CHAPTER 7

PHARMACOPHORE MODELING ON EGFR KINASE INHIBITORS: A NOVEL STRATEGY FOR LIGAND BASED VIRTUAL SCREENING

7.1 Introduction

Both the reliability and utility of specific virtual screening (VS) tools grow with the knowledge available on a particular drug target and the underlying ligand–receptor interaction pattern. A widely used form of VS employs protein–ligand docking [7.1, 7.2]. Most docking approaches keep the receptor rigid and the ligand flexible with respect to conformation, orientation and position in the active site during docking. Considering protein flexibility in docking simulations drastically increases the complexity of the problem [7.3, 7.4]. Recent methods try to relax a part of the protein (side chains or in some cases even the backbone) [7.5].

In cases where there is no structural information available on the target or if it is present but difficult to model, ligand based pharmacophore models used with chemical and other filters, [7.6–7.11] can provide powerful tools for VS. There have been continuing efforts to design methods based on pharmacophore model for ligand based virtual screening (LBVS) [7.12–7.21]. The aim of the present chapter is to model a LBVS method based on pharmacophore modeling for epidermal growth factor receptor (EGFR) kinase inhibitors. To supplement the LBVS model a detailed validation of pharmacophore model was carried out by activity prediction, and mining of EGFR and non-EGFR kinase inhibitor databases. Further structure based pharmacophore modeling and docking study was carried out on eight known inhibitors. The structural basis of ligand selectivity on active and inactive form of receptor was revealed through docking study.

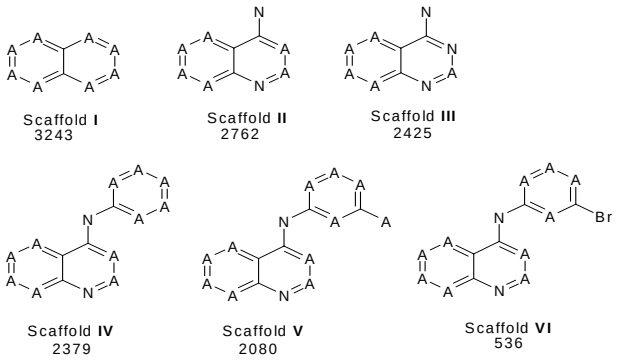
To explore the possible substitutions to the anilinoquinazoline moiety that could enhance the potency of the inhibitor, 3D-QSAR studies of this class of molecules has been done earlier from our group and by others [7.22–7.24]. Further, a study to select an appropriate combination of docking and scoring functions was carried out for EGFR kinase domain [7.25]. EGFR is frequently over-expressed in a broad range of human cancers (e.g. squamous cell carcinoma, breast, ovarian, non-small cell lung cancer etc.). EGFR is therefore a potential target for various cancer therapies [7.26–7.30]. Ligands binding EGFR are divided into two main categories, extracellular and intracellular, depending on the

receptor region targeted [7.30–7.32]. A number of compounds have shown both selectivity towards EFGR kinase domain, and potency. Many selective inhibitors of EGFR kinase are presently in the clinical trials at different phases or even marketed [7.33–7.40]. Ligands selected in this study are targeted to the kinase domain of EGFR which is intracellular.

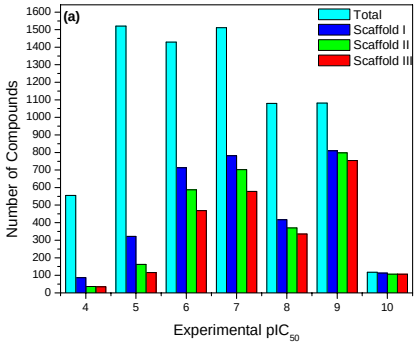
7.2 Materials and methods

7.2.1 *Database profile*

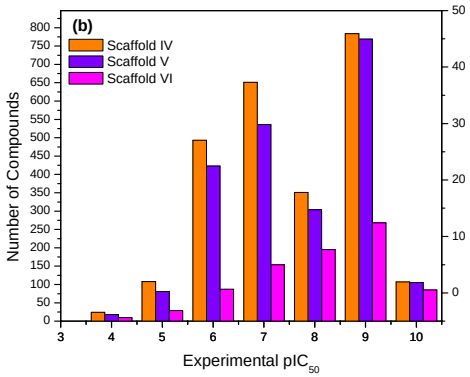
The GVK bio, kinase database contains total 1,38,805 inhibitors [7.41]. Nearly 9700 diverse molecules which have been screened for EGFR inhibitory activity are present in this maintained database, were used in this study. Among these, 3243 compounds contain scaffold **I** (Scheme 7.1). Within this set, subsets of 2762 and 2425 compounds contain scaffolds **II** and **III** respectively, that is, n(**III**) ⊂ n(**II**) ⊂ n(**I**) ⊂ n(total). Figure 7.1 is a histogram of activities for compounds containing these three scaffolds and also for all the 9490 compounds. As seen in Figure 7.1*b*, scaffolds **IV** and **V** consist of many active molecules. Many inhibitors belonging to scaffold **IV** and **V**, and having substitution at the 4-amino group are presently in various phases of clinical trials.



Scheme 7.1: Scaffolds representing database molecule.



(a)

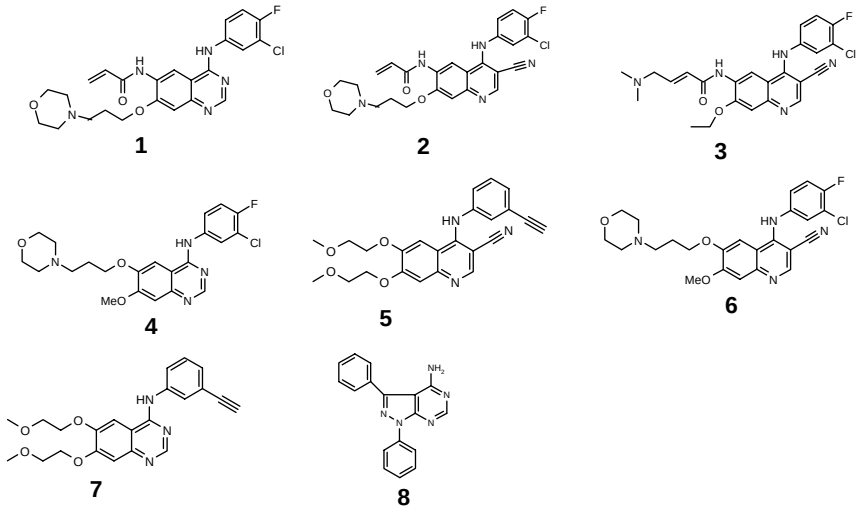


(b)

Figure 7.1: Frequency distribution of EGFR inhibitors (9940) based on different scaffolds, **I** to **VI** [**I**–**III** in (a) and **IV**-**VI** in (b)] The effect of phenyl substitution on the 4-amino group is evident from Figure 1b (scaffolds **IV**, **V**, **VI**). A large percentage of scaffold **VI** compounds (*meta* Br-substituted) have high activity in the nM range. To summarize, n(**VI**)⊂n(**V**)⊂ n(**IV**)⊂n(**II**).

