Journal of Molecular Graphics and Modelling 28 (2010) 683–694

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

A comprehensive structure–activity analysis of protein kinase B-alpha (Akt1) inhibitors

Subhash Ajmani *, Avantika Agrawal, Sudhir A. Kulkarni [1]

NovaLead Pharma Pvt. Ltd., Pride Purple Coronet, 1st floor, S No. 287, Baner Road, Pune 411045, India

ARTICLE INFO

Article history:
Received 11 June 2009
Received in revised form 12 January 2010
Accepted 16 January 2010
Available online 25 January 2010

Keywords:
Protein kinase B
Akt inhibitors
QSAR
Anticancer
Kinase inhibitors

ABSTRACT

Protein kinase B (PKB, also known as Akt) belongs to the AGC subfamily of the protein kinase superfamily. Akt1 has been reported as a central player in regulation of metabolism, cell survival, motility, transcription and cell-cycle progression, among the signalling proteins that respond to a large variety of signals. In this study an attempt was made to understand structural requirements for Akt1 inhibition using conventional QSAR, k-nearest neighbour QSAR and novel GQSAR methods. With this intention, a wide variety of structurally diverse Akt1 inhibitors were collected from various literature reports. The conventional QSAR analyses revealed the key role of Baumann's alignment independent topological descriptors along with other descriptors such as the number of hydrogen bond acceptors, hydrogen bond donors, rotatable bonds and aromatic oxygen (SaaOcount) along with molecular branching (chi3Cluster), alkene carbon atom type (SdsCHE-index) in governing activity variation. Further, the GQSAR analyses show that chemical variations like presence of hetero-aromatic ring, flexibility, polar surface area and fragment length present in the hinge binding fragment (in the present case fragment D) are highly influential for achieving highly potent Akt1 inhibitors. In addition, this study resulted in a k-nearest neighbour classification model with three descriptors suggesting the key role of oxygen (SssOE-index) and aromatic carbon (SaaCHE-index and SaasCE-index) atoms electro-topological environment that differentiate molecules binding to Akt1 kinase or PH domain. The developed models are interpretable, with good statistical and predictive significance, and can be used for guiding ligand modification for the development of potential new Akt1 inhibitors.

1. Introduction

Kinases are the key players in cell signaling pathways that regulate cell growth, proliferation, and apoptosis. They transduct signals from growth factor receptors for cell growth or apoptosis by phosphorylation of their substrates which are mostly downstream kinases involved in cell signaling processes themselves [1]. Consequently, unregulated kinase activity can result in uncontrolled cellular growth and inappropriate regulation of apoptosis; a key mechanism in oncogenesis [2]. Amongst various kinases, an aberrant activation of Akt, has been recognized to be responsible for a wide range of proliferative and anti-apoptotic processes in many human tumors.

Akt is a part of AGC (cAMP-dependent (A), cGMP-dependent (G), and phospholipid-dependent (C)) family of kinases. Because it bears high homology to protein kinase A (PKA) and protein kinase C (PKC), Akt is also referred to as protein kinase B (PKB). Akt consists of three different cellular isoforms, namely, Akt1 (PKBα), Akt2 (PKBβ), and Akt3 (PKBγ). Being approximately 80 percent identical, these isozymes have a high degree of overall homology [3].

Akt1 is composed of a kinase domain, a N-terminal pleckstrin homology (PH) domain, and a short carboxy terminal tail region. This protein is activated when Thr308 and Ser473 are phosphorylated. Once activated, Akt1 inhibits apoptosis and stimulates cell-cycle progression by phosphorylating numerous targets in various cell types, including cancer cells. Numerous studies have shown that dysregulation of Akt is a major contributor to tumorigenesis [4–5]. Literature reveals that Akt1 is mostly involved in breast cancer, human prostate, ovarian carcinomas and in gastric adenocarcinomas; Akt2 is amplified in ovarian, pancreatic, and breast cancers; and Akt3 is amplified in breast cancer and prostate cell lines [6]. Consequently, the development of molecules capable of blocking protein kinase B activity is a valuable route for anticancer drug discovery [7]. However, the development of inhibitors of Akt as small molecule therapeutics for the treatment of cancer has been hindered by a lack of Akt specific inhibitors (versus the AGC family of kinases mostly PKA and PKC) and isozyme selective (Akt1, Akt2, and Akt3) Akt inhibitors due to high sequence identity and similarity of corresponding targets [8].

* Corresponding author. Tel.: +91 20 27291590; fax: +91 20 27291590.
E-mail address: subhasha@novaleadpharma.com (S. Ajmani).
[1] Also at VLife Sciences Technologies Pvt. Ltd.

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.01.007

Recently Lindsley et al. reported few Akt1 selective molecules and attributed the observed selectivity to be the PH-domain dependent (allosteric) mode of inhibition [9]. Additionally, few Akt inhibitors are currently being tested in clinical trials [8].

A number of *in-silico* and experimental approaches have been mentioned for assisting in the design of novel and more effective Akt inhibitors [10–12]. However, there is a small number of quantitative structure activity relationship (QSAR) studies reported in the literature and they mainly focus on a particular chemical class of molecules [13–14]. Recently Dong et al. reported a QSAR study of Akt1 inhibitors using support vector machine [15]. However, application of this model is limited to virtual screening as the QSAR model descriptors are difficult to interpret and would not be helpful in the design of new molecules. Therefore, a QSAR study which can provide understanding of the structural requirement of Akt1 inhibitor and aid in the design of new inhibitors is needed. Also, a QSAR study using a wide variety of chemically diverse Akt1 inhibitors would allow application of the developed model on a variety of designed new chemical entities.

For the present study, we gathered a chemically diverse set of Akt1 inhibitors available in various literature reports [9,10,18–26]. To gain deeper insights into the structural requirements for Akt1 inhibition and develop quantitative models for the development of new Akt1 inhibitors, we utilized the recently developed method of Group-Based QSAR [16]. In addition, we also developed a conventional quantitative structure activity relationship using regression methods as well as k-nearest neighbour (k-NN) method. Moreover, it was thought worthwhile to develop a classification model that is helpful in highlighting new molecules on the basis of their ability to interact with kinase or PH binding domain of Akt1.

Since a wide variety of structurally diverse Akt1 inhibitors have been considered in this study, it provides the ability to understand the structural and feature requirements for Akt1 inhibition, thus aiding in the design of novel and potent Akt1 inhibitors.

2. Methodology

All computations and molecular modeling studies were carried out on a Windows XP workstation using the molecular modeling software package VLife Molecular Design Suite (VLifeMDS) version 3.5 [17].

2.1. Dataset

The binding affinity (IC_{50} ranging 0.16 nm to 126 μm) data of 265 molecules was collected from the various literature reports [9,10,18–26] and converted to pIC_{50} for the QSAR analysis. Additionally, the literature also reports the binding domain (PH or kinase) for all these molecules. The three dimensional structures of the 265 molecules were downloaded from bindingDB database [27]. The dataset can be classified into 12 different chemical classes *i.e.* azepane, cyanopyridine, indazole-pyridine, isoquinoline-pyridine, oxindole-pyridine, pyrazinone, pyrazolopyridine-pyridine, pyridine-base, quinoxaline, tetrazolylpyridine, trans-bispyridinylethylene, and miscellaneous.

2.2. Molecular descriptors

A total of 634 two dimensional descriptors were calculated using VLifeMDS software [17]. These included various physicochemical, structural, topological, electro-topological, Baumann alignment independent topological descriptors [28] and Merck molecular force field (MMFF) atom type count descriptors. Preprocessing of the independent variables (*i.e.* descriptors) was done by removing the invariables (*i.e.* descriptor with a constant value for more than 95 percent molecules), which resulted in 373 descriptors in the descriptor pool.

2.3. Creation of training and test set

Optimal training and test sets were generated using sphere exclusion algorithm [29]. This algorithm allows construction of the data sets by using descriptor space occupied by the representative points, such that the test set molecules represent a range of biological activities similar to the training set; thus, the test set is truly a representative of the training set. In order to assess the similarity of the distribution pattern of the molecules in the generated sets, statistical parameters (with respect to the biological activity) *i.e.* mean, maximum, minimum and standard deviation were calculated for the training and test sets.

2.4. Simulated annealing as variable selection method

In order to select a subset of descriptors (variables) from the descriptor pool, a variable selection method like simulated annealing is required. Simulated annealing mimics the physical process of annealing, which involves heating a system to a high temperature and then gradually cooling it to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Boltzmann distribution so that at equilibrium, low energy states are the most populated [30].

At each non-zero temperature, configurations with higher energy are allowed. In current context, energy is reflected in terms of the objective function r^2 (determined by regression model) and a configuration with the minimum energy is the optimal solution with a subset of descriptors (V). The temperature becomes a control parameter and by allowing configurations with higher energy (V's with higher cost) the search method is less likely to get trapped into local minima. A Metropolis method can be used to simulate the evolution of the system. At each step the algorithm selects V^{new} from V^{old}. Notice that at each move, V is changed by swapping descriptor(s) from a pool of remaining descriptors.

