was performed using EGFR inhibitory activity (IC_{50}) as dependent variable and the calculated distances (F1-F2, F2-F3 and F1-F3) as independent variables. Three equations were exported using regression analysis. The quality of each equation was assessed using the statistical parameters viz., correlation coefficient (r), squared predictive correlation coefficient (r^2) and standard error of estimate (s). Significant distances were obtained as shown in Table 3. Among the three equations reported, equation 3 had the highest correlation coefficient (r), the highest squared predictive correlation coefficient (r^2) and the lowest standard error of estimate (s) (Table 4). The model exhibited excellent linearity with R value, i.e., 0.943, and squared predictive correlation coefficient (r^2) of 0.889 was also observed between experimental and predicted activity values of test set molecules. Thus this pharmacophore can predict over 94.3% of the true activity. According to equation 3, all three distances were found to be essential for EGFR inhibition in the pharmacophore model, F1–F2, F2–F3 and F2–F4.

Table 3 Measured distances between the test set compounds structural features

C	F1-F3	F2-F3	F1-F2
1	2.78	5.66	3.35
1	3.82	6.36	3.53
1	3.77	7.03	3.84
1	3.82	6.79	3.59
1	3.78	6.83	3.77
1	3.73	6.55	3.14
1	2.84	5.48	3.79
2	3.86	6.01	3.35
2	3.86	6.01	3.35
2	3.05	5.79	3.41
2	3.69	6.21	3.30
2	3.00	6.58	3.34
2	3.14	6.58	3.36

Table 4 Significance of the distances between the pharmacophoric features

Equation	% Inhibition	r	r^2	s
1	-133.477+18.295(F2-F3)+9.107(F1-F2)	0.941	0.885	3.881
2	-106.817+20.308(F2-F3)-2.232(F1-F3)	0.925	0.855	4.205
3	-132.353+19.240(F2-F3)+8.866(F1-F2)-1.804(F1-F3)	0.943	0.889	3.874

Applying equation 3 to the tested compounds, their EGFR inhibitory activity was predicted (Table 5). The correlation plot generated from the regression analysis showed a linear relationship between the actual and predicted activities among the training set of 25 compounds (Figure 7). In this study, test set compounds were classified by their activity as highly active (0.09-3.40µM, +++), moderately active (3.43–11.29µM, ++), and least active (>11.29µM+).The model successfully predicted the activities of most compounds. The error value (residual) was calculated as the difference between the predicted and experimental activities. A positive error value indicates that the predicted activity is higher than that obtained experimentally, while a negative error value indicates that the predicted activity is lower than that obtained experimentally. The activities of only two compounds were either overestimated or underestimated. Compound **21** was actually highly active (+++) but was underestimated as moderately active (++), compound **24** was moderately active (++) but was estimated as least active (+).However, the differences between the actual and estimated activities showed an error value of less than ten, hence these values shows that the generated model is statistically significant.

Table 5 Evaluated and predicted inhibitory activity (IC_{50}) of the test set compounds using equation (3).

Compound no.	Actual Activity(μM)	Estimated Activity(μM)	Residual[a]	Actual Activity Scale	Estimated Activity Scale
13	.09	1.23523	-1.14523	+++	+++
14	12.11	14.42364	-2.31364	+	+
15	29.07	30.15332	-1.08332	+	+
16	24.83	23.22890	1.60110	+	+
17	27.05	25.66660	1.38340	+	+
18	18.15	14.78371	3.36629	+	+
19	2.94	1.56499	1.37501	+++	+++
20	3.43	6.02149	-2.59149	++	++
21	2.17	6.02149	-3.85149	+++	++
22	6.74	3.78148	2.95852	++	++
23	13.73	9.73281	3.99719	+	+
24	11.29	18.45076	-7.16076	++	+
25	21.84	18.37559	3.46441	+	+