7.2.2 *HipHop*

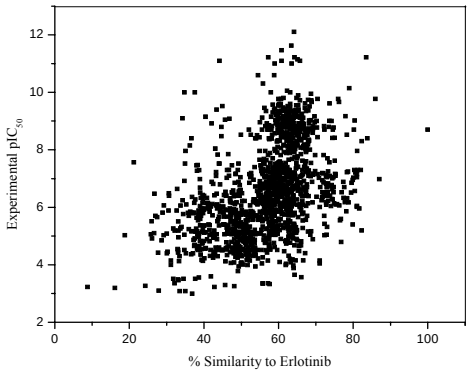
A set of eight active inhibitors **1-8** (Scheme 7.2) mostly belonging to scaffold **IV** and **V** were used to identify a qualitative pharmacophore model [7.42–7.43] using the *HipHop* module in Catalyst 4.7 [7.44]. Among these eight compounds (**1-8**), a few are in various phases of clinical trials [7.35, 7.36, 7.38]. A *HipHop* model identifies important features present in a set of active molecules. The model is built by aligning the common features present in the input training set structures, usually a small number of active compounds. *HipHop* models do not predict inhibitory activity.



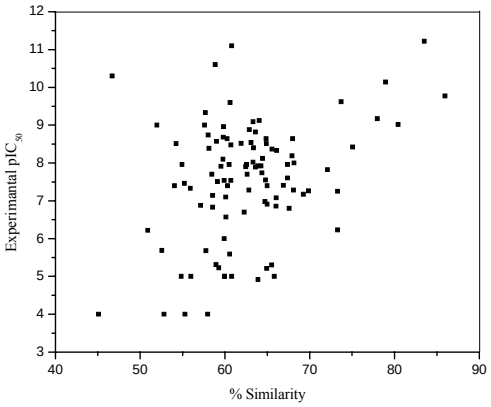
Scheme 7.2: Compounds considered in generating *HipHop* model and docking study.

7.2.3 *HypoGen*

A quantitative model was developed using *HypoGen* in Catalyst software. The *HypoGen* model was generated selecting 3800 EGFR kinase inhibitors from the GVK bio kinase database [7.41]. Ligands were selected considering diversity in structure and activities with consistent assay technique. For all these molecules ISIS structural keys [7.45, 7.46] were generated. Erlotinib (**7**), which is structurally similar to several compounds in clinical trials (Appendix II, Scheme 1), was used as the reference molecule for calculating the Tanimoto similarity coefficient. The correlation between the experimental pIC_{50} values and the Tanimoto similarity coefficients for all these inhibitors of EGFR (3800) is given in Figure 7.2.



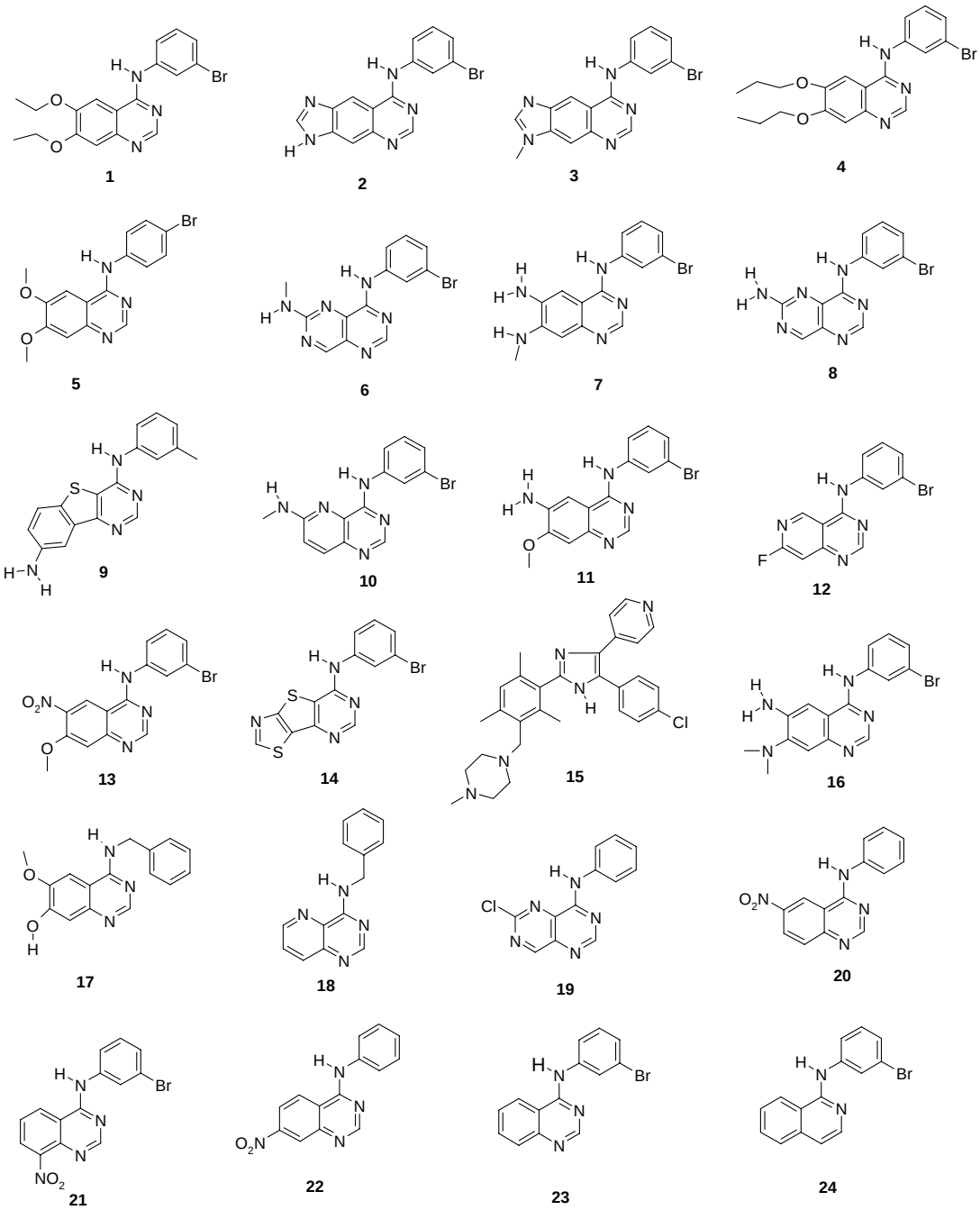
(a)



(b)

Figure 7.2: Percentage structural similarity with respect to erlotinib versus experimental activity for ligands (a) Diversity of the EGFR database molecules, (b) Subset 100 molecules, considered for *HypoGen* model building.

A subset of 100 molecules were selected for *HypoGen* model building by considering structural diversity and activity (Figure 7.2*b*). The training set comprises 24 molecules (Scheme 7.3) and the test set 76 (Appendix II, Table 1). The inhibitory activities for the training set compounds are listed in Table 7.1.



Scheme 7.3: Training set compounds in *HypoGen* model.

Table 7.1: Training set molecules considered in *HypoGen* model. Also shown are the experimental and predicted activities (pIC_{50}).