At each step $d = r^2(V^{new}) - r^2(V^{old})$ is calculated. If $d > 0$, V^{new} is accepted, else, it is accepted with probability $\exp(-d/T)$, where T stands for temperature control parameter. The overall idea is to start with a high value of T, so that all steps are accepted and then gradually reduce T as the simulation progresses, so that eventually only steps that improve the solutions are accepted. At every m step, an acceptance ratio is calculated. The process stops when acceptance ratio $\leq \beta$ (a set convergence criterion), giving a vector V with minimum cost.

2.5. Partial least squares regression (PLSR) method

PLS was developed in the 1960's by Herman Wold as an econometric technique, but some of its most avid proponents (including Wold's son Svante) are chemical engineers and chemometricians. PLS is an effective technique for finding the relationship between the properties of a molecule and its structure. In mathematical terms, PLS relates a matrix Y of dependent variables to a matrix X of molecular structure descriptors [31]. PLS has two objectives: to approximate the X and Y data matrices, and to maximize the correlation between them. Whereas the extraction of PLS components is performed stepwise and the importance of a single component is assessed independently, a regression equation relating each Y variable with the X matrix is created. PLS decomposes the matrix X into several latent variables that correlate best with the activity of the molecules [32].

2.6. k-nearest neighbour classification

In order to predict binding domain of Akt1 (PH or kinase), a k-nearest neighbour (k-NN) classification model was developed. The k-NN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbours in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard k-NN method follows the following steps: (1) calculate the distances between an unknown object (u) and all the objects in the training set; (2) select k objects from the training set most similar to object u according to the calculated distances; and (3) classify object u to the group which the majority of the k objects belongs [33]. An optimal k value is selected by optimization through the classification of a test set of samples or by leave-one-out cross-validation. The variables and optimal k values were selected by using simulated annealing as variable selection method.

2.7. k-nearest neighbour QSAR (k-NN weighted average method)

The k-NN method was also used to develop a QSAR model using continuous variable *i.e.* using activity as pIC_{50} values. In this case, by using a developed k-NN QSAR model the activity of a molecule can be predicted using weighted average activity (Eq. (1)) of the k most similar molecules in the training set.

$$\hat{y}_i = \sum w_i y_i \tag{1}$$

Where y_i and $\hat{y}_i$ are the actual and predicted activity of the ith molecule respectively, and w_i are weights calculated using (Eq. (2)).

$$w_i = \frac{\exp(-d_j)}{\sum_{j=1}^{k} \exp(-d_j)} \tag{2}$$

The similarities were evaluated as the inverse of Euclidean distances (d_j) between molecules (Eq. (3)) using only the subset of descriptors corresponding to the model. Where, k is number of nearest neighbours in the model.

$$d_{i,j} = \left[\sum_{m=1}^{Vn} (X_{i,m} - X_{j,m})\right]^{1/2} \tag{3}$$

Where, X is the matrix of selected descriptors (Vn) for the k-NN QSAR model.

2.8. Model evaluation and validation

This is done to test the internal stability and predictive ability of the QSAR models.

2.8.1. Internal validation

Internal validation was carried out using leave-one-out (q^2, LOO) method. To calculate q^2, each molecule in the training set was sequentially removed, the model refit using same descriptors, and the biological activity of the removed molecule predicted using the refit model. The q^2 was calculated using Eq. (4).

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \tag{4}$$

Where y_i, $\hat{y}_i$ are the actual and predicted activity of the ith molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

2.8.2. External validation

For external validation, activity of each molecule in the test set was predicted using the model generated from the training set. The $pred\_r^2$ value is calculated as follows (Eq. (5))

$$pred\_r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{\text{mean}})^2} \tag{5}$$

Where y_i, $\hat{y}_i$ are the actual and predicted activity of the ith molecule in the test set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Both summations are over all molecules in the test set. Thus the $pred\_r^2$ value is indicative of the predictive power of the current model based on the external test set.

2.8.3. Randomization test

To evaluate the statistical significance of a QSAR model for an actual data set, one tail hypothesis testing was used [34]. The robustness of the models for training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging the activities of the molecules in the training set and the significance of the models was derived based on the calculated *Zscore*; Eq. (6).

$$Zscore = \frac{(q^2_{org} - q^2_a)}{q^2_{std}} \tag{6}$$

Where q^2_{org} is the q^2 value calculated for the actual data set, q^2_a is the average q^2, and q^2_{std} is the standard deviation of q^2, calculated for various iterations using different random data sets.

The probability (α) of significance of the randomization test is derived by using calculated *Zscore* value as given in the literature [35].

2.8.4. Evaluation of the quantitative models

Developed quantitative models were evaluated using following statistical measures: n, number of observations (molecules); k, number of variables (descriptors); Number of components, number of optimum PLS components in the model; Number of nearest neighbours, number of k-nearest neighbour in the model; r^2, coefficient of determination; q^2, cross-validated r^2 (by leave one out); $pred\_r^2$, r^2 for external test set; F-test, F-test value for statistical significance; *Zscore*, *Zscore* calculated by the randomization test; best_ran_q^2, highest q^2 value in the randomization test; α, statistical significance parameter obtained by the randomization test; SEE, standard error of estimate of the model; cv_SE, standard error of cross-validation and pred_SE, standard error of external test set prediction.

The r^2 and q^2 values were used as deciding factors in selecting the optimal models.

2.9. Group-based QSAR (GQSAR) method

GQSAR is a recent QSAR method developed in our laboratory, which addresses the challenges of QSAR model interpretation and the inverse QSAR problem [16]. GQSAR method comprises of three steps: (1) generation of molecule fragments using a set of predefined chemical rules, (2) calculation of descriptors for the generated fragments, (3) build statistical models using the calculated fragment descriptors and their interactions. GQSAR thus allows establishing a correlation of chemical group/fragment variation at different molecular sites of interest with the biological activity.

Fragmentation is done by applying specific chemical rules for breaking the molecules along specific bonds and/or bonds on ring fusion and/or any pharmacophoric feature such as hydrogen bond acceptor, hydrogen bond donor, hydrophobic group, charged group

etc. Thus, the GQSAR method deals with molecular fragments instead of the molecule as a whole. The fragment descriptors and their interactions are related to biological activity, resulting in model(s) that highlight important substitution site(s) along with their chemical nature and interactions. The suggested important fragments can be used as the building blocks to design novel molecules.

Here in, molecules were divided into four fragments based on the fragmentation rules derived in light of the specific molecular structure requirement for active site interactions [36]. In order to consider the environment of the neighbouring fragment(s), the attachment point atoms were also included in the fragments. The scheme of molecular fragmentation is shown in Fig. 1a. Also, several representative molecules, with their corresponding fragments considered in the present study, are shown in Fig. 1b.

Fragment A: It is constituted of the terminal aromatic ring at position 1 with respect to the fragment C. Methyl groups were considered part of fragment A, for molecules lacking this substituent.

Fragment B: This is constituted of a linker (with or without ring) between fragment A and C.

Fragment C: It is formed by a substituted aromatic ring present in the middle of molecule.

Fragment D: It is constituted of the substituent at position 3 with respect to the fragment C. For molecules lacking this substituent, the methyl group was considered as fragment D.

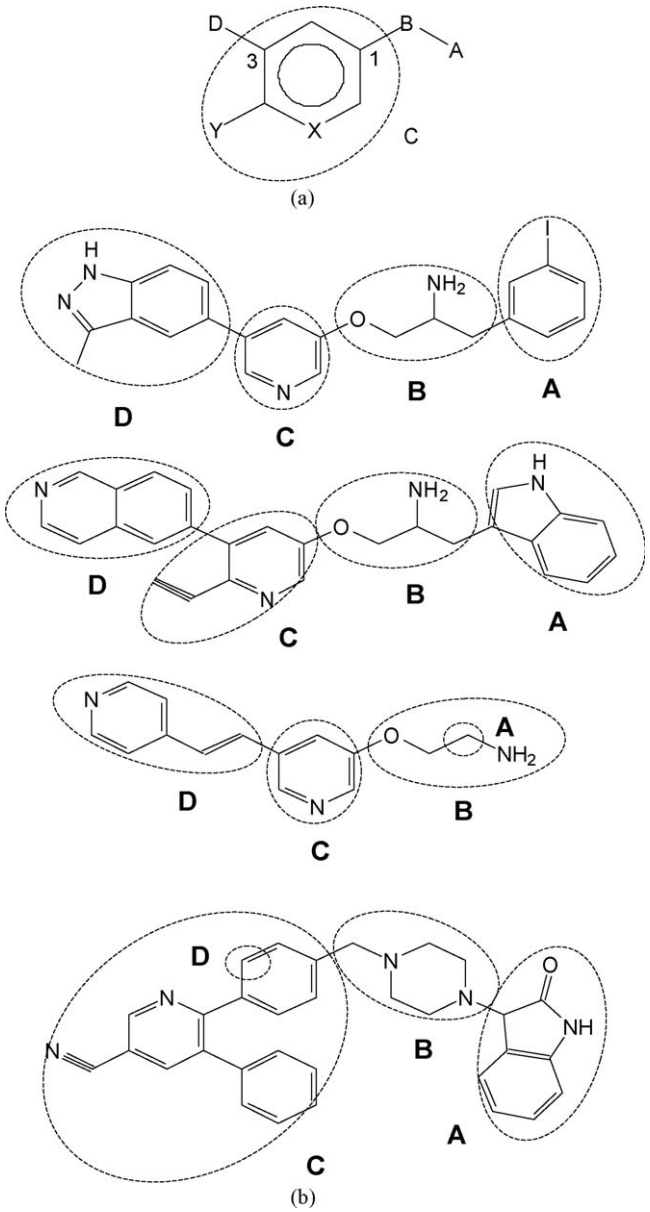

Fig. 1. (a) Molecular fragmentation scheme for Akt1 inhibitors. (b) Representative Akt1 inhibitors with their corresponding fragments collected from various literature reports [refs. [9,10,18–26].

