Journal of Applied Pharmaceutical Science Vol. 4 (06), pp. 016-024, June, 2014
Available online at http://www.japsonline.com
DOI: 10.7324/JAPS.2014.40603
ISSN 2231-3354

Novel Insight into the Structural Requirements of P70S6K Inhibition Using Group-based Quantitative Structure Activity Relationship (GQSAR)

Abubakar Danjuma Abdullahi[1], Abdualrahman Mohammed Abdualkader[2], Nadia Hanis Abdul Samat[3], Farahida Mohamed[3], Bala Yauri Muhammad[4], Helaluddin Abulbashar Mohammed[2], Ahmed Aljarbou[4] and Abdulrazaq Kasmuri[1]

[1] Department of Basic Medical Sciences, Kulliyyah of Pharmacy, International Islamic University Malaysia, [2]Department of Pharmaceutical Chemistry, Kulliyyah of Pharmacy International Islamic University Malaysia, [3]Department of Pharmaceutical Technology, Kulliyyah of Pharmacy, International Islamic University Malaysia, [4]College of Pharmacy, Al-Qassim University, Buraidah, Saudi Arabia.

ARTICLE INFO

Article history:
Received on: 07/02/2014
Revised on: 28/03/2014
Accepted on: 19/04/2014
Available online: 28/06/2014

Key words:
Descriptors; electrotopological indices; P70 ribosomal S6 kinase; quantitative structure activity relationship; regression analysis; simulated annealing, sphere exclusion.

ABSTRACT

Dysregulation of P70 ribosomal S6 kinase (P70S6K) has been observed in many cancers; therefore, the design of new molecules targeting p70S6K of paramount importance in cancer therapy. The current study employed a group-based quantitative structure-activity relationship (GQSAR) to develop global QSAR models capable of predicting the bioactivity of P70S6K inhibitors. A wide variety of chemical structures and biological activities (half maximal inhibitory concentration) of P70S6K inhibitors were collected from the binding database website. Compounds were classified into various chemical groups and then fragmented into R1, R2, and R3 fragments based on certain pharmacophoric features required for ligand-target biointeractions. Different two-dimensional fragment-based descriptors were calculated for each fragment. The dataset was then divided into a training set (n=40) and a test set (n=10) using a sphere exclusion algorithm. Multiple linear regressions coupled with simulated annealing or stepwise regression resulted in model A (r^2=0.92) and model B (r^2=0.87), respectively. Leave-one-out validation showed that models A and B have internal predictive abilities of 72% and 61%, respectively. External validation indicated that both models are robust, with squared cross-correlation coefficients of the training set (pred-r^2) of 0.87 and 0.89, respectively. The developed GQSAR models indicate that fragment R3 plays a key role in activity variation (65%) with sound contribution of five-membered rings (5 chain count), aromatic carbons (SaaaCE-index), and aromatic nitrogens (SaaNcount). In contrast, fragments R1 and R2 together contribute 35% of activity variation, suggesting that sulfur atoms (Sulfur count) and hydrophobic three-membered rings (chi3 chain) at R1 are preferable for inhibitory activity.

INTRODUCTION

The P70 ribosomal S6 kinase (P70S6K) is a serine/threonine kinase that belongs to the protein kinase A/protein kinase G/protein kinase C family. In humans, there are two isoforms of P70, P70 ribosomal S6 kinase 1 (S6K1) and p70 ribosomal S6 kinase 2 (S6K2). Activation of P70S6K1 occurs through insulin and growth factor stimulation of PI3K (phosphoinositide 3-kinase) and mTOR (mammalian target of rapamycin) signaling pathways by insulin and growth factors (Vivanco and Sawyers, 2002). P70S6K, in turn, increases protein synthesis, growth, proliferation, and longevity through translation of mRNA that possesses five polypyrimidine tracts. P70S6K has been implicated to promote malignant transformation of cancers. Fluorescence *in situ* hybridization of breast cancer tissues revealed amplification of P70S6K together with other genes on 17q22-q24. This is accompanied by its increased expression, which contributes to more aggressive clinical outcomes in patients (Bärlund *et al.*, 2000b). Other studies showed P70S6K is amplified and over expressed in MCF-7 cells compared with normal mammary epithelium (Bärlund *et al.*, 2000a) and is an essential part of neoplasia in many cancerous cell lines, such as A549 (Bussenius *et al.*, 2012) and Sf21(Davies *et al.*, 2000). Considering the role of P70S6K1 in cancer, development of P70S6K1 inhibitors would be of paramount importance in cancer therapy. Generally, kinases have similar catalytic domains, and their activation requires phosphorylation of an important residue in the activation loop (T-loop). In P70S6K1, phosphorylation of residue T229 in the T-loop

* *Corresponding Author*
A. D. Abdullahi , Email: bagaruwa@gmail.com
Tel/ Fax: +60199175042/+609-5716775

depends on the phosphorylation cascades of serine/threonine residues in the C-terminal regulatory domain. Other kinases, such as extracellular signal-regulated kinases 1/2, c-Jun N-terminal kinases1/2 and cyclin-dependent kinase 1, have been involved in these phosphorylation series (Mukhopadhyay *et al.,* 1992). It was reported that residues 236-263 in P70S6K1 comprised the activation loop, corresponding to the DFG and (A)PE residues of other kinases. In its active state, P70S6K1 binds the β-phosphate group of the ATP molecule via the Asp residue located at the beginning of the activation loop (Sunami *et al.,* 2010). Many chemically diverse P70S6K inhibitors, including pyrazolopyrimidines (Bussenius *et al.,* 2012), thiophenes, and thiophene-ureas (Ye *et al.,* 2011), have been reported.

Fragment-based lead design has shown promise in current drug discovery and lead optimization efforts. The main concept of this method is driving new molecules by combining fragments determined based on ligand-target interaction information (Schulz and Hubbard, 2009). Recently, many successful applications of group-based quantitative structure activity relationship (GQSAR) for lead optimization have been reported (Ajmani *et al.,* 2010). In GQSAR, molecules are fragmented, based on specific molecular sites, bonds, rings, and interaction keys with the target, and fragment-based descriptors are calculated. In contrast with three-dimensional (3D) QSAR techniques that provide information on the whole molecule, GQSAR-developed models give hints on the impact of each fragment on activity variation and ligand-target bio-interactions. Thus, the interpretation of GQSAR models into new molecules is a more practical and achievable task compared to 3D QSARs. In addition, 3D QSAR is only applicable on congeneric conformers, which must be properly aligned to match their pharmacophoric sites, whereas conformational analysis and molecular alignment are not required to perform GQSAR (Ajmani *et al.,* 2009). A comprehensive search in the literature revealed that there is a paucity of GQSAR studies on p70S6 kinase inhibitors. In the current study, we report two GQSAR models for the inhibition of p70S6K. The models were generated and validated using different statistical techniques, and the descriptors used in the models are discussed and interpreted. Our data highlight important requirements for designing innovative and powerful p70S6K inhibitors.