Sl. No.	Actual pIC_{50}	Predicted pIC_{50}
1	11.22	10.55
2	11.10	10.29
3	9.77	10.20
4	10.14	9.07
5	9.00	8.43
6	8.82	8.67
7	8.51	8.76
8	8.42	7.76
9	7.89	7.76
10	8.68	8.49
11	9.12	8.69
12	10.60	10.23
13	7.82	7.79
14	7.40	8.18
15	6.24	6.14
16	6.23	7.60
17	6.22	6.40
18	6.11	5.58
19	5.59	5.95
20	5.30	5.03
21	5.00	5.06
22	4.92	5.03
23	4.00	5.00
24	4.00	5.09

In *HypoGen*, the features are correlated with the biological activity of molecules (active and inactive) and a quantitative feature activity relationship is established. The minimum number of compounds in the training set should be 16 [7.47]. The default uncertainty value of 3 was used for each compound. The uncertain factor for each compound represents the ratio range of uncertainty in the activity value based on the expected statistical straggling of biological data collection. *HypoGen* generates pharmacophore models that are a set of features in 3D space, each containing a certain tolerance and weight, that fits to the features of the training set, and further correlates to the activity. A *HypoGen* run includes three phases: (1) pharmacophore features that are common among the most active compounds are elaborated in the constructive phase; (2) those models that fit the inactive training set members are abolished in the subtractive phase and; (3) the remaining hypotheses are refined in the optimization phase. The optimization is done by random transition of features, rotation of vectored features and removal or addition of features from the preferred models. The statistical parameters like cost values also determine the significance of the model. For a statistically significant model, (i) the

difference between the fixed cost and null cost (cost difference) should be above 40 indicating an activity correlation above 75%; (ii) the configuration cost should be less than 17; (iii) the root mean square (RMS) deviation between predicted and actual activities for the training set molecules should be low and (iv) the difference between total cost and the fixed cost should be as low as possible. The configuration cost quantifies the entropy of the pharmacophore space. Any model with a configuration cost greater than 17 leads to a chance correlation of the model. The total cost is the sum of error cost, weight cost and configuration cost. The fixed cost is the lowest possible cost representing a simple model that fits all the data perfectly. The null cost represents the maximum cost of a featureless model, which estimates the activity of all the training set molecules to be the average activity.

Ligands used for pharmacophore modeling were built and minimized in Catalyst using the 3D-minimizer module with CHARMM force field parameters. The conformations were generated within Catalyst ConFirm method using the "Poling" algorithm [7.48]. A maximum of 250 conformations were generated for each molecule within an energy threshold of 20.0 kcal/mol above the global energy minimum.

7.2.4 *LigandScout and GOLD*

The X-ray structures of human EGFR tyrosine kinase domain (1M17) and (1XKK) complexed with the inhibitor erlotinib and lapatinib respectively were obtained from the Protein Data Bank (PDB) [7.49]. Crystal structure 1M17.pdb is the active form of EGFR kinase domain, whereas 1XKK.pdb represent the inactive form [7.29, 7.50]. H-atoms were added to the proteins in the hydrogen addition option in MOE software [7.51]. Both the complexes were subjected to minimization keeping the heavy atoms rigid. This exercise was carried out in MOE with the MMFFx force field [7.52]. Structure based pharmacophore models were built for these minimized structures. This study was performed with LigandScout software [7.53]. LigandScout perform pharmacophore modeling for available protein–ligand complexes in PDB. LigandScout generates a pharmacophore model through a set of chemical features and volume constraints present in the active site. The 3-D pharmacophore models are constructed, which describe the binding mode of a ligand to the receptor.

The minimized structures were exported to GOLD for docking. Molecules considered in *HipHop* model were docked separately into the 1M17 and 1XKK crystal structures using Chemscore, GOLD [7.54]. A region of 10 Å radius around the bound

ligands, for each structure was defined as the active site. The default parameters in all the calculations were used. For each of the 10 independent Genetic Algorithm (GA) runs, with a selection pressure 1.1 and 100,000 GA operations were performed on a set of 5 islands. The population size of 200 individuals was specified. Default operator weights were used for crossover, mutation, and migration of 95, 95 and 10 respectively. To further speed up the calculation, the GA docking was stopped when the top three solutions were within 1.5 Å RMSD of each other. All the other values were set to the default. While docking, a limited flexibility is allowed for the hydrogen atoms in the –OH, and $–NH_3^+$ substituents of the side chains of Ser, Thr and Lys residues.

7.3 Results and discussion

7.3.1 *HipHop*

A set of 10 *HipHop* models that were generated with features like hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic aliphatic (HYA), hydrophobic aromatic (HYR) are ranked in Table 7.2. Model 1 (shown in Figure 7.3), was used to mine EGFR kinase inhibitors in the database [7.41]. Table 7.3 shows various measures for hit quality; the underlying equations to measure the quality of hits have been mentioned in the Appendix II, Scheme 2. The results suggest poor mining performance of this model. The database mining resulted in 300 hits. However, the *HipHop* model has provided the list of important features in these inhibitors. These features were specified as input to generate a *HypoGen* model.

Table 7.2: Top 10 pharmacophore models generated by *HipHop*. The features corresponding to each model are listed.

Model Rank	Features	Score
1	HYA, HYA, HBD, HBA, HBA, HBA	107.169
2	HYA, HYA, HBD, HBA, HBA, HBA	103.302
3	HYA, HYA, HBD, HBA, HBA, HBA	102.478
4	HYA, HYA, HBD, HBA, HBA, HBA	101.598
5	HYA, HYA, HBD, HBA, HBA, HBA	101.396
6	HYA, HBD, HBA, HBA, HBA	88.4952
7	HYA, HBD, HBA, HBA, HBA	88.4950
8	HYA, HBD, HBA, HBA, HBA	87.3170
9	HYA, HBD, HBA, HBA, HBA	86.4055
10	HYA, HBD, HBA, HBA, HBA	86.2553

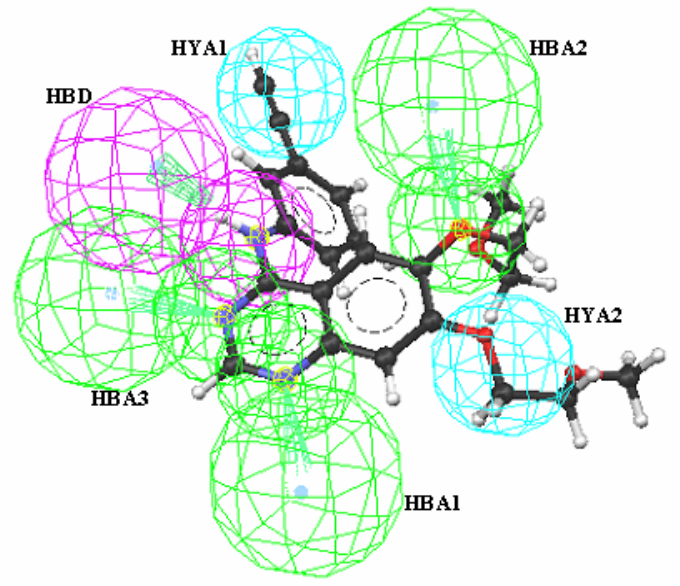


Figure 7.3: *HipHop* model mapping onto Erlotinib: The green contours represent the positioning of hydrogen bond acceptors (HBA 1, 2, 3) and their projected points on the receptor are indicated by conical projections. Blue contours represent hydrophobic aliphatic (HYA1, 2). Magenta contours represent the hydrogen bond donor and the projected point on the receptor is indicated by conical projections (HBD).