For GQSAR analysis, various two dimensional descriptors (as discussed in Section 2.2), were calculated for various groups present at different substitution sites of the molecules (*i.e.* Fragment A, B, C and D). The removal of the invariable group descriptors resulted in a total of 311 group descriptors which can be used further. Since the same descriptors are calculated for various groups at different sites, the following nomenclature is used for naming a descriptor at a particular position, for example A_smr represents the molar refractivity of the group present at substitution site A. The following formula was used for the calculation of interaction/cross terms of the various group descriptors at different substituent sites *e.g.*:

'A_slogp*B_smr' interaction descriptor means the mathematical product of A_slogp with B_smr (A_slogp × B_smr).

Where, A_slogp corresponds to the value of slogP of the group at substitution site A and B_smr is the value of molar refractivity of the group at substitution site B. The GQSAR model developed using these fragment interaction/cross terms are referred as GQSAR_IT model.

Since a large pool of group descriptors is now available for building a quantitative model and not all of the group descriptors are important for the activity, one needs a method to pick optimal subset of group descriptors that explains variation in the activity. In this study, a simulated annealing variable selection method was coupled with a variety of statistical methods available for building quantitative GQSAR models.

2.9.1. Role of GQSAR in solving inverse QSAR problem

A methodology is said to address inverse QSAR problem when it provides a systematic way/method to design molecules that satisfy QSAR requirements and thereby design active molecules.

In the GQSAR method, the following steps address the inverse QSAR problem:

- Identification of important molecular part(s) and their corresponding properties (along with interaction terms) found to play key role in activity determination based on developed GQSAR model(s).
- Derive ranges of properties that are found to be important in GQSAR model(s) by using active molecules and their corresponding properties ranges in the dataset used for building model.
- Obtain similar fragments by searching fragment in databases that satisfy derived ranges for all the fragment properties that are found to be important in GQSAR model(s).
- Generation of new molecules by combining fragments that satisfy ranges mentioned above.

3. Results and discussion

3.1. k-nearest neighbour classification model

A classification model was generated using k-nearest neighbour method in VLifeMDS, for predicting binding domain (PH or kinase) of designed molecules. The classification model was built from a training set of 211 molecules (21 PH and 190 kinase domain

Table 1a
Confusion matrix of k-NN classification model for training and test sets.

	Training set	
	PH domain (Predicted)	Kinase domain (Predicted)
PH domain (Actual)	$A = 21$	$B = 0$
Kinase domain (Actual)	$C = 0$	$D = 190$
	Test set	
PH domain (Actual)	$A = 13$	$B = 0$
Kinase domain (Actual)	$C = 0$	$D = 41$

Table 1b
Parameters used to interpret confusion matrix of k-NN classification model.

Parameter	Significance/formula
Accuracy (AC) Training AC = 1 Test AC = 1	The proportion total number of correct predictions = $[(A+D)/(A+B+C+D)]$
True positive rate (TP)/sensitivity Training TP = 1 Test TP = 1	The proportion of correctly identified kinase domain molecules = $[D/(C+D)]$
False positive rate (FP) Training FP = 0 Test FP = 0	The proportion of incorrectly identified PH-domain molecules = $[B/(A+B)]$
True negative rate (TN)/specificity Training TN = 1 Test TN = 1	The proportion of correctly identified PH-domain molecules = $[A/(A+B)]$
False negative rate (FN) Training FN = 0 Test FN = 0	The proportion of incorrectly identified kinase domain molecules = $[C/(C+D)]$

inhibitors), while the test set included 54 molecules (13 PH and 41 kinase domain inhibitors). This study led to the development of statistically significant k-NN model with three descriptors *i.e.* SssOE-index, SaaCHE-index, SaasCE-index using two nearest neighbours. These molecular descriptors suggest the key role of oxygen (SssOE-index) and aromatic carbon (SaaCHE-index and SaasCE-index) atoms electro-topological environment to differentiate between the binding domains of the molecule (kinase or PH). The confusion matrix, Table 1a, was generated to evaluate performance of this classification model on training and test sets. Table 1b reports the parameters used to interpret the confusion matrix.

All the parameters reported in Table 1b indicate that the model is able to classify the overall dataset correctly with respect to their binding domain (*i.e.* kinase or PH). The prediction of binding domain is accurate for both the training and test sets.

3.2. QSAR models

Using the sphere exclusion method, the data set was divided into a training set (217 molecules) and test set (48 molecules). The statistical parameters for assessing the distribution of activity in the training and test sets have been listed in Table 2. As can be seen from Table 2, the minimum 'Akt1 inhibition' (activity) of test set is greater than the minimum activity of training set and the maximum activity of the test set is less than the maximum activity of the training set, this indicates that the test set is within the activity domain of the training set. The comparable standard deviation and the mean values (as shown in Table 2) of training and test sets show that there is a similar distribution of training and test set molecules with respect to the activity.

Table 2
Statistical parameters for Akt1 inhibitory activity distribution in training and test sets.

Parameters	Training set	Test set
Maximum	0.770	−0.080
Minimum	−5.100	−4.370
Mean	−2.287	−1.754
Standard deviation	1.333	1.309

All the calculated descriptors remaining after preprocessing (373) were subjected to simulated annealing variable selection coupled, separately, with multiple regression (MR), PLS regression (PLSR) and k-nearest neighbour (k-NN) methods for building three different QSAR models (MR, PLSR, and k-NN based) based on the same training set. This study led to various statistically significant models and their statistical parameters are reported in Table 3. Table 4 reports selected descriptors with their regression coeffcient and percentage contribution in each of the reported QSAR models.

Fig. 2 shows the comparison of percentage contribution of 18 descriptors common between MR and PLS QSAR models. One alignment independent topological descriptor TNO8 from the MR model and two descriptors in PLSR model *i.e.* SdsCHE-index and TCC5 were uncommon between the two models. The percent contribution of these three descriptors in their respective models was TNO8 (1.98), SdsCHE-index (2.26) and TCC5 (2.02). Table 5 reports the list of important descriptors found in various reported QSAR models along with their descriptor category and definition.

Fig. 3 shows the observed versus predicted 'Akt1 inhibition' (activity) plot of training and test set molecules by QSAR PLSR model. The randomization tests suggest that all the proposed QSAR models have a probability of less than 0.0001 of being generated by chance (as can be seen in Table 3).

It was observed (as shown in Table 4) that the activity variation is explained in terms of approximately 75 percent and 25 percent by the Baumann's alignment independent topological descriptors and other basic descriptors, respectively. Also, descriptors influencing the activity in favourable and unfavourable ways were found to be near 55 percent and 45 percent, respectively. This information suggests that there is almost equal opportunity to optimize both the favourable and unfavourable descriptors in the design of new molecules.

The appearance of the basic molecular descriptors like number of hydrogen bond acceptor (HBA), number of hydrogen bond donor (HBD), rotatable bond count, SaaOcount and chi3Cluster provides the following explanation towards activity variation:

- Presence of only hydrogen bond acceptor is detrimental for the activity, since HBA and HBD are inversely and directly proportional respectively to the variation in activity.
- If only a hydrogen bond acceptor is required, it is best to have an oxygen atom connected to two aromatic atoms, as descriptor SaaOcount (*i.e.* oxygen present in an aromatic ring or connecting two aromatic rings) is directly proportional to the activity variation.
- Since rotatable bond is inversely proportional to activity, it is more favourable if the hydrogen bond donor is added at the terminal side to prevent the increase in the number of rotatable bonds. Additionally, this also suggests designing rigid molecules.
- Descriptor chi3Cluster is directly proportional to activity, indicating the importance of fused rings or branched molecules are favourable for the activity.

In general, a Baumann's alignment independent topological descriptor TXYZ can be defined as a count of fragments formed with atom types X and Y separated by topological distance of Z bonds. The symbols = and # (for atom types) correspond to double bonded (sp2 hybridized) and triple bonded (sp hybridized) atoms, respectively.

Careful examination of the descriptors selected in both the QSAR MR and QSAR PLSR models reveals that 10 out of 14

Table 3
Statistical parameters of various QSAR and GQSAR models.

