Journal of Molecular Graphics and Modelling 34 (2012) 76–88

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

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

Exploring the structure requirement for PKCθ inhibitory activity of pyridinecarbonitrile derivatives: an *in silico* analysis

Yan Li[a,*,1], Ming Hao[a,1], Hong Ren[b,c], Shuwei Zhang[a], Xia Wang[d], Ming Ma[d], Guohui Li[c], Ling Yang[e]

[a] *Department of Materials Science and Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116023, China*
[b] *Department of Ophthalmology, Qi Lu Hospital, Medical School of Shandong University, Jinan 250012, China*
[c] *Laboratory of Molecular Modeling and Design, State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*
[d] *Center of Bioinformatics, Northwest A&F University, Yangling, Shaanxi 712100, China*
[e] *Laboratory of Pharmaceutical Resource Discovery, Dalian Institute of Chemical Physics, Graduate School of the Chinese Academy of Sciences, Dalian, Liaoning 116023, China*

ARTICLE INFO

Article history:
Received 29 July 2011
Received in revised form
20 December 2011
Accepted 27 December 2011
Available online 3 January 2012

Keywords:
Protein kinase Cθ
CoMFA
CoMSIA
3D-QSAR
Molecular docking
Molecular dynamics

ABSTRACT

Presently, an *in silico* modeling was carried out on a large series of 263 PKCθ inhibitors using 3D-QSAR, molecular docking and molecular dynamics (MD) simulations for the first time. Based on different alignment rules, several computational models were established with their statistical results compared. The resultant models derived from the database alignment exhibit satisfying internal and external predictive capabilities with q^2 of 0.503, 0.616 and r^2_{pred} of 0.568, 0.602 for CoMFA and CoMSIA, respectively. The consistency of conclusion among 3D contour maps of CoMFA and CoMSIA, molecular docking and molecular dynamics proves the reliability of the developed models. The analysis of the 3D contour plots permits interesting conclusions about the effects of different substituent groups at different positions of the common scaffold. In addition, Leu461 and Asn509 have been identified as the key amino acid residues to form H-bond interaction with the ligand compound. The developed models will provide a clue to the design of novel PKCθ inhibitors.

1. Introduction

Protein kinase Cθ (PKCθ), a Ca^{2+}-independent novel PKC subfamily member, is a key T-cell receptor (TCR) signaling molecule [1]. PKCθ is expressed primarily in the lymphocytes, skeletal muscle and platelets, which is independently cloned and characterized by three different groups [2–4]. It has been reported that the key translocation of PKCθ results in an activation of several transcription factors required for the T cell activation. PKCθ plays a central role in the TCR-mediated activation of transcription factors which, in turn, upregulate the interleukin-2 (IL-2) gene expression [5]. The previous experiment illustrates PKCθ knockout mice have shown diminished responses in various T-cell mediated disease models including the type II collagen-induced arthritis model [6], the experimental autoimmune encephalomyelitis model of multiple sclerosis [7] as well as the ovalbumin challenge model of asthma [8], etc. It is interesting that, in spite of their defective T cell activation pathway, PKCθ knockout mice have been shown to have normal Th1 cell response in the lung and normal viral clearance [9]. For these reasons, the inhibition of PKCθ has recently become an attractive target for treatment of autoimmune and inflammatory diseases. Up to date several kind of small molecular PKCθ inhibitors have been reviewed [10] including the Boehringer Ingleheim-, Millennium-, Altana- and Wyeth-based PKCθ inhibitors [11–22] and so forth. However, among them, a large number of PKCθ inhibitors are from the patent literatures. Therefore, for many of these compounds, no absolute value for the inhibition of PKCθ or selectivity against other PKC isoforms is disclosed except those PKCθ inhibitors reported by Wyeth Research available to the public. The SAR assessment of these Wyeth-based PKCθ inhibitors, based on the common pyridinecarbonitrile structure, illustrates both their potential against PKCθ and the excellent selectivity over a variety of PKC isoforms.

However, the production of such a large quantity of SAR information using a trial and error method is time-consuming and costly. The application of quantitative structure–activity relationship (QSAR) methodologies to the development of new leads, thus, has the potential to decrease substantially the time and effort required to discover new medicines or improve current ones in terms of their efficacy [23]. Therefore, *in silico* approaches have been successfully applied to various biochemical fields [24–26].

* Corresponding author. Tel.: +86 411 84986063.
E-mail address: yanli@dlut.edu.cn (Y. Li).
[1] Co-first author.