METHODOLOGY

Data mining and preparation

The biological activities and the two dimensional structures of 68 p70S6K inhibitors were collected from different literature sources (Bamford *et al.,* 2005, Bandarage *et al.,* 2009, Charrier *et al.,* 2011, Davies *et al.,* 2000, Lin *et al.,* 2010, Nittoli *et al.,* 2010, Okuzumi *et al.,* 2009, Tao *et al.,* 2007, Wang *et al.,* 2010, Ye *et al.,* 2011). The SDF files of the inhibitors were obtained from the BindingDB website, a public online database of measured binding affinities, with special focus on interactions of druggable target proteins and drug-like small molecules. Although the BindingDB is derived mainly from enzyme inhibition and kinetics assays, isothermal titration colorimetry, nuclear magnetic resonance, and radioligand and competition assays, our study used only enzyme inhibition assays in the form of *in vitro* biological activities (Chen *et al.,* 2002)(Chen *et al.,* 2002, Liu *et al.,* 2007) (Wikipedia). The *in vitro* biological activities are expressed as the concentration of the compound required to inhibit 50% of the enzyme activity (half maximal inhibitory concentration [IC_{50}]), which were converted into the negative logarithm (pIC= -log IC_{50}) for the QSAR analysis. The two-dimensional (2D) structures were converted into the corresponding 3D conformers and the energy was batch-minimized by Merck Molecular Force Field using convergence criterion (root mean square [RMS] gradient) of 0.01 kcal/mol and a maximum number of cycles of 100 (Halgren, 1996). The dataset was filtered based on drug-like physicochemical properties described by Lipinski et al. (Lipinski *et al.,* 1997), which uses molecular weights less than 500 Da as the criterion of filtration.

GQSAR modeling

Software

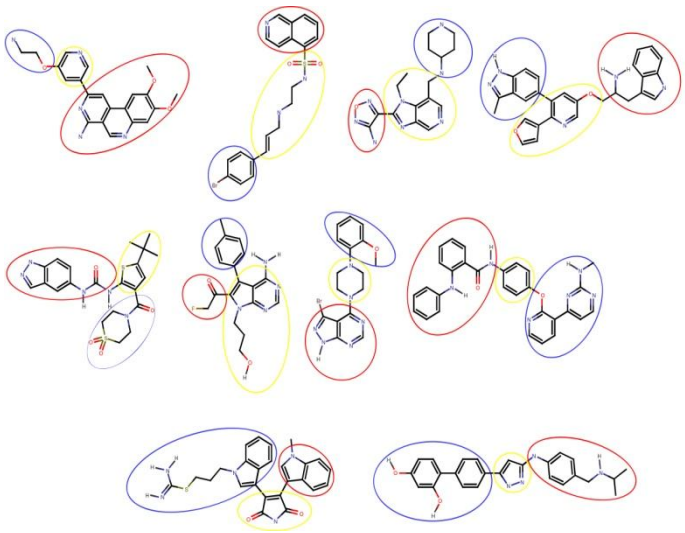
GQSAR modeling was performed using the Molecular Design Suite (VLifeMDS software package, version 4.1, from Vlife Sciences Technologies Pvt. Ltd., India) on a Windows 7 operating system.

Fragmentation pattern

The batch-minimized-generated 3D conformers were grouped into different chemical scaffolds, including:

- Pyridine
- Imidazo[4,5-c]pyridine
- 2-(furan-e-yl)pyridine
- Pyrrolo[2,3-d]pyrimidine
- Peperazine
- Benzene
- Pyrrole-2,5-dione
- Pyrazole
- Sulfonamide

Thereafter, they were divided into three molecular fragments (R1, R2, and R3). The fragmentation described by Akritopoulou-Zanze et al. was executed based on pharmacophoric features such as molecular weight, fragment size, hydrophobic regions, and hydrogen bond donors/acceptors, including heteroatoms (O, N, S) and heterocycles (indole, indazole, imidazoles), all of which are required for ligand-target interactions (Akritopoulou-Zanze and Hajduk, 2009). Examples of the fragmentation pattern of the dataset are presented in **Scheme 1**.Fragment R2 consisted of the scaffolds, whereas fragments R1 and R3 comprised the chemical groups substituted with respect to the main scaffold. R1 contains unsaturated rings or long aliphatic chains and R3 consists mainly of hetero-aromatic rings. For molecules that did not contain such structures, a methyl group or hydrogen atom was considered the corresponding fragment.



Scheme. 1: Fragmentation patterns of P70S6K inhibitors. Compounds were fragmented into three fragments considering the scaffolds as a fragment R2 (yellow circle), and fragment R1 (blue circle) and fragment R3 (red circle) are the substituents with respect to the scaffolds. The scheme was sketched using MarvinSketch 5.9.4.

Calculation of fragment-based descriptors

For the generated fragments, a pool of 464 two-dimensional descriptors was calculated using the VLifeMDS software package. These descriptors include retention index (chi), atomic valence connectivity index (chiv), path count, chi chain, chiv chain, path cluster, kappa, element count, estate numbers, and polar surface area. All descriptors with constant values among the dataset were deleted, resulting in 316 different descriptors (independent variables) which were used in the QSAR analysis.

Selection of the training and test sets

In order to compare the biological activities of the set of compounds which have a wide range of chemical structures (*i.e.*, different fragment-based descriptors), the dataset was divided into representative training and test sets using a dissimilarity-based compound selection method called sphere-exclusion algorithms (Snarey *et al.*, 1997).

In this algorithm, each compound will be represented by one point and the total volume (V) occupied by this point will be defined in the multidimensional descriptor space (*K*) as described by Golbraikh (2000). Exclusion starts by constructing a sphere whose center is the nearest representative point to the center of the dataset having a radius $R = c(V|N)^{1/K}$, where c and N denote the dissimilarity value and the number of compounds in the dataset, respectively. Test set will comprise all compounds (representative points) included within this sphere apart from the center. The latter compounds will be excluded from the dataset and the process will be repeated with a new sphere until all points are exhausted (Golbraikh and Tropsha, 2003, Snarey *et al.*, 1997). To determine the best representing test set, different dissimilarity values were used and different statistical parameters for both sets were calculated (average, maximum, minimum, and standard deviation).