Model Parameters	QSAR MR	QSAR PLSR	QSAR k-NN	GQSAR MR	GQSAR_IT MR	GQSAR k-NN
Training(n)/test	217/48	217/48	217/48	217/48	217/48	217/48
r^2	0.689	0.697	*	0.654	0.741	*
q^2	0.628	0.632	0.654	0.576	0.693	0.646
F-test	22.98	68.79	*	18.51	29.59	*
SEE	0.778	0.746	*	0.824	0.711	*
cv_SE	0.852	0.823	0.784	0.912	0.774	0.793
pred_r^2	0.711	0.734	0.690	0.653	0.706	0.698
pred_SE	0.761	0.730	0.787	0.834	0.768	0.778
Zscore_cv	11.86	18.96	*	11.47	11.01	*
best_rand_q^2	−0.045	−0.016	*	−0.124	−0.095	*
α_rand_cv	<0.0001	<0.0001	*	<0.0001	<0.0001	*
Number of descriptors (k)	19	20	21	20	19	18
Number of components/nearest neighbours	*	7	4	*	*	6

* NA – not applicable.

descriptors contain an atom pair, either nitrogen atoms or a nitrogen atom along with a carbon, oxygen or fluorine atom. This suggests the overall impact a nitrogen atom has on the activity.

- The descriptors TCF6 and TNF13 are found to be directly proportional to the activity variation indicating the presence of a fluorine atom (like trifluoro methyl, trifluoro methoxy, fluoro, di or mono-substituted methyl) to be favourable for activity.
- The descriptor TOO7 shows the importance of two carbonyl groups at the para-position to each other in a phenyl ring (as present in most of the azepane derivatives) to be conducive for the activity.
- The descriptor T=#10 indicates that in general the presence of a sp hybridized carbon atom (*i.e.* cyano or ethyene group) in any part of the molecule is detrimental to the activity except for the pyrazinone or pyrazolopyridine-pyridine class, here these groups, if present at the 2 position in pyridine, are found to be beneficial for activity.
- The presence of the descriptor TNO17 shows the relative importance of long azepane derivatives which have 4-pyridine and o-hydroxy phenyl substitution on the two opposite terminals.
- Investigation of the family of alignment independent descriptors with C–N pair and N–N pair led to the suggestion that these descriptors form an optimal structural requirement for activity and have to be considered collectively for new molecule design.

The above mentioned descriptors of QSAR MR model were also found to be important in QSAR PLSR model. In addition to these, the descriptors SdsCHE-index and TCC5 were found to be directly and inversely proportional to the activity in QSAR PLSR model, respectively. Thus, the SdsCHE-index indicates the importance of substituted ethylene (–CH=CH–) or a double bonded carbon atom (–CH=) to increase the activity. The descriptor TCC5 is difficult to interpret but may be an indicator of the molecular size or enhancement in partition coefficient of molecule.

The MR and PLSR QSAR models reported above captured a linear relationship of descriptors with the activity. In order to capture a non-linear relationship of the descriptors to activity, a k-NN QSAR model (derived by a distance based learning approach) was developed. The analysis led to a statistically significant k-NN QSAR model using four nearest neighbours with 21 descriptors, and is comparable to regression models. The statistical parameters of the developed k-NN QSAR model are reported in Table 3. Fifteen out of 21 descriptors in the model were Baumann's alignment independent topological descriptors *i.e.* T==4, T=#12, T=#4, T=Cl4, T=F7, T=N13, T=N16, T=O6, T=O9, TCC13, TCC9, TCF11, TCN13, TCO9, and TNN11 and the remaining six were basic molecular descriptors *i.e.*

Table 4
Selected descriptors with their regression coeffcient and percentage contribution in QSAR MR and QSAR PLSR models.

Descriptor	Multiple regression coefficient (±Error)	Percent contribution	PLS regression coefficient	Percent contribution
H-AcceptorCount	−0.2955(±0.0472)	−6.49	−0.2622	−5.42
H-DonorCount	0.2969(±0.0877)	3.88	0.2900	3.56
RotBondCount	−0.1337(±0.0293)	−3.08	−0.1735	−3.76
chi3cluster	0.8437(±0.1607)	4.77	1.2330	6.55
SaaOcount	1.4777(±0.2439)	5.02	1.4929	4.77
SdsCHE-index	*	*	0.0935	2.16
T=#10	−0.1495(±0.0494)	−2.31	−0.1166	−1.70
TCC5	*	*	−0.0160	−2.02
TCN11	0.3543(±0.0166)	9.93	0.3561	9.39
TCN18	−0.2870(±0.0226)	-7.87	−0.2848	−7.34
TCO12	−0.3279(±0.0119)	−8.61	−0.3306	−8.16
TCF6	0.2758(±0.0767)	4.11	0.2772	3.88
TNN4	−0.5369(±0.0833)	−6.68	−0.5453	−6.38
TNN11	0.7056(±0.0885)	9.34	0.8533	10.61
TNN13	−0.6730(±0.1098)	−5.28	−0.6496	−4.79
TNO5	0.2472(±0.0488)	3.53	0.2781	3.73
TNO6	−0.3898(±0.0916)	−4.17	−0.3749	−3.77
TNO8	0.1372(±0.0653)	1.98	*	*
TNO17	1.1181(±0.2271)	4.38	1.0336	3.81
TNF13	0.3869(± 0.1331)	2.62	0.3974	2.53
TOO7	0.8437(±0.1331)	5.93	0.8592	5.68
Regression constant	−4.3472	*	−4.5447	*

* NA – not applicable.

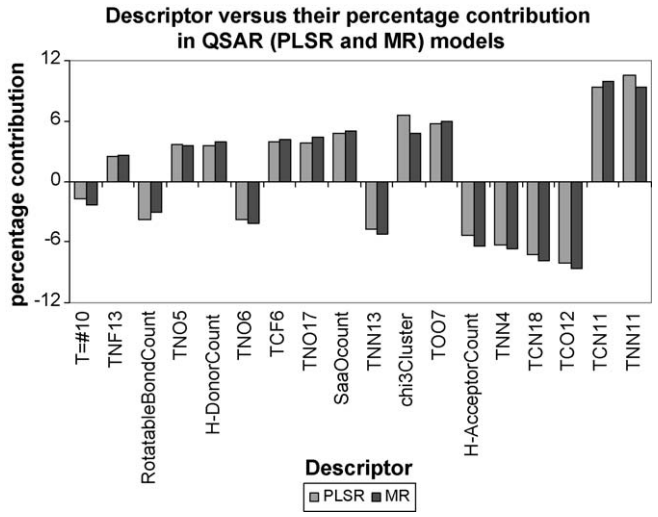

Fig. 2. Plot of percentage contribution of the descriptors common in both QSAR MR and QSAR PLS regression models towards Akt1 inhibitory activity.

XlogP, SdsCHE-index, SssNHE-index, chiV2, SsNH2E-index, and SaaOE-index (as shown in Table 5). The limitation of the k-NN method (as it is not a curve fitting method) does not allow estimation of the relative importance of individual descriptors in the model.

The developed QSAR models allow for an understanding the molecular properties/features that play an important role in governing the variation in the activities. In addition, this QSAR study allowed investigating the influence of simple and easy to compute descriptors in determining biological activities that could highlight the key factors and may aid in the design of novel and potent molecules.

3.3. GQSAR models

As reported above, the conventional QSAR models generated are statistically significant and indicate the importance of basic molecular properties such as HBA, HBD, and various alignment independent topological descriptors. However, based on the information of these descriptors, these models do not exactly indicate the part of the molecule where the modifications are required to improve the activity, thus posing a hurdle in the complete structural interpretation. Therefore, in order to gain insight to the influential molecular part(s), in terms of their chemical information responsible for the variation in activity, GQSAR models involving fragment descriptors and their interactions (cross terms) were developed.

Table 3 reports the statistical parameters of the developed GQSAR MR and GQSAR_IT MR (model with cross terms) models. The GQSAR_IT MR model, using simple two dimensional descriptors, resulted in statistically improved parameters in comparison to the regression based QSAR models as can be seen from Table 3. Fig. 4 shows the observed versus predicted Akt1 inhibition (activity) plot of training and test set molecules by the GQSAR_IT MR model. Tables 6 and 7 report selected descriptors with their regression coefficient and percentage contribution in the GQSAR MR and GQSAR_IT MR models, respectively. Table 8 contains a list of important descriptors found in the GQSAR MR and GQSAR_IT MR models along with their descriptor category and definition.

It can be seen from Figs. 5 and 6, that substitutions at part-D are most influencing with highest percentage contribution in both GQSAR MR and GQSAR_IT MR models. This is also supported by the fact that the largest amount of variation in the chemical substituents is contained in fragment D followed by fragment A then fragments B and C. This is also in line with the recently reported study where it has been reported that hinge binding fragment (in the present case fragment D) is critical for achieving highly potent kinase inhibitors [35].

