

## Original Research Article

# Application of Group-Based QSAR and Molecular Docking in the Design of Insulin-Like Growth Factor Antagonists

Abubakar Danjuma Abdullahi<sup>1</sup>, Abdulrahman Mohammed Abdulkader<sup>2,4</sup>,  
Nadiahanis binti Abdulsamat<sup>3</sup> and Kundan Ingale<sup>5</sup>

<sup>1</sup>Department of Basic Medical Sciences, <sup>2</sup>Department of Pharmaceutical Chemistry, <sup>3</sup>Department of Pharmaceutical Technology, Faculty of Pharmacy, International Islamic University Malaysia, Jalan Istana, 25200, Kuantan, Pahang, Malaysia, <sup>4</sup>Department of Pharmaceutical Chemistry and Quality Control, Aleppo University, Syria, <sup>5</sup>VLife Sciences Technologies Pvt Ltd, Pune, India

\*For correspondence: **Email:** [bagaruwa@gmail.com](mailto:bagaruwa@gmail.com); **Tel:** 0060 19 917 5042

Received: 16 May 2014

Revised accepted: 14 April 2015

## Abstract

**Purpose:** To identify the structural requirements for designing a lead key for insulin-like growth factor (IGF-1R) inhibition using group-based quantitative structure activity relationship (GQSAR) and molecular docking.

**Methods:** GQSAR method requires fragmentation of molecules. The molecules in the current dataset were fragmented into three (R1, R2 and R3) by applying common fragmentation pattern, and fragment-based 2D descriptors were then calculated. GQSAR models were derived by applying various methods including multiple linear regressions and partial least square or k-nearest neighbour.

**Results:** Four generated GQSAR models were selected based on the statistical significance of the model. It was found that the presence of flexible and non-aromatic groups on fragment R1 was conducive for inhibition. Additionally, the existence of amino groups as hydrogen bond donors at fragments R2 and R3 was fruitful for inhibition. Docking studies revealed the binding orientation adopted by the active compounds at several amino acid residues, including Met 1126, Arg, 1128, Met 1052, GLU 1050, Met 1112, Leu 1051, Met 1049, Val 1033, and Val 983 at ATP binding sites of IGF-1R kinase domain.

**Conclusion:** The generated models provide a site-specific insight into the structural requirements for IGF-1R inhibition which can be used to design and develop potent inhibitors.

**Keywords:** Insulin-like growth factor 1 (IGF-1) receptor, Quantitative structure-activity relationship, Adenosine triphosphate, Competitive inhibitors, Electrotopological state index.

Tropical Journal of Pharmaceutical Research is indexed by Science Citation Index (SciSearch), Scopus, International Pharmaceutical Abstract, Chemical Abstracts, Embase, Index Copernicus, EBSCO, African Index Medicus, JournalSeek, Journal Citation Reports/Science Edition, Directory of Open Access Journals (DOAJ), African Journal Online, Bioline International, Open-J-Gate and Pharmacy Abstracts

## INTRODUCTION

Insulin-like growth factor-1 receptor (IGF-1R) is a tyrosine kinase of the insulin receptor family with hetero-tetrameric  $\alpha_2\beta_2$  subunits. Alpha subunit is located extracellularly and contains the binding domain, while the  $\beta$  subunit is mainly intracellular with a small portion spanning across the membranes. The latter subunits are responsible

for intracellular signal transduction due to tyrosine kinase domain and associated motifs [1]. IGF-1 receptors are located in all brain tissues such as median eminence, hippocampus, and cortex [2].

Physiological activation of IGF-1R, which activates PI3K/Akt/mTOR, and Ras/Raf/MEK/ERK pathways only occurs upon binding of IGF-1

or IGF-2 ligands to the receptor. Conversely, ligand-dependent activation of IGF-1R is bypassed in many solid and hematological cancers leading to continuous activation of PI3K/Akt/mTOR and Ras/Raf/MEK/ERK pathways that promote proliferation and inhibition of apoptosis [3]. IGF-1R has been widely targeted to decrease cellular proliferation and promote apoptosis in cancers [4,5]. Crystallographic structures of IGF-1R co-crystallized with various ligands have been identified [3,6-8]. The biological activity of these ligands depends on the interactions with certain residues in the ATP binding sites of IGF-1R kinase domain, which serve as hydrogen bond donors, acceptors, or hydrophobic pockets. Despite the success in the design of IGF-1R inhibitors, there is a need to optimize the lead compounds to efficient drug compounds.

Fragment-based lead discovery (FBLD) has shown promise in current drug discovery and lead optimization. The main concept of this method is to drive new molecules by combining fragments that are determined based on ligand-target interaction information [9]. Recently, successful application of group-based quantitative structure activity relationship (GQSAR) for lead optimization have been reported [10]. In GQSAR, molecules are fragmented, based on specific molecular sites, bonds, and rings or even interactions with the target, and then the 2D fragment-based descriptors are calculated. The models generated by GQSAR provide hints on the impact of fragment variation on activity [11].