Optimized variable selection

Due to the fact that it is tedious and impractical to investigate all possible combinations of the descriptor pool, simulated annealing and stepwise regression, which simplify the process and reduce the time required to execute algorithms, were implemented (Kirkpatrick *et al.*, 1983, Scior *et al.*, 2009).

Simulated annealing

Simulated annealing (SA) is a global and iterative combinatorial optimization method that does not concur with the first encountered variable configuration. The main concept of SA depends on the physical process of annealing, during which the system is melted at a high temperature and cooled slowly until it reaches the steady state. In SA, the configuration of system points (descriptors) and the cost function (configuration energy) are the parameters to be optimized.

The configuration of system points is defined by Boltzmann probability factor of distribution that is in turn correlated to the energy of the configuration (E) and the applied temperature (T). Lowering T will lead the system towards lower E states. At a given temperature, a population of problem configurations (subset of descriptors) will be generated and the process will be iterated searching for better solution that can be effectively defined by Metropolis algorithms. In the Metropolis procedure, one variable (descriptor) will be randomly displaced, and the difference in the energy states ($\Delta E = E_{new} - E_{old}$) between the two configurations is calculated. If $\Delta E \leq 0$, the new configuration will be accepted and used as a starting point for the next iteration. If $\Delta E > 0$, a new displacement will be executed using different descriptor (Kirkpatrick *et al.*, 1983).

In the context of QSAR analysis, the squared correlation coefficient of the regression (r^2) will serve as the cost function, whereas the descriptors involved in the final regression equation will represent the system configuration (*i.e.*, the process aims to give the descriptor combination with the best r^2).

Stepwise regression

In stepwise regression, the response (pIC values) regress with each term (variables ordescriptors) and the model with the highest r^2 will be selected. Thereafter, a new variable will be added to the most significant model and the bi-term model giving the highest r^2 will be used in the following step. The process will be repeated until none of the terms left out of the model would give statistically significant improvement if added to the model (Armstrong, 2006).

Model building by multiple linear regression

To quantitatively describe the relationship between the dependent variable (*y;* the activity) and the selected independent variables (x_1, x_2,....x_n,;the calculated values of the descriptors), linear regression techniques were employed as follows (1):

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots . + \alpha_n x_n \qquad (1)$$

Where $\alpha_1, \alpha_2 \ldots . + \alpha_n$ are the regression coefficients and α_0 is the regression intercept that represents the predicted value when all variables have zero value. To assess the regression relationship, Fisher's test (*F*-test), the degree of freedom, and the coefficient of determination (r^2) were calculated from the following equations (2, 3) (Armstrong, 2006):

$$degree\ of\ freedom = n - (k + 1) \quad (2)$$

Where *n* and *k* are the number of compounds in training set and the number of variables in the final equation, respectively.

$$r = \frac{\sum xy - \sum x \sum y/n}{\sqrt{[\sum x^2 - \sum x^2/n]}\sqrt{[[\sum y^2 - \sum y^2/n]]}} \quad (3)$$

Where *r* is the correlation coefficient, *n* is the number of compounds, *x* and *y* are the actual and predicted activities.

Model validation

Internal validation of training set

To evaluate the robustness of the generated GQSAR models, internal validation was performed on the training set using the leave-one-out (LOO) method (Tropsha *et al.,* 2003, Zheng and Tropsha, 2000). The compounds from the training set were removed individually, and the activity of each was predicted using the model fitted to the remaining molecules. The process is repeated until all compounds in the training set are exhausted, and the cross-validated coefficient of determination (q^2) was found from the equation (4):

$$q^2 = 1 - \frac{\sum_{i=1}^{training}(y_i - \hat{y}_i)}{\sum_{i=1}^{training}(y_i - \bar{y})} \quad (4)$$

Where y_i, $\hat{y}_i$, and $\bar{y}$ denote the actual, predicted, and average activity of training set molecules.

External validation of test set

The predictive power of the developed models was further validated using the squared correlation coefficient (pred-r^2) of the test set (Tropsha *et al.,* 2003). The model was generated from training set data, and the pIC values of test set compounds were predicted from the model and pred-r^2 was determined from the following equation:

$$pred - 1^2 = 1 - \frac{\sum_{i=1}^{test}(y_i - \hat{y}_i)}{\sum_{i=1}^{test}(y_i - \bar{y})} \quad (5)$$

Where y_i and $\hat{y}_i$ are the actual and predicted activities of test set compounds and $\bar{y}$ represents the average activity of training set molecules.

Randomization and Z-scores test

Randomization tests have been reported as essential validation techniques to evaluate the robustness of the generated models (Tropsha *et al.,* 2003, Zheng and Tropsha, 2000). Using this method, many models were generated by reorganizing the pIC values (the dependent variable) of the entire dataset and the training set compounds randomly. A robust and reliable GQSAR model should have significantly higher r^2 and q^2values than any of the randomly generated models. To assess this assumption, the average coefficient of determination values of the random sets (r_{av}^2 , q_{av}^2)and their standard errors (r_σ^2 , q_σ^2) were calculated. Z-score values were calculated from the equations (5, 6). Thereafter, standard Z-score tables were used to determine the corresponding probability of significance value (α):

$$Z - scores - q^2 = \frac{q^2 - q_{av}^2}{q_\sigma^2} \quad (5)$$

$$Z - scores - r^2 = \frac{r^2 - r_{av}^2}{r_\sigma^2} \quad (6)$$

Model evaluation criteria

According to published reports (Ajmani *et al.,* 2010, Tropsha *et al.,* 2003), the acceptable and significant GQSAR model is determined when: $r^2 > 0.6$, $q^2 > 0.5$, and $0.85 \leq k \leq 1.15$. Other statistical parameters used to evaluate the generated models include: N (number of compounds), F (Fisher's test for statistical significance), r^2 (coefficient of determination), q^2 (cross-validated r^2 by leave-one-out method), pred-r^2 (squared correlation coefficient of external validation of the test set), RSM (root mean squared error), k (slope of regression line), σ (standard error), Z-scores (Z-score value of randomization test), best r and (highest r^2 and q^2 values of the randomization test), α-rand- (the tabular value of statistical significance of Z-scores randomization test), and K (number of descriptors.).