[a] *Residual= Actual Activity-Estimated Activity*

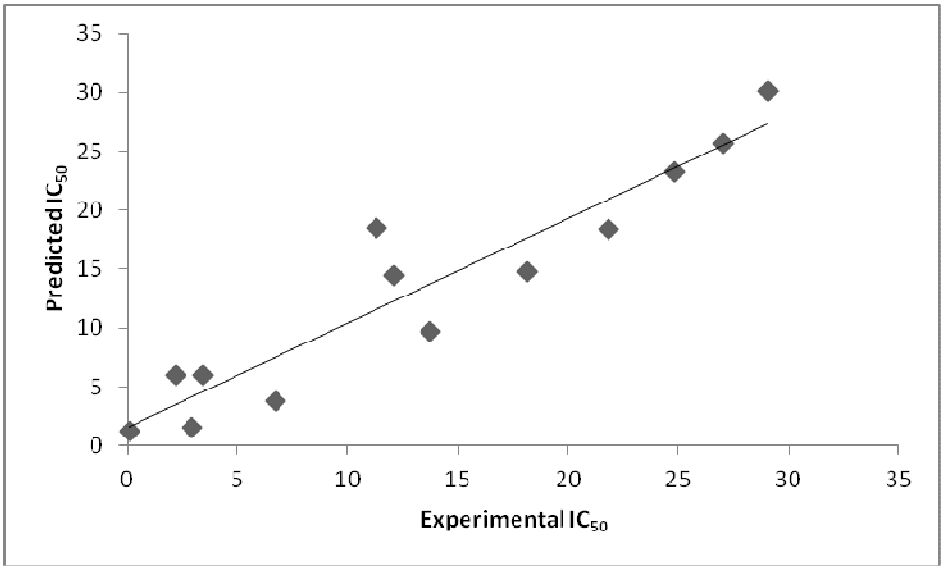


Fig. 7 Relation between experimental and predicted EGFR inhibitory activity values of test set molecules using our predicted pharmacophore model

Pharmacophore mapping of test set compounds

The pharmacophore model was also validated for its predictive power by mapping test set compounds onto the model and comparing the resultant root mean square deviation (RMSD) values. RMSD value refers to the root of the mean square distance between the query features and their matching ligand target points. To be considered as hit, the compound has to fit all the features of the pharmacophore. The program expresses the degree of mapping of a given compound to a generated hypothetical model in terms of RMSD value, which in turn is correlated with its activity. The higher the RMSD value, the higher the expected activity against EGFR. To obtain an accurate prediction for RMSD values of test set compounds, known EGFR inhibitor Erlotinib (PDB ID: 1M17, IC_{50} 0.03) was docked into the ATP binding site of EGFR using our proposed pharmacophore model and an RMSD value of 0.1360 was obtained. Next, all compounds in the test set were docked into the active site using our pharmacophore as the query. Results indicated that the RMSD values for the most active compounds in the test set were close to that of Erlotinib. Our model was able to identify 4 (4/6 or 67.7%) of the compounds of medium and high inhibitory activities and misclassify only 2(2/7 or 28.5%) of the compounds with low inhibitory activities. Results listed in Table 6 revealed that the lower RMSD values were obtained for the highly active compounds. Figure 8 shows how our model maps

to representative highly active compound **13** and least active compound **15** from the test set. Compound **13** has RMSD value of 0.1656 and IC_{50} value of 0.09.The NH group fitted the region of the F1 donor feature, while the pyridine ring fitted the F2 aromatic/hydrophobic/acceptor feature. In addition, the oxygen atom belonging to the SO_2 moiety, aligned well with the F3 feature. On the other hand, compound **15** (RMSD 0.3266, IC_{50} 29.07) was selected as a representative example of the compounds that partially matched our pharmacophore and showed fitting to only F1 and F2 features of the pharmacophoric model.