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.12.010

However, up to now, to our best knowledge, there is still no report of *in silico* modeling on the interactions between PKCθ and its inhibitors except the recent 2D-QSAR study using the random forest (RF) algorithm based on a series of 208 structurally diverse PKCθ inhibitors [27]. By using the RF built-in measure of the relative importance of the descriptors, an important predictor – the number of group donor atoms for H-bonds (with N and O) – has been identified to play a crucial role in PKCθ inhibitory activity. Although the results of the RF method indicate the importance of hydrogen bonding but do not precisely locate the site of hydrogen bonding. 3D-QSAR, if successful, might provide that additional information. Consequently, using an enlarging data set consisting of 263 molecules, we performed 3D-QSAR analyses, as a supplement for our previous report [27] to gain further insights into the structural and chemical features required for the PKCθ inhibitory activity.

CoMFA [28] and CoMSIA [29] are powerful and versatile tools to build and design an activity model (QSAR) for a given set of molecules in rational drug design and related applications, which have already been successfully applied to different classes of compounds. As potential-based approaches, CoMFA and CoMSIA methods correlate the biological data with the field properties generated around the aligned molecules using certain probes at a space grid, and can offer information of interaction between ligand and and putative receptors. The success of a CoMFA and CoMSIA study is usually determined by the quality of the alignment rule, or the choice of superimposition of the molecules in the study. Many successful 3D-QSAR models have been generated employing a wide variety of alignment procedures. Thus, in this work, both the ligand-based (i.e., database alignment and field fit alignment) and receptor-based (docking alignment) superimposition schemes were employed to build 3D-QSAR models, for purpose of comparing and seeking for the optimal one between the two alternatives for the 3D-QSAR model development.

In the present study, the calculated octanol/water coefficient ($\log P$) parameter was included as an additional descriptor to determine if its adding improves the statistical quality of the standard CoMFA and CoMSIA models and, if so, to what degree. Additionally, to maximize the predictive ability of the present 3D-QSAR model, we also explored several variations in establishing the CoMFA and CoMSIA models with respect to the alignment scheme, the partial charge formalism, region focusing, etc. We also employed molecular dynamics calculations to validate the reliability of the docking model. The developed models will be of help for predicting the activity of new PKCθ inhibitors.

2. Experimental

2.1. Datasets

A large dataset of 263 PKCθ inhibitors and experimental IC_{50} values were collected from the literatures [11–22]. These large series of compounds were synthesized by Wyeth Research continuous efforts. The PKCθ inhibitory activity of all these molecules was measured by the same assay on human PKCθ by employing a modified IMAP protocol from molecular devices [11]. Here, the converted molar pIC_{50} ($-\log IC_{50}$) values, spanning 4 log units, were used as the dependent variables in the QSAR regression analysis to improve the normal distribution of the experimental data points. The whole data set was divided into a training set of 197 compounds to develop the models and a test set of 66 compounds to evaluate the models in an approximate ratio of 3:1. To diversify the training and test sets, all compounds were grouped so that molecules possessing diverse functionalities and biological activity of wide range were included in both sets. The mean biological activity of the chosen training and test set molecules was 7.63 and 7.62, respectively. All information about the structures and activity of the 263 compounds are provided in Table S1 (Supplementary Information), with several representative ones together with their activity shown in Table 1.

2.2. 3D-QSAR modeling and structure alignment

The three-dimensional structures of all compounds in the dataset were obtained using the SYBYL 6.9 program (SYBYL 6.9 Tripos Inc.). Partial atomic charges calculated using the Gasteiger–Hückel method were assigned to each atom and the energy minimization of each molecule was performed using Powell method and Tripos standard force field with a distance-dependent dielectric function. The minimization was terminated when the energy gradient convergence criterion of 0.05 kcal/mol was reached or when the 1000-step minimization cycle limit was exceeded.

Molecular alignment of the compounds is a crucial step in the development of CoMFA and CoMSIA models, since the prediction accuracy of the model depends greatly on the structural alignment of the molecules. In this work, for comparison and development of the optimal models, three different alignment methods including two ligand-based, i.e., the database and field fit alignments, and one receptor-based (the docking-based alignment) superposition approaches were performed.

2.2.1. Database alignment

The superimposition of the molecules was based on an attempt to minimize the root-mean-squares differences in the fitting of the selected common framework with a template molecule. Presently, compound 174 who has the highest activity was used as the template, and all other molecules were superimposed to the common scaffold on this template using the database alignment option available in SYBYL. All molecules were aligned based on the assumption that the common substructure contributes equally to the biological activity of the molecules and can be assumed together as a constant [30]. Conformations which exhibited minimum of root-mean-squares after superimposition procedure were selected and stored in the database for the next step. Fig. 1A shows the common part for the database alignment and the resulting alignment model is shown in Fig. 1B.