- The importance of rotatable bond count was reflected in the conventional QSAR models, in addition, the GQSAR MR model has indicated the part (Region; *i.e.* part-D) of the molecule where the rotatable bond count descriptor has to be varied to get stronger inhibitors (as shown in Fig. 5).
- The presence of A_Mol.Wt. shows that increase in molecular weight of fragment A may lead to an increase in the activity. Descriptor A_MMFF64 (beta carbon in 5-membered hetero-aromatic ring) indicates the importance of indole or benzimidazole nucleus in part-A to be conducive for the activity. Descriptors A_SdOE-index and A_T=C7 are directly and inversely proportional to the activity, respectively. This indicates the role of the substituted carbonyl (C=O) groups to be conducive and mono-substituted aromatic ring to be detrimental for the activity.
- The three descriptors in fragment B (B_4PathCount, B_PSAExP&S, and B_T=C6) in GQSAR MR model shows the variation is mainly contained to the azepane derivatives. Two descriptors B_PSAExP&S (polar surface area) and B_T=C6 show that fragment B in the azepane derivatives is favourable for the activity.
- Fragments A and B contribute equally (approximately 16 percent) in the GQSAR MR model, while fragment C was found to be the least contributing (approximately 8 percent) of all four fragments in GQSAR MR model. The descriptors C_SaasCcount and C_TCO7 were found to be directly and inversely proportional to the activity, respectively. This indicates that increasing in number of substituted aromatic carbon atoms is conducive as where those substituted with oxygen atom containing groups (i.e. –OH, COOH, –CO etc.) is detrimental to the activity.
- In the GQSAR MR model, fragment D contains the major variation (60 percent). The most contributing descriptor in part-D is D_4PathCount (15 percent) and suggests the role of substituted oxindole nucleus (most bulky and lengthy) to be conducive for the activity. The next most influential (approximately 9 percent) fragment D descriptor is D_T=N5 and suggests that increasing the nitrogen atom substituent is detrimental to the activity. The presence of the descriptor D_SaaNHE-index (directly proportional contributing around 6 percent) shows the importance of aromatic rings containing nitrogen atoms (for example pyrrole, indole, and indazole) in fragment D is conducive to the activity.
- The descriptors D_TCC7, D_Rotbonds, and D_PSAInP&S are inversely proportional to the activity and suggest that increasing the length by adding carbon atoms, increasing number of rotatable bonds and/or increasing polar surface area of fragment D could be detrimental for the activity.
- The descriptor D_SdsCHE-index is directly contributing to the activity, suggesting that presence of substituted –CH=CH– (vinyl group) or =CH– (methylene group) in fragment D is conducive for the activity.
- The descriptors D_TCN1 and D_TCN5 are directly proportional to the activity and indicates that the presence of nitrogen atoms either in ring or 1,4-disubstituted aromatic ring could be conducive for the activity.
- The GQSAR_IT MR model was built to find the importance of the interaction descriptors. As seen in Table 7, the interaction between all pairs of fragments A, B, C and D were found to be important and impacted the activity. The contributions of the interaction descriptors, as seen in Fig. 6, revealed that the interaction between fragments C and D of the molecules have most impact in the activity variation, which was followed by interaction between fragments B and C. Discussed below are

Table 5
List of important descriptors (found in various QSAR models) along with their category and definition.

Descriptor	Category	Definition
H-AcceptorCount	Structural	Number of hydrogen bond acceptor in a molecule
H-DonorCount	Structural	Number of hydrogen bond donor in a molecule
RotatableBondCount	Structural	Number of rotatable bond in a molecule
XlogP	Physicochemical	Log of octanol/water partition coefficient
chiV2	Topological	Molecular connectivity index of fragment made of two consecutive edges
chi3Cluster	Topological	Molecular connectivity index of cluster (branched fragment) made of three edges
SaaOcount	Electro-topological	Count of oxygen atom in an aromatic ring
SaaOE-index	Electro-topological	Electro-topological state index of oxygen atom in an aromatic ring
SdsCHE-index	Electro-topological	Electro-topological state index of carbon atom connected with three atoms viz. a heavy atom by single bond, a heavy atom by double bond and a hydrogen atom
SsNH2E-index	Electro-topological	Electro-topological state index of nitrogen atom connected with three atoms viz. one heavy atom by single bond and two hydrogen atoms.
SssNHE-index	Electro-topological	Electro-topological state index of nitrogen atom connected with three atoms viz. two heavy atoms by single bond and a hydrogen atom.
T==4	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atoms separated by 4 bonds
T=#4	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any triple bonded (sp hybridized) atom separated by 4 bonds
T=#10	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any triple bonded (sp hybridized) atom separated by 10 bonds
T=#12	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any triple bonded (sp hybridized) atom separated by 12 bonds
T=Cl4	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and chlorine atom separated by 4 bonds
T=F7	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and fluorine atom separated by 7 bonds
T=N13	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 13 bonds
T=N16	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 16 bonds
T=O6	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any oxygen atom separated by 6 bonds
T=O9	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any oxygen atom separated by 9 bonds
TCN11	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 11 bonds
TCN13	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 13 bonds
TCN18	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 18 bonds
TCO9	Alignment independent topological	Count of pair of any carbon atom and any oxygen atom separated by 9 bonds
TCO12	Alignment independent topological	Count of pair of any carbon atom and any oxygen atom separated by 12 bonds
TCF6	Alignment independent topological	Count of pair of any carbon atom and fluorine atom separated by 6 bonds
TCF11	Alignment independent topological	Count of pair of any carbon atom and fluorine atom separated by 11 bonds
TCC5	Alignment independent topological	Count of pair of any two carbon atoms separated by 5 bonds
TCC9	Alignment independent topological	Count of pair of any two carbon atoms separated by 9 bonds
TCC13	Alignment independent topological	Count of pair of any two carbon atoms separated by 13 bonds
TNN4	Alignment independent topological	Count of pair of any two nitrogen atoms separated by 4 bonds
TNN11	Alignment independent topological	Count of pair of any two nitrogen atoms separated by 11 bonds
TNN13	Alignment independent topological	Count of pair of any two nitrogen atoms separated by 13 bonds
TNO5	Alignment independent topological	Count of pair of any nitrogen atom and any oxygen atom separated by 5 bonds
TNO6	Alignment independent topological	Count of pair of any nitrogen atom and any oxygen atom separated by 6 bonds
TNO8	Alignment independent topological	Count of pair of any nitrogen atom and any oxygen atom separated by 8 bonds
TNO17	Alignment independent topological	Count of pair of any nitrogen atom and any oxygen atom separated by 17 bonds
TNF13	Alignment independent topological	Count of pair of any nitrogen atom and fluorine atom separated by 13 bonds
TOO7	Alignment independent topological	Count of pair of any two oxygen atoms separated by 7 bonds

several dominating interaction descriptors found in the GQSAR_IT MR model.

- The two highly influential descriptors in GQSAR_IT MR model are D_SaaNHE-index*D_T=N1 and D_SaaNHE-index*D_TCN2 with 12.06 and 18.16 percent contributions respectively. The descriptor D_SaaNHE-index, common in both the interaction descriptors, shows the role of aromatic ring containing nitrogens (for example pyrrole, indole, and indazole) in fragment D to be important in determining the activity. The other two descriptors D_T=N1 and D_TCN2 collectively show that methyl substituted pyrazolopyridine-pyridine derivatives to be preferred over unsubstituted pyrazolopyridine-pyridine derivatives and indazole-pyridine derivatives for Akt1 inhibition.
- The interaction descriptors D_T=N5*D_MMFF64 and D_TCO6*D_MMFF64 having MMFF64 (beta carbon in 5-membered hetero-aromatic ring)) as common descriptor indicate the importance of substituted 5-memebered hetero-aromatic ring in the part-D to be conducive for the activity. The two descriptors D_T=N5 and D_TCO6 together reveal the importance of furan hetero-aromatic ring to be preferred over thiophene and pyrrole ring in the D fragment.
- The descriptor C_SaasCcount* D_4PathCount suggests that increasing substitution on aromatic carbon atoms of fragment C along with increasing length of fragment D would be favourable for activity. In addition, the descriptor C_SaasCcount* D_RotBondCount was found to be inversely proportional to activity. This suggests that the increasing substitution on aromatic carbon atoms of fragment C, should be in such a way that number of rotatable bonds in fragment D should not be increased, *i.e.* the substitution should made on rings or as branching in existing aliphatic chain.

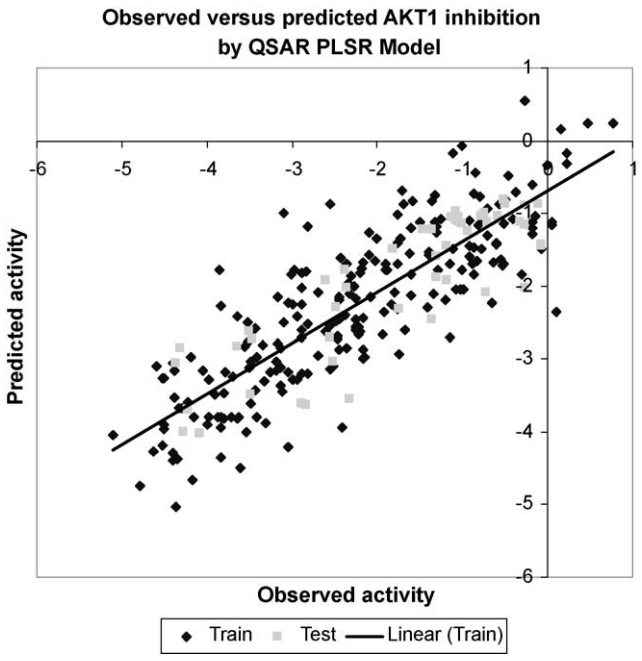

Fig. 3. Plot of observed versus predicted Akt1 inhibition by QSAR PLSR model.