Table 6 RMSD values for test set compounds

Compound no.	IC_{50}	RMSD	Compound no.	IC_{50}	RMSD
13	.09	0.1656	**20**	3.43	0.1656
14	12.11	0.2493	**21**	2.17	0.1656
15	29.07	0.3266	**22**	6.74	0.3469
16	24.83	0.3127	**23**	13.73	0.3465
17	27.05	0.2149	**24**	11.29	0.3465
18	18.15	0.2575	**25**	21.84	0.3469
19	2.94	0.1656	**Erlotinib**	0.03	0.1360

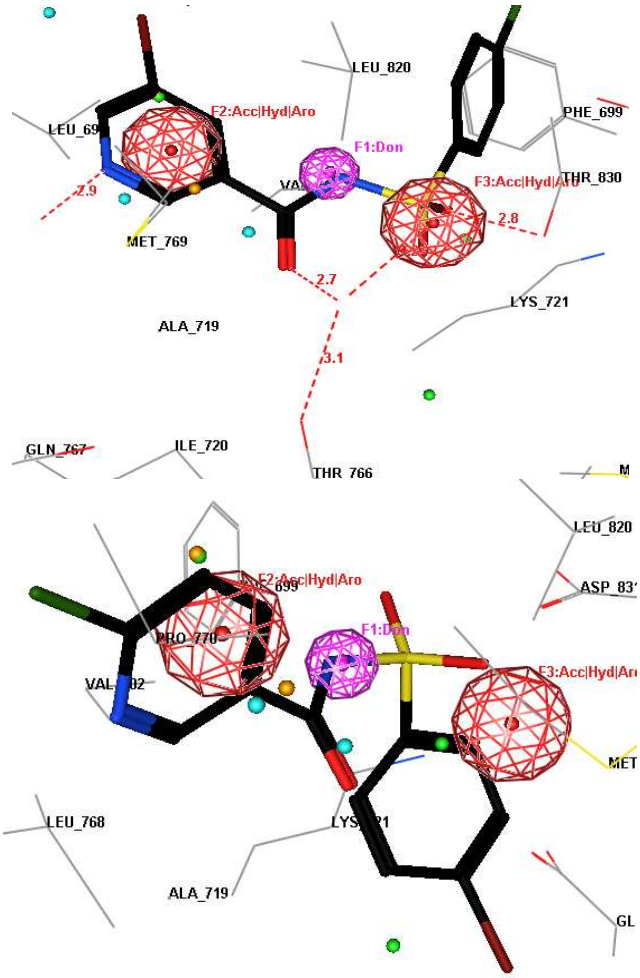


Fig.8 Mapping of highly active compound 13(IC_{50} 0.09µM, RMSD 0.1656) (left) and least active compound 15 (IC_{50} 29.07µM, RMSD 0.3266) (right) onto the pharmacophore model

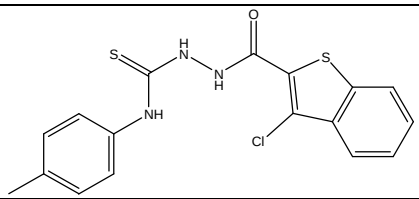
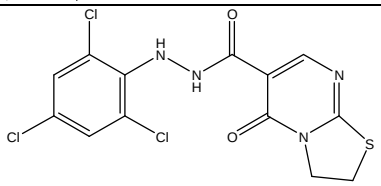
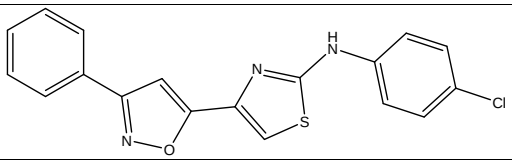
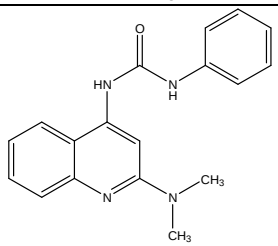
To understand the relative importance of each individual pharmacophore feature on the sensitivity and specificity of the model, the following test was performed. Each pharmacophore feature was removed individually and the test set was scanned using the reduced pharmacophore model consisting of only the remaining two pharmacophore features. The test result is shown in Table 7. The removal of any one pharmacophore feature caused a minor decrease in the number of true positives but a dramatic increase in the number of false positives. Specifically, removal of either F1 or F2 resulted in a false positive rate of 71.4% and 57.1% respectively. Therefore, the F1 aromatic/Hydrophobic/Acceptor feature and F2 donor feature residue were the most critical pharmacophore features in ensuring the specificity of the model.

