European Journal of Pharmaceutical Sciences 76 (2015) 110–118

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

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

QSAR based docking studies of marine algal anticancer compounds as inhibitors of protein kinase B (PKBβ)

G. Dicky John Davis [a,*], A. Hannah Rachel Vasanthi [b]

[a] Department of Bioinformatics, Sri Ramachandra University, Porur, Chennai 600116, India
[b] Department of Biotechnology, Pondicherry University, Kalapet, Puducherry 605014, India

ARTICLE INFO

Article history:
Received 31 January 2015
Received in revised form 22 April 2015
Accepted 29 April 2015
Available online 29 April 2015

Chemical compounds studied in this article:
Callophycin A (PubChem CID: 57399101)
Prevezol C (PubChem CID: 10046593)
Caespitol (PubChem CID: 14314413)

Keywords:
Marine algae
QSAR
Anticancer activity
Docking
Protein kinase B
In silico ADMET
Callophycin A

ABSTRACT

Marine algae are prolific source of bioactive secondary metabolites and are found to be active against different cancer cell lines. QSAR studies will explicate the significance of a particular class of descriptor in eliciting anticancer activity against a cancer type. Marine algal compounds showing anticancer activity against six different cancer cell lines namely MCF-7, A431, HeLa, HT-29, P388 and A549 taken from Seaweed metabolite database were subjected to comprehensive QSAR modeling studies. A hybrid-GA (genetic algorithm) optimization technique for descriptor space reduction and multiple linear regression analysis (MLR) approach was used as fitness functions. Cell lines HeLa and MCF-7 showed good statistical quality ($R^2 \sim 0.75$, $Q^2 \sim 0.65$) followed by A431, HT29 and P388 cell lines with reasonable statistical values ($R^2 \sim 0.70$, $Q^2 \sim 0.60$). The models developed were interpretable, with good statistical and predictive significance. Molecular descriptor analyses revealed that Baumann's alignment-independent topological descriptors had a major role in variation of activity along with other descriptors. Incidentally, earlier QSAR analysis on a variety of chemically diverse PKBα inhibitors revealed Baumann's alignment-independent topological descriptors that differentiated the molecules binding to Protein kinase B (PKBα) kinase or PH domain, hence a docking study of two crystal structures of PKBβ was performed for identification of novel ATP-competitive inhibitors of PKBβ. Five compounds had a good docking score and Callophycin A showed better ligand efficiency than other PKBβ inhibitors. Furthermore *in silico* pharmacokinetic and toxicity studies also showed that Callophycin A had a high drug score (0.85) compared to the other inhibitors. These results encourages discovering novel inhibitors for cancer therapeutic targets by screening metabolites from marine algae.

1. Introduction

Cancer is a complex disease of global concern. In 2008, about 70% of cancer deaths took place in countries with low- and middle income. The American Cancer Society estimated a total of 1,665,540 new cancer cases and 585,720 cancer deaths are projected to occur in the United States alone in 2014 (Siegel et al., 2014). In analogy, Cancer deaths continue to keep rising worldwide and in 2030 it is projected to be a cause for 13.1 million deaths (Ferlay et al., 2010). Cancer-relevant genes have been intensively studied and numerous therapeutic targets which play a vital role in cancer pharmacology have been elucidated.

Protein kinase B (PKB/Akt) is a foremost protein mediating the proliferation and involving several pathways that are essential for the growth of cancer, which makes it a unique therapeutic target for treating cancers of multiple origins. Akt consists of three different cellular isoforms, namely, Akt1 (PKBα), Akt2 (PKBβ), and Akt3 (PKBγ). All three Akt isoforms have the ability to transform into cancerous cells *in vitro*. However, Akt2 is the major isoform found to be amplified or over expressed in human cancer which has been observed in pancreatic tumours (10%), hepatocellular carcinomas (40%) and colorectal cancers (57%) (Hers et al., 2011). Several functionally important regions are present in the protein; specifically the central kinase domain has a classical kinase ATP-binding site that can be explored as binding pockets for small molecule inhibitors. Virtual screening have been employed in previous studies to identify potent inhibitors using the crystal structure of PKBβ in complex with glycogen synthase kinase-3β peptide (GSK-3β) and 5′-adenylylimidodiphosphate (AMP–PNP) (Vyas et al., 2013). Medina-Franco et al. (2009) reported the discovery of a novel competitive inhibitor for ATP binding site by search in a different area of chemical space for selective and potent inhibitors.

* Corresponding author.

http://dx.doi.org/10.1016/j.ejps.2015.04.026
0928-0987/

Marine organisms constitute an important source of novel molecules for new drug discovery and drug development research of which 25% are from algae. Seaweeds have a distinct evolution on their biosynthetic pathways that frequently yield complex molecules with no counterparts in the terrestrial environment. Secondary metabolites from these algae are predominantly sesquiterpenes, diterpenes, triterpenes and C15-acetogenins characterized by the presence of halogen atoms in their chemical structures. An online database – Seaweed Metabolite Database (SWMD) was developed by us to share organized information about marine algal secondary metabolites and their geographical location obtained by literature data mining (www.swmd.co.in). It presently has 1055 compounds from red, green and brown macroalgae describing the method of compound extraction, chemical description of the compound and biological activity of the compound i.e., antibacterial, antimalarial, antioxidant and cytotoxic activities (Davis and Vasanthi, 2011).

Quantitative Structure–Activity Relationship (QSAR) modeling is a ligand-based drug design method for both exploring and exploiting the relationship between chemical structure and its biological action (Liao et al., 2011). To predict the activities of anticancer compounds, quantum chemical descriptors like molecular orbital, dipole moment, charge, etc. and molecular property descriptors like hydrophobic, steric coefficient, etc. have been applied to develop 2D QSAR models (Chen et al., 2007; Zhang et al., 2007). In-vitro evaluation of biological activity can be performed on a number of cell lines for a specific cancer type, but the results of the evaluation would vary based on the cell line employed for the assay. Consequently to design novel and potent anticancer compounds, all the experimental data has to be considered from all the cell lines, such analyses involving all the QSAR descriptors would pave way for predictive models highlighting the importance of a particular class of descriptor in modeling anticancer activity against a specific cancer type (Bohari et al., 2011). Such extensive QSAR studies against many diverse cancer cell lines which are statistically robust are warranted. Hence, a comprehensive 2D QSAR modeling study was performed in the present study using the compounds in SWMD that have cytotoxic activity.