In this study, GQSAR was performed using various compounds from different literature sources with the aim of designing potent IGF-1R inhibitors. To further elucidate the binding modes of the generated compounds, high-throughput virtual screening (HTVS) and extra-precision docking (XP) of compounds from the developed GQSAR models were carried out using the Glide module of Schrodinger LLC.

## EXPERIMENTAL

### Group-based quantitative structure activity relationship (GQSAR)

The Hansch method of QSAR assumes that groups at different substitution sites of congeneric series do not interact with each other; on the contrary, conventional molecular descriptors currently in use capture this interaction by calculating them for the whole

molecule. However, for this method, interpretation of generated QSAR equation poses significant challenges. In view of this, the new G-QSAR approach, which uses established 2-D/3-D descriptors for the fragments and also includes interaction of fragments using cross terms, can provide promising results. Cross terms added are descriptors evaluated as a product of vectors of fragment descriptors. G-QSAR method was very recently proposed, [11] and differs from conventional fragment based QSAR methods in two ways:

1. Fragmentation of each molecule in the dataset is done with a set of predefined rules.
2. G-QSAR method considers cross/interaction terms as descriptors to account for the fragment interactions.

### Software

GQSAR modeling was executed using the Molecular Design Suite (VLifeMDS software package, version 4.1, from Vlife Sciences Technologies Pvt Ltd, India) on a Microsoft Windows 7 operating system.

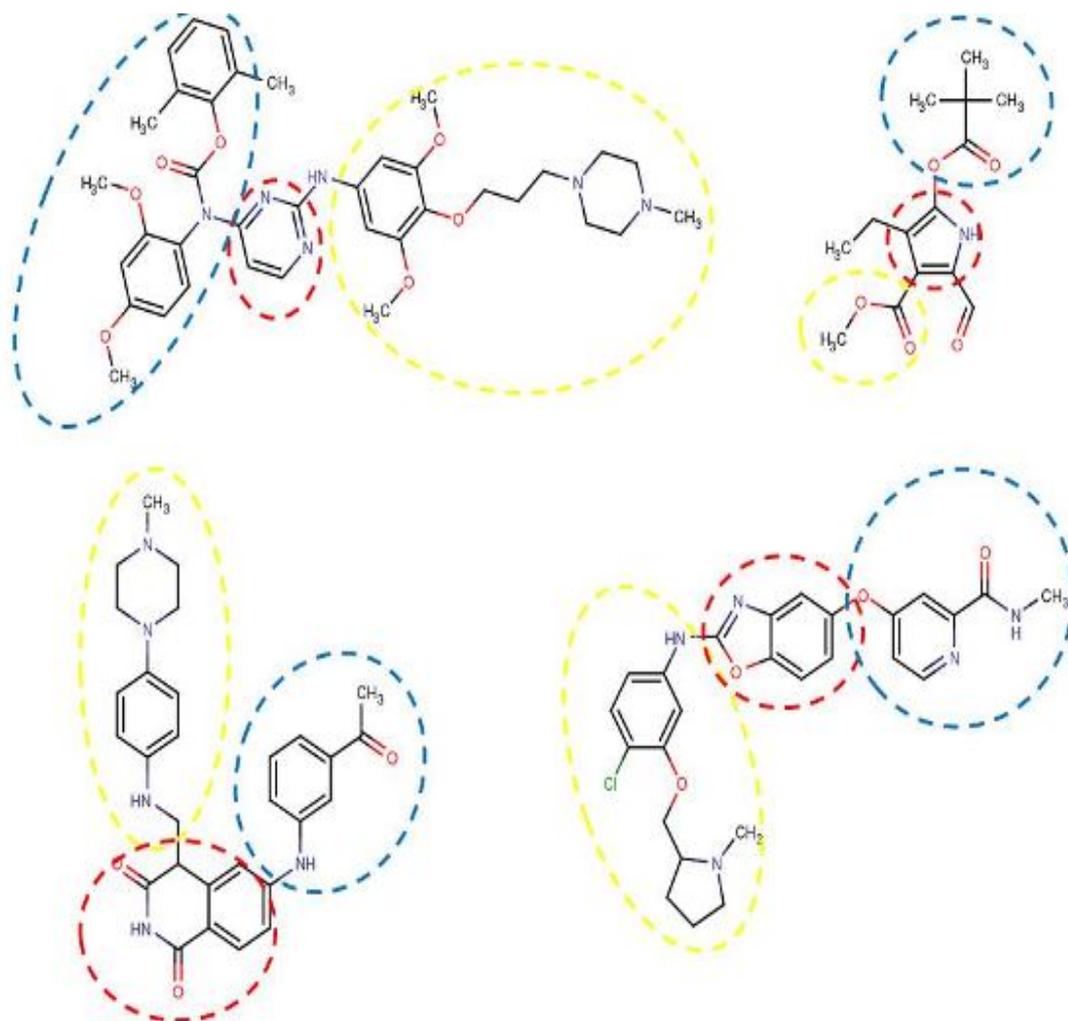
### Dataset collection and fragmentation

A dataset of 164 IGF-1R inhibitors with their  $IC_{50}$  values were collected from a Binding-DB database [8,12,13]. The biological activity values ( $IC_{50}$ ) were converted into the negative logarithmic  $pIC$  ( $-\log IC_{50} + 6$ ). The structures were energy minimized using Merck Molecular Force Field (MMFF) using convergence criterion (RMS gradient) of 0.01 kcal/mol and maximum number of cycles of 100 [14].