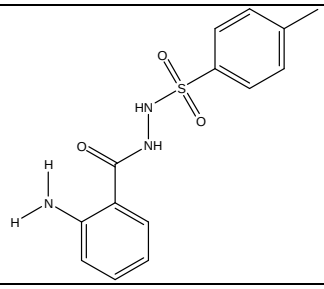
Table 7 The relative importance of each individual pharmacophore feature on the sensitivity and specificity of the model, as indicated by the success rate and the false positive rate after the feature was removed

ID	Feature	Success Rate after the feature was removed[a]	False positive rate after the feature was removed[b]
F1	Acceptor/Hydrophobic/Aromatic	4/6(66.7)	5/7(71.4)
F2	Donor	3/6(50)	4/7(57.1)
F3	Acceptor/Hydrophobic/Aromatic	4/6(62.5)	2/7(28.5)

[a] Study conducted among compounds of high and moderate activities
[b]Study conducted among compounds with low activities
Values expressed in percentage are given in parentheses

Table 8 Result of 3D search of Maybridge database

Compound	Structure	RMSD	Predicted Activity	Lipinski Violation
BTB00810		0.0355	0.632	0
SEW01394		0.0346	1.271	0
SPB04883		0.0367	2.707	0
MWP01055		0.0453	2.727	0

BTB02067		0.0422	2.989	0

Database screening

Pharmacophore models are useful to formulate a query to search chemical collections in search of mechanistically homogeneous but structurally diverse scaffolds. In the present study, the validated three feature pharmacophore was used to screen molecules from the Mini Maybridge (56,000 compounds) .Out of the 548 hits obtained from the Mini Maybridge database, five structurally diverse leads, **SPB04883**, **MWP01055**, **BTB02067**, **SEW01394** and **BTB00810** were selected on the basis of RMSD values, potencies and druggable properties (see next sub heading). The structures and estimated activities of selected leads are shown in Table 8. All five compounds showed a good fit with our proposed model (Fig 9). The most active compound, compound **BTB00810**, of Mini Maybridge database, showed good fit with all the three features. In this case, the F3 (hyd/acc/aromatic) feature is mapped by sulfur atom, the F1 donor feature is mapped by an NH group and the F2 (hyd/acc/aro) feature is mapped by the oxygen of the carbonyl group. The second most active compound, compound **SEW01394**, also showed a best fit with all features of our pharmacophore hypothesis. In this case, one of the chlorine atoms fits the F3 (hyd/acc/aro) feature, one the NH atoms maps well to the F1 donor feature and the oxygen atom of the carbonyl group maps to the F2 (acc/hyd/aro) feature. The third most active compound **SPB04883,** also mapped well with our pharmacophore. In this case, the F2 (acc/hyd/aro) feature is fitted by the benzene ring, The F1 donor feature is fitted by an NH group and the F3 (acc/hyd/aro) feature is fitted by the thiazole ring.