RESULTS AND DISCUSSION

GQSAR modeling

Training and test set selection

Fifty compounds satisfied the filtration criteria (MW <500 Da) and were used for fragment-based descriptor calculation. After removing invariable descriptors, a pool of 317 two-dimensional molecular descriptors remained and their contribution to activity variation was evaluated. Sphere exclusion algorithms with a dissimilarity value of +0.5 resulted in a reasonable rational division of the data into a training set (n=40) and test set (n=10). The calculated unicolumn statistics of both sets (Table 1) show that activity was evenly distributed within both tests and the selected sets fulfilled the main characteristics of valid data selection (Scior *et al.,* 2009). Two GQSAR models were generated and validated based on the selected sets as discussed below.

Table. 1: Statistical parameters of activity distribution within the selected training and test sets.

	N	Average	Max	Min	SD	Sum
Training set	40	2.7920	5.1550	0.4210	1.4327	111.6810
Test set	10	3.4397	4.1140	0.7080	1.0709	30.9570

Sphere exclusion algorithm was applied to divide the data into training and test sets using a dissimilarity value of +0.5.

Statistical evaluation and validation of the developed GQSAR models

Model A

The model was generated by simulated annealing algorithms followed by multiple linear regression using forty compounds as the training set and nine compounds as the test set (one compound from the original test set, n=10, was excluded to

improve the model). The statistical parameters of model A are shown in Table 2. The regression equation of the developed model A explains ~92.71% (r^2=0.9271) of the total variance in the training set and has an internal and external predictive ability of approximately 87% (q^2=0.8782) and approximately 90% (pred_r^2=0.9061), respectively. From the plot (Figure 1A), it can be seen that the model was able to predict activity similar to the actual activity with relatively small-calculated residuals and a low RMS value of 0.3915.

Table. 2: Statistical parameters for the developed GQSAR models of P70S6K inhibitors.

	Model A (SA/MLR)	Model B (STP/MLR)
N (training/test)	40/9	40/9
Degree of freedom	31	31
F-test	49.2570	28.0084
r^2	0.9271	0.8785
q^2	0.7282	0.6109
pred-r^2	0.8707	0.8900
RSM	0.3915	0.4771
k	0.9432	0.8888
r^2- σ	0.4340	0.5602
q^2-σ	0.8379	1.0024
pred-r^2- σ	0.4574	0.4219
Z-score-q^2	3.6620	2.7425
Z-score-r^2	6.20085	10.06522
Best-rand-r^2	0.48751	0.26731
Best-rand-q^2	0.1185	0.0670
α-rand-r^2	<0.00001	<0.00001
α-rand-q^2	<0.001	<0.01
K(number of components)	8	8

Model A was generated using a multiple linear regression method coupled with a simulated annealing algorithm.
Model B was generated using a multiple linear regression method coupled with a stepwise regression algorithm by forward backward elimination.
N: number of compounds, F: Fisher's test for statistical significance, r^2: coefficient of determination, q^2: cross-validated r^2 by leave-one-out, pred- r^2: squared correlation coefficient of external validation of the test set, RSM: root mean squared error, *k:* slope of regression line, σ: standard error, Z-scores: Z-score value of randomization test, Best-rand-: highest r^2 and q^2 values of the randomization test, α-rand- r^2: α-rand-cv: statistical significance of Z-scores randomization test, K: number of descriptors.

Model B

Stepwise regression using forward-backward elimination coupled with multiple linear regression resulted in model B with a significant predictive ability of approximately 87% (r^2=0.8785). Cross-validation using the leave-one-out method showed a significant internal stability of approximately 61% (q^2=0.6109). External validation of the test set displayed pred_r^2=0.8534. All statistical parameters of the developed model are shown in Table 2. Plotting the actual pIC values against the predicted ones from model B displays how well the model predicts the activity because the majority of the points are close to the regression line (Figure 1B).

Compared with the calculated parameters of the previous model, it can be presumed that variable selection by simulated annealing (model A) was capable of developing a more robust and predictive GQSAR model than stepwise regression (model B) in terms of r^2, q^2, and pred-r^2. However, both models have statistically significant robustness and predictive power, and therefore can be used to explain the structural requirements for P70S6K inhibition and to construct a library of potential P70S6K inhibitors.

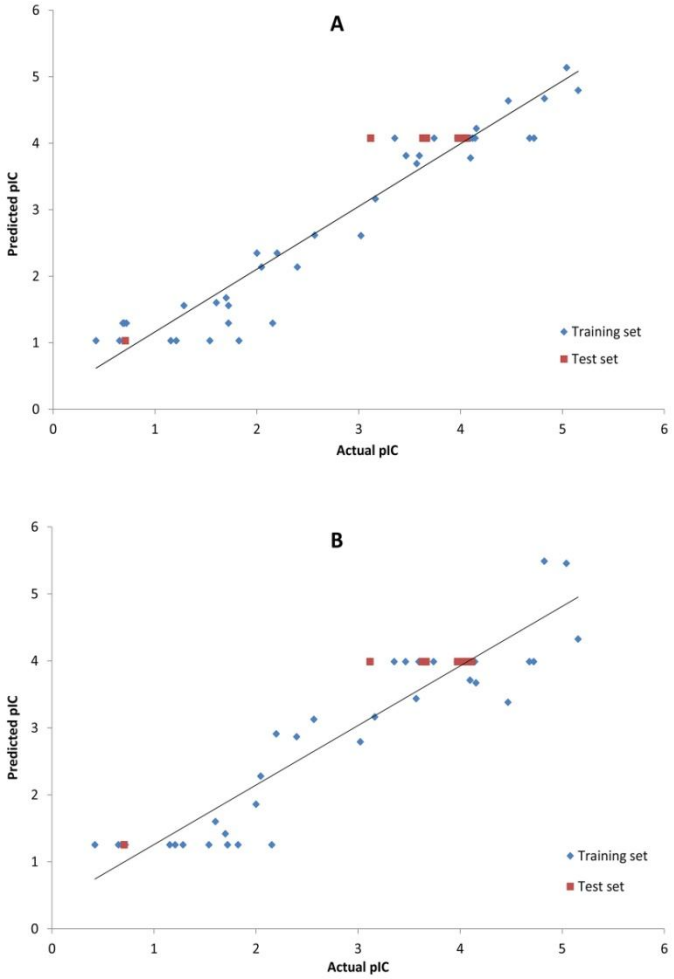


Fig. 1A,B: Regression plots of the generated GQSAR models for p70S6K inhibitors. Activity is expressed as the negative logarithm of IC_{50}. (A) Fitness plot of model A that was generated by simulated annealing algorithms coupled with multiple linear regression, where the squared correlation coefficient r^2= 0.9487 and the slope of regression line *k*= 0.9468. (B) Fitness plot of model B that was generated by stepwise regression coupled with multiple linear regression, where the squared correlation coefficient r^2= 0.9336 and the slope of regression line *k*=0.9361.