The success of GQSAR methodology depends mainly on the fragmentation of molecules in dataset, thus the compounds were grouped into different chemical scaffolds including: isoquinoline-1, 3-dione, pyrimidine, 3-cyano-6-hydroxy quinolone, benzoxazole, pyrrole, benzimidazole, 7-amino pyrolopyrimidine, and pyrrolopyrimidine.

All molecules were then fragmented into three different fragments as shown in Figure 1.

1. **Fragment R1:** Aromatic substitution on the scaffold, i.e., Fragment R2.
2. **Fragment R2:** It is formed by aromatic heterocyclic ring (chemical scaffold) structure after Fragment R1.
3. **Fragment R3:** Substitution of the scaffold, i.e., Fragment R2.



**Figure 1:** Fragmentation pattern. The dataset was grouped into different chemical structures based on common chemical parts that were mainly aromatic heterocyclic ring named as the scaffold or fragment R2 (red circle). The 3D structures were fragmented into three fragments R1 (blue circle), R2 (red circle), R3 (yellow circle)

### Fragment-based molecular descriptors calculation

Individual physicochemical descriptors like molecular weight, hydrogen bond donors and acceptors, retention index (chi), path count, estate numbers, atomic valence connectivity index (chiv), polar surface area, etc were calculated for all fragment-based descriptors. The important descriptors found fruitful according to the developed model can be defined as follows:

1. chi1: The retention index (first order) derived directly from gradient retention times.
2. chi3Cluster: Third order cluster chi index.
3. chiV6chain: Atomic valence connectivity index for six membered ring.
4. chiV4pathCluster: Valence molecular connectivity index of 3rd order path cluster.
5. H-AcceptorCount: Number of hydrogen bond acceptor atoms.
6. H-DonorCount: Number of hydrogen bond donor atoms.
7. IdwAverage: Information-based descriptors.
8. k3alpha: Third alpha modified shape index.
9. kappa2: Second kappa shape index.
10. OxygensCount: Number of oxygen atoms in a compound.
11. RotatableBondCount: Number of rotatable bonds.
12. SaaCHcount: The total number of carbon atoms connected with a hydrogen along with two aromatic bonds.
13. SaaCHE-index: Electrotopological state indices for number of  $-CH$  group connected with two aromatic bonds.
14. SaaNHE-index: Electrotopological state indices for number of  $-NH$  group connected with two aromatic bonds.

15. SaasCE-index: Electrotopological state indices for number of carbon atom connected with one single bond along with two aromatic bonds.
16. SaasN(Noxide)count: The total number of nitro oxide group connected with one single bond along with two aromatic bonds.
17. SdOE-index: Electrotopological state indices for number of oxygen atom connected with one double bond.
18. SdsCHE-index: Electrotopological state indices for number of –CH group connected with one double and one single bond.
19. SdssCcount: Total number of carbon connected with one double and two single bond.
20. Smr: This descriptor evaluates molecular refractivity (including implicit hydrogen) which also measure molecular size (number). This property is an atomic contribution model that assumes the correct protonation state (washed structures).
21. SsNH<sub>2</sub>count: The total number of –NH<sub>2</sub> group connected with one single bond.
22. SsNH<sub>2</sub>E-index: Electrotopological state indices for number of –NH<sub>2</sub> group connected with one single bond.
23. SssNHcount: The total number of –NH group connected with two single bonds.
24. SssNHE-index: Electrotopological state indices for number of –NH group connected with two single bonds.
25. SsssNcount: Total number of nitrogen connected with three single bonds.
26. SssSE-index: Electrotopological state indices for number of sulphur atom connected with two single bonds.
27. Volume: volume of a compound.
28. XlogP: This descriptor signifies ratio of solute concentration in octanol and water and generally termed as octanol water partition coefficient.

### Dividing the dataset into training and test sets

Data was divided into training and test sets using sphere exclusion algorithms with a dissimilarity level of (+1) [15]. To ensure the goodness of data distribution across the two sets, uni-column statistics parameters of each set (maximum value, minimum value, average and standard deviation) were evaluated.

### Model building

A simulated annealing algorithm (SA) was applied for variable subset selection [16]. Thereafter, multiple model equations were generated using multiple linear regression

(MLR), [17] partial least squares regression (PLSR) [18], and k-nearest neighbor (kNN) [19]. The model was considered to have a significant predictivity when the squared correlation coefficient ( $r^2$ ) between descriptors and activity (pIC) was more than 0.6. Similarly, the models were considered to possess significant internal and external predictivity when the cross-validated correlation coefficient of the leave-one-out method ( $q^2$ ) > 0.6 and the correlation coefficient of the training set ( $\text{pred}_r^2$ ) > 0.5 [10].