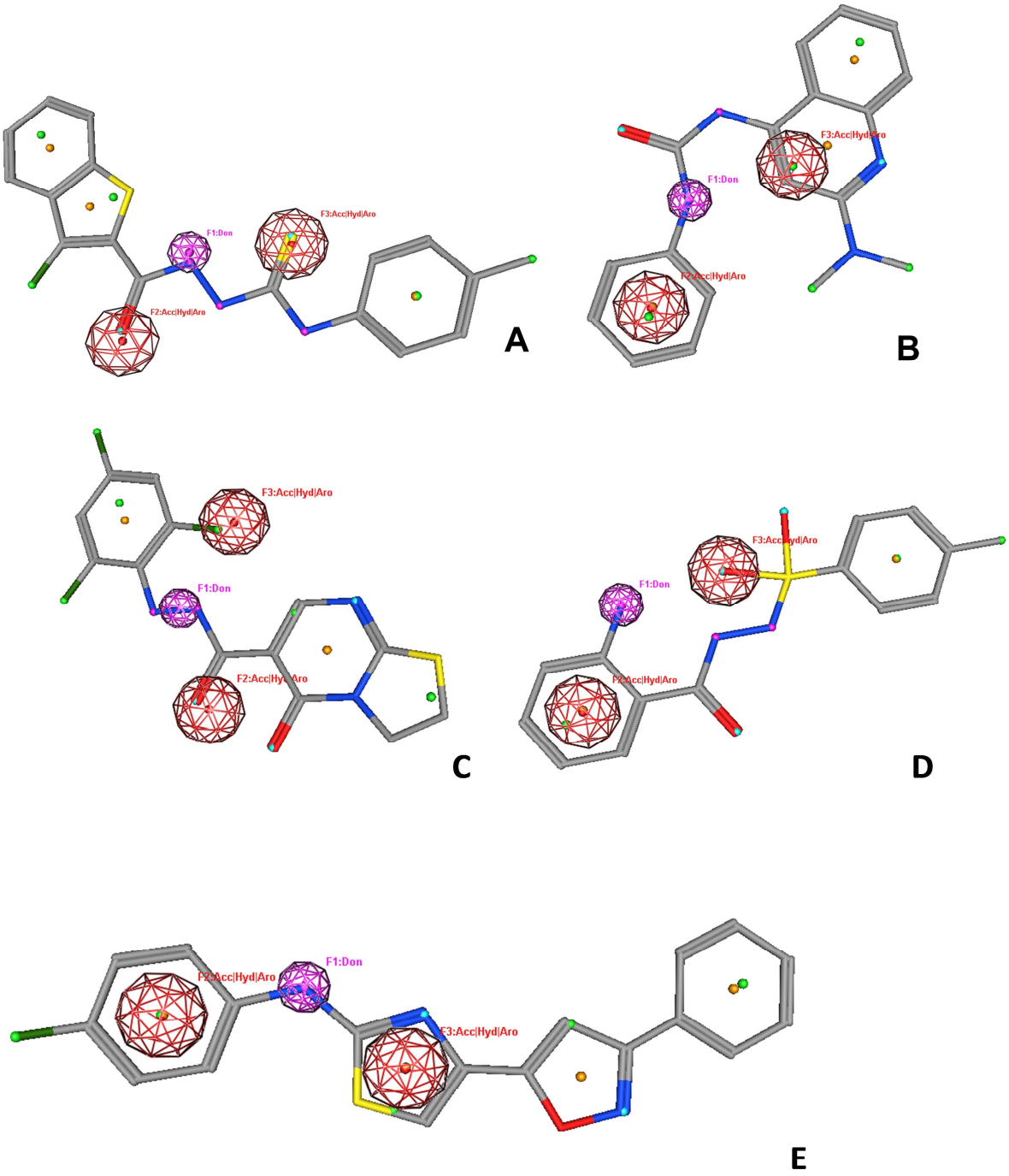


Fig 9A-E Mapping of **BTB00810, MWP01055**, **SEW01394**, **BTB02067** and **SPB04883** onto the pharmacophoric model respectively

Drug –Like property calculation

In recent years, one of the tools for predicting drug likeness, which discriminates between drug-like and non drug-like compounds, is the Lipinski's rule of five which takes into consideration, the compound's molecular weight, hydrophilicity (cLogP), number of hydrogen bond donors, and number of hydrogen bond acceptors. [22] According to the results obtained using the Molinspiration software, none of the compounds violate any of the Lipinski's criteria, an important characteristic for future drug development. Additional SAR parameters were calculated using the Osiris program such as solubility (LogS), drug likeness and drugscore. Osiris program druglikeness values are calculated from 15000 Fluka compounds and from 3300 traded drugs. A positive value states that your molecule contains predominantly fragments which are frequently present in commercial drugs. The drug score combines druglikeness and overall potential to qualify for a drug. A high solubility and a low hydrophilicity contribute to a compound's absorption or permeation ability. Compound **BTB00810** which displayed highest predicted activity possessed a reasonable hydrophilicity, a reasonably high solubility and the highest drug score compared to all active compounds (Table 9). Thus there is a good correlation between calculated SAR parameters and the predicted activity of compounds screened from Maybridge database.