Table 7.3: Results of database mining tabulated for *HipHop* and *HypoGen* models.

Sl. No	Parameter	*HipHop*	*HypoGen*
1	Total compounds in database (D)	3800	3800
2	Total Number of actives in database (A)	1083	1083
3	Total Hits (H_t)	300	1324
4	Active Hits (H_a)	101	710
5	% Yield of Actives	33.3	50.09
6	% Ratio of actives in the Hit list	9.32	65.59
7	Enrichment or Enhancement (E)	1.18	1.88 (Maximum value of 2.87)
8	False Negatives	982	373
9	False Positives	199	614
10	GH score (Goodness of Hit list)	0.255	0.178

7.3.2 *HypoGen*

After 30 iterations, 10 models were selected (Table 7.4). The features found to be important are: hydrogen bond acceptor (HBA), hydrophobic aliphatic (HYA), hydrophobic aromatic (HYR). Based on various statistical analyses e.g. cost difference, RMS deviation, correlation coefficient, configuration cost, model 1 was found to be the best. It has the lowest RMSD (1.269) and highest correlation coefficient (r = 0.958). This model was validated by predicting the activities of 76 test set compounds (Appendix II, Table 1). The correlation between experimental and predicted biological activity data is shown in Figure 7.4.

Table 7.4: Ranking of generated *HypoGen* models. Also mentioned are statistical values associated with each model and pharmacophore features.

Model number	Total cost	Cost difference	RMS deviation	Correlation (r)	Features[a]
1	121.874	186.657	1.269	0.958	HBA, HBA, HYA, HYR
2	121.889	186.642	1.360	0.950	HBA, HBA, HYA, HYR
3	126.526	182.005	1.518	0.938	HBA, HBA, HYA, HYR
4	127.396	181.135	1.550	0.934	HBA, HBA, HYA, HYR
5	128.398	180.133	1.527	0.937	HBA, HBA, HYA, HYR
6	131.685	176.846	1.539	0.937	HBA, HBA, HYA, HYR
7	133.527	175.004	1.720	0.919	HBA, HBA, HBA, HYA
8	135.243	173.288	1.747	0.916	HBA, HBA, HYA, HYR
9	135.5	173.031	1.769	0.914	HBA, HBA, HBA, HYA
10	135.5	173.031	1.847	0.905	HBA, HBA, HYA, HYA, HYR

Null cost = 308.531, Fixed cost = 93.41, Configuration = 11.559, All cost units are in bits. [a]HBA hydrogen bond acceptor, HYA hydrophobic aliphatic, HYR hydrophobic aromatic.

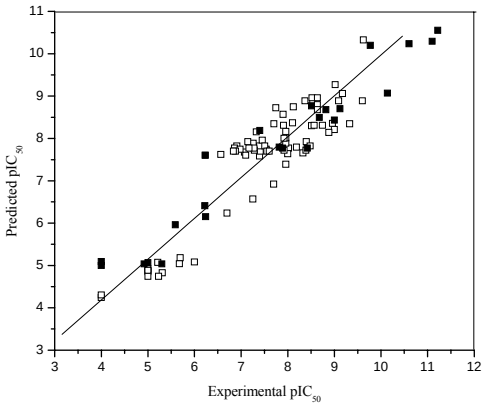


Figure 7.4: Scatter plots of actual versus predicted activities of training set (■) and test set (□) molecules used in the *HypoGen* model.

The model contained four features (Figure 7.5): 1 HBA, 2 HYA and 1 HYR. The most active molecule **1** had a fitness score of 12.04 when mapped to the pharmacophore (Figure 7.5*a*), while the least active molecule **24** maps to a value of 6.58 (Figure 7.5*b*). In molecule **1** the HYA1 feature corresponds to *meta* substitution on the aniline ring. Of the two hydrogen bond acceptors, HBA1 and HBA2 correspond to the N(1) nitrogen and substitution at 6th position of the quinazoline ring respectively while HYR corresponds to the aniline ring. In the inactive molecule, HBA1 is mapped to the N(3) position of the quinazoline ring, while HYA1 and HYR are mapped to the same bromine substitution and the aniline ring. Noteworthy is the absence of the HBA2 feature at the 6th position in the inactive molecule. These features are also mapped with Erlotinib in Figure 7.5*c*. The distance matrix (in Å) of the generated hypotheses is provided in Table 7.5.

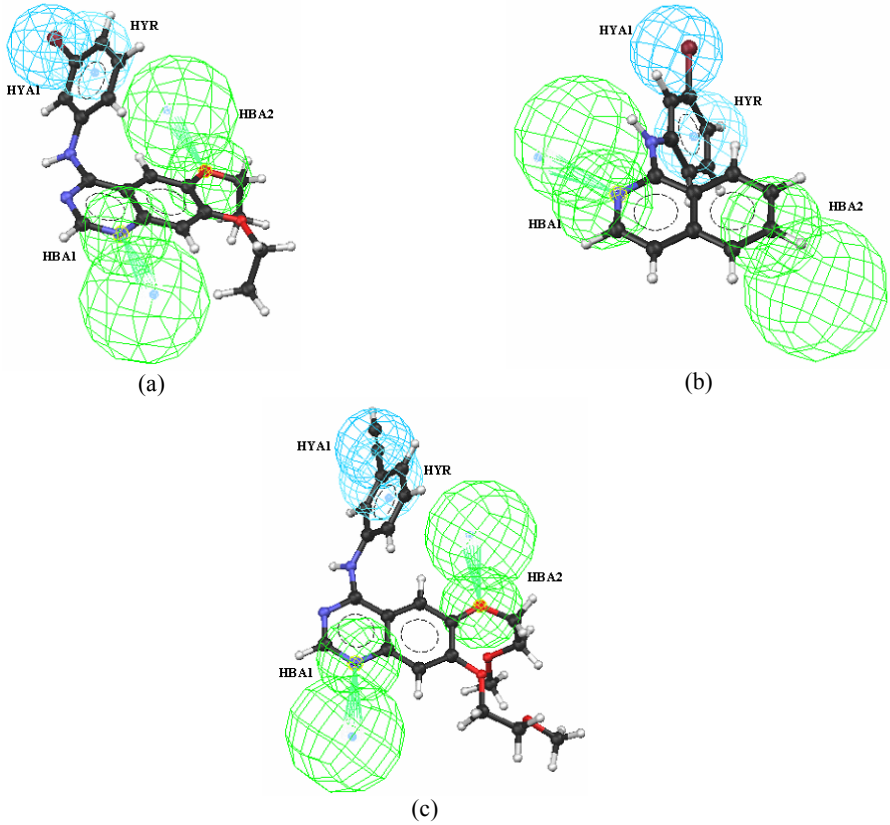


Figure 7.5: *HypoGen* model map on to: (a) most active molecule **1,** (b) least active molecule **24**, (c) Erlotinib. The green contours represent the positioning of hydrogen bond acceptors (HBA1, 2) with the projected points on the receptor indicated by conical projections. Blue contours represent the hydrophobic aliphatic (HYA1) and hydrophobic aromatic (HYR1).