2. Methodology

SWMD which has 1055 entries, 245 compounds (23%) has documented anticancer activity against 43 different cell lines. For continuous response variable (activity), the total minimum number of compounds should be no less than 40 as the number of compounds in the training set should be at least 20, and about 10 compounds should be in each of the test and external evaluation sets (Tropsha, 2010). Therefore, the dataset taken for the study has 157 compounds having cytotoxic activity against six different cancer cell lines namely MCF-7 (Human breast adenocarcinoma), A431 (Human epithelial carcinoma), HeLa (Human cervical adenocarcinoma), HT-29 (Human colon adenocarcinoma grade II), P388 (Murine leukemia) and A549 (Human lung epithelial adenocarcinoma) cells, each having more than 40 compounds (Table 1). The dataset consists of chemical diverse compounds which include sesquiterpenes, diterpenes, triterpenes, sterol and acetogenins that are usually characterized by the presence of one or more halogen atoms in their structures (Fig. 1). The structure of all the 157 compounds and their experimental cytotoxic activity against six cell lines listed in SWMD along with its accession numbers are presented in Additional file (Tables S1).

Table 1
Cell lines against which their anticancer activity was reported in SWMD along with the number of molecules in each cell lines.

S. No.	Cell lines	Cancer type	# of compounds
1	A549	Human lung epithelial adenocarcinoma	72
2	MCF7	Human breast adenocarcinoma	65
3	HT29	Human colon adenocarcinoma grade II	62
4	P388	Murine leukemia	54
5	HeLa	Human cervical adenocarcinoma	50
6	A431	Human epithelial carcinoma	41

Measurement of cytotoxic activity is expressed as half maximal (50%) inhibitory concentration of a substance (IC_{50}) and values are expressed in nanomolar (nM – 10^{-9}) and micromolar (μM – 10^{-6}) levels. The values were converted to the pIC_{50} scale ($-\log IC_{50}$) to predict the narrow value wherein higher values indicate exponentially greater potency. The formula for micromolar conversion of IC_{50} values to pIC_{50} values is

$$pIC_{50} = -\log(IC_{50} * 10^{-6})$$

The pIC_{50} values were used as the dependant variables to construct the QSAR model.

2.1. Calculation of molecular descriptor

QSAR modeling involves the use of software to sketch chemical structures and calculate the descriptors to build predictive models. The molecules to be used in the study were taken from SWMD and 3D structures were generated using MarvinSketch 2011 (v.5.4.1.1) from ChemAxon. The resultant structures are crudely optimized using a molecular mechanics method within MarvinSketch (Csizmadia, 2000). Since some molecular descriptors also require information about the electronic environment of the molecule, the molecules were also optimized for electronic properties with Molecular Orbital PACkage (MOPAC) using WinMopac 7.21 software (Stewart, 1990). Descriptors for the present study were obtained using Vlife MDS 3.5; 239 descriptors based on the physicochemical properties of the molecule and 391 alignment independent descriptors considering topology of the molecule was used (VLife sciences technologies).

2.2. Selection of relevant descriptors

To select the most important descriptors and to decrease the overfitting/overtraining risk, Feature selection techniques were applied to decrease the model complexity using open source software – WEKA (Waikato Environment for Knowledge Analysis) from The University of Waikato (v.3.6.6) (Hall et al., 2009). The dataset consists of 157 anticancer compounds from SWMD for six cancer cell lines models with 630 descriptors each. Filters were applied in the first step to remove compounds and descriptors with missing or null values. Either one of the correlated descriptors having more than 90% correlation in their values were identified and removed. To search the descriptor subspace, linear forward selection was used wherein; correlation-based feature selection (CFS) algorithm in WEKA was employed to evaluate the descriptors. Efforts taken to keep the descriptors employed in the models virtually orthogonal to each other include changing and refining the descriptors that were highly inter-correlated in a given model. This pre-screening gave a quality-assured dataset of compounds which are used for further analysis.

Selection of descriptors was done using the hybrid optimization technique developed by Rogers and Hopfinger wherein a wrapper of genetic algorithms (GA) for feature selection and multiple linear regression (MLR) as regression technique was employed (Rogers and Hopfinger, 1994). The GA was employed for searching the descriptor subspace, whereas the MLR was used for fitness evaluation to determine the significant descriptors. GeneticSearch in

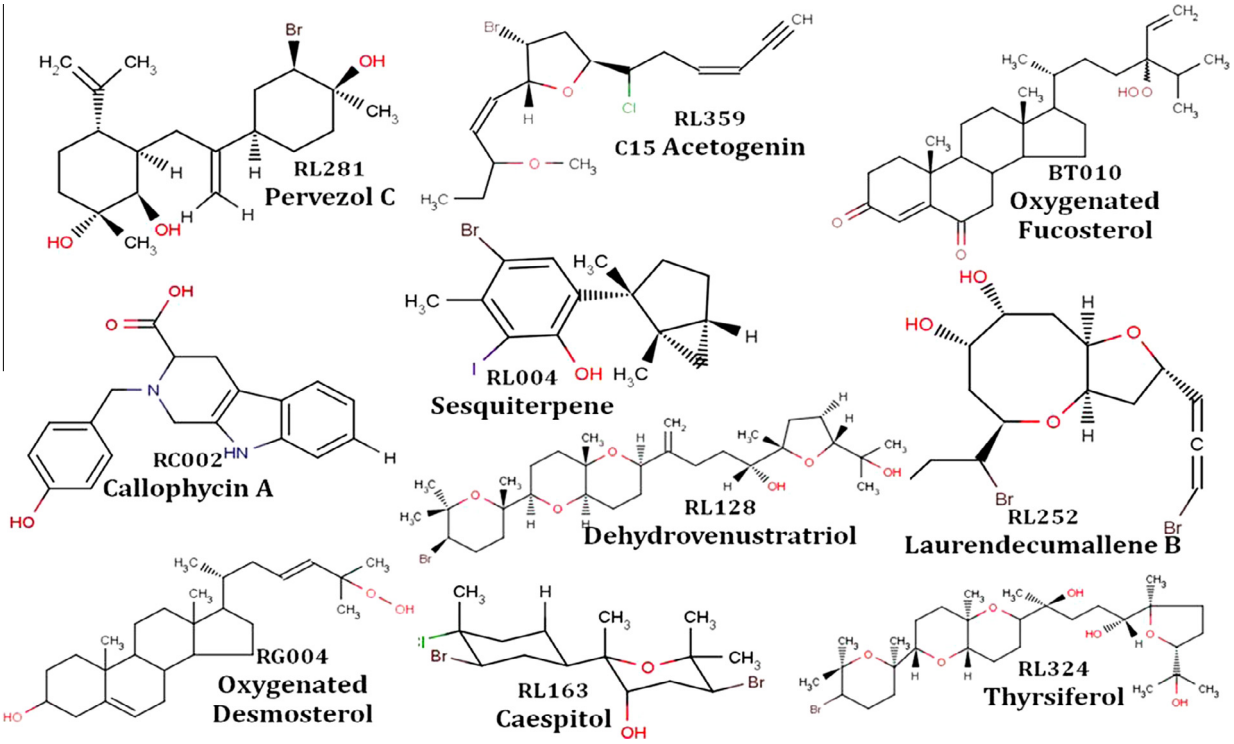

Fig. 1. Structurally diverse cytotoxic compounds in Seaweed Metabolite Database (SWMD).