Table 9 Physico-chemical and absorption properties for the most active compounds

Compound	cLogP[a]	MW[b]	n-OHNH donors[c]	n-ON acceptors[d]	Lipinski's violations	Log S[e]	Drug-likeness	Drug Score
BTB00810	4.495	375.9	3	2	0	-3.81	2.93	0.67
SEW01394	2.919	391.666	2	3	0	-4.57	4.27	0.25
SPB04883	5.862	353.8	1	2	0	-6.01	1.66	0.36
MWP01055	3.945	306.4	2	2	0	-4.91	-0.16	0.47
BTB02067	1.201	305.3	3	3	0	-1.93	-6.16	0.13

a: cLogP = logarithm of compound partition coefficient between n-octanol and water
b: MW= molecular weight
c: n-OHNH number of hydrogen bond donors
d: n-ON number of hydrogen bond acceptors
e: aqueous solubility

Docking Studies

To compare the binding mode of most active compound **BTB00810** to the known EGFR inhibitor Erlotinib, a docking study into the ATP binding site of epidermal growth factor receptor (EGFR) was performed. Structure coordinates for the crystal structure of Erlotinib-EGFR inhibitor were used to define the binding site and were obtained from RCSB Protein Data Bank (PDB ID: 1M17). Docking study of the designed compound into EGFR tyrosine kinase was performed using MOE. Prior to docking, Erlotinib was energy-minimized with a MMFF94 force-field till the gradient convergence 0.01 kcal/mol was reached.

The ATP binding site of EGFR has the following features; Adenine region: contains two key hydrogen bonds formed by the interaction of the purine base of ATP and the protein backbone between amino acids Gln767 and Met769. Hydrophobic pockets: Though not used by ATP but plays an important role in inhibitor selectivity. Phosphate binding region: This is used for improving inhibitor selectivity .[23] Erlotinib binds to the adenine region of EGFR by mimicking the binding of ATP with the protein backbone via an interaction between the N-1 of the quinazoline which accepts an H-bond from the backbone Met-769 amide nitrogen existing in the hinge region of EGFR. Another significant interaction comes from the presence of a water mediated hydrogen bond between N-3 of quinazoline ring and a water molecule near Thr-766 side chain of EGFR molecule.[24] Our docking studies on **BTB00810** have revealed the presence of two hydrogen bonding interactions with Met769 residue. The first interaction is the donation of a hydrogen from the NH group of **BTB00810** to Met769 residue present in the backbone of EGFR (3.0A°). The second interaction is the donation of a hydrogen atom from NH group of Met 769 to the oxygen atom of carbonyl group present in **BTB00810** (2.9°) (Fig 10). Comparing the docking mode of our compound with Erlotinib, it could be postulated that the designed compound might bind even better than Erlotinib to EGFR since it forms two Hydrogen bonding interactions with Met 769 residue in the ATP binding site.

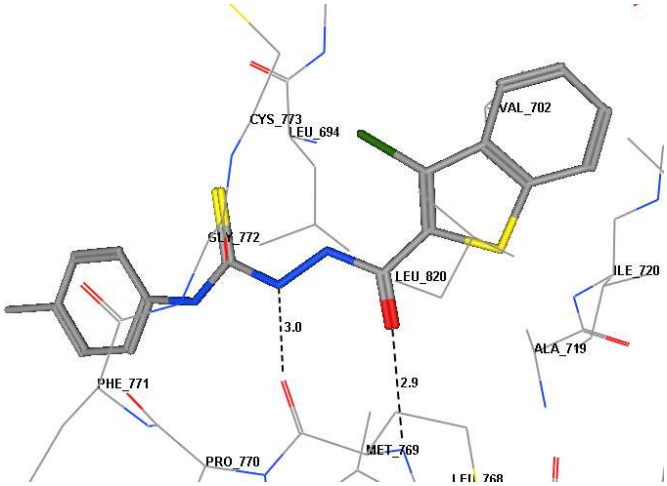


Fig. 10 Docking of compound BTB00810 into the binding site of EGFR

CONCLUSION

In this study, three diverse groups of EGFR inhibitors were used to propose a good pharmacophore model , which shed light on the important role of the donor, acceptor, aromatic and hydrophobic moieties for recognition and binding to receptor sites and the significance of the distances between such features. The generated model was validated by two methods, test set prediction, and mapping test set compounds onto the model and comparing the resultant RMSD values. The sensitivity and specificity of the model was tested by determining the success rate and false positive rate after each feature was removed separately. The validated pharmacophore model was then used for searching new lead compounds. Five structurally diverse compounds from Maybridge database with µM activity against EGFR were retrieved. Undoubtedly, the identified leads have the potential for their development as EGFR inhibitors. To our knowledge, this is the first proposed pharmacophore model that attempted to correlate activities with distances between pharmacophoric features. Thus this model can be helpful in future investigations regarding the design of more potent and specific drugs for EGFR inhibition and can be utilized to predict the activity of a wide variety of chemical scaffolds.