### Molecular docking of the QQSAR-generated compounds

In addition to QQSAR analysis of the generated compounds, we explored the use of High throughput virtual screening (HTVS) and extra-precision (XP) docking to compare the binding mode as well as docking scores of the highly active BMS1112 and poorly active compounds BMS1605 in comparison to the native ligand BMS\_754805 in the 3I81 crystal structure.

### Protein preparation and minimization

The coordinate of 3I81 crystal structure was obtained from the RCSB Data Bank (PDB) [8] and prepared using Protein Preparation Wizard, which is a part of the Maestro software package (Maestro, v9.3, Schrodinger, LLC, New York, NY, 2012). Briefly, all formal charges and bond orders were added for heteroatoms. Hydrogen atoms were also added to all atoms of the protein crystal structure. All water molecules were removed and the protein was then minimized using the refinement module in Prime (Prime, v3.1, Schrodinger, LLC, New York, NY, 2012).

### Ligands preparation, refinement, and docking

Using LigPrep (LigPrep, v2.5, Schrödinger, LLC, New York, NY, 2012), all ligands were converted into low energy 3D structures by adding hydrogen atoms, neutralizing charged groups, and generating ionization and tautomeric states. A rectangular box was used to define the binding site of the ligands in 3I81 receptor and glide energy grid was generated using Glide Receptor Grid Generation workflow (Glide, v5.8 Schrödinger, LLC, New York, NY, 2012). High throughput virtual screening (HTVS) and extra-precision (XP) docking were carried out using the default Glide settings. The virtual screening workflow allows screening of a large number of ligands in three stages. The grid-based ligand docking utilized here allows the approximation of positions, orientations, and conformations of the ligand in the 3I81-binding pocket through many hierarchical filters. The scored docking

parameters include total GlideScore (GScore): sum of XP terms excluding Epik state penalties, total docking score (Dockscore): sum of XP terms, lipophilicEvdW and Hydrogen bonding (HBond): ChemScore H-bond pair term. The docking mode of the most active compounds BMS1112 and an inactive compound BMS1605 in the active site of 3I81 were studied in terms of functional groups interactions with amino acid residues in the ATP binding site. The docking result of these compounds was compared to the docking results of the original co-crystallized ligand BMS\_754807 in 3I81.

## RESULTS

### Group-based quantitative structure relationship (QQSAR)

#### Selection of training and test sets

Fragment-based molecular descriptor calculations resulted in a pool of 325 different two-dimensional descriptors divided into 110, 100, and 115 descriptors for fragment R1, R2, and R3, respectively. The sphere exclusion method with a dissimilarity value of +1 resulted in a training set of 117 compounds and a test set of 47 compounds. The statistical parameters of both sets show that activity was evenly distributed within both sets; i.e. the average  $pIC_{50}$  values for the training and test sets are 4.25 and 4.73, respectively. The test set (maximum and minimum  $pIC_{50}$  values are 6.39 and 2.23) represents the whole training set (maximum and minimum  $pIC_{50}$  values are 6.52 and 1.86). Multiple models were built using various model-building and variable-selection methods. Table 1 shows the statistical parameters for each model.

### QQSAR models

#### Model 1 (SA/MLR)

$$pIC_{50} = -0.5019 (\pm 0.1931) R1_{SssNHcount} - 0.1485 (\pm 0.0484) R1_{XlogP} - 0.9183 (\pm 0.0614) R2_{OxygensCount} - 1.1604 (\pm 0.5860) R2_{SssSE-index} - 2.2208 (\pm 0.4239) R2_{SaaN(Noxide)count} - 0.1028 (\pm 0.0093) R3_{SaaCHcount} + 0.2638 (\pm 0.0801) R3_{IdwAverage} + 0.0476 (\pm 0.0040) R3_{SssNH2E-index} + 3.4422.$$

The simulated annealing variable selection method coupled with multiple linear regression using 117/47 compounds as the training/test set resulted in a statistically significant model having a predictive activity of  $r^2 = 72.80\%$ . Internal and external validation revealed a significant predictive power of 61.70% and 81.24%, respectively. The regression plot of predicted versus actual activity and the percentage contribution plot of descriptors are displayed in Figure 2A. Table 2 shows the percentage contribution of each 2D descriptor found in model 1.