WEKA uses a simple genetic algorithm (Goldberg et al., 1993) and default values were selected for the GA parameters, such as 20 generations, population size of 20, crossover probability of 0.6 and mutation probability of 0.033. The attribute selection in WEKA employed for MLR uses the M5's method where the attributes are removed stepwise given by the Akaike information criterion; the ones with the smallest standardised coefficient is removed until no improvement in the estimate of the error is observed (Witten and Frank, 2005). The extent of the coefficients discloses the degree of influence of the corresponding molecular descriptors on the given property.

2.3. Validation of QSAR models

The regression models are developed by dividing the dataset into multiple chemically diverse training and test sets with a rational approach based on Sphere Exclusion (SE) algorithm (Golbraikh and Tropsha, 2002). To evaluate the performance of the QSAR model, Leave-one-out cross validation (Q^2) is carried out to obtain the optimal number of components (N) and the correlation coefficient R^2. To evaluate the performance without any bias, two independent test set was made and the remaining compounds were used for model development. Models were developed using the training set while the test set was not used but serves to test the extrapolative ability of final models. For each compound in the training set, a correlation equation was derived with descriptors. The observed and predicted activity with residuals and descriptor values for all the developed models are presented in Additional file (Tables S2–S13). The predicted biological activities of untested compounds from their molecular structures are also presented in the above said tables. The two independent test set are presented at the end of the table marked with asterisks. The MLR regression equations for each of the table are also presented in Additional file (Tables S2–S13).

2.4. Docking of protein kinase (PKB β) inhibitors

Structures of PKBβ were taken from Protein databank that were quite similar; (PDB ID: 2UW9 and 2JDR). The backbone RMSD for the whole protein was only 0.36 Å and 0.3 Å for the binding site residues that are within 5 Å of ligand (Medina-Franco et al., 2009). Bioactive conformation was simulated for 2UW9 and 2JDR using Molegro Virtual Docker (MVD v3.0.0) which implements evolutionary algorithms for molecular docking simulations and docking was performed by Moldock function (Thomsen and Christensen, 2006). For both PDB structures, during docking water molecules were removed and co-crystal inhibitors were ignored. From the docking wizard, ligands were selected from SWMD; 157 cytotoxic compounds that were used in the QSAR study.

The protein and ligands molecules were prepared at first and bonds, bond orders, explicit hydrogens, charges and flexible torsions, were assigned if they were missing using the MVD program. MVD was used for active site (pocket) detection on PKBβ protein. The ATP binding site was defined as active site box having volumes of 359 $Å^3$ and 388 $Å^3$ for 2UW9 and 2JDR, respectively. Further, the binding site was delineated by choosing all atoms within 10 Å of the analogous crystallographic ligand with the cavity detection mode active and using default parameters. The Ignore distant atoms option was applied to ignore atoms far-off from the binding site. The search algorithm taken was Moldock SE and the number of runs taken as 10 and max iterations were 2000 with population size of 50 and with an energy threshold of 100. The energy penalty was set to 100, RMSD threshold was 2.00 and RMSD calculation by atom ID (fast) were set. After the docking simulation was completed the poses generated were sorted by rerank score (De Azevedo, 2010). Results of the top ligands whose rerank score >−100 were selected.

2.5. In silico ADMET analysis

Successful drug discovery necessitate good pharmacokinetics and no toxicity of lead structures which would make it more drug-like. *In silico* tools on absorption, distribution, metabolism, excretion and toxicity (ADMET) screening of candidate drugs help to reduce the risk of late-stage attrition and optimize the most promising compounds. Therefore, the ligands with top docking score were further screened for druglikeliness, drug score and toxicity characteristics using the program OSIRIS property explorer (v.2.0) (Sander et al., 2009).

3. Results and discussion

The predictive QSAR models were built for six different cancer cell lines with experimental data from 157 compounds, using independent and with minimum number of descriptors. The distribution of IC_{50} values among the six cell lines was seen to differ from one compound to another and is shown Fig. 2. The criterion for selecting the best model was based on the correlation coefficient values taken from the correlation of approximately 630 descriptors in different combinations. To evaluate the effect of the number of descriptors on the correlation coefficient values, all the models were tested on training set by correlating 1–10 descriptors separately and presented in Fig. 3.

It was observed that in various models, four descriptors were adequate for obtaining a good correlation and in most of the models using more than four descriptors made only a little variation on the statistical quality. While in most of the cases, seven or more descriptor-based models gave high correlation and cross-validation coefficient values, nevertheless this may be false and so will not be accurate in prediction of IC_{50} values further.

Various statistical measures were adapted in the present study to assess and compare the predictive power and the stability of QSAR models, which is also widely applied for evolution of a significant model. As four descriptors were sufficient for getting a good correlation, all the training set had more than 20 numbers of molecules. Square of the correlation coefficient (R^2) represents the statistical significance of the model, herein all the models in the study were inferred significant if R^2 is greater than 0.7. While Q^2 is the cross-validated R^2, a measure of the quality of the QSAR model was inferred significant if Q^2 is greater than 0.5. Fischer statistics (F) is the ratio between explained and unexplained variance for a given number of degrees of freedom. This indicates a factual correlation or the significance level for QSAR models. Higher the F-test more significant is the model. The average of absolute difference between experimental and predicted IC_{50} values is average residual (AE). Lower the AE, more significant is the model.

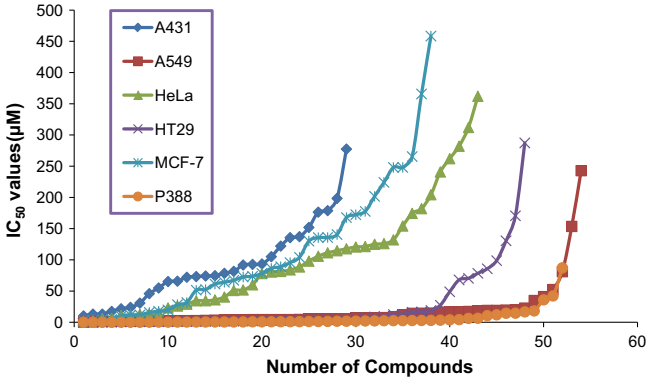

Fig. 2. Distribution of IC_{50} value among cell lines.