Definition of important 2D molecular descriptors found in the GQSAR models

A pool of 465 two-dimensional fragment-based molecular descriptors was calculated. The descriptors that are constant for all molecules do not contribute to GQSAR models and therefore were removed. Finally, a total 316 variable descriptors were used to build the models. As shown in Table 3, each developed GQSAR model has eight descriptors in the regression equation. These descriptors include atom-type count, atom-type E-state indices, electrotopological state indices, chromatographic descriptors, partition coefficient, and topological descriptors (molecular sub-graph descriptors).

Table. 3: Definition of important 2D descriptors found in GQSAR models for P70S6K inhibitors.

Descriptor	Sub-class	Definition
E-state indices		
Nitrogen count	Atom type descriptors	The total number of nitrogen atoms
Sulfur count	Atom type descriptors	The total number of sulfur atoms
E-state indices		
SaaaCE-index	E-state contributions	The electrotopological state indices for number of carbon atoms connected with three aromatic bonds
SddssS (sulfate) E-index	E-state contributions	The electrotopological state indices for number of sulfate groups connected with two single bonds and two double bonds
SsssNE-index	E-state contributions	The electrotopological state indices for the number of nitrogen atoms connected with three single bonds
SaaN count	E-state numbers	The total number of nitrogens connected with two aromatic bonds
SdssC count	E-state numbers	The total number of carbon atoms connected with one double and two single bonds
SssNH count	E-state numbers	The total number of –NH groups connected with two single bonds
SsssN count	E-state numbers	The total number of nitrogens connected with three single bonds.
Lipophilicity-related descriptors		
chi3 chain	Chromatographic descriptors	The retention index for three-membered ring
slogp		log of the octanol/water partition coefficient
Molecular graph		
5 chain count	Molecular sub-graph (chain count)	The total number of five-membered rings

Element count descriptors

Element count or atomic descriptors represent the total number of atoms *A* from the same type. They depend on atomic properties such as atomic number (Z), atomic mass, atomic charges, and the van der Waals radius (r_w). They are zero dimensional descriptors (0D-molecular descriptor) and describe the general molecular structure of the molecule/fragment. Atom-type molecular descriptors give global information on the contribution of each atom type in activity and physicochemical properties (Consonni and Todeschini, 2010).

E-state indices

Table 3 shows that GQSAR models are involved two types of E-state indices descriptors in activity variation among the data set, including E-state contribution (electrotopological state indices) (Kier and Hall, 1990, Kier and Hall, 1997) and E-state number (atom-type E-state indices) (Hall and Kier, 1995, Hall *et al.*, 1995).

E-state contribution indices (S_i) of the atom i^{th}in the molecule/fragmentprovide information on the electronic (intrinsic) and topological states of the atom and can be calculated using equation (7) (Kier and Hall, 1997):

$$S_i = I_i + \Delta I_i \quad (7)$$

Where I_i *is* the intrinsic state of the atom $i^{th}$and $\Delta I_i$is the perturbation of the intrinsic state of the atom i^{th} due to the interaction with another atom in the molecule.

The intrinsic state (I_i) of the atom $i^{th}$is related to the number of valence electrons (non-σ electrons) and σ-bonded electrons and thus provides information on the availability of valence electrons (π and lone pair electrons) for ligand-target interactions. The intrinsic state of the i^{th}atom in an H-depleted molecular graph in the molecule/fragment can be identified using equation (8) (Kier and Hall, 1997):

$$I_i = \frac{\delta^v + 1}{\delta} \quad (8)$$

Where δ^v and δ values are the number of valence electrons and sigma electrons on the atom i^{th}, respectively.

The topological state gives information on the effects of all atoms in the molecule/fragment on the intrinsic state of each atom, and is mainly related to the electronegativity and topological distance descriptors, as shown in equation (9) (Kier and Hall, 1997):

$$\Delta I_i = \sum_{j=1}^{N} \frac{I_i - I_j}{r_{ij}^2} \quad (9)$$

Where $\Delta I_i$is the perturbation of the intrinsic state of the atom i^{th} due to the interaction with the atom j^{th} in the molecule/fragment. $I_i$and $I_j$are the intrinsic state values of the atoms *i* and *j*, respectively. The r_{ij} value is the count of atoms in the shortest molecular path between the atoms *i* and *j*.

E-state number descriptors are the total number of a specific type of atoms having the same atomic number (Z), valence state (δ^v, δ), and bonding type. The value of the E-state number can be defined as the summation of the electrotopological state values of all atoms of the same type in the molecule/fragment (Hall and Kier, 1995, Hall *et al.*, 1995, Todeschini and Consonni, 2000).

Lipophilicity-related descriptors

Partition coefficient (LogP)

The logarithm of octanol-water partition coefficient ($LogP_{ow}$) has been known as the most important quantitative representation of the lipophilicity/hydrophobicity properties of a molecule and can be defined using equation (10) (Patrick, 2001):

$$LogP_{o/w} = log \frac{drug\ concentration\ in\ octanol\ (organic\ phase)}{drug\ concentration\ in\ water\ (aqueous\ phase)} \quad 10$$

Chromatographic descriptors

Chromatographic descriptors are closely related to the lipophilicity descriptor and octanol/water partition coefficient (*LogP*). They are obtained by various chromatographic techniques and can be described using different descriptors. The capacity factor (*k*) is the quantitative representation of the retention index of a substance in HPLC column and can be defined using equation (11). The retention indices obtained from TLC and paper

chromatography are described by Bate-Smith-Westall retention index (R_M) as shown in equation (12) (Todeschini and Consonni, 2000).

$$Logk = \frac{t_R - t_M}{t_M} \quad (11)$$

Where $t_R$and $t_M$are the retention times of the solute and the un-retained solute, respectively.

$$R_M = \log\left(\frac{1}{R_f} - 1\right) \quad (12)$$

Where R_f is the ratio of a distance passed by the solute to that of the solvent front.