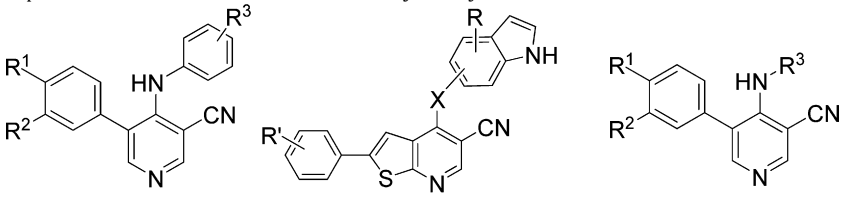
2.2.2. Field fit

The field fit procedure was used as the second alignment criterion to increase field similarity within a series of molecules. In the field fit operation, the root-mean-squares difference in the sum of steric and electrostatic interaction energies averaged across all (possibly weighted) lattice points between molecules in the training set molecule and template molecule was minimized to find the best fit. The same template molecule (compound 174) was also used in field fit alignment. All molecular conformations obtained from superimposition were used to calculate the steric and electrostatic field around the molecules to find the best field fit.

2.2.3. Molecular docking

To explore the interaction and illustrate the accurate binding model for the active site of PKCθ with ligands, molecular docking was performed by using the Surflex-dock module (V 2.51) of another advanced version of SYBYL package (X 1.1). This module utilizes a so-called "whole" molecule alignment algorithm based on the morphological similarity between the ligand and the target [31]. This docking approach aligns the ligand to a "protomol" (called also idealized ligand) in the active site of the target. For our studies, the X-ray crystal structure of PKCθ with high resolution (2.0 Å) was retrieved from RCSB Protein Data Bank (PDB entry code: 1XJD). Prior to docking, all original ligands and water molecules in 1XJD were extracted from the crystal structure. Hydrogen atoms were added in standard geometry using the Biopolymer module

Table 1
Representative chemical structures and inhibitory activity of the PKCθ inhibitor data set.



(1-5) (18-20, 25, 26) (119-121)

No.	Substituent				pIC_{50}	Ref[b]
	R^1	R^2		R^3		
1	OMe	OMe		3-Br	5.34	[11]
2	OMe	OMe		–	5.80	[11]
3[a]	OMe	OMe		3-Cl	5.41	[11]
4	OMe	OMe		3-F	5.85	[11]
5	OMe	OMe		4-Cl	5.55	[11]
	Indole isomer	X	R	R′		
18	4	NH	H	H	6.64	[12]
19	5	$NHCH_2$	H	H	5.43	[12]
20	5	NH	2-Me	H	5.74	[12]
25[a]	5	NH	H	4-CH_2-N-Me-piperazine	6.70	[12]
26	5	NH	H	4-CH_2-NMe_2	6.89	[12]
	R^1	R^2		R^3		
119	H	$O(CH_2)_2$-N-morpholine		4-indolyl	7.77	[16]
120[a]	H	$O(CH_2)_2$-N-(4-Me-piperazine)		4-indolyl	7.74	[16]
121	H	OCH_2CH_2-N-pyrrolidine		4-indolyl	6.72	[16]

[a] Test set.
[b] From corresponding reference.

as implemented in SYBYL. In this study, to generate the protomol in Surflex-dock program the Automatic Mode was adopted with two parameters that significantly affect the size and extend of the protomol generated, i.e., the threshold and the bloat values, set to 0.5 and 0, respectively. In addition, to improve the docking accuracy, the additional starting conformations per molecule were set to 10. During the docking process, the protein was considered to be rigid while the ligand molecules flexible, with all other parameters adopted by default values in the software. Finally, the molecules with the best conformations were automatically