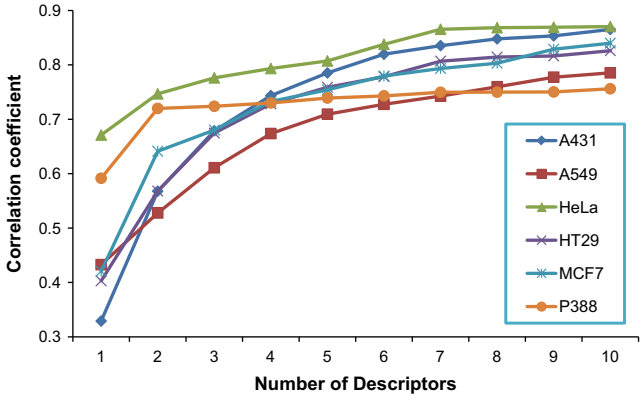

Fig. 3. Effect of number of descriptors on the correlation coefficient.

Regression equation for the QSAR models is presented in the regression summary along with name of the cell lines and types of cancer (Table 2). To identify that the developed models are valuable, AE was calculated for both training and test set of compounds and it was found that all the models have a lower AE values (0.15–0.35), thus have the capability to establish the relationship between the structure and activity. The QSAR models for HeLa (Cervical) and MCF-7 (Breast) cell lines can be used for the prediction as it exhibits good statistical quality ($R^2 \sim 0.75$, $Q^2 \sim 0.65$) and considered valuable for the available class of compounds. The statistical quality of A431 (Epithelial), HT29 (Colon) and P388 (Leukemia) cell lines are also reasonable ($R^2 \sim 0.70$, $Q^2 \sim 0.60$), and extra care is required before utilizing these models for the prediction. However, the statistical quality of A549 (Lung) cell lines cannot be used for the prediction because of the insignificant statistical results obtained for this model ($R^2 = 0.67$, $Q^2 = 0.50$). The reason for poor result in A549 cell lines is probably due to the involvement of diverse compound types in this model. The increase in the number of descriptors for A549 does not improve the quality of the model (with 10 descriptors $R^2 \sim 0.78$) and thus indicates that the currently used four descriptors are good enough for developing the structure–activity relationship for this model.

Statistical quality of QSAR depends on the distribution of IC_{50} values of compounds among the cell lines where IC_{50} values are distributed in broad range would show good statistical quality and vice versa. IC_{50} values for HeLa and MCF-7 cell lines are for a broad range wherein the predictions showed good statistical quality (Fig. 2). The range of IC_{50} values for A431, HT29, P388 and A549 cell lines are reasonable and so also the QSAR prediction. The quality of the QSAR model was poor for A549 cell lines as the compounds had diverse structures as shown in Additional file (Table S14). Compounds that do not fit in the developed QSAR model are called as outliers. Compound RL018 (2-tridecyl-2-heptadecenal) is an outlier in A431, HeLa and MCF7 QSAR models as it has an aliphatic hydrocarbon whereas the others are aromatic compounds.

The experimental and predicted IC_{50} values for training and test set (1) and the number compounds in both the sets along with their average residual for all QSAR models are represented in Fig. 4. This figure clearly demonstrates that the compounds of the test set are closer to the line compared with the compounds of training set. The applicability of generated QSAR models was rigorously validated by dividing another independent test set. The statistical performance of the second test set was similar to that of the first test set. Both the test sets revealed similar statistical performance indicating that the developed models were satisfactory. The observed and predicted activities with residuals and number compounds in training and test set are presented in Fig. 5, for all the developed models for the second set of test compounds.

Table 2
Regression summary for all the QSAR models.

Cell line (Type)	# compounds				Regression equation	R^2	Q^2	AE	F
	TR	TS	PD	O					
A431 (Epithelial)	24	4	12	1	=−0.2967 * 5ChainCount + 0.2105 * SsssCHE-index	0.74	0.60	0.15	13.76
					−0.1817 * T_2_Br_5 + 0.3143 * T_O_Br_4 + 4.1884	0.73	0.51	0.16	12.67
A549 (Lung)	46	9	17	0	=−0.1969 * SaaCHE-index + 1.1969 * T_2_Cl_1 + 0.5973 * chiV3Cluster	0.67	0.50	0.23	21.16
					−3.2011 * SAMostHydrophilic + 4.5436	0.65	0.47	0.25	18.95
HeLa (Cervical)	34	7	8	1	=0.0039 * 5PathCount −0.0929 * *X*log*P* + 0.5874 * T_O_Br_6 + 0.2666 * 3ChainCount + 4.0602	0.79	0.67	0.21	27.83
						0.78	0.71	0.19	25.74
HT29 (Colon)	42	8	12	0	=0.3328 * SdssCE-index	0.73	0.58	0.35	25.39
					−9.9514 * SAAverageHydrophilicity + 0.1929 * T_2_2_4 + 9.7350 * chiV6chain + 3.5298	0.78	0.61	0.32	32.46
MCF7 (Breast)	31	6	26	1	=−4.8683 * chiV5chain + 0.2718 * 6ChainCount + 0.1071 * T_O_Br_4 + 0.0674 * T_2_2_5 + 3.9220	0.74	0.67	0.22	18.40
						0.77	0.69	0.22	21.16
P388 (Leukemia)	39	9	6	0	=18.4947 * chi6chain −0.2150 * T_O_O_2 −0.0535 * T_2_O_4 −0.2514 * 3ChainCount + 3.4853	0.73	0.67	0.29	22.98
						0.72	0.66	0.32	21.52

Cell line with type of cancer in parenthesis, regression summary (regression equation, correlation coefficient (R^2), cross validation coefficient (Q^2), average residual (AE) and number of outliers (O) and number of compounds (training set (TR), test set (TS) and predicted set (PD)) in various cell lines based QSAR models for both the test set.