Table 7.5: Distance matrix (in Å) of the generated quantitative model.

	HBA1 (IP)	HBA1 (PP)	HBA2 (IP)	HBA2 (PP)	HYA	HYR
HBA1 (IP)						
HBA1 (PP)	3.0					
HBA2 (IP)	5.4	7.4				
HBA2 (PP)	7.7	10.1	3.0			
HYA	7.7	5.4	9.1	12.1		
HYR	6.3	4.8	7.0	9.9	3.1	

IP, initial point on ligand, PP, projected point on the receptor.

7.3.3 *Model validation*

The quantitative model has mixed success in predicting the activity of the inhibitors considered in the qualitative model generation Table 7.6. The predicted activities are close to experimental activities, except for molecule **8**. This is because two of the pharmacophore features: HBA2 and HYA1 do not map properly as shown in Figure 7.6.

Table 7.6: Predicted activity for eight inhibitors considered in the qualitative model.

Molecule No.	Actual (pIC_{50})	Predicted (pIC_{50})	Fitness score
1	7.13	6.07	11.59
2	7.09	6.65	12.23
3	7.08	5.18	12.17
4	6.28	8.01	12.83
5	6.07	7.74	11.55
6	5.98	6.13	12.88
7	5.83	6.65	11.09
8	5.80	2.20	6.51

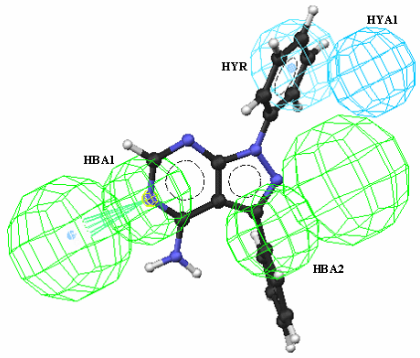


Figure 7.6: *HypoGen* model map on to molecule **8**. Notice the incomplete mapping of pharmacophore features HYA1 and HBA2.

Therefore there is a large deviation between experimental activity and predicted activity. Model validation was carried out in an external dataset of 128 EGFR kinase inhibitors from the previous study ([7.25], Appendix II, Table 2). These include simple quinazolines (class **A**: 45 compounds), pyrido[3,2-d]pyrimidines (**B**: 7 compounds), pyrido[4,3-d]pyrimidines (**C**: 49 compounds), pyrido[3,4-d]pyrimidines (**D**: 5 compounds), pyrido[2,3-d]pyrimidines (**E**: 6 compounds), pyrimido[5,4-d]pyrimidines (**F**: 16 compounds). The activities were predicted for 128 ligands (r = 0.6) shown in Figure 7.7*a*. Poor correlation was attributed to 32 (25 %) molecules, (Table 7.7).

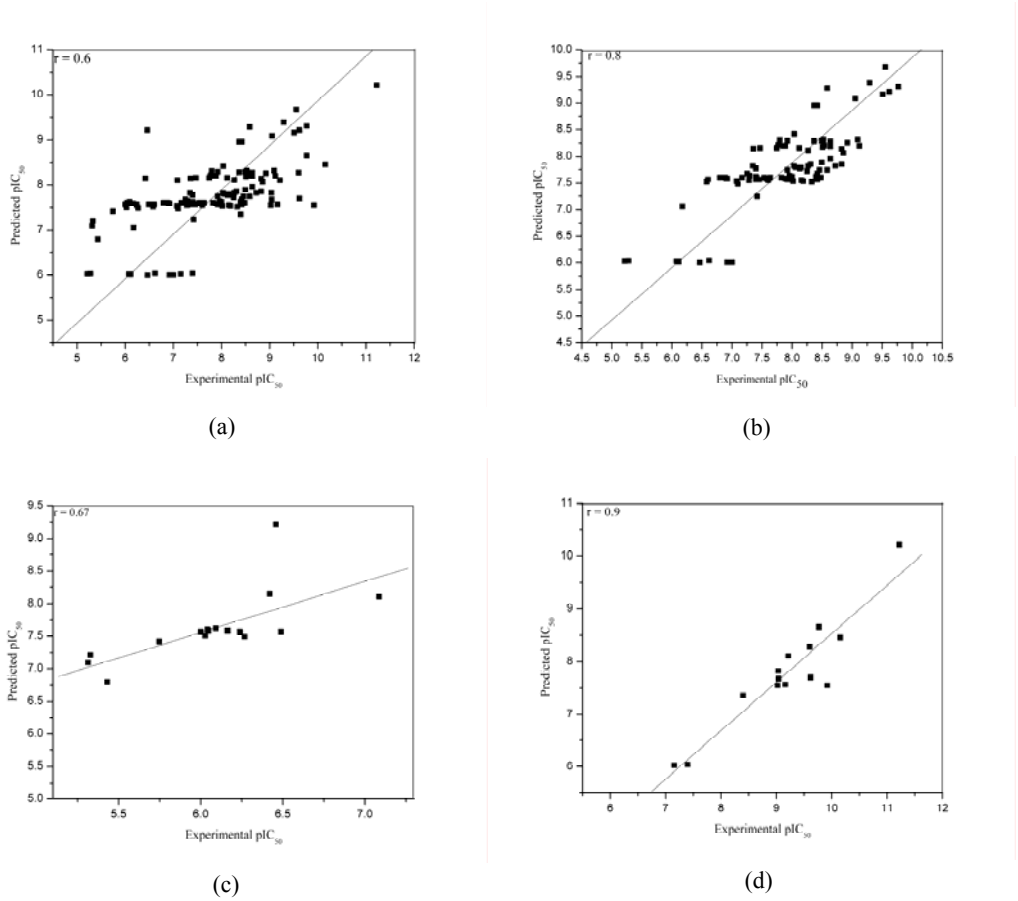
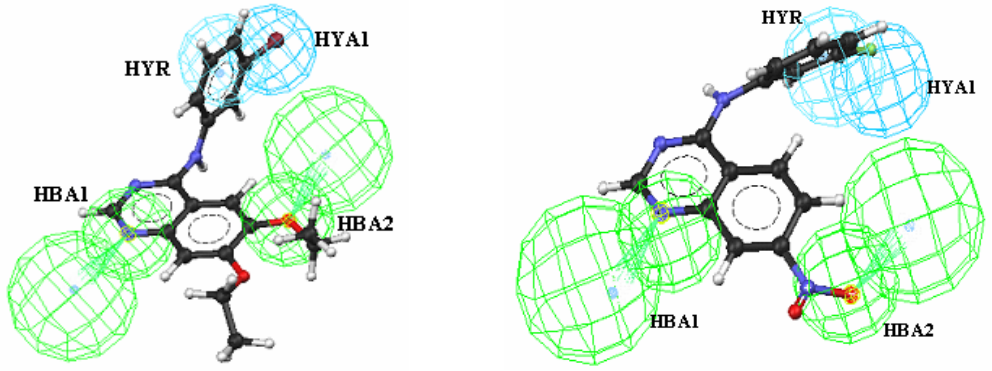


Figure 7.7: Scatter plot of actual versus predicted activities of 128 molecules from the external EGFR kinase inhibitor database. (a) Correlation for 128 molecules, (b) Correlation for 96 compounds, (c) False positives, (d) False negatives.