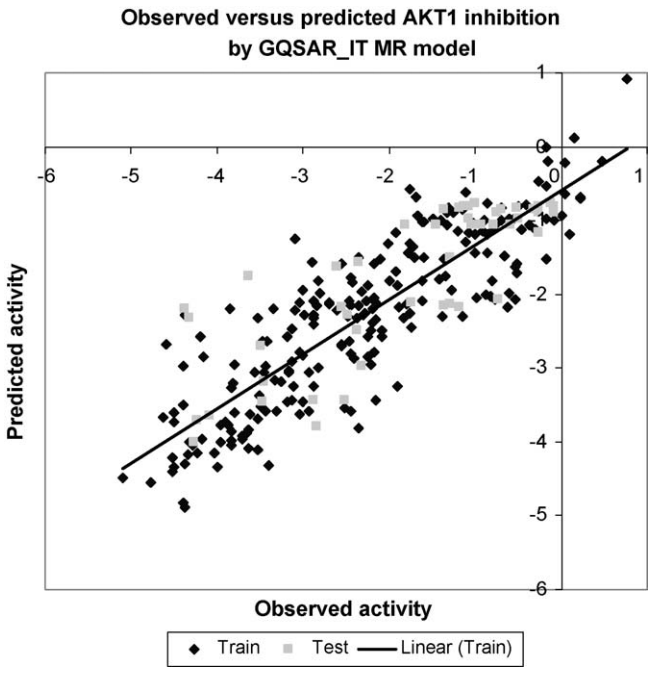

Fig. 4. Plot of observed versus predicted Akt1 inhibition by GQSAR_IT MR model.

- The interaction descriptors D_T==6*D_TCO3 and D_TCC7*D_TCO3 reveal the importance of oxygen atoms with a ring (oxindole-pyridine derivatives) as compared to oxygen atom between two aromatic rings (pyridine base derivatives and trans-bispyridinylethylene derivatives) to be favourable for Akt1 inhibition.
- The presence of descriptor B_PSAExPandS*C_TCN5 shows the importance of interaction of fragments B and C in such a way that the polar surface area in fragment B should be reduced along with the 1,3-disubstituted aromatic ring of fragment C; this would be an optimal requirement for favourable activity.

Table 7 shows that the GQSAR_IT MR model (which contains interaction descriptors *i.e.* non-linear second order) has improved statistical parameters as compared to GQSAR MR model. Hence, we have also developed GQSAR k-NN model by subjecting all the calculated fragment descriptors to the simulated annealing variable selection coupled with k-NN method, to capture nonlinearity in terms of individual fragment descriptors. This study has resulted in a k-NN GQSAR model which was found to be comparable to above reported k-NN QSAR and GQSAR MR models but has lower statistical significance (with respect to q^2 and $pred\_r^2$) as compared to GQSAR_IT model. The descriptors that were found to be important in the k-NN GQSAR model are: A_1PathCount, A_SdOE-index, A_T=F4, A_TCN3, B_PSAExPandS, B_SssNHE-index, B_SsNH2E-index, C_SaasCE-index, C_T=N3, C_T=N6, C_TCN5, D_SaaaCE-index, D_SaaNHE-index, D_T==3, D_T=N1, D_T=N2, D_T=O3 and

Table 6
Selected descriptors with their regression coeffcient and percentage contribution in GQSAR MR model.

Descriptor	Multiple regression coefficient (±Error)	Percent contribution
A_Mol.Wt.	0.0057(±0.0001)	2.73
A_SdOE-index	0.1021(±0.0138)	4.39
A_T=C7	−0.2094(±0.0492)	−3.40
A_MMFF64	0.3794(±0.0707)	5.07
B_4PathCount	−0.0803(±0.0072)	−4.84
B_PSAExPandS	0.0131(±0.0011)	1.89
B_T=C6	0.5186(±0.0668)	9.16
C_SaasCcount	0.2417(±0.0430)	4.92
C_TCO7	−0.1912(±0.0710)	−2.89
D_RotBondCount	−0.2007(±0.0766)	−3.32
D_4PathCount	0.0710(±0.0004)	15.13
D_SdsCHE-index	0.1124(±0.0272)	2.46
D_SaaNHE-index	0.3425(±0.0594)	6.00
D_PSAInPandS	−0.0121(±0.0003)	−3.01
D_T=N5	−0.4552(±0.0793)	−8.92
D_TCC7	−0.1289(±0.0205)	−5.51
D_TCN1	0.1723(±0.0516)	3.58
D_TCN5	0.1729(±0.0834)	3.44
D_TCO3	−0.2501(±0.0473)	−5.07
D_TCO6	0.2884(±0.0627)	4.28
Regression constant	−5.3234	*

* NA – not applicable.

Table 7
Selected descriptors with their regression coeffcient and percentage contribution in GQSAR_IT MR model.

Descriptor	Multiple regression coefficient (±Error)	Percent contribution
A_MMFF64	0.3133(±0.0519)	2.71
B_T=C6^2	0.0441(±0.0021)	3.31
D_T==6^2	−0.0109(±0.0003)	−4.17
D_TCN5^2	0.0366(±0.0040)	2.58
A_Mol.Wt.*B_PSAExPandS	0.0001(±0.0000)	1.83
A_SdOE-index*C_T=N3	0.0126(±0.0002)	3.70
A_T=C7*B_4PathCount	−0.0446(±0.0050)	−2.34
A_MMFF64*D_SdsCHE-index	0.0833(±0.0160)	1.89
B_4PathCount*D_PSAInPandS	−0.0027(±0.0000)	−2.22
B_PSAExPandS*C_TCN5	−0.0095(±0.0001)	−5.90
C_SaasCcount*D_RotBondCount	−0.1928(±0.0207)	−4.31
C_SaasCcount*D_4PathCount	0.0302(±0.0002)	8.61
C_SaasCE-index*C_T=N3	0.0208(±0.0007)	3.84
D_SaaNHE-index*D_T=N1	−0.2617(±0.0157)	−12.06
D_SaaNHE-index*D_TCN2	0.2299(±0.0045)	18.16
D_T==6*D_TCO3	0.0684(±0.0060)	6.22
D_T=N5*D_MMFF64	−0.2231(±0.0291)	−4.75
D_TCC7*D_TCO3	−0.0956(±0.0085)	−7.19
D_TCO6*D_MMFF64	0.2613(±0.0330)	4.23
Regression constant	−3.9983	*

* NA – not applicable.

Table 8
List of important descriptors (found in GQSAR and GQSAR_IT models) along with their category and definition.

Descriptor	Category	Definition
Fragment A		
A_MMFF64	Merck molecular force field (MMFF) atom type	Count of beta carbon in 5-membered hetero-aromatic ring
A_Mol.Wt.	Structural	Molecular weight of fragment A
A_1PathCount	Topological	Count of bonds in fragment A
A_SdOE-index	Electro-topological	Electro-topological state index of oxygen atom connected with a heavy atom by a double bond
A_TCN3	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 3 bonds
A_T=F4	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and fluorine atom separated by 4 bonds
A_T=C7	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any carbon atom separated by 7 bonds
Fragment B		
B_4PathCount	Topological	Count of fragment made of four consecutive (unbranched) edges
B_PSAExPandS	Structural	Polar surface area of fragment B excluding phosphorous and sulphur atoms
B_SsNH2E-index	Electro-topological	Electro-topological state index of nitrogen atom connected with three atoms viz. one heavy atom by single bond and two hydrogen atoms
B_SssNHE-index	Electro-topological	Electro-topological state index of nitrogen atom connected with three atoms viz. two heavy atoms by single bond and a hydrogen atom
B_T=C6	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any carbon atom separated by 6 bonds
Fragment C		
C_SaasCcount	Electro-topological	Count of carbon atom connected with three atoms viz. two heavy atoms by aromatic bonds and a heavy atom by single bond
C_SaasCE-index	Electro-topological	Electro-topological state index of carbon atom connected with three atoms viz. two heavy atoms by aromatic bonds and a heavy atom by single bond
C_T=N3	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 3 bonds
C_T=N6	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 6 bonds
C_TCN5	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 5 bonds
C_TCO7	Alignment independent topological	Count of pair of any carbon atom and any oxygen atom separated by 7 bonds
Fragment D		
D_3PathCount	Topological	Count of fragment made of three consecutive (unbranched) edges
D_4PathCount	Topological	Count of fragment made of four consecutive (unbranched) edges
D_MMFF64	Merck molecular force field (MMFF) atom type	Count of beta carbon in 5-membered hetero-aromatic ring
D_PSAInPandS	Structural	Polar surface area of fragment D including phosphorous and sulphur atoms
D_RotBondCount	Structural	Number of rotatable bond in fragment D
D_SaaNHE-index	Electro-topological	Electro-topological state index of nitrogen atom connected with three atoms viz. two heavy atoms by aromatic bonds and a hydrogen atom
D_SdsCHE-index	Electro-topological	Electro-topological state index of carbon atom connected with three atoms viz. a heavy atom by single bond, a heavy atom by double bond and a hydrogen atom
D_SaaaCE-index	Electro-topological	Electro-topological state index of carbon atom connected with three heavy atoms by three aromatic bonds
D_T==6	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atoms separated by 6 bonds
D_T=N1	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 1 bond
D_T=N2	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 2 bonds
D_T=N5	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any nitrogen atom separated by 5 bonds
D_T=O3	Alignment independent topological	Count of pair of any double bonded (sp2 hybridized) atom and any oxygen atom separated by 3 bonds
D_TCC7	Alignment independent topological	Count of pair of any carbon atoms separated by 7 bonds
D_TCN1	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 1 bond
D_TCN2	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 2 bonds
D_TCN5	Alignment independent topological	Count of pair of any carbon atom and any nitrogen atom separated by 5 bonds
D_TCO3	Alignment independent topological	Count of pair of any carbon atom and any oxygen atom separated by 3 bonds
D_TCO6	Alignment independent topological	Count of pair of any carbon atom and any oxygen atom separated by 6 bonds