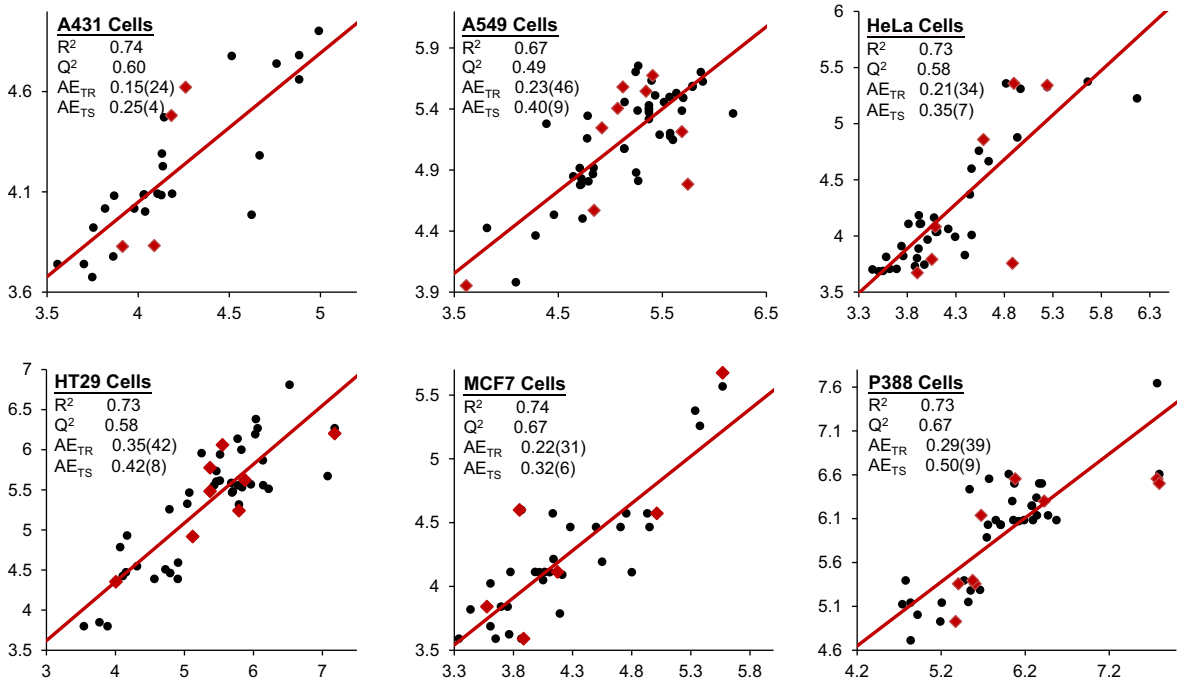

Fig. 4. Plot between experimental and predicted IC_{50} values for training and test set (1) QSAR models. Correlation coefficient (R^2), Cross-validation coefficient (Q^2), Average residual Training set (AE_{TR}), Average residual Test set (AE_{TS}) & number of compounds in training and test set are shown in brackets.

In the developed QSAR models, 22 descriptors (14 Physicochemical and 8 Baumann's Alignment independent) were used in different combinations. The particulars of all the 22 descriptors, its type and incidence in the models are depicted in Fig. 6. The details of the descriptors involved in the study and their occurrence in the QSAR models is shown in Additional file (Table S14). The inter-correlation of the descriptors appeared in all the developed models were taken into account, and the descriptors were found to be reasonably orthogonal and presented in Additional file (Table S15). All the models have identified alignment independent descriptors as vital descriptors. The 'atom and bond count' descriptors especially number of Oxygen, Bromine and Chlorine atoms is chosen in most models. It is well known that the seaweed metabolites are biologically active due to the high degree of halogenations thereby exhibiting antibacterial, antifungal, antiviral, anti-inflammatory, antiproliferative, cytotoxic, antifouling, antifeedant, ichthyotoxic, and/or insecticidal activity (Lhullier et al., 2010). The same is reflected in the obtained descriptors as halogenation of molecules increases the cytotoxic activity and is also proved. Chain path count descriptors (such as 3ChainCount, 5ChainCount, 6ChainCount), retention index and atomic valence connectivity index also were identified in the models. The maximum six membered rings positively influence the activity and percentage contribution in most of the models (Fig. 7). The Hydrophobicity SlogpA descriptors which is the hydrophilic value on the Van der Waals surface of molecules decreases the anticancer activity of compounds. The other descriptors include Estate contributions which are the Electrotopological state indices of valency of C atoms and bond order.

QSAR analyses on a wide variety of structurally diverse Akt1 inhibitors revealed the central role of Baumann's alignment independent topological descriptors beside additional descriptors such

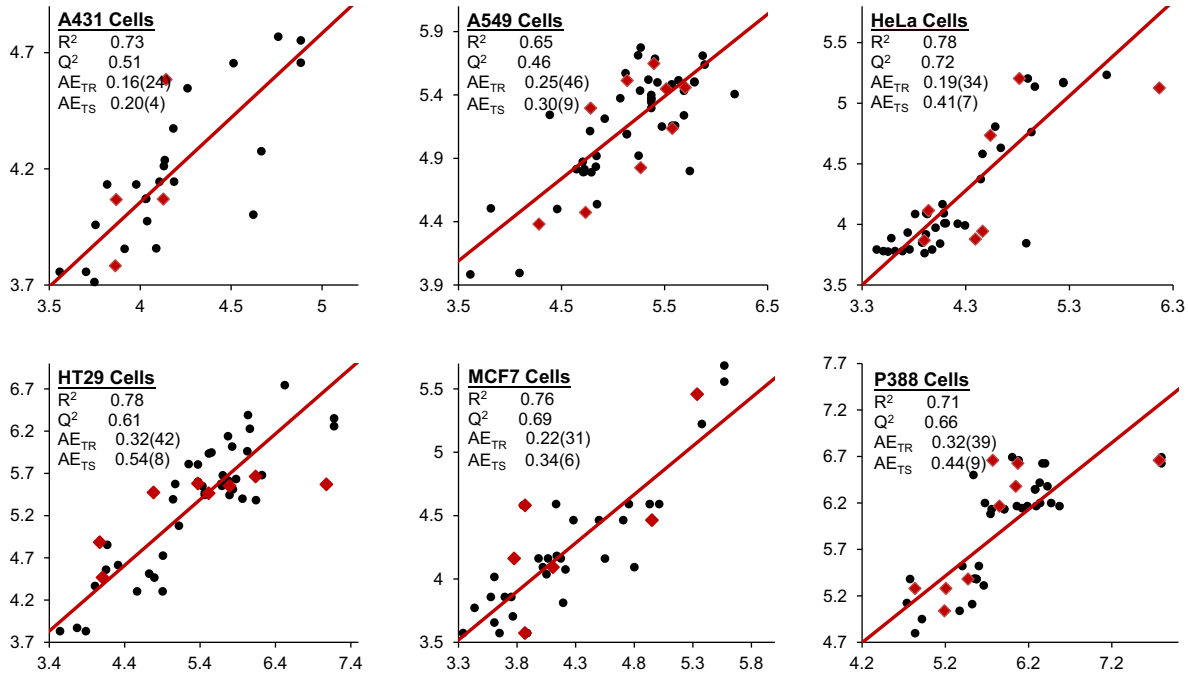

Fig. 5. Plot between experimental and predicted IC_{50} values for training and test set (2) QSAR models. Correlation coefficient (R^2), Cross-validation coefficient (Q^2), Average residual Training set (AE_{TR}), Average residual Test set (AE_{TS}) & number of compounds in training and test set are shown in brackets.