Table 7.7: Activity prediction for external 128 EGFR kinase inhibitors.

Compound class	Scaffold	Total Number of compounds	False positive	False negative	False prediction
quinazolines	**A**	45	8	6	31%
pyrido[3,2-d]pyrimidines	**B**	7	-		
pyrido[4,3-d]pyrimidines	**C**	49	4	6	20%
pyrido[3,4-d]pyrimidines	**D**	5	-	-	
pyrido[2,3-d]pyrimidines	**E**	6	4	-	66%
pyrimido[5,4-d]pyrimidines	**F**	16	2	2	25%
Total		*128*	*18*	*14*	25%

Among these, 18 are false positives and 14 are false negatives. The predicted activities for molecules of the pyrido[3,2-d]pyrimidine and pyrido[3,4-d]pyrimidine classes are better compared to all other compound classes. After removing the false positives and negatives the correlation between experimental and predicted activity was improved (r = 0.8) for 96 compounds as shown in Figure 7.7*b*. Interestingly, the correlation coefficients for false positive and negatives among themselves are 0.7 and 0.9 respectively. This is shown in Figures 7.7*c* and *d*. This suggests that the model is able to rank the outliers to a reasonable limit. The outliers are usually misoriented conformations of molecules (compared to crystal structure conformation of Erlotinib/Lapatinib). Also, the pharmacophore feature HBA2 is not mapped by any of these compounds. Sometimes in these outliers the pharmacophore features HBA1, HBA2 are assigned to other functional groups instead of N(1) and substitution at 6th position. The active (reference molecule **43** in Appendix II, Table 2) and inactive (reference molecule **20** in Appendix II, Table 2) molecules collected from this data set are mapped onto the generated pharmacophore model and is shown in Figure 7.8.



(a) (b)

Figure 7.8: Pharmacophore mapping of the compounds (a) most active molecule (**43**) (b) least active molecule (**20**) in the external 128 EGFR kinase inhibitors. Note the improper mapping of the hydrophobic aliphatic (HYA1) and hydrogen bond (HBA2) in the inactive molecule.

7.3.4 *Mining*

The best *HypoGen* model was used to screen the EGFR kinase inhibitor database with the best flexible search method implemented in Catalyst. The performance of *HypoGen* is better, compared to that of *HipHop* in terms of the number of hits and active hits (Table 7.3). A non-EGFR kinase inhibitor database of 4590 compounds was prepared using various kinase inhibitors. These include CDK2, FGFR kinase, IGFR kinase, MAP kinase, PDGFR kinase, PKA, Src kinase, and VEGFR kinase. The non EGFR kinase dataset molecules are spiked with 128 EGFR kinase inhibitors of the earlier study [7.25]. The database mining resulted in a smaller number of hits (129 out of 4590). Out of 129 hits, 95 belong to EGFR kinase inhibitors (i.e. 95 out of 128 EGFR kinase inhibitors were successfully retrieved). This suggests that the generated model can selectively pick EGFR kinase inhibitors from other kinase inhibitors.

7.3.5 *Active site and structure based pharmacophore*

The X-ray structures of EGFR kinase domain active site in 1M17 and 1XKK was manually superposed (Figure 7.9). A comparison of the two structures reveals that the shape of the active site is closed and truncated in 1XKK while it is open in 1M17. This is because the A-loop and G-loop shrink towards each other in 1XKK while they are far apart in 1M17. This results in a variation in the size of the active sites, 199 $Å^3$ and 145 $Å^3$ for 1M17 and 1XKK respectively. Further 1XKK has an extra large pocket to accommodate the 3-fluorobenzyl-oxy group, whereas no such pocket is observed in the EGFR/Erlotinib complex. This pocket is where two hydrophobic aliphatic (HYA) and an aromatic (HYR) features are seen. The active site in the crystal structure of 1XKK is more hydrated than in 1M17. There are 20 water molecules within a 10 Å radius around the bound ligand in 1XKK compared to 3 waters in 1M17. In both crystal structures, the quinazoline N(1) forms a hydrogen bond to the NH of MET769, while N(3) is hydrogen bonded to WAT 10 in 1M17 and to WAT 4 in 1XKK. However this water interacts differently with residue THR766 in 1M17 and 1XKK respectively. These hydrogen bonds are shown with HBA at N(1) and N(3) positions as pharmacophore points. Quinazoline NH is hydrogen bonded to another water molecule (WAT 22) in the EGFR/GW572016 complex. This hydrogen bond donor (HBD) is unique to 1XKK. Apart from these, two hydrophobic aromatic (HYR) pharmacophore features (quinazoline and furan substitution at 6^{th} position) are observed in 1XKK. The orientation of the aniline ring is similar for both structures in the hydrophobic pocket, created by residues VAL702, LYS721, LEU764 (Figure 7.9, 7.10). The above

mentioned differences in active site environment are also reflected in structure based pharmacophore model generated with LigandScout software (shown in Figure 7.10).

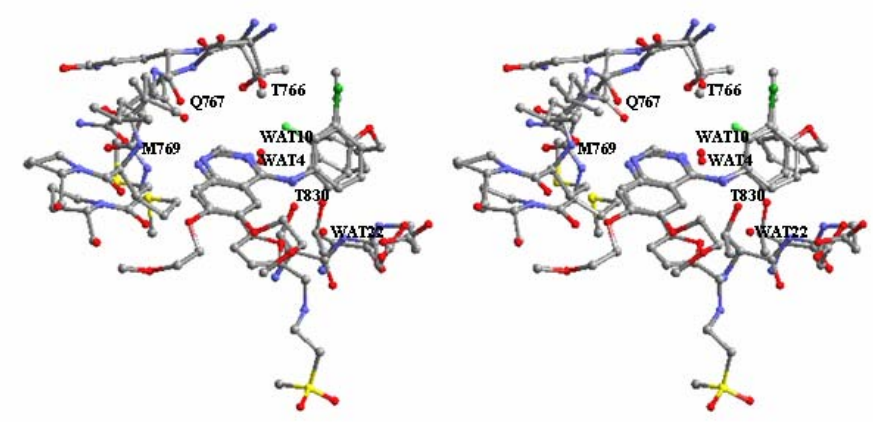


Figure 7.9: Stereo view of superposed crystal structures of EGFR/Erlotinib (1M17) and EGFR/GW572016 (1XKK).