#### Model 2 (SA/MLR)

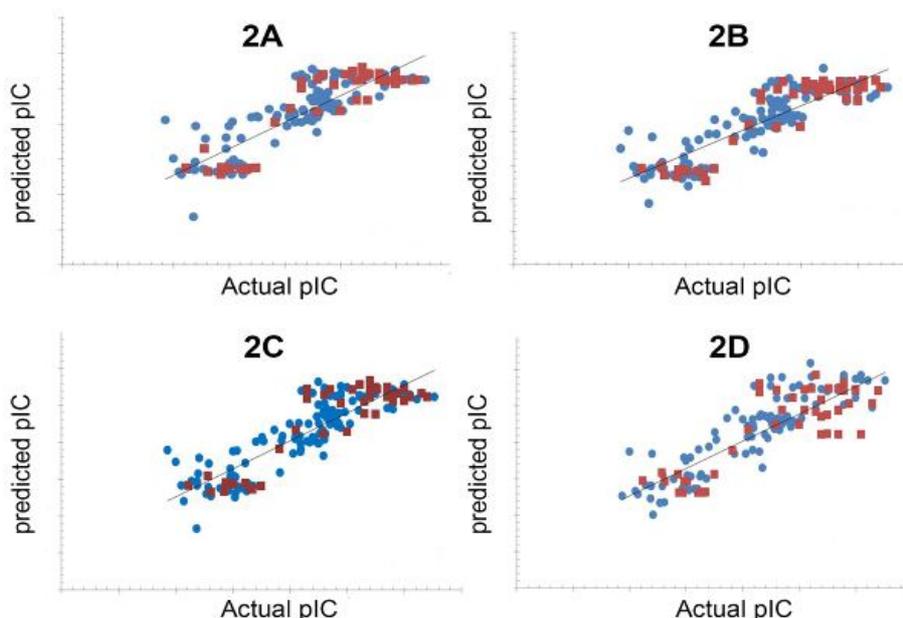
$$pIC = 0.1080 (\pm 0.0371) R1_{chr1} - 0.0819 (\pm 0.0022) R1_{SaaCHE-index} - 1.6774 (\pm 0.3544) R1_{SdsCHcount} + 0.4275 (\pm 0.0158) R2_{SaaNHE-index} - 0.1671 (\pm 0.0060) R3_{SaaCHcount} + 0.0067 (\pm 0.0001) R3_{smr} + 0.7122 (\pm 0.1197) R3_{SssNH2count} - 0.0943 (\pm 0.0328) R3_{H-AcceptorCount} + 3.2704$$

Model 2 shows an improved correlation coefficient ( $r^2 = 0.7306$ ) compared to model 1. Both  $q^2$  (0.6817) and  $pred_r^2$  (0.8445) values indicate that the model is statistically significant in terms of internal and external predictivity.

**Table 1:** Statistical parameters for the developed QQSAR models

Parameter	Model 1 (SA/MLR)	Model 2 (SA/MLR)	Model 3 (SA/PLSR)	Model 4 (SA/kNN)
n (Test/training)	47/117	47/117	47/117	47/117
Degree of freedom	108	108	113	108
F-test (Fischer statistics)	36.1235	36.6188	105.6782	NA
$r^2$	0.7280	0.7306	0.7372	NA
$q^2$	0.6170	0.6817	0.6688	0.7828
$pred_r^2$	0.8124	0.8445	0.8220	0.6152
$r^2_{se}$	0.5975	0.5945	0.5741	NA
$q^2_{se}$	0.7090	0.6463	0.6645	0.5151
$pred_r^2_{se}$	0.5620	0.5117	0.5574	0.8050
Number of contributing descriptors (k)	8	8	8	8

*F-test:* Fisher test;  $r^2$ : squared correlation coefficient;  $q^2$ : cross-validated correlation coefficient of leave-one-out method (internal validation);  $pred_r^2$ : correlation coefficient of the training set (external validation); *se:* standard error; SA: simulated annealing; MLR: multiple linear regression; PLSR: partial least square regression; kNN: k nearest neighbor



**Figure 2:** Fitness plot between the actual vs predicted activity values (pIC = log of the inhibitory concentration) of the training and test sets. (A) Model 1 where  $r^2 = 0.7280$ , (B) Model 2 where  $r^2 = 0.7306$ , (C) Model 3 where  $r^2 = 0.7372$ , (D) Model 4 where  $q^2 = 0.7828$

**Table 2:** The percentage contribution of important 2D descriptors found in the developed G-QSAR models

Model	Fragment	2D descriptor	% Contribution
Model 1	R1	SssNHcount	-7.5
		XlogP	-7
	R2	OxygensCount	-35
		SaasN(Noxide)count	-18
		SssSE-index	-7
	R3	ldwAverage	+8.5
SaaCHcount		-10	
SsNH2E-index		+7	
Model 2	R1	Chi1	+7.5
		SaaCHE-index	-11.5
		SdsCHcount	-12.5
	R2	SaaNHE-index	+29
	R3	H-AcceptorCount	-4.5
		SaaCHcount	-14.5
Smr		+3.5	
Model 3	R1	SsNH2count	+17
		ChiV4pathCluster	+5
		SdOE-index	-5.5
		SssNHE-index	-6
	R2	XlogP	-12
		OxygenCount	-31.5
R3	SaasN(Noxide)count	-13.5	
	ldwAverage	+9	
		SaaCHcount	-13.5

Figure 2B represents the regression plot of the actual activity against the predicted values by model 2. The percentages contributions of the important descriptors were shown in Table 2.

### Model 3 (SA/PLSR)

$$\text{pIC} = -0.1561 \mathbf{R1}_{\text{SssNHE-index}} - 0.2765 \mathbf{R1}_{\text{XlogP}} + 0.3384 \mathbf{R1}_{\text{ChiV4pathCluster}} - 0.0188 \mathbf{R1}_{\text{SdOE-index}} -$$

$$0.8690 \mathbf{R2}_{\text{OxygensCount}} - 1.7626 \mathbf{R2}_{\text{SaasN(Noxide)count}} - 0.1522 \mathbf{R3}_{\text{SaaCHcount}} + 0.3072 \mathbf{R3}_{\text{ldwAverage}} + 3.2957$$

Statistics suggests that the model generated using PLSR is comparable to the earlier models. Regression plot and the percentage contribution were described in Figure 2C and Table 2.