D_3PathCount (see descriptors Table 8). It can be seen that most of the descriptors in the model are from fragment D (7 descriptors) followed by fragments A and C (4 descriptors each). Thus, like the above reported GQSAR models, k-NN GQSAR model also indicates the dominance of fragment D in determining Akt1 inhibition. An advantage of the k-NN method is that it can provide ranges (minimum and maximum, derived from the k-nearest neighbours of the most active molecule) for each fragment descriptor. These ranges can be used as a reference when searching for similar fragments in a fragment database during the design of new molecules.

Thus, unlike traditional QSAR models, the developed GQSAR MR and GQSAR_IT MR models provide information about the important substitution site(s) along with their chemical nature and their interactions which could prove useful for designing of new molecules.

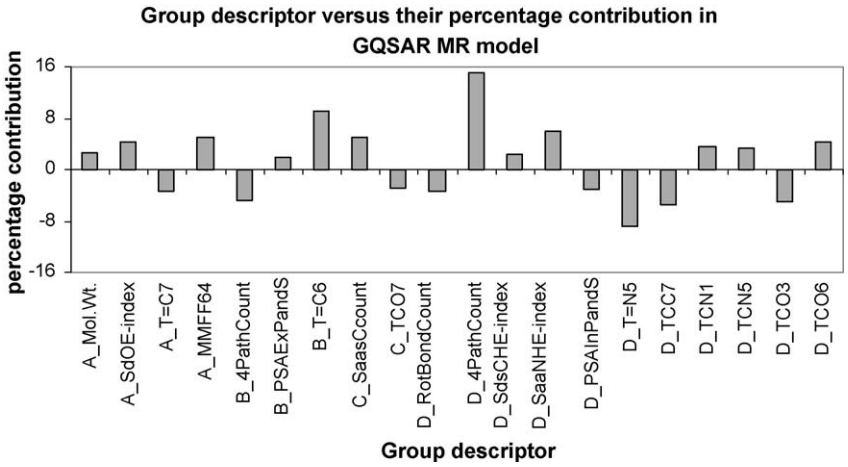

Fig. 5. Plot of percentage contribution of various individual fragment descriptors in GQSAR MR model.

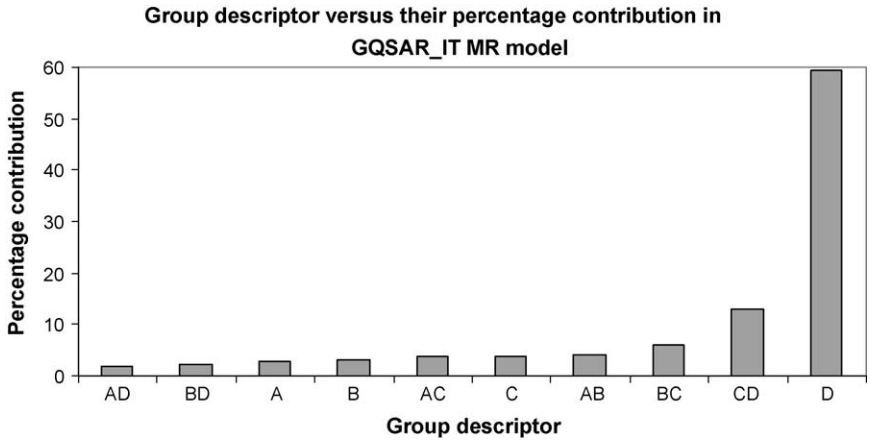

Fig. 6. Plot of percentage contribution of various individual fragment descriptors and interaction descriptors in GQSAR_IT MR model.

4. Summary and conclusions

The present study unveils key structural requirements for Akt1 inhibition utilizing various QSAR methods. A wide variety of structurally diverse Akt1 inhibitors collected from various literature reports were used in this study [9,10,18–26]. The conventional QSAR analyses revealed the major importance of Baumann's alignment independent topological descriptors along with other descriptors such as number of hydrogen bond acceptors, number of hydrogen bond donors, number of rotatable bonds, molecular branching (chi3Cluster), number of aromatic oxygen (SaaOcount), and alkene carbon atom type (SdsCHE-index) in determining Akt1 inhibition activity.

Further, the GQSAR analyses emphasize that the hinge binding fragment (in the present case fragment D) is highly dominant (approximately 60 percent), followed by fragments A and B (approximately 16 percent each), whereas fragment C was found to be the least contributing (approximately 8 percent) of the four fragments in governing activity variation. In addition, the GQSAR analyses shows that chemical variations such as the presence of hetero-aromatic ring, flexibility, polar surface area, fragment length, etc. at fragment D are critical for achieving highly potent kinase inhibitors. These important fragment features can form the building blocks to design new molecules with their corresponding highlighted features being optimized.

Also, in the present study a k-NN classification model was developed using two dimensional electro-topological descriptors: SssOE-index, SaaCHE-index, and SaasCE-index. These featurs can be used for predicting the ability of the newly designed molecules to interact with Akt1 binding domain (kinase or PH).

Combination of the above methods is useful in understanding the structural requirements for design of novel, potent, and selective Akt1 inhibitors. Also the proposed models, due to the good predictive ability, offer a useful alternative for determining Akt1 inhibition of newly designed molecules. To the best of our knowledge, this is the first comprehensive study which provides information on molecular attributes responsible for Akt1 inhibitory activity.

Acknowledgements

Authors would like to thank journal referees for their valuable suggestions to improve the manuscript. Authors thank to VLife team for providing their support during this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.01.007.

References

[1] P. Cohen, Protein kinases – the major drug targets of the twenty-first century? Nat. Rev. Drug Discov. 1 (2002) 309–315.