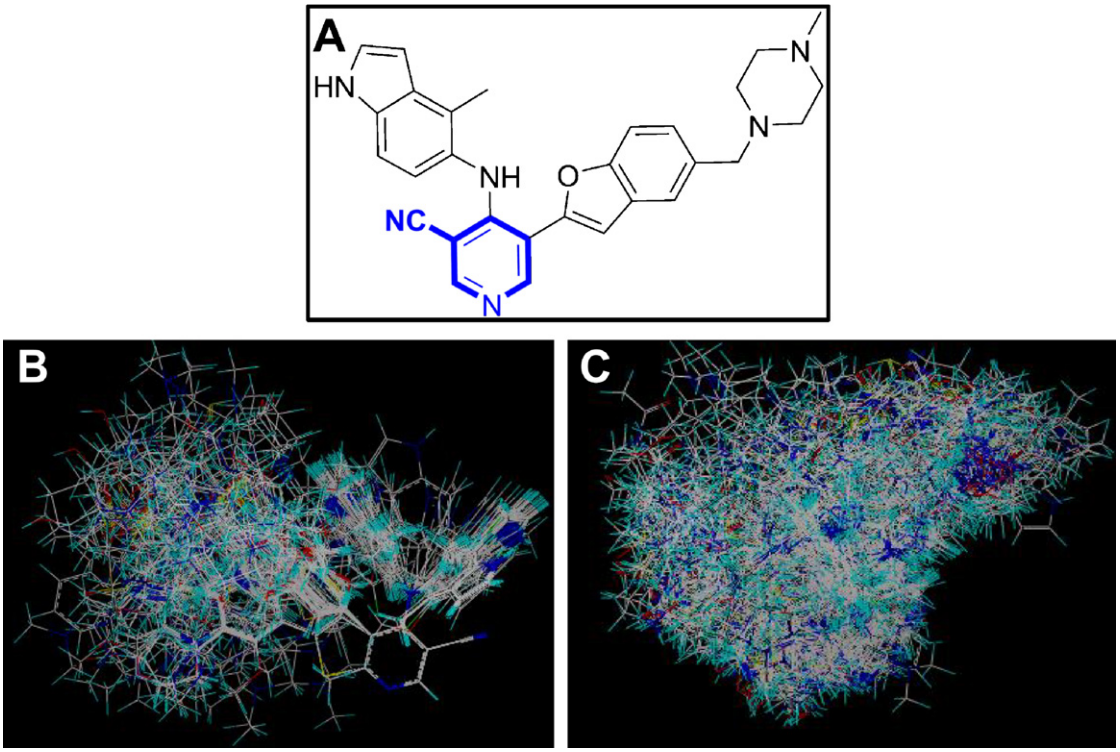


Fig. 1. Molecular alignments of PKCθ inhibitors. (A) Compound 174 used as the template molecule for database alignment. The common scaffold is shown in bold. (B) Database alignment and docking-based alignments of all the compounds in the dataset are shown in panels (B) and (C), respectively.

superimposed together to directly perform further QSAR researches. The receptor-based alignment is shown in Fig. 1C.

2.3. CoMFA and CoMSIA analyses

To generate the 3D-QSAR models, for current alignments the CoMFA and CoMSIA methods were used. All CoMFA calculations were performed with the SYBYL standard setup, i.e., the steric and electrostatic fields described by Lennard–Jones and Coulomb potentials, respectively, dielectric constant $1/r$, and cutoff 30 kcal/mol using an sp^3 carbon atom with a charge of +1 as a probe atom. After the alignment, a lattice of 2 Å was generated to surround the whole molecular aggregates. The probe atom was placed at each lattice point, and their steric and electrostatic interactions with each atom in the molecule were computed using the CoMFA standard scaling.

CoMSIA descriptors were derived according to Klebe et al. [29] with the same lattice box as employed for the CoMFA calculations, and a grid spacing of 2 Å, a C^+ probe atom with a radius of 1.0 Å as implemented in SYBYL. To evaluate the mutual distance between the probe atom and each molecule atom, a Gaussian function is used. CoMSIA similarity indices (A_F) for molecule j with atom i at grid point q are calculated by Eq. (1):

$$A^q_{F,k}(j) = -\sum \omega_{\text{probe},k}\omega_{ik}e^{-\alpha r^2_{iq}} \quad (1)$$

where k represents the steric, electrostatic, hydrophobic, H-bond donor or acceptor descriptor; $\omega_{\text{probe},k}$ is the probe atom with radius 1.0 Å, charge +1, hydrophobicity +1, H-bond donating +1, H-bond accepting +1; ω_{ik} is the actual value of the physicochemical property k of atom i; r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test molecule. The attenuation factor α was set to 0.3. In CoMSIA, the steric indices were related to the third power of the atomic radii, the electrostatic descriptors were derived from partial atomic charges, the hydrophobic fields were derived from atom-based parameters, and the hydrogen bond donor and acceptor atoms within a putative protein environment were derived from the experimental values [32].

2.4. Partial least square (PLS) analysis

For deducing the 3D-QSAR models, in PLS regression analysis the CoMFA/CoMSIA descriptors served as the independent variables and pIC_{50} values as the dependent ones. As an extension of multiple regression analysis, PLS is useful in many cases, especially in those where the number of descriptors is greater than the number of samples as is the case of CoMFA and CoMSIA. In PLS analysis, the original variables are replaced by a small set of their linear combinations [33], and these latent variables generated are used for multivariate regression, maximizing the commonality of explanatory and response variable blocks.

The predictive power of the obtained models was evaluated first by the leave-one-out (LOO) cross-validation procedure, where the cross-validated coefficient of q^2 (also called r^2_{cv}) was calculated using Eq. (2):

$$q^2 = 1 - \frac{\sum^{\text{train}}_{i=1}(y_i - \hat{y}_i)^2}{\sum^{\text{train}}_{i=1}(y_i - \bar{y}_{tr})^2} \quad (2)$$

in which the y_i, $\hat{y}_i$, and $\bar{y}_{tr}$ are the observed, predicted, and mean values of the target property (pIC_{50}) for the training set, respectively, and the $\sum^{\text{train}}_{i=1}(y_i - \hat{y}_i)^2$ is the predictive residual sum of squares (PRESS). Then the optimal number of components determined by the cross-validated PLS analysis was used to derive the final QSAR model using the compounds in the training set without cross-validation. Thereafter, a non-cross-validation analysis was carried out with the Pearson coefficient (r^2) and standard error of estimates (SEE) calculated. Finally, the CoMFA/CoMSIA results were graphically represented by field contour maps, where the coefficients were generated using the field type "StDev*Coeff".