### Model 4 (SA/kNN)

The classification method of k-nearest neighbour was applied in combination with simulated annealing algorithms for generating a new model for IGF-1R inhibitors. Distance-based weighted average method was used for the prediction of biological activities of the molecules. This study led to a statistically significant QQSAR model with eight descriptors using the five nearest neighbours. The k-NN-generated model provided ranges (maximum and minimum) for eight different descriptors, as presented in Table 3.

### Docking studies of best active compound and moderately active compound

The docking results showed two QQSAR generated compounds 608 and 1112 with docking scores of -10.10 and -9.72 respectively while the native ligand BMS\_754807 scored -9.29 which is slightly lower score. Other compounds such as BMS\_1126 and BMS\_1148 scored -9.22 and 8.72 respectively. These scores are comparable to the native ligand's score. In order to understand the binding mode of the most active compound (BMS\_1112) in comparison to the native ligand (BMS\_754807) and to obtain the binding information for further structural optimization, we conducted molecular docking using Glide. We utilized Glide XP extra precision mode which allows flexible docking of ligands. As shown in Figure 3, the docking poses of compound 1112 at the 3I81 receptor active site (IGF-1R) shows interaction between fragment R1 and MET 1126 and ARG 1128 and between fragment R3 and MET 1052 and GLU 1050 on the hinge. The amide group of fragment R1 donates hydrogen to Met 1126 and Arg 1128. No hydrophobic interactions were observed between R1 and the active site of IGF-1R. Binding mode of moderately active compound BMS-1605 shows hydrophobic interactions

between the phenyl ring of fragment R1 with VAL 983, ME1049, and VAL 1033 on the IGF-1R. Lastly the docking pose of the co-crystallize BMS-754807 ligand forms hydrogen bonds to GLU 1050 and MET 1052 at the hinge. No hydrogen bonds are formed with the other residues in the active site.

## DISCUSSION

### Group-based quantitative structure activity relationship (GQSAR)

#### Model 1 (SA/MLR)

As shown in Figure 2A and Table 2, model 1 indicates that structural modifications on fragment R2 are critical for generating potent IGF-1R inhibitors. Negatively contributing features like oxygen count (-35 %), N-oxides (-18 %) and single bonded sulfur atoms are detrimental for activity.

Remarkably, all of the above structures (O, S and N oxides) can be considered as hydrogen bond acceptors since they have lone pair electrons. Thus, we can deduce that the actual role of the scaffolds in ligand-target interaction is most likely different from being hydrogen bond acceptors.

On the other hand, the current model attributes about a quarter of the variation in biological activity to structural variation on fragment R3. Number of carbon atoms connected to one hydrogen atom and two aromatic bonds (SaaCHcount) are detrimental to the activity leading to the conclusion that aromatic hydrocarbons in fragment R3 are harmful to the inhibitory activity. Therefore, multi-nitrogen-containing heteroaromatic rings like triazine are expected to provide better inhibition than aromatic rings such as benzene, pyridine and pyrimidine.

**Table 3:** Contribution range for fragment descriptors found in model 4

Descriptor	Range	
	Minimum	Maximum
R1-k3alpha	1.413	1.654
R3-chi3Cluster	1.015	1.441
R2-SsssNcount	0	0
R1-SaaCHcount	2	4
R3-SdssCcount	0	1
R3-SaasCE-index	2.167	4.25
R2-Volume	99.962	100.953
R2-chiV6chain	0.017	0.017



Accordingly, we can surmise that fragment R3 of the ligand is crucial for binding to IGF-1R. The XlogP (-7 %) and NH group connected to two single bonds (7.5 %) of fragment R1 suggest that hydrophobic interactions, along with hydrogen donor of R1, are not favorable for inhibition. Consequently, aromatic rings containing nitrogen atoms in fragment R1 are more preferable for activity than their saturated counterparts.

### Model 2 (SA/MLR)

As shown in Figure 2B and Table 2, model 2 indicates that the three fragments are almost equally contributing; Fragment R1 contributes to ~31 %, fragment R2 contributes to ~29 % and fragment R3 contributes to ~39 % of variation within the biological activity. According to the current model, fragment R3 contributes to ~39 % of the variation of the biological activity, molecular refractivity index (smr) (~3.5 %) and SsNH<sub>2</sub>count (~17 %) are conducive for activity while H-acceptor Count (~-4.5 %) and SaaCHcount (~-14.5 %) are deleterious towards activity. The descriptor (SaaCHcount) has a negative contribution towards activity indicates that the presence of substitutions on the aromatic ring can increase activity. The descriptor SaaNHEindex (29 %) at fragment R2 signifies electro topological states indices of -NH group connected to two aromatic bonds and is the largest contributor towards biological activity.