Molecular graph descriptors

An H-depleted molecular graph (*G*)is a topological descriptor of a molecule/fragment where all chemical atoms and covalent bonds are presented as vertices (*V*) and edges (*E*), excluding all hydrogen atoms. The *path* is a walk within the molecular graphfrom one vertex to another without any repeated vertices and the number of the *E* encountered within the *path* is called the *path length* (Consonni and Todeschini, 2010). A group of atoms and a bonding system within *G* constitutes a molecular sub-graph which can be classified into: *path* (vertex degree equals to 2 or 1; each atom is adjacent to 2 or 1 atom), *cluster* (vertex degree is greater than 2 or equal to 1), *path-cluster* (other vertex degree values), and *chain* (cyclic sub-graphs) (Todeschini and Consonni, 2000).

Percentage contributions of the fragment-based descriptors by GQSAR models and their interpretation

Model A

The generated model indicates that structural modifications on fragment R3 are critical for producing potent P70S6K inhibitors because fragment R3-based descriptors contribute approximately 63.81% of activity variation and fragment R1 (~26.47%) and fragment R2 (~9.72%) have lower contributions. The percentage contribution of fragment-based descriptors is presented in Figure 2A. As a negative contributor at fragment R1, the electrotopological state indices of the sulfate group connected to two double bonds and two single bonds [R1-SddssS(sulfate)E-index] are detrimental to activity and contribute approximately -16.73% of activity variation. Consequently, to produce potent P70S6K inhibitors, the thiomorpholine-1,1-dione ring at R1 in LIG3 and LIG66 should be replaced by thiomorphine, which has two sigma-bonded electrons and a lower $S_i$value. It was determined that nitrogen atoms in fragment R1 are unfavorable for inhibition because the descriptor (R1-Nitrogen count) was shown to negatively contribute to the inhibitory activity by approximately -4.20%. For example, LIG3 contains three nitrogen atoms in fragment R1 due to the presence of an indole ring and propylsulfanyl(methanimidamide), which are suggested to be replaced by non-containing structures such as an indene ring and 1-(methylsulfanyl)propane, respectively. On the other hand, the descriptor (R1-chi3 chain) is directly proportional to activity and contributes approximately 5.55% of activity variation within the dataset, indicating that increasing the retention index value of three-membered rings at fragment R1 could be conducive for activity as in LIG27. Consequently, a cyclopropane ring is more preferable for inhibition than its nitrogen-containing counterpart, aziridine. In fragment R2, the E-state contribution of nitrogen atoms of sp3 hybridization (connected with three single bonds) is inversely proportional to activity (~-9.72%), indicating that better inhibition requires a lower value of the R2-SsssNE-index descriptor. For instance, the value of R2-SsssNE-index descriptor in LIG53 could be reduced by using a pyrrolidine ring instead of 2,5-dihydro-1H-pyrrol-2,5-dione at fragment R2. The total number of five-membered rings (R3-5 chain count) and the E-state contribution of carbon atoms connected to three aromatic bonds (R3-SaaaCE-index) at fragment R3 were directly proportional to activity and contributed approximately 15.24% and 11.32% of activity variation, respectively. Thus, replacing indole by benzofuran at fragment R3 in LIG3 is thought to increase inhibition. In contrast, the total number of sp2-hybridized carbon atoms (R3-SdssCcount) and the E-state indices of sp3-hybridized nitrogen atoms (R3-SsssNE-index) were shown to be deleterious to inhibitory activity with a percentage contribution of approximately -33.37% and -3.88%, respectively. This indicates that the urea group in LIG57-65 is unfavorable and might be one of the reasons for its low activity (pIC ranged from 1.284 to 0.421).

Model B

The model confined activity variation to eight molecular descriptors with a sound effect of fragment R2-based descriptors that contribute about approximately 62.141% of the inhibitory activity variation. Fragment R1 and R2 are contribute almost equally, approximately 18.13% and 19.17%, respectively. The percentage contribution of fragment-based descriptors found in model B is presented in Figure 2B. Consistent with the previous model, it was found that R1-chi3 chain and R1-sulfur count are directly proportional to activity and contribute approximately 5.40% and 12.75%, respectively. Therefore, to design potent P70S6K inhibitors, it is recommended to increase the total number of sulfur atoms and the retention index value of three-membered rings by using cyclopropane-containing thioether and thioester derivatives at fragment R1. In fragment R2, two negatively contributing descriptors were found. First, the E-state indices of carbon atoms connected to three aromatic bonds contribute approximately -14.72% of activity variation. For example, decreasing the value of the electrotopological state indices of the bridgehead atoms of the indazole ring at R2 in LIG34 would result in better inhibition. Second, the total number of nitrogen atoms connected to three single bonds was found to inhibit activity (~-5.00%), suggesting that replacing the 2,5-dihydro-1H-pyrrol-2,5-dione in LIG53 by cyclopent-4-ene-1,3-dione is favorable for activity. Model B suggests that the majority of the R3-based descriptors are related to nitrogen atoms. It was found that secondary amines (R3-SssNHcount) and tertiary amines (R3-SsssNcount) are adversely correlated to activity, explaining

approximately -31.10% and -12.53% of pIC value variation within the dataset. In contrast, aromatic nitrogen atoms (R3-SaaNcount) are conducive to inhibitory activity (~5.67%). Consequently, indazole (LIG27), quinazoline, quinoxaline, and benzotriazine are highly recommended for better inhibition. The SlogP (~-12.48%) of fragment R3 suggests that hydrophobic interactions are not favorable for inhibition and more effective inhibitors should contain a less hydrophobic Fragment R3.