To test the utility of the model as a predictive tool, an external set of compounds with known activity but not used in the model generation (the test set) were predicted. Based on the test set molecules, the predictive r^2, also called r^2_{pred}, was calculated using the following equation to evaluate the external predictive power of the CoMFA and CoMSIA models.

$$r^2_{\text{pred}} = 1 - \left(\frac{\text{"PRESS"}}{\text{SD}}\right) \quad (3)$$

where SD is the sum of the squared deviations between the experimental activity of the compounds in the test set and the mean activity of the training set molecules, and "PRESS" is the sum of the squared deviations between predicted and experimental activity for every compound in the test set. Using the predict property command the activity of the test set was predicted by the final models.

2.5. Molecular dynamics simulation

Simulation parameters (topologies) were developed under the CHARMM force field parameter scheme, the protein was represented using the CHARMM27 force field [34] with CMAP correlations [35]. Parameters for ligands were prepared using SwissParam (http://www.swissparam.ch). Protein atom types and charges already described in the CHARMM27 force field were assigned using the existing parameters. The new complete parameters for protein with the ligands were obtained using the PARATOOL program [36] implemented in VMD [37].

The 3D structure of the docked complex with compound 174 was analyzed in detail and served as a starting structure for MD simulations. In that complex, the missing residues were reconstructed by using SYBYL software package. The C-terminal fragment (Gln688–Phe696) was truncated. All molecular dynamics simulations were performed with VMD version 1.9 and NAMD version 2.8 [38]. The complex was placed in a 62.5 Å × 85 Å × 67 Å diameter rectangular box, surrounded by TIP3P water that has been pre-equilibrated at 300 K. Nine sodium ions were then added to yield a neutral system by placing the ions at random positions in the asymmetric unit that was ~3 Å from any crystallographic non-hydrogen atom. The systems were minimized for 3000 integration steps and equilibrated for 5 ns with 2 fs time stepping. Constant temperature (T = 310 K) was enforced using Langevin dynamics with a damping coefficient of 5 ps. Constant pressure (p = 1 atm) was enforced through the Nosé–Hoover Langevin piston method with a decay period of 200 fs and a damping time constant of 50 fs. Van der Waals interaction cutoff distances were set at 12 Å (smooth switching function beginning at 10 Å), and long-range electrostatic forces were computed using the particle-mesh Ewald (PME) [39] with a grid size of less than 1 Å. Trajectory frames were saved every 10 ps for subsequent analysis.

3. Results and discussion

3.1. Split of the training and test set

Rational selection of training and test sets is one of the crucial and challenging steps for the development of validated QSAR models. Principal component analysis (PCA), as a kind of multivariate statistical technique, can be employed for data survey due to its visualization properties which has been successfully applied to dataset split [40]. In the present work, to investigate the descriptor space, totally 929 two-dimensional descriptors were

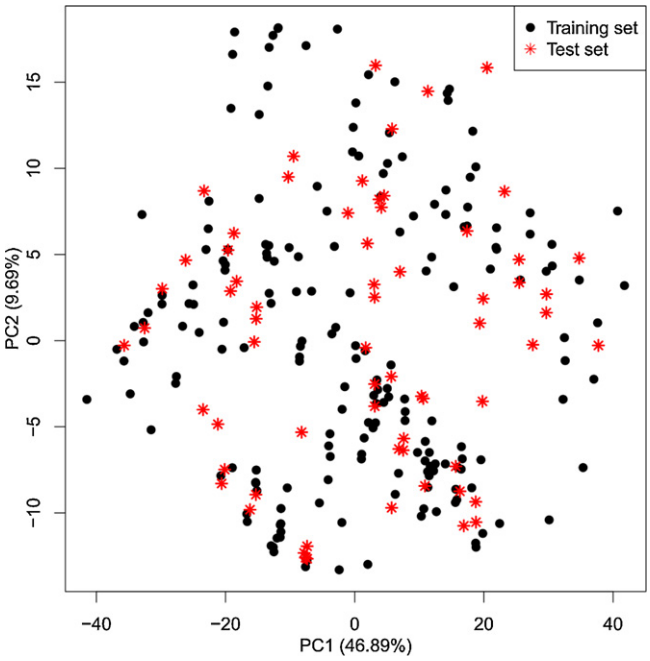


Fig. 2. Principal components analysis of the structural descriptors for PKCθ inhibitors. Black dot and red asterisks denote the training and test set compounds, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