The positive contribution of this descriptor states that nitrogen-containing five-membered heteroaromatic rings (pyrrole, imidazole and triazole) are preferable as scaffolds. The terms found to be contributing to fragment R1, SaaCHE-index (~-11.5 %) and SdsCHcount (~-12.5 %), are unfavorable contributors while Chi1 (7.5 %) is a positive contributor towards activity. Collectively, the descriptors indicate double bonded CH group is not recommended in this fragment.

### Model 3 (SA/PLSR)

As shown in Figure 2C and Table 2, model 3 indicates that fragment R2 (~45 %) has the highest contribution towards biological activity while Fragment R1 (~29 %) and Fragment R3 (~23 %) have lower contribution. As shown in the previous models, OxygensCount (~-32 %) and SaasN (Noxide) count (~-14 %) suggest the importance of reducing both oxygen atoms and aromatic N-oxide at fragment R2. Moreover, SdOE-index (~-6 %), SssNHE-index (~-6 %) XlogP (~-12) and ChiV4PathCluster (~5 %) are terms of fragment R1 found correlating with biological activity. The electro-topological index of double

bonded oxygen and single bonded NH group seem to have negative impacts on the biological activity. Taken together, N-mono substituted sulfonamides (R-NH-SO<sub>2</sub>), nitrosamides (R-NH-NO), and nitroamines (R-NH-NO<sub>2</sub>) would result in low inhibitory activity. The molecular connectivity indices of path cluster 3rd order (chiV4pathCluster) are parallel with ligand activity having a positive contribution, which suggests that R1 should contain a higher number of molecular subgraphs of 3-edge cluster, such as 2-methylpentane and 3-methylpentane.

### Model 4 (SA/kNN)

As shown in Table 2, the molecular descriptors suggest a key role of aromatic carbon (SaaCHcount and SaasCE-index) at fragment R1 and R3 along with terms like K3alpha at fragment R1. In addition, nitrogen count, volume, and chiV6 chain at R2, and chi3cluster and double bonded carbon count at R3 were classified as contributors. The ranges could be of a critical value when searching in the fragment library for designing new inhibitors. A fitness plot of the actual and predicted activities by this model is displayed in Figure 2D.

### Molecular docking of the QQSAR-generated compounds

The docking results show some QQSAR generated compounds including BMS\_608 and BMS\_1112 scoring better than the native ligand BMS-754807. BMS\_1112 formed three hydrogen bonds with the hinge in addition to forming three hydrogen bonds with ARG 1054, MET 1126 and ARG 1128, while the co-crystallize BMS-754807 ligand formed three hydrogen bonds to GLU 1050 and MET 1052 at the hinge without hydrogen bonding with other residues. This result revealed the importance of the BMS\_1112's two hydroxyl groups, which donated two hydrogens to GLU1050 and accepted hydrogen from MET1052. The BMS\_1112 also donated hydrogen atoms in the formation of non-hinge hydrogen bonds with ARG1128, MET1126 and ARG 1054 highlighting the availability of two amides in donating hydrogens to form many hydrogen bonds.

The co-crystallized BMS-754807 ligand lacks the non-hinge hydrogen bonds. Furthermore as shown in figure 3A, the docking analysis of fragment R1 of the BMS\_1112 at the IGF-1R active site shows the amide group of fragment R1 donates hydrogen to Met 1126 and Arg 1128. No hydrophobic interactions were observed between R1 and the active site of IGF-1R. This is in agreement with the results of QQSAR models.

Furthermore, the analysis of fragment R3 of BMS\_1112 shows interactions between hydroxyl groups on R3 with Met 1052 and GLU 1050, which are strategically located on the hinge region. There were also hydrophobic interactions between R3 and Met 1112, Leu 1051, Ala 1001, Met 1052, Met 1049, Val 1033, and Val 983. The large hydrophobic surfaces and three hydrogen bonding can increase the entropic interaction field of R3 that is related to information based descriptors (ldwAverage) mentioned in the QQSAR models.

## CONCLUSION

All generated QQSAR models were statistically significant, and the consensus of all models provide site-specific clues for design of new IGF-1R inhibitors. Non-hydrophobic substituents with branching and low NH count at fragment R1 are found to be essential for better activity. Scaffolds of the molecules that form the fragment R2 can be modified by decreasing oxygen atoms and increasing hydrogen bond donors. QQSAR studies suggest that the major contribution towards activity is from substituents at R3 that have less aromatic carbon count and more NH<sub>2</sub> groups.

In searching for novel and potent compounds, fragments that fit within the requirements of the QQSAR models can be selected and grown to generate a library of compounds which can be synthetically evaluated for IGF-1R inhibition. Molecular docking results of the QQSAR-generated compounds showed that the molecules of a high predicted activity possess interaction terms in good agreements with the developed models (hydrogen bonding at R1 and hydrophobic interaction and hydrogen bonding at R3). Some of the generated molecules rewarded high docking scores and efficient interaction features with the hinge, which can be strong candidates for future investigation.

## ACKNOWLEDGEMENT

Authors acknowledge the International Islamic University Malaysia for providing funds for this research under research matching grant RMGS 10-003-0013, and also thank Vlife Sciences Technologies, Pune, India for providing assistance during this work.