[2] D.C. Lev, L.S. Kim, V. Melnikova, M. Ruiz, H.N. Ananthaswamy, J.E. Price, Dual blockade of EGFR and ERK1/2 phosphorylation potentiates growth inhibition of breast cancer cells, Br. J. Cancer 91 (2004) 795–802.
[3] C.C. Kumar, R. Diao, Z. Yin, Y. Liu, A.A. Samatar, V. Madison, L. Xiao, Expression, purification, characterization and homology modeling of active Akt/PKB, a key enzyme involved in cell survival signaling, Biochim. Biophys. Acta 1526 (2001) 257–268.
[4] J.H. Hsu, Y. Shi, L.P. Hu, M. Fisher, T.F. Franke, A. Lichtenstein, Role of the AKT kinase in expansion of multiple myeloma clones: effects on cytokine-dependent proliferative and survival responses, Oncogene 21 (2002) 1391–1400.
[5] C. Page, H. Lin, Y. Jin, V.P. Castle, G. Nunez, M. Huang, J. Lin, Overexpression of Akt/AKT can modulate chemotherapy induced apoptosis, Anticancer Res. 20 (2000) 407–416.
[6] J. Okano, I. Gaslightwala, M.J. Birnbaum, A.K. Rustgi, H. Nakagawa, Akt/protein kinase B isoforms are differentially regulated by epidermal growth factor stimulation, J. Biol. Chem. 275 (2000) 30934–33942.
[7] S.A. Beresford, M.A. Davies, G.E. Gallick, N.J. Donato, Differential effects of phosphatidylinositol-3/Akt-kinase inhibition on apoptotic sensitization to cytokines in LNCaP and PC-3 prostate cancer cells, J. Interferon Cytokine Res. 21 (2001) 313–322.
[8] Q. Li, Recent progress in the discovery of Akt inhibitors as anticancer agents, Expert Opin. Ther. Patents 17 (2007) 1077–1130.
[9] C. Lindsley, Z. Zhao, Leister, H. William, R. Robinson, S. Barnett, D. Defeo-Jones, R. Jones, G. Hartman, J. Huff, H. Huber, M. Duggan, Allosteric Akt (PKB) inhibitors: discovery and SAR of isozyme selective inhibitors, Bioorg. Med. Chem. Lett. 15 (2005) 761–764.
[10] C.B. Breitenlechner, W.G. Friebe, E. Brunet, G. Werner, K. Graul, U. Thomas, K.P. Künkele, W. Schäfer, M. Gassel, D. Bossemeyer, R. Huber, R.A. Engh, B. Masjost, Design and crystal structures of protein kinase B-selective inhibitors in complex with protein kinase A and mutants, J. Med. Chem. 48 (2005) 163–170.
[11] C.B. Breitenlechner, T. Wegge, L. Berillon, K. Graul, K. Marzenell, W.G. Friebe, U. Thomas, R. Schumacher, R. Huber, R.A. Engh, B. Masjost, Structure based optimization of novel azepane derivatives as PKB inhibitors, J. Med. Chem. 47 (2004) 1375–1390.
[12] Y. Luo, A.R. Shoemaker, X. Liu, K.W. Woods, S.A. Thomas, R. de Jong, E.K. Han, T. Li, V.S. Stoll, J.A. Powlas, A. Oleksijew, M.J. Mitten, Y. Shi, R. Guan, T.P. McGonigal, V. Klinghofer, E.F. Johnson, J.D. Leverson, J.J. Bouska, M. Mamo, R.A. Smith, E.E. Gramling-Evans, B.A. Zinker, A.K. Mika, P.T. Nguyen, T. Oltersdorf, S.H. Rosenberg, Q. Li, V.L. Giranda, Potent and selective inhibitors of Akt kinases slow the progress of tumors in vivo, Mol. Cancer Ther. 4 (2005) 977–986.
[13] F. Deanda, E.L. Stewart, M.J. Reno, D.H. Drewry, Kinase-targeted library design through the application of the PharmPrint methodology, J. Chem. Inf. Model. 48 (2008) 2395–2403.
[14] M. Muddassar, F.A. Pasha, M.M. Neaz, Y. Saleem, S.J. Cho, Binding mode elucidation and three dimensional quantitative structure activity relationship studies on novel series of protein kinase B/Akt inhibitors, J. Mol. Mod. 15 (2009) 183–192.
[15] X. Dong, C. Jiang, H. Hu, J. Yan, J. Chen, Y. Hu, QSAR study of Akt/protein kinase B (PKB) inhibitors using support vector machine, Eur. J. Med. Chem. 44 (2009) 4090–4097.
[16] S. Ajmani, K. Jadhav, S.A. Kulkarni, Group-Based QSAR (GQSAR): mitigating interpretation challenges in QSAR, QSAR Comb. Sci. 28 (2009) 36–41.
[17] VLifeMDS, Version 3.5, VLife Sciences Technologies Pvt. Ltd., Pune, India, (2008).
[18] Z. Zhao, W.H. Leister, R.G. Robinson, S.F. Barnett, D. Defeo-Jones, R.E. Jones, G.D. Hartman, J.R. Huff, H.E. Huber, M.E. Duggan, C.W. Lindsley, Discovery of 2,3,5-trisubstituted pyridine derivatives as potent Akt1 and Akt2 dual inhibitors, Bioorg. Med. Chem. Lett. 15 (2005) 905–909.
[19] Q. Li, K.W. Woods, S. Thomas, G.D. Zhu, G. Packard, J. Fisher, T. Li, J. Gong, J. Dinges, X. Song, J. Abrams, Y. Luo, E.F. Johnson, Y. Shi, X. Liu, V. Klinghofer, R. de Jong, T. Oltersdorf, V.S. Stoll, C.G. Jakob, S.H. Rosenberg, V.L. Giranda, Synthesis and structure–activity relationship of 3,4′-bispyridinylethylenes: discovery of a potent 3-isoquinolinylpyridine inhibitor of protein kinase B (PKB/Akt) for the treatment of cancer, Bioorg. Med. Chem. Lett. 16 (2006) 2000–2007.
[20] M. Forino, D. Jung, J.B. Easton, P.J. Houghton, M. Pellecchia, Virtual docking approaches to protein kinase B inhibition, J. Med. Chem. 48 (2005) 2278–2281.
[21] G.D. Zhu, J. Gong, A. Claiborne, K.W. Woods, V.B. Gandhi, S. Thomas, Y. Luo, X. Liu, Y. Shi, R. Guan, S.R. Magnone, V. Klinghofer, E.F. Johnson, J. Bouska, A. Shoemaker, A. Oleksijew, V.S. Stoll, R. De Jong, T. Oltersdorf, Q. Li, S.H. Rosenberg, V.L. Giranda, Isoquinoline-pyridine-based protein kinase B/Akt antagonists: SAR and in vivo antitumor activity, Bioorg. Med. Chem. Lett. 16 (2006) 3150–3155.
[22] G.D. Zhu, V.B. Gandhi, J. Gong, Y. Luo, X. Liu, Y. Shi, R. Guan, S.R. Magnone, V. Klinghofer, E.F. Johnson, J. Bouska, A. Shoemaker, A. Oleksijew, K. Jarvis, C. Park, R.D. Jong, T. Oltersdorf, Q. Li, S.H. Rosenberg, V.L. Giranda, Discovery and SAR of oxindole-pyridine-based protein kinase B/Akt inhibitors for treating cancers, Bioorg. Med. Chem. Lett. 16 (2006) 3424–3429.
[23] G.D. Zhu, J. Gong, V.B. Gandhi, K. Woods, Y. Luo, X. Liu, R. Guan, V. Klinghofer, E.F. Johnson, V.S. Stoll, M. Mamo, Q. Li, S.H. Rosenberg, V.L. Giranda, Design and synthesis of pyridine-pyrazolopyridine-based inhibitors of protein kinase B/Akt, Bioorg. Med. Chem. 15 (2007) 2441–2452.
[24] N. Foloppe, L.M. Fisher, G. Francis, R. Howes, P. Kierstan, A. Potter, Identification of a buried pocket for potent and selective inhibition of Chk1: prediction and verification, Bioorg. Med. Chem. 14 (2006) 1792–1804.
[25] Q. Li, T. Li, G.D. Zhu, J. Gong, A. Claibone, C. Dalton, Y. Luo, E.F. Johnson, Y. Shi, X. Liu, V. Klinghofer, J.L. Bauch, K.C. Marsh, J.J. Bouska, S. Arries, R. De Jong, T. Oltersdorf, V.S. Stoll, C.G. Jakob, S.H. Rosenberg, V.L. Giranda, Discovery of trans-3,4′-bispyridinylethylenes as potent and novel inhibitors of protein kinase B (PKB/Akt) for the treatment of cancer: synthesis and biological evaluation, Bioorg. Med. Chem. Lett. 16 (2006) 1679–1685.
[26] G.D. Zhu, V.B. Gandhi, J. Gong, S. Thomas, K.W. Woods, X. Song, T. Li, R.B. Diebold, Y. Luo, X. Liu, R. Guan, V. Klinghofer, E.F. Johnson, J. Bouska, A. Olson, K.C. Marsh, V.S. Stoll, M. Mamo, J. Polakowski, T.J. Campbell, R.L. Martin, G.A. Gintant, T.D. Penning, Q. Li, S.H. Rosenberg, V.L. Giranda, Syntheses of potent, selective, and orally bioavailable indazole-pyridine series of protein kinase B/Akt inhibitors with reduced hypotension, J. Med. Chem. 50 (2007) 2990–3003.
[27] T. Liu, Y. Lin, X. Wen, R.N. Jorissen, M.K. Gilson, BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities, Nucleic Acids Res. 35 (2007) D198–D201. http://www.bindingdb.org/bind/index.jsp.
[28] K. Baumann, An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features, J. Chem. Inf. Comput. Sci. 42 (2002) 26–35.
[29] A. Golbraikh, A. Tropsha, QSAR modeling using chirality descriptors derived from molecular topology, J. Chem. Inf. Comput. Sci. 43 (2003) 144–154.
[30] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[31] S. Wold, PLS for multivariate linear modelling, in: H. van de Waterbeemd (Ed.), QSAR-Chemometric Methods in Molecular Design, vol. 2, Wiley–VCH, Weinheim, Germany, 1995, pp. 195–218.
[32] S. Wold, A. Ruhe, H. Wold, W.J. Dunn, The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses, SIAM J. Sci. Stat. Comp. 5 (1984) 735–743.
[33] M.A. Sharaf, D.L. Illman, B.R. Kowalski, Chemometrics, Chemical Analysis Series, Wiley, New York, 1986.
[34] W. Zheng, A. Tropsha, Novel variable selection quantitative structure-property relationship approach based on the k-nearest neighbor principle, J. Chem. Inf. Comput. Sci. 40 (2000) 185–194.
[35] M. Shen, Y. Xiao, A. Golbraikh, V.K. Gombar, A. Tropsha, An in silico screen for human S9 metabolic turnover using k-nearest neighbor QSPR method, J. Med. Chem. 46 (2003) 3013–3020.
[36] I. Akritopoulou-Zanze, P.J. Hajduk, Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors, Drug Discov. Today 14 (2009) 291–297.