calculated using the Dragon software package [41] for each PKCθ inhibitors. After removing zero- and near zero-variance predictors, the remaining 672 descriptors were further studied. Fig. 2 demonstrates the distribution of the compounds in the first two principal components, where the black dot denotes the training set and red asterisk stands for the test set, respectively. As seen from this figure, on the one hand, representative points of the test set are close to those of the training set and on the other hand, the training and test sets uniformly fill the whole chemical space, indicating a rational selection of the training and test compounds in the current work. In addition, it should be noted that the biological activity distribution is also a crucial element to the selection of training and test sets. In this article, the training and test sets both present a uniform distribution with average activity of 7.63 and 7.62, respectively. All these results reveal that the split of the data set is rational.

The training set was applied for the development of the model and the external prediction set was used for the assessment of the system. The training and independent test sets include 197 and 66 compounds, respectively. Table S1 (Supplementary Information) lists the structures of all the molecules.

3.2. Results of CoMFA

Out of the various CoMFA models generated, only the best one with statistically significant results is selected for further analysis of the importance of each field on the model individually. All analyses were performed with steric and electrostatic fields calculated at each grid point simultaneously. Table 2 shows the CoMFA results obtained from the three different alignments using the 197 training set compounds.

By database-based alignment I, a CoMFA model with statistical results of $q^2 = 0.503$ with five optimal components, non-cross-validated $r^2 = 0.757$, F-value = 119.038, and $r^2_{pred} = 0.568$, as well as, respectively, 59.7% and 40.3% of the steric and electrostatic contributions is obtained. The CoMFA model generated from field fit alignment II presents $q^2 = 0.485$ with nine principal components (PCs), non-cross-validated $r^2 = 0.925$, F-value = 254.827, but $r^2_{pred} = 0.187$ with 42.6% steric and 57.4% electrostatic contributions, respectively. For the docking-based alignment III, the optimal CoMFA model established yields $q^2 = 0.211$ with three PCs, non-cross-validated $r^2 = 0.575$, F-value = 86.925, and $r^2_{pred} = 0.093$, and 42.6% and 57.4% of the steric and electrostatic contributions, respectively.

Table 2
Summary of CoMFA results based on different alignments.

Parameters[a]	Alignments		
	I[b]	II[c]	III[d]
q^2	0.503	0.485	0.211
SEP	0.657	0.676	0.824
PCs	5	9	3
SEE	0.459	0.259	0.605
r^2	0.757	0.925	0.575
F	119.038	254.827	86.925
Fraction			
S	0.597	0.426	0.426
E	0.403	0.574	0.574
r^2_{pred}	0.568	0.187	0.093

[a] q^2, LOO cross-validation correlation efficient derived from SAMPLS; SEP, standard error of prediction; PCs, principal components; SEE, non-cross-validated standard error of estimate; r^2, regression coefficient in the training set; F, Fisher's F-value; r^2_{pred}, predictive r^2 in the test set; S, steric field; and E, electrostatic field.
[b] I, alignment by database fit.
[c] II, alignment by field fit.
[d] III, docking based alignment.

Thus, the model generated with alignment I with a good internal statistics ($q^2 = 0.503$) is selected as the best model to carry out further analysis. To maximize the predictive ability of the 3D-QSAR model, various efforts were also performed to explore several variations in this best CoMFA model. Firstly, the influence of different grid spacing ranging from 1.0 to 3.0 with a step of 0.5 Å was investigated, ending in a result that the grid spacing affected the model only to certain extent. If the grid spacing is too large, the grids become sparse which may cause some important molecular field information being unable to be well expressed. Presently, a smaller step size was found to increase the q^2 value from 0.503 at grid spacing 2.0 Å to 0.522 at grid spacing 1.0 Å during the modeling process. However, its r^2_{pred} falls by 0.025 than that observed from the default one (grid spacing 2.0 Å). Moreover, smaller step size dramatically increases the computation effect with much more time required to convergence. Thus in the following modeling process the grid spacing is set to 2.0 Å. Secondly, since the previous report [42] have revealed that different charges may have an effect on the CoMFA results, in the present work, four charge calculation methods including the Gasteiger–Marsili, Pullman, Gasteiger–Hückle and MMFF94 were compared individually in terms of the predictive power. Both Gasteiger–Hückle and MMFF94 illustrate the comparative but better results than the other two approaches. However, MMFF94 charge brings a relative larger SEE (0.487) than Gasteiger–Hückle charge (0.459), and therefore the Gasteiger–Hückle charge is utilized. Thirdly, usually, log P, the octanol/water coefficient related to the hydrophobicity or lipophilicity of the molecule, is found correlated to many biochemical or biological properties of the molecules. Thus, using as additional descriptors, M log P and A log P, two hydrophobic parameters calculated by using dragon software were added to the CoMFA model to see their respective contributions to the existing model. Unfortunately, as a result, none of them exhibited significant influence on the CoMFA results. In addition, region focusing is the application of weights to the lattice points in a CoMFA region to enhance or attenuate the contribution of those points to subsequent analyses. Often, the application of region focusing can enhance greatly the statistical performance. Keeping this in

Table 3
Summary of CoMSIA results based on different alignments.