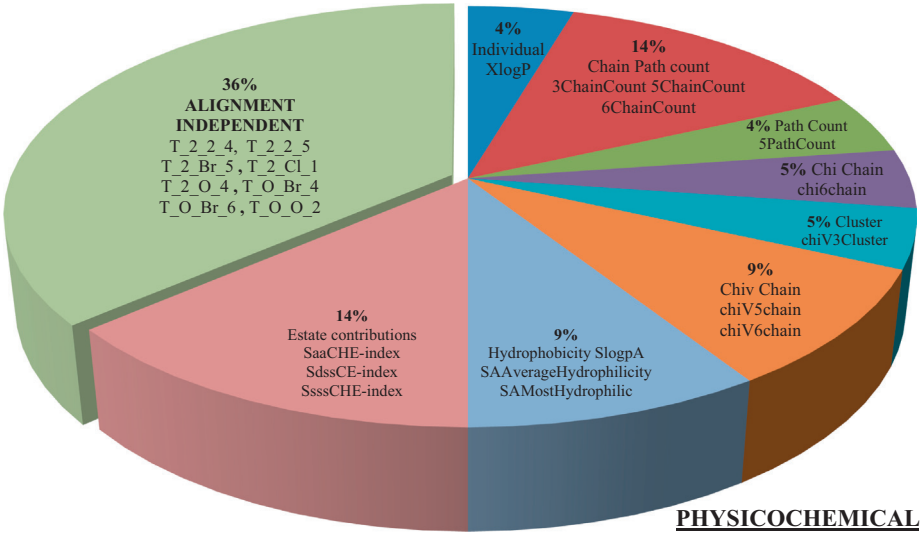

Fig. 6. Classification of various descriptors involved in QSAR model.

as the number of hydrogen bond donors, hydrogen bond acceptors, rotatable bonds and aromatic oxygen (SaaOcount) together with alkene carbon atom type (SdsCHE-index) and molecular branching (chi3Cluster), in accomplishing differential activity (Ajmani et al., 2010). Further in the same study, Group-based QSAR analyses showed that for achieving highly potent Akt1 inhibitors, chemical variations such as presence of hetero-aromatic ring, flexibility, polar surface area and fragment length present in the hinge binding fragment are highly influential. In addition, the study also reported a three descriptors model using *k*-nearest neighbor classification to differentiate molecules binding to Akt1 kinase or PH domain and reported the main role of oxygen (SssOE-index) and aromatic carbon (SaaCHE-index and SaasCE-index) atoms electro-topological environment in binding to Akt1 kinase. Structure–activity analysis of Akt1 kinase reported by Ajmani et al. (2010) correlated well with the descriptor attributes of our present QSAR study of anticancer compounds from seaweeds. Wherein, the anticancer activity was contributed by Baumann's alignment independent topological descriptors along with Oxygen, Bromine and Chlorine atoms and aromatic carbon (SaaCHE-index) atoms. Also as Akt isozymes are approximately 80 percent identical and have a high degree of overall homology, the QSAR study was further extended to identify novel Akt2 inhibitors.

The 157 cytotoxic compounds from SWMD were docked at the ATP binding site of 2JDR and 2UW9 wherein several molecules showed a better Moldock score than the co-crystal inhibitor. A low degree of consensus of Moldock score was observed with each crystal structure (2UW9 and 2JDR) between the top ranked scoring molecules. In general among the top 10 ranked compounds docked

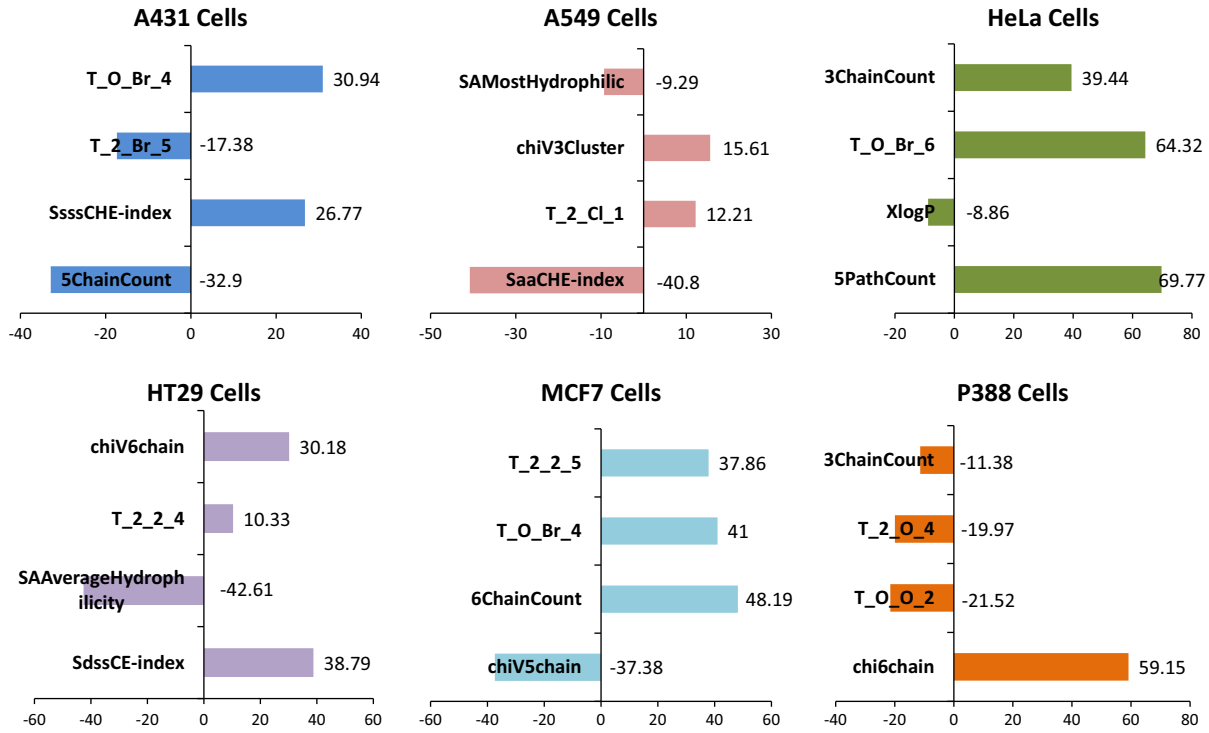

Fig. 7. Percentage contribution of each descriptor in developed QSAR model explaining variation in the activity.

Table 3
Docking results of PKBβ inhibitors.