## REFERENCES

1. Butler AA, Yakar S, Gewolb IH, Karas M, Okubo Y, LeRoith D. *Insulin-like growth factor-I receptor signal*

*transduction: at the interface between physiology and cell biology. Comp Biochem Physiol B Biochem Mol Biol* 1998; 121(1): 19-26.

2. Lackey BR, Gray SL, Henricks DM. *The insulin-like growth factor (IGF) system and gonadotropin regulation: actions and interactions. Cytokine Growth Factor Rev* 1999; 10(3): 201-217.
3. Buchanan JL, Newcomb JR, Carney DP, Chaffee SC, Chai L, Cupples R, Epstein LF, Gallant P, Gu Y, Harmange JC et al. *Discovery of 2, 4-bis-arylamino-1, 3-pyrimidines as insulin-like growth factor-1 receptor (IGF-1R) inhibitors. Bioorg Med Chem Lett* 2011; 21(8): 2394-2399.
4. Velaparthy U, Wittman M, Liu P, Stoffan K, Zimmermann K, Sang X, Carboni J, Li A, Attar R, Gottardis M et al. *Discovery and initial SAR of 3-(1H-benzo[d]imidazol-2-yl)pyridin-2(1H)-ones as inhibitors of insulin-like growth factor 1-receptor (IGF-1R). Bioorg Med Chem Lett* 2007; 17(8): 2317-2321.
5. Sabbatini P, Rowand JL, Groy A, Korenchuk S, Liu Q, Atkins C, Dumble M, Yang J, Anderson K, Wilson BJ et al. *Antitumor activity of GSK1904529A, a small-molecule inhibitor of the insulin-like growth factor-I receptor tyrosine kinase. Clin Cancer Res* 2009; 15(9): 3058-3067.
6. Mayer SC, Banker AL, Boschelli F, Di L, Johnson M, Kenny CH, Krishnamurthy G, Kutterer K, Moy F, Petusky S et al. *Lead identification to generate isoquinolinedione inhibitors of insulin-like growth factor receptor (IGF-1R) for potential use in cancer treatment. Bioorg Med Chem Lett* 2008; 18(12): 3641-3645.
7. Chamberlain SD, Wilson JW, Deanda F, Patnaik S, Redman AM, Yang B, Shewchuk L, Sabbatini P, Leesnitzer MA, Groy A et al. *Discovery of 4, 6-bis-anilino-1H-pyrrolo [2, 3-d] pyrimidines: Potent inhibitors of the IGF-1R receptor tyrosine kinase. Bioorg Med Chem Lett* 2009; 19(2): 469-473.
8. Wittman MD, Carboni JM, Yang Z, Lee FY, Antman M, Attar R, Balimane P, Chang C, Chen C, Discenza L et al. *Discovery of a 2, 4-disubstituted pyrrolo [1, 2-f][1, 2, 4] triazine inhibitor (BMS-754807) of insulin-like growth factor receptor (IGF-1R) kinase in clinical development. J Med Chem* 2009; 52(23): 7360-7363.
9. Schulz MN, Hubbard, RE. *Recent progress in fragment-based lead discovery. Curr Opin Pharmacol* 2009; 9(5): 615-621.
10. Ajmani S, Agrawal A, Kulkarni SA. *A comprehensive structure-activity analysis of protein kinase B-alpha (Akt1) inhibitors. J Mol Graph Model* 2010; 28(7): 683-694.
11. Ajmani S, Jadhav K, Kulkarni SA. *Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR. QSAR Comb Sci* 2009; 28(1): 36-51.
12. Sampognaro AJ, Wittman MD, Carboni JM, Chang C, Greer AF, Hurlburt WW, Sack JS, Vyas DM. *Proline isosteres in a series of 2, 4-disubstituted pyrrolo [1, 2-f][1, 2, 4] triazine inhibitors of IGF-1R kinase and IR*

- kinase. *Bioorg Med Chem Lett* 2010;20(17): 5027-5030.
13. Lesuisse D, Mauger J, Nemecek C, Maignan S, Boiziau J, Harlow G, Hittinger A, Ruf S, Strobel H, Nair A et al. Discovery of the first non ATP competitive IGF-1R kinase inhibitors: advantages in comparison with competitive inhibitors. *Bioorg Med Chem Lett* 2011; 21(8): 2224-2228.
  14. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 1996; 17(5): 490-519.
  15. Golbraikh A1, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* 2003; 17(2): 241-253.
  16. Kirkpatrick S, Gelatt CD, Vecchi, MP. Optimization by simulated annealing. *Science* 1983; 220(4598): 671-680.
  17. Luco JM, Ferretti FH. QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. *J Chem Inf Comput Sci* 1997; 37(2): 392-401.
  18. Scior T, Medina-Franco JL, Do QT, Martínez-Mayorga K, Yunes-Rojas JA, Bernard P. How to recognize and workarround pitfalls in QSAR studies: a critical review. *Curr Med Chem* 2009; 16(32): 4297-4313.
  19. Sharaf MA, Illman DL, Kowalski BR. *Chemometrics*. New York: Wiley; 1986; p 332.