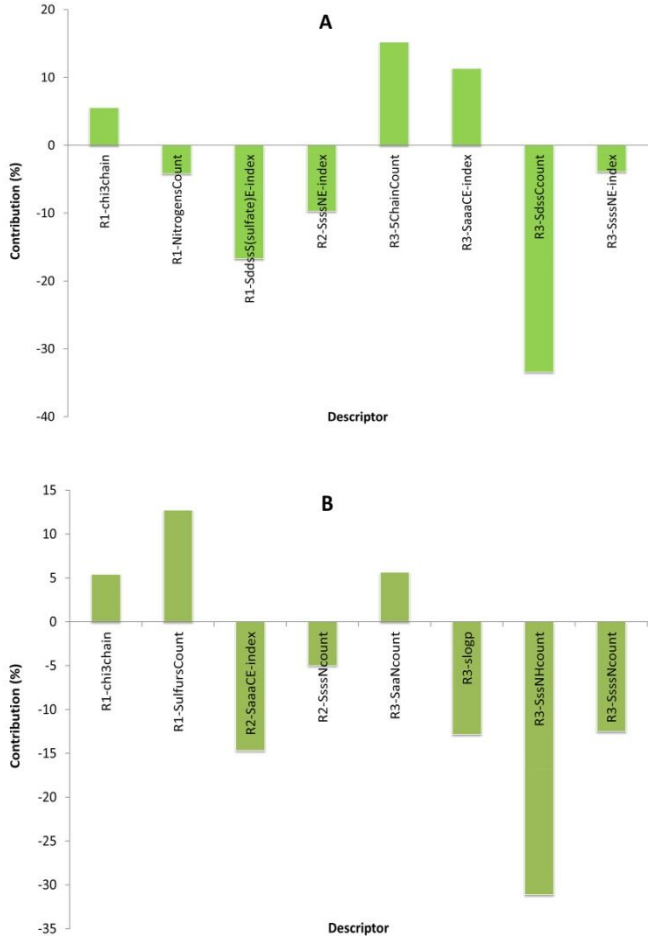


Fig. 2A,B: Percentage contribution of important 2D descriptors found in the GQSAR models for P70S6K inhibitors. (A) Contributing descriptors according to model A. (B) Contributing descriptors according to model B. Descriptor definitions can be found in Table 2.

Importantly, the current GQSAR results, and the interpreted assumptions, are in good agreement with previous reports on SARs of P70S6K inhibitors. Models A and B stated that fragment R1 and R2 are less important for inhibitory activity (Figure 3) with unfavorable roles for the nitrogen-containing structures at R2, suggesting that the role of these fragments in inhibitory activity might be different from hydrogen bonding or a direct interaction with the kinase hinge. Ye and colleagues reported that substituents at C3 of the thiophene ring (corresponding to fragment R1) take part in drug solvation with no interaction with the binding pocket. In addition, *tert*-Butanol (*t*Bu) at C5 of the thiophene ring (fragment R2) is essential for hydrophobic interactions with P70S6K (Ye *et al.,* 2011). On the other hand, the majority of activity variation was confined to fragment R3 in both GQSAR models. The models revealed that (i) nitrogen-based aromatic rings are more preferable than their non-aromatic or non-nitrogen-containing counterparts and (ii) hydrophobic interactions with R3 are not recommended. Bussenius et al. highlighted that pyrozolopyrimidine-containing derivatives are potent inhibitors having IC_{50} ranging from 107-13 nM. They proposed that -NH at the pyrazole ring and N7 at the pyrimidine moiety (corresponding to fragment R3 in the GQSAR model) form hydrogen bonds with Glu173 and Leu175 of the hinge region, respectively (Bussenius *et al.,* 2012). Moreover, it was proposed that an indazole ring linked to thiophene-urea (fragment R3) resulted in stronger binding to the hinge compared to heteroaryl/aryl ether (Ye *et al.,* 2011).

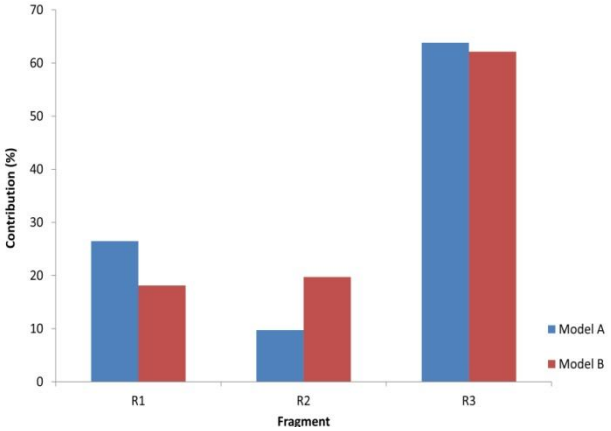


Fig. 3: Percentage contribution of fragments R1, R2, and R3 on GQSAR models.

CONCLUSION

To the best of our knowledge, this is the first study on the structural requirements for P70S6K inhibitors using the GQSAR approach. The current study explored the global chemical space of P70S6K datasets and used a wide range of structurally diverse scaffolds to generate GQSAR models with significant and robust predictive power. The generated models describe the role of each molecular fragment in ligand-target bio-interactions. It was found that fragment R3 has a key role in governing the inhibitory activity of the molecule by the formation of hydrogen bonds with the kinase hinge via aromatic nitrogen-containing structures. In contrast, nitrogen atoms and aromaticity are not preferred at R2, which is expected to form hydrophobic interactions with the binding pocket. The retention index of the cyclopropane ring at R1 is favourable, suggesting an important role of R1 in solvation. In summary, the present investigation indicates the principal requirements for designing innovative and powerful P70S6K inhibitors.

ACKNOWLEDGMENTS

We acknowledge International Islamic University Malaysia who provided funding of this research under research matching grant RMGS 10-003-0013.

REFERENCES

Ajmani S, Agrawal A, Kulkarni SA. A comprehensive structure-activity analysis of protein kinase B-alpha (Akt1) inhibitors. J Mol Graph Model, 2010; 28:683-94.

Ajmani S, Jadhav K, Kulkarni SA. Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR. QSAR Comb Sci, 2009; 28: 36-51.

Akritopoulou-Zanze I, Hajduk PJ. Kinase-targeted libraries: the design and synthesis of novel,potent, and selective kinase inhibitors. Drug Discov Tod, 2009; 14(5): 291-297.

Armstrong A. 2006. Pharmaceutical experimental design and interpretation.CRC Press: Taylor &Francis.

Bamford MJ, Bailey N, Davies S, Dean DK, Francis L, Panchal TA, Parr CA, Sehmi S, Steadman JG, Takle AK, Townsend JT, Wilson DM. (1H-Imidazo[4,5-c]pyridin-2-yl)-1,2,5-oxadiazol-3-ylamine derivatives: Further optimisation as highly potent and selective MSK-1-inhibitors. Bioorg Med Chem Lett, 2005; 15:3407-3411.

Bandarage U, Hare B, Parsons J, Pham L, Marhefka C, Bemis G, Tang Q, Moody CS, Rodems S, Shah S, Adams C, Bravo J, Charonnet E, Savic V, Come JH, Green J. 4-(Benzimidazol-2-yl)-1,2,5-oxadiazol-3-ylamine derivatives: potent and selective p70S6 kinase inhibitors.Bioorg Med Chem Lett, 2009; 19:5191-5194.