Ligand name	Molecular formula	MolDock score	Rerank score	No of H bond	H bond energy	Ligand efficiency		Interacting residues
						LE1	LE2	
PDB: 2UW9								
RL378	$C_{24}H_{35}BrO_6$	−139.13	−104.23	8	−9.39	−4.49	−3.36	Ala232, Glu230, Glu230, Thr213, Thr292, Asp293, Lys181, Lys277
RG009	$C_{27}H_{42}O_4$	−135.71	−103.21	8	−14.15	−4.38	−3.33	Ala232, Ala232, Glu230, Glu230, Thr213, Asn280, Asp293, Lys277
RG004	$C_{27}H_{44}O_3$	−130.26	−103.01	7	−10.42	−4.34	−3.43	Ala232, Ala232, Glu230 Glu230, Thr213, Lys277, Glu279
PDB: 2JDR								
RL078 Laurenmariannol	$C_{30}H_{53}BrO_7$	−156.93	−100.91	8	−12.91	−4.13	−2.66	Ala232, Glu230, Glu230, Thr213, Thr292, Thr292, Asp293, Asp293
RC002 Callophycin A	$C_{19}H_{18}N_2O_3$	−123.00	−106.14	5	−9.81	−5.13	−4.42	Ala232, Glu230, Asp293, Glu200, Lys181

Ligand Efficiency 1 (LE1) – Moldock score divided by Heavy Atoms count and Ligand Efficiency 2 (LE2) – rerank Score divided by Heavy Atoms count.

with Moldock only one molecule was found in common in both crystal structures. The selection of molecules were based on a high docking score with rerank score >−100 and a characteristic feature observed in several PKBβ inhibitors which is the ability to make hydrogen bonds with Glu230 and Ala232 (Saxty et al., 2007; Vyas et al., 2013; Medina-Franco et al., 2009).

Docking results in the present study suggested that compound from marine red algae RL378, RG004 and RG009 had a good docking score for 2UW9 whereas RL078 and RC002 had a good docking score for 2JDR (Table 3). RC002 (Callophycin A), tetrahydro-β-carboline was isolated from the methanol extract of red algae *Callophycus oppositifolius* and was shown to mediate anticancer and cytotoxic effects on a series of human tumour cell lines and a normal mammalian cell line (Ovenden et al., 2011). Five H-bonds were formed between ligand RC002 and protein PKBβ (2JDR) with a rerank score of 106.14 (Fig. 8). The oxygen atom (hydroxyl) of ligand RC002 formed two H-bonds, one with $-NH_2$ of Ala232 and another carbonyl oxygen atom of Glu230. The oxygen atom of carboxylic acid moiety formed two H-bond with NH_2 of Lys181 and another with carboxylic acid moiety oxygen atom of Glu200. Oxygen atom (hydroxyl) of carboxylic acid formed H-bonds with $-NH_2$ of Asp293. RC002 which has a good docking score and ligand efficiency better than the other ligands studied to be an active PKBβ inhibitor hit and confirms the affinity with Glu230 and Ala232.

Pharmacokinetic properties and toxicities were predicted for all the five ligands that showed good docking results using OSIRIS property explorer and are shown in Table 4. A compound's hydrophilicity is measured as the logarithm of its partition coefficient between n-octanol and water $\log(c_{octanol}/c_{water})$, given as $\log P$ value of a compound. Calculated Log P (cLog P) is given by OSIRIS. Drug absorption and distribution characteristics depend on its aqueous solubility, estimated as $\log S$ value. The program calculates the drug score which is the compound's overall potential to qualify for a drug by adding the total values of drug-likeliness, $c\log P$, $\log S$, molecular weight, and toxicity risks. Thus the drug score judges the compound's overall potential as a drug candidate wherein RC002 had the highest drug score (0.85) among the five compounds. Docking studies at the ATP-binding site of PKBβ and the *in silico* ADMET properties suggest RC002 (Callophycin A) as a potential anticancer drug candidate.

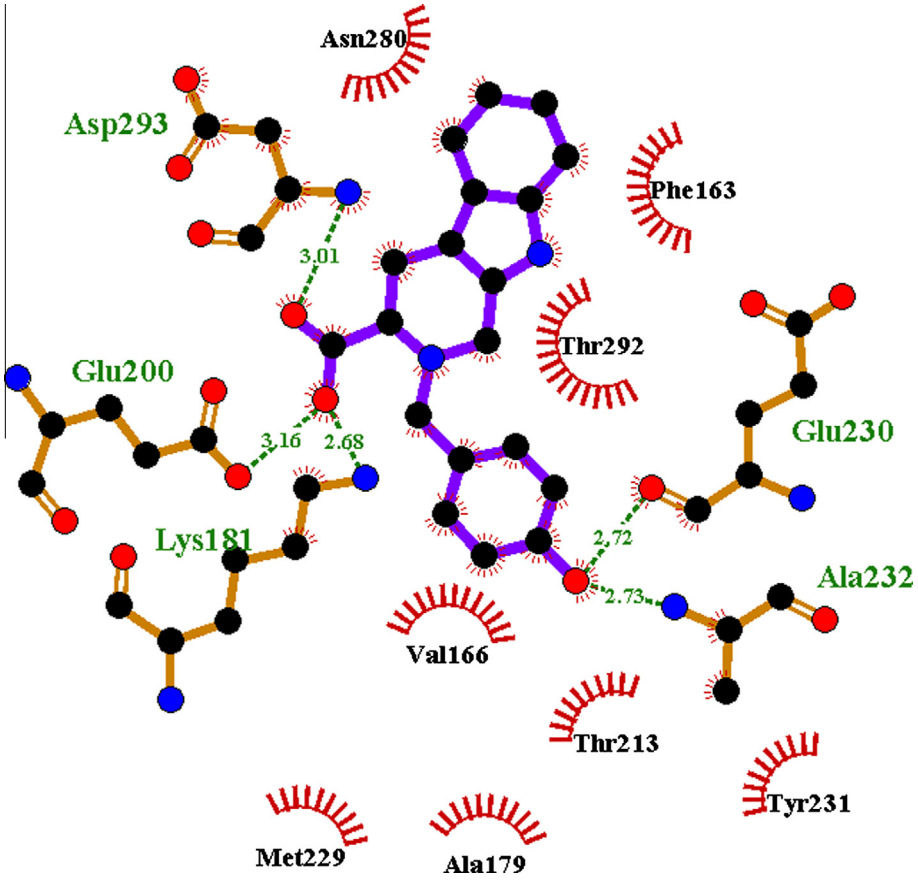

Fig. 8. Interaction of Callophycin A (RC002) in the ATP site of PKBβ (PDB Code 2JDR). Docking studies showing 5 hydrogen bond interactions of Callophycin A with 2JDR at Ala232, Glu230, Asp293, Glu200, Lys181. Hydrogen bonds are shown in green. Atom colors: black, carbon; red, oxygen; blue, nitrogen. The diagram was created with the program LigPlot+ (v.1.4.5) Laskowski and Swindells, 2011. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
In silico ADMET prediction of PKBβ inhibitors using OSIRIS property explorer.