Parameters[a]	Alignments		
	I[b]	II[c]	III[d]
q^2	0.622	0.546	0.319
SEP	0.574	0.628	0.767
PCs	6	5	4
SEE	0.368	0.409	0.500
r^2	0.845	0.808	0.710
F	172.91	160.502	117.673
Fraction			
S	0.137	0.099	0.146
E	0.207	0.210	0.233
H	0.202	0.180	0.188
D	0.323	0.328	0.247
A	0.131	0.183	0.185
r^2_{pred}	0.559	0.429	0.406

[a] q^2, LOO cross-validation correlation coefficient derived from SAMPLS; SEP, standard error of prediction; PCs, principal components; SEE, non-cross-validated standard error of estimate; r^2, regression coefficient in the training set; *F*, Fisher's *F*-value; r^2_{pred}, predictive r^2 in the test set; S, steric field; E, electrostatic field; H, hydrophobic field; D, H-bond donor field; and A, H-bond acceptor field.
[b] I, alignment by database fit.
[c] II, alignment by field fit.
[d] III, docking based alignment.

mind, we have also applied the region focusing that is weighted by StDev*Coefficient values ranging from 0.3 to 1.5, with grid spacing varying from 0.5 to 1.5, respectively. Consequently, the best statistical results among all models are obtained with region focusing weighted by a StDev*Coefficient value of 0.3 and a grid spacing of 1.0, with certainly the q^2 value increased to 0.571 from the original 0.503 but a r^2_{pred} decreased to 0.532 compared to the beginning value (0.568).

3.3. Results of CoMSIA

Using all five fields, i.e., the steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields, the CoMSIA analysis was performed, with results summarized in Table 3.

As seen from this table, clearly alignment I (by database alignment) has the best results. The CoMSIA model with combination of all the fields yields a cross-validated $q^2 = 0.622$ with six components, non-cross-validated $r^2 = 0.845$, *F*-value = 172.91, and $r^2_{pred} = 0.559$, suggesting both proper reliability and predictive power of the model. The contributions of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields of this model are 13.7%, 20.7%, 20.2%, 32.3%, and 13.1%, respectively. Also, alignment III presents the worst statistics. As for alignment II, though it achieves a good internal predictivity illustrated by a cross-validated $q^2 = 0.546$ with five PCs, non-cross-validated $r^2 = 0.808$, *F*-value of 160.502, but the bad external prediction results showed by the r^2_{pred} of 0.429, prevent its application on the dataset for further analysis. Thus, our further investigation is restricted only on the model derived by alignment I.

In addition, the five different descriptor fields in CoMSIA (steric, electrostatic, hydrophobic, H-bond-donor and -acceptor, represented by S, E, H, D and A, respectively) are not completely independent of each other and such dependencies of individual fields usually decrease the statistical significance and generalization of results. Thus, all 31 possible combinations of fields, in this work, were calculated with their respective q^2 value and optimum number of components determined by the SAMPLS analysis implemented in SYBYL, and the r^2_{pred} values for estimating which of the five CoMSIA fields are actually needed for a predictive model. Table 4 summarizes the representative results of different CoMSIA descriptor combinations. And Fig. 3 shows the obtained q^2 values for all 31 combinations of the five property fields considered in CoMSIA analysis. The first five models, using a single CoMSIA field, indicate that the H-bond donor field is more important than the other four fields.

As shown in Table 4, the CoMSIA model with a combination of steric, electrostatic, hydrophobic and hydrogen bond donor (SEHD) fields yields the highest cross-validated $q^2 = 0.625$ with six components, non-cross-validated $r^2 = 0.847$, *F*-value of 175.788, and r^2_{pred} of 0.589. The steric, electrostatic, hydrophobic, and hydrogen bond donor fields contributions are 15.8%, 25.3%, 23.7%, and 35.2%, respectively. Combinations of steric, electrostatic, and hydrogen bond donor (SED) fields yield a CoMSIA model with a cross-validated $q^2 = 0.616$ with six components, non-cross-validated $r^2 = 0.820$, *F*-value 144.672, and $r^2_{pred} = 0.602$. The steric, electrostatic, and hydrogen bond donor contributions are 21.7%, 34.1%, and 44.2%, respectively. As noted that although the SEHD combination gives slightly higher q^2 value than that of SED one, the predictive r^2_{pred} in the test set is lower (0.589) than the corresponding SED one (0.602). Therefore, the CoMSIA model of SED combination is chosen for further analysis. Like the CoMFA investigations, several efforts were also carried out presently with attempt to enhance the statistical performance of the model, with unfortunately none succeeded.