Bärlund M, Forozan F, Kononen J, Bubendorf L, Chen Y, Bittner ML, Torhorst J, Haas P, Bucher C, Sauter G, Kallioniemi OP, Kallioniemi A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J Natl Cancer Inst, 2000; 92:1252-1259.

Bärlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi P, Kallioniemi A.Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. Cancer Res, 2000; 60:5340-5344.

Bussenius J, Anand NK, Blazey CM, Bowles OJ, Bannen LC, Chan DS, et al. Design and evaluation of a series of pyrazolopyrimidines as p70S6K inhibitors. Bioorg Med Chem Lett, 2012; 22:2283-2286.

Charrier JD, Miller A, Kay DP, Brenchley G, Twin HC, Collier PN, et al. Discovery and structure-activity relationship of 3-aminopyrid-2-ones as potent and selective interleukin-2 inducible T-cell kinase (Itk) inhibitors. J Med Chem, 2011; 54:2341-2350.

Chen X, Lin Y, Liu M, Gilson MK.The Binding Database: data management and interface design.Bioinformatics, 2002;18:130-139.

Consonni V, Todeschini, R.Molecular descriptors. Rec Adv QSAR Studies, 2010; 29-102.

Davies SP, Reddy H, Caivano M, Cohen P.Specificity and mechanism of action of some commonly used protein kinase inhibitors. Biochem J, 2000; 351:95-105.

Golbraikh A.Molecular dataset diversity indices and their applications to comparison of chemical databases and QSAR analysis. J Chem Inf Comput Sci, 2000; 40:414-425.

Golbraikh A, Tropsha A.QSAR modeling using chirality descriptors derived from molecular topology. J Chem Inf Comput Sci 2003; 43:144-154.

Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comp Chem, 1996; 17:490-519.

Hall LH, Kier LB. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. J Chem Inf Comp Sci, 1995; 35.6:1039-1045.

Hall LH, Kier LB, Brown BB. Molecular similarity based on novel atom-type electrotopological state indices. J Chem Inf Comp Sci, 1995; 35:1039-1045.

Kier LB, Hall LH. An electrotopological-state index for atoms in molecules. Pharma Res, 1990; 7:801-807.

Kier LB,Hall LH. The E-state as an extended free valence. J Chem Inf Comp Sci, 1997;37:548-552.

Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science, 1983; 220:671-680.

Lin H, Yamashita DS, Zeng J, Xie R, Verma S, Luengo JI, Choudhry AE. 2, 3, 5-Trisubstituted pyridines as selective AKT inhibitors. Part II: Improved drug-like properties and kinase selectivity from azaindazoles. Bioorg Med Chem Lett, 2010; 20:679-683.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev, 1997; 23:3-25.

Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res, 2007; 35:D198-D201.

Mukhopadhyay NK, Price DJ, Kyriakis JM, Pelech S, Sanghera J, Avruch J. An array of insulin-activated, proline-directed serine/threonine protein kinases phosphorylate the p70 S6 kinase. J Biol Chem, 1992; 267:3325-3335.

Nittoli T, Dushin RG, Ingalls C, Cheung K, Floyd MB, Fraser H, et al. The identification of 8,9-dimethoxy-5-(2-aminoalkoxy-pyridin-3-yl)-benzo[c][2,7]naphthyridin-4-ylamines as potent inhibitors of 3-phosphoinositide-dependent kinase-1 (PDK-1). Eur J Med Che, 2010; 45:1379-1386.

Okuzumi, T., Fiedler, D., Zhang, C., Gray, D.C., Aizenstein, B., Hoffman, R., and Shokat, K.M. Inhibitor hijacking of Akt activation. Nat Che Biol, 2009; 5:484-493.

Patrick G. 2001. An introduction to medicinal chemistry. Oxford University Press.

Schulz MN,Hubbard RE. Recent progress in fragment-based lead discovery. Cur Opinion Pharm, 2009; 9:615-621.

Scior T, Medina-Franco J, Do QT, Martínez-Mayorga K, Yunes Rojas J, Bernard P. How to recognize and workaround pitfalls in QSAR studies: a critical review. Cur Med Chem, 2009; 16:4297-4313.

Snarey M, Terrett NK, Willett P, Wilton DJ. Comparison of algorithms for dissimilarity-based compound selection. J Mol Graph Model, 1997; 15:372-385.

Sunami T, Byrne N, Diehl RE, Funabashi K, Hall, DL, Ikuta M, et al. Structural basis of human p70 ribosomal S6 kinase-1 regulation by activation loop phosphorylation. J Biol Chem, 2010; 7:285, 4587.

Tao ZF, Chen Z, Bui MH, Kovar P, Johnson E, Bouska J, et al. Macrocyclic ureas as potent and selective Chk1 inhibitors: An improved synthesis, kinome profiling, structure–activity relationships, and preliminary pharmacokinetics. BioorgMed Chem Let, 2007; 17:6593-6601.

Todeschini R,Consonni V. 2000. Handbook of molecular descriptors. WILEY-VCH Verlag GmbH.

Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR & Comb Sci, 2003; 22:69-77.

Vivanco I,Sawyers CL. The phosphatidylinositol 3-kinase–AKT pathway in human cancer. Nat Rev Cancer, 2002; 2:489-501.

Wang S, Midgley CA, Scaërou F, Grabarek JB, Griffiths G, Jackson W, et al. Discovery of N-Phenyl-4-(thiazol-5-yl)pyrimidin-2-amine Aurora Kinase Inhibitors. J Med Chem, 2010; 53:4367-4378.

Ye P, Kuhn C, Juan M, Sharma R, Connoll, B, Alton G, et al. Potent and selective thiophene urea-templated inhibitors of S6K. Bioorg Med Chem Lett, 2011; 21:849-852.

Zheng W, Tropsha A. Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci, 2000; 40:185-194.

How to cite this article:

Abubakar Danjuma Abdullahi, Abdualrahman Mohammed Abdualkader, Nadiahanis Abdullahi, Mohamed Farahidah, Abdul Razak Kasmuri, A. B. M. Helal Uddin. Novel Insight into the Structural Requirements of P70S6K Inhibition Using Group-based Quantitative Structure Activity Relationship (GQSAR). J App Pharm Sci, 2014; 4 (06): 016-024.

Conflict of Interest: None.