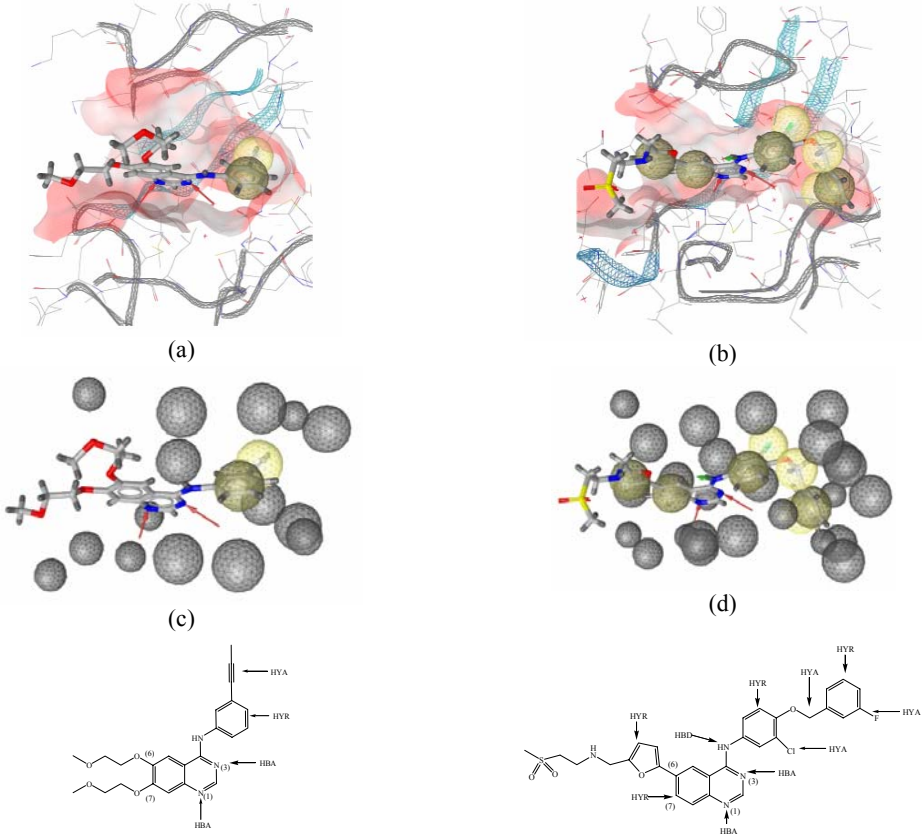
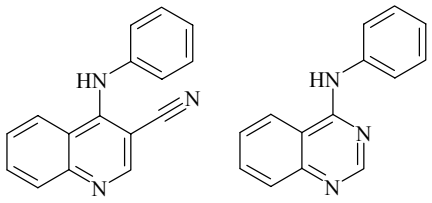


Figure 7.10: Structure based pharmacophore model for EGFR/Erlotinib (1M17) and EGFR/GW572016 (1XKK). (a)-(b) Pharmacophore features along with polarity of binding sites. (c)-(d) Pharmacophore with excluded volume.

7.3.6 *Docking*

Docking studies were carried out for molecules in Scheme 7.2 using Chemscore in GOLD [7.54]. These molecules contain the cyanoquinoline (molecules **2**, **3**, **5)** and anilinoquinazoline (molecules **1**, **4**, **7**) fragment and are shown in Scheme 7.4.



Scheme 7.4: Scaffold for cyanoquinoline and anilinoquinazoline

The weak interaction was modeled with a parameter which includes the weak C–H···O hydrogen bond. The initial docking runs were aimed at choosing the form of the receptor suitable to carry out the docking study. Docking poses were filtered through structure based pharmacophore criteria. Ligand poses satisfying the pharmacophore criteria were selected for further analysis. Proper docking poses and scores for molecules having the cyanoquinoline fragment were achieved only in 1M17, whereas for the anilinoquinazolines, docking poses and scores are consistent in 1M17 and 1XKK. Interestingly, the desired poses of cyanoquinoline were achieved only after removing WAT 10 from the crystal structure of 1M17. The hydrogen bonds distances are listed in Table 7.8.

Table 7.8: Strong and weak hydrogen bonds for docked solutions of cyanoquinoline (in 1M17) and anilinoquinazoline (in 1XKK) inhibitors. Molecule 8 is a different class of compound.

Molecule	MET769 [N–H···N(1)]	THR766 [O–H···N]	GLN767 [C–H···O]	WAT 4 [Ow–H···N(3)]	WAT 22 [N–H···Ow]
Cyanoquinoline					
2	1.85	1.71	2.17		
3	1.85	1.97	2.29		
5	1.8	2.25	2.46		
6	1.92	1.75	2.27		
Anilinoquinazoline					
1	1.84	-	2.04	1.95	2.22
4	2.05	-	2.25	1.89	2.01
7	2.11	-	2.35	1.85	1.68
8	1.61	-	1.99	2.08	

Possible reasons for selective binding (docking scores) of cyanoquinolines to the active conformations are: (1) Open conformation of the active site in 1M17. This facilitates accessibility of the polar hydroxyl group of THR766 to the 3-cyano group in the ligand (Figure 7.11); (2) The water molecule (WAT 10 in 1M17 or WAT 4 in 1XKK) sterically hinders the correct positioning of the cyano group, (3) Crystal structure of 1XKK shows that the G-loop may also hinder interactions to THR766, The orientation of OH^{γ} of THR766 is far from the 3-cyano group. In anilinoquinazolines, water (WAT 10 in 1M17 or WAT 4 in 1XKK) forms a Ow–H···N(3) interaction as shown in Figure. 7.12 (shows docking solution of molecule **7** in 1XKK). In 1M17, WAT 10 forms a hydrogen bond with the OH^{γ} of the THR766. Similarly in 1XKK, WAT 4 interacts with THR830 through a O–H···O hydrogen bond. Therefore WAT 10 and the 3-cyano group are supramolecular isosteres in the present context. Quinazolines also show favorable binding pose in 1M17. Quinazolines form N–H···N(1) and Ow–H···N(3) interactions with MET769 and WAT 10, but the additional water (WAT 22) bonded to the quinazoline NH through a N–H···Ow hydrogen bond and the greater contact surface, discussed later in the chapter, provide extra stabilization to the quinazoline binding in 1XKK. These interactions are consistent over all the quinazolines. Sometimes water (WAT 22) interacts with the 6^{th} position substituent with an Ow–H···O hydrogen bond.

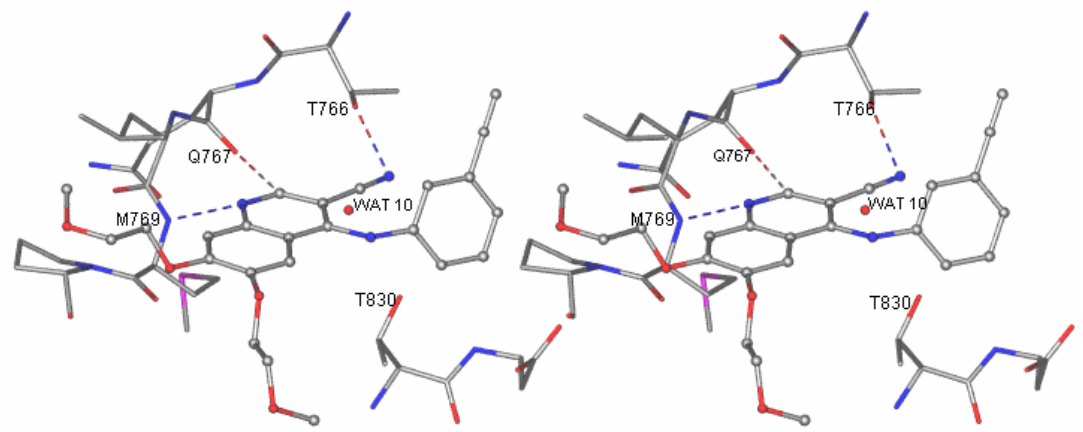


Figure 7.11: Stereo view of GOLD best solution for molecule **5** (cyanoquinoline) in 1M17 crystal structure.