Physicochemical and ADMET parameters/properties	RL378	RG004	RG009	RL078	RC002
Mutagenic	No	Yes	Yes	No	No
Tumorigenic	No	Yes	Yes	No	No
Irritant	Yes	Yes	Yes	No	No
Reproductive effective	No	No	No	No	No
*c*Log*P*	2.65	5.98	5.32	3.22	0.67
Solubility	−4.35	−5.89	−5.54	−5.58	−2.91
Molecular weight	498	416	430	604	322
Drug likeness	−1.31	−2.52	−1.79	−10.96	2.44
Drug score	0.22	0.04	0.05	0.19	0.85

Solubility measured in mol/liter is estimated as log*S* value.

4. Conclusion

A novel PKBβ inhibitor with a unique scaffold compared to existing published PKBβ inhibitors was identified by virtual screening from marine algae. This would be a starting point for an optimizing the molecule as PKBβ inhibitor. Further development of Callophycin A will include exploring the structure–activity relationship required to obtain the desired PKB selectivity. These results further encourages discovering newer PKBβ inhibitors for the treatment of cancer and screening metabolites of marine algae for particular beneficial biological effects which will undoubtedly pay off in the future. The present study has shown a roadmap for further exploiting the chemoinformatics approach in cancer drug discovery using various molecular targets for the development of novel anticancer agents from marine algae. The same approach can further be used for drug discovery and development from the ocean for other diseases as well.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejps.2015.04.026.

References

Ajmani, S., Agrawal, A., Kulkarni, S.A., 2010. A comprehensive structure-activity analysis of protein kinase B-alpha (Akt1) inhibitors. J. Mol. Graph. Model. 28, 683–694.
Bohari, M.H., Srivastava, H.K., Sastry, G.N., 2011. Analogue-based approaches in anti-cancer compound modelling: the relevance of QSAR models. Org. Med. Chem. Lett. 1, 3–15.
Chen, J., Shen, Y., Liao, S., Chen, L., Zheng, K., 2007. DFT-based QSAR study and molecular design of AHMA derivatives as potent anticancer agents. Int. J. Quantum Chem. 107, 1468–1478.
Csizmadia, F., 2000. JChem: Java applets and modules supporting chemical database handling from web browsers. J. Chem. Inf. Comput. Sci. 40, 323–324.
Davis, G.D.J., Vasanthi, A.H.R., 2011. Seaweed metabolite database (SWMD): a database of natural compounds from marine algae. Bioinformation 5, 361–364.
De Azevedo, W.F., 2010. MolDock applied to structure-based virtual screening. Curr. Drug Targets 11, 327–334.
Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M., 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer 127, 2893–2917.
Golbraikh, A., Tropsha, A., 2002. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. Mol. Divers. 5, 231–243.

Goldberg, D.E., Deb, K., Kargupta, H., Harik, G., 1993. Rapid, accurate optimization of difficult problems using messy genetic algorithms. In: Proceedings of the Fifth International Conference on Genetic Algorithms, S. Forrest, Ed. pp. 56–64.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newslett. 11, 10–18.

Hers, I., Vincent, E.E., Tavaré, J.M., 2011. Akt signalling in health and disease. Cell. Signal. 23, 1515–1527.

Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 51, 2778–2786.

Lhullier, C., Falkenberg, M., Ioannou, E., Quesada, A., Papazafiri, P., Horta, P.A., Schenkel, E.P., Vagias, C., Roussis, V., 2010. Cytotoxic halogenated metabolites from the Brazilian red alga Laurencia catarinensis. J. Nat. Prod. 73, 27–32.

Liao, C., Sitzmann, M., Pugliese, A., Nicklaus, M.C., 2011. Software and resources for computational medicinal chemistry. Future Med. Chem. 3, 1057–1085.

Medina-Franco, J.L., Giulianotti, M.A., Yu, Y., Shen, L., Yao, L., Singh, N., 2009. Discovery of a novel protein kinase B inhibitor by structure-based virtual screening. Bioorg. Med. Chem. Lett. 19, 4634–4638.

Ovenden, S.P., Nielson, J.L., Liptrot, C.H., Willis, R.H., Tapiolas, D.M., Wright, A.D., Motti, C.A., 2011. Callophycin A, a cytotoxic tetrahydro-β-carboline from the red alga Callophycus oppositifolius. Phytochem. Lett. 4, 69–71.

Rogers, D., Hopfinger, A.J., 1994. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J. Chem. Inf. Comput. Sci. 34, 854–866.

Sander, T., Freyss, J., von Korff, M., Reich, J.R., Rufener, C., 2009. OSIRIS, an entirely in-house developed drug discovery informatics system. J. Chem. Inf. Model. 49, 232–246.

Saxty, G., Woodhead, S.J., Berdini, V., Davies, T.G., Verdonk, M.L., Wyatt, P.G., Boyle, R.G., Barford, D., Downham, R., Garrett, M.D., Carr, R.A., 2007. Identification of inhibitors of protein kinase B using fragment-based lead discovery. J. Med. Chem. 50, 2293–2296.

Siegel, R., Ma, J., Zou, Z., Jemal, A., 2014. Cancer statistics. Cancer J. Clin. 64, 9–29.

Stewart, J.J.P., 1990. MOPAC: a semiempirical molecular orbital program. J. Comput. Aided Mol. Des. 4, 1–103.

Thomsen, R., Christensen, M.H., 2006. MolDock: a new technique for high-accuracy molecular docking. J. Med. Chem. 49, 3315–3321.

Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Mol. Infor. 29, 476–488.

VLife Sciences technologies: MDS 3.5 Molecular Design Suite. Pvt. Ltd., Pune, India. <http://www.vlifesciences.com/>.

Vyas, V.K., Ghate, M., Goel, A., 2013. Pharmacophore modeling, virtual screening, docking and in silico ADMET analysis of protein kinase B (PKB β) inhibitors. J. Mol. Graph. Model. 42, 17–25.

Witten, I.H., Frank, E., 2005. Data Mining, Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufmann.

Zhang, S., Wei, L., Bastow, K., Zheng, W., Brossi, A., Lee, K.H., Tropsha, A., 2007. Antitumor agents 252. Application of validated QSAR models to database mining: discovery of novel tylophorine derivatives as potential anticancer agents. J. Comput. Aided Mol. Des. 21, 97–112.