REFERENCES

[1] S Eckhardt. *Curr Med Chem Anti-Canc Agents* , **2002**, 2, 419.
[2] CW Lee, DH Hong, SB Han, SH Jong, HC Kim, RL Fine, SH Lee, HM H. Kim. *Biochem Pharmacol*, **2002**, 64,473.
[3] P Blume-Jensen,T Hunter. *Nature*, **2011**, 411, 355.
[4] J Mendelsohn, J Baselga. *Oncogene*, **2000**, 19, 6550.
[5] BP Ceresa, SL Schmid. *Curr Opin Cell Biol*, **2000**, 12, 204.
[6] R Sordella, DW Bell, DA Haber, Settleman. *J Science*, **2004**, 305, 1163.
[7] H Zhang, A Berezov, Q Wang, G Zhang, J Drebin, R Murali, MI Greene. *J Clin Invest*, **2007**, 117, 2051.
[8] J Brabenber, KD Danenberg, R Metzger, PM Schneider, J Park, D Salonga, D Holscher, P Danenberg. *Clin Cancer Res*, **2001**, 7, 1850.
[9] JR Woodburn .*Pharmacol Ther*, **1999**, 82, 241.
[10] L Smith, L Evgueni, EL Piatnitski, AS Kiselyov, X Ouyang, X Chen, S Burdzovic-Wizemann, Y Xu, Y Wang, RL Rosler, SN Patel, HH Chiang, DL Milligan, J Columbus, WC Wong, JF Doody, YR Hadari. *Bio Med Chem Lett*, **2006**, 16, 1643.
[11] R Lin, SG Johnson, PJ Connolly, SK Wetter, E Binnun, TV Hughes, WV Murray, NB Pandey, SJ Moreno-Mazza, M Adams, AR Fuentes-Pesquera, SA Middleton. *Bioorg Med Chem Lett*, **2009**, 19, 2333.

[12] SM Abou-Seri, NA Farag, GS Hassan .*Chem Pharm Bull* , **2011**, 59,1124.
[13] JD Ha, SK Kang, K-D Kim, J-K Choi, J-Y Kong, CH Ahn. *Bull Korean Chem Soc*, **2005**, 26, 959.
[14] G Wolber, T Seidel, F Bendix, T Langer. *Drug Discov Today*, **2008**, 13, 23.
[15] MJ McGregor.*J Chem Inf Model* , **2007**,47, 2374.
[16] P Traxler, G Bold, J Frei, M Lang, N Lydon, H Mett, E Buchdunger, T Meyer, M Mueller, P Furet. *J Med Chem*, **1997**, 40, 3601.
[17] H Zhang, X Lu, LR Zhang, JJ Liu, XH Yang, XM Wang, HL Zhu. Bioorg Med Chem, **2012**, 20, 1411.
[18] Y Luo, Y Li, KM Qui, X Lu, J Fu, HL Zhu. *Bioorg Med Chem*, **2011**, 19, 6069.
[19] A Zielesny. *J Chem Inf Model*, **2005**, 45, 1474.
[20] TA Halgren. *J Comp Chem*, **1996**, 17, 490.
[21] S Kearsley. *Tetrahedron Comput Methodol* , **1990**, 3, 615.
[22] A Lipinski, F Lombardo, FW Dominy , PJ Feeney . *Adv Drug Delivery Rev*, **2001**, 46, 3.
[23] D Fabbro et al. *Pharmacol Ther*, **2002**, 93, 79.
[24] J Stamos, MX Sliwkowski, C Eigenbrot. *J Biol Chem*, **2002**, 277,46265.