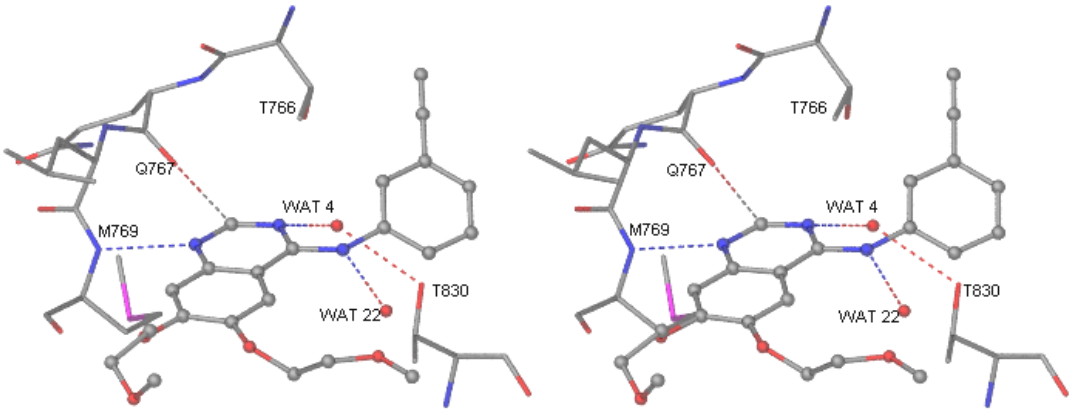


Figure 7.12: Stereo view of GOLD best solution for molecule **7** (anilinoquinazoline) in X-ray structure of 1XKK.

7.3.7 *Ligand based pharmacophore model versus structure based model*

Ligand receptor interaction is based on mutual rearrangement of both ligand and protein in order to attain complementary shapes required for binding. Ligand and protein independently may have various local or global energy minima conformations, but the complex state energy minima are not the same [7.55, 7.56]. In the present study, the ligand conformation generated in Catalyst is generally different from the conformation present in the crystal structure. This is because the crystal structure conformation is not found among the lowest energy conformation pool. For instance comparison of the crystal structure conformation of Erlotinib with that generated by Catalyst reveals that the energy of crystal structure conformation is 6 kcal/mol. This is because pharmacophore modeling in Catalyst is usually carried out on conformations generated within a threshold limit of 20 kcal/mol. The model is built on sample conformations which are local energy minima for each structure. In the present situation the sampled conformations used for model generation have energy minima well below the crystal structure conformation. In another words, the crystal structure conformation is a higher energy conformation (in vacuum) and has higher strain energy in its complexed form. To verify this hypothesis that the crystal structure conformation is in a state of higher energy, the strain energies were calculated in MOE for both ligands present in crystal structures using the E strain descriptor. For Erlotinib and Lapatinib the strain energies are 60 kcal/mol and 78 kcal/mol respectively. Similarly for eight molecules, the strain energies of the docked conformations are listed in Table. 7.9. However this strain energy is counterbalanced through enthalpy gains of various interactions like hydrogen bonds with water molecules and hydrophobic interactions.

Table 7.9: Docking scores and E_strain energies for eight inhibitors. Interaction energies for anilinoquinazoline and cyanoquinoline are separately mentioned for active (1M17), inactive form (1XKK) of the EGFR kinase domain.

Molecule	pIC_{50}	Fitness Score	Chemscore ΔG	E_strain
1M17		Quinoline		
2	7.09	32.3853	-34.4174	78677.91
3	7.08	32.2471	-33.3677	49342.44
5	6.07	27.4658	-32.1522	1400.799
6	5.98	27.2096	-32.1036	1242.748
		Quinazoline		
1	7.13	27.7393	-31.5134	117.428
4	6.28	27.632	-30.4159	66.24467
7(Erlotinib)	5.83	27.3265	-30.1183	45.28573
8	5.80	27.1163	-27.6344	31.3697
1XKK		Quinazoline		
1	7.13	28.518	-37.1237	59.36667
4	6.28	28.4873	-34.0155	44.67819
7(Erlotinib)	5.83	27.5014	-36.5353	29.54187
8	5.80	27.3567	-31.7945	43.93411
		Quinoline		
2	7.09	22.6237	-36.9178	1086.931
3	7.08	21.913	-34.3228	3226.395
5	6.07	26.2797	-34.3154	1411.588
6	5.98	31.0039	-34.5998	1602.754

Despite differences in conformations, the ligand based pharmacophore model has generated features similar to the structure based pharmacophore model. The pharmacophore features observed in *HipHop* is similar to those in the receptor based model. HBA1 and HBA3 are shown having hydrogen bonds to MET769 and WAT 10. The docking result support HBA2 can exist, wherein WAT 22 in 1XKK is an important counterpoint to the acceptor for substitution at the 6^{th} position. Similarly HBD is exhibited by quinazoline NH to water acceptor (WAT 22) in 1XKK. The hydrophobic aliphatic and aromatic interactions are well in agreement with the pharmacophore features of the crystal structures. The *HypoGen* model on the other hand contains features which are a subset of the features

observed in the structure based and *HipHop* models. Features like HBA1, HBA2, HYA1, HYR present in this model are able to predict well the activities of EGFR inhibitors. Presumably, features like HBA3, HBD, HYA2 which are absent in this model, negotiate the ligand strain energy by providing stabilizing interactions in the active site.

7.4 Conclusions

This chapter describes a method of LBVS for EGFR kinase inhibitors through pharmacophore modeling. A qualitative pharmacophore model was designed which describes the important pharmacophore features present in a set of active compounds. Another quantitative pharmacophore model was modeled through the selection of a structural similarity driven training set, taking into structure activity correlation. It was demonstrated that the quantitative pharmacophore model has better success in database mining of EGFR kinase inhibitors. This validation was carried out using both EGFR and spiked EGFR kinase inhibitors in a non-EGFR kinase inhibitor database. The predictive power of the quantitative model was validated in an external dataset of 128 EGFR kinase inhibitors. A four feature pharmacophore model, with two hydrogen bond acceptors, one hydrophobic aliphatic and a hydrophobic aromatic features, best describes the activity of EGFR kinase inhibitors.

A structure based pharmacophore model was designed for two crystal structures, representing active and inactive form of the EGFR kinase domain. The ligand based pharmacophore model and the structure based pharmacophore model complement each other in various pharmacophore features. A docking study of 8 known inhibitors in the active and inactive form of the receptor reveals selectivity of the receptor toward anilinoquinazoline and cyanoquinoline. The anilinoquinazoline type of ligands have a similar binding mode irrespective of the active and inactive form of receptor, whereas cyanoquinolines have a preferred pose in the active form of the kinase domain. Ligands of the cyanoquinoline type displace a water molecule (WAT 10) to have a preferred pose in the active site. The same water molecule bridges the anilinoquinazoline ligand and THR766 (in 1M17)/THR830 (in 1XKK). In summary a VS model for large scale screening of ligand libraries against EGFR kinase has been successfully developed and validated. Additionally, the structural basis of ligand selectivity for the active and inactive forms of the EGFR kinase domain was demonstrated through docking.