Besides, another interesting phenomenon is observed that among all models generated by various combinations of the CoMSIA fields (Table 4) with moderate to high internal and external

Table 4
Representative CoMSIA results for different descriptors combinations.

Parameters[a]	S	E	H	D	SD	SED	SEHD	SEHDA
q^2	0.474	0.445	0.449	0.553	0.597	0.616	0.625	0.622
SEP	0.678	0.696	0.69	0.623	0.592	0.579	0.572	0.574
PCs	6	6	4	5	5	6	6	6
SEE	0.534	0.446	0.566	0.521	0.469	0.396	0.365	0.368
r^2	0.673	0.772	0.630	0.687	0.747	0.820	0.847	0.845
F	65.213	107.476	81.569	83.953	112.683	144.672	175.788	172.91
Fraction								
S	–	–	–	–	0.338	0.217	0.158	0.137
E	–	–	–	–	–	0.341	0.253	0.207
H	–	–	–	–	–	–	0.237	0.202
D	–	–	–	–	0.662	0.442	0.352	0.323
A	–	–	–	–	–	–	–	0.131
r^2_{pred}	0.444	0.399	0.428	0.493	0.566	0.602	0.589	0.559

[a] q^2, cross-validation correlation coefficient derived from SAMPLS; SEP, standard error of prediction; PCs, principal components; SEE, non-cross-validated standard error of estimate; r^2, regression coefficient; *F*, Fisher's *F*-value; r^2_{pred}, predictive r^2 in the test set; S, steric field; E, electrostatic field; H, hydrophobic field; D, H-bond donor field; and A, H-bond acceptor field.

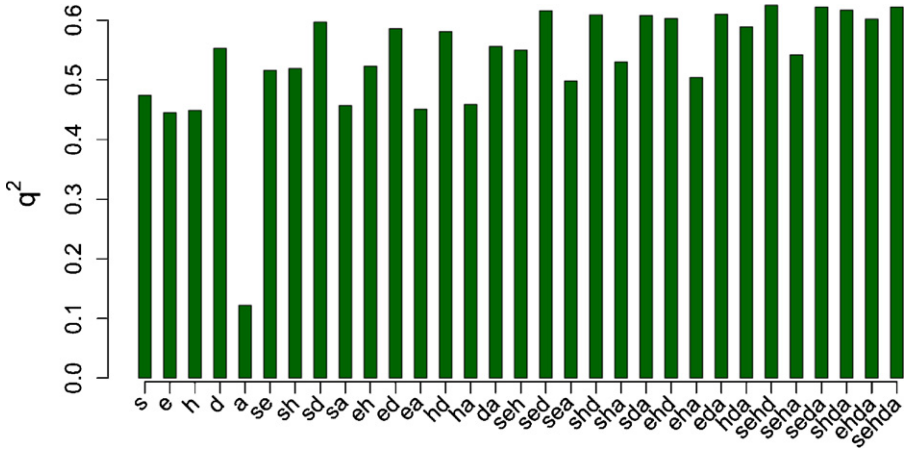


Fig. 3. Obtained q^2 values for 31 combinations of the five property fields in CoMSIA analysis. The *s*, *e*, *h*, *d*, and *a* represent the steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor property fields, respectively.

predictions, the contribution of hydrogen bond donor field is predominant over other fields, indicating the importance of H-bond donor interactions for PKCθ and its inhibitors.

3.4. Validation of the CoMFA and CoMSIA models

Any QSAR modeling should ultimately lead to statistically robust models capable of making accurate and reliable predictions of biological activity of compounds [43]. In general, validation can be divided into two parts: internal and external validations. In this work, all chosen CoMFA and CoMSIA models have undertaken strict internal and external checks. Generally speaking, in 3D-QSAR CoMFA and CoMSIA studies, models with a cross-validation q^2 value above 0.3 are sought, since at this value the probability of chance correlation is less than 5% [44], but of course a q^2 value of 0.4 is usually considered much better. In view of it, the CoMFA and CoMSIA models having $q^2 > 0.5$ are considered highly statistically significant [30]. In addition, high r^2 and F values along with low SEE values should also be considered as an important factor of reliable QSAR models. Furthermore, it has been also noted that, only accepted LOO cross-validated q^2 value is insufficient to assess the predictive power of the QSAR models [45]. Thus, presently to validate the predictive power of the 3D-QSAR models derived by the training set, other 66 inhibitors that were not included in the training set were used as an external test set.

In addition to LOO cross-validation, the cross-validation in groups using ten groups repeating the procedure 10 times was also carried out. The mean of 10 readings was given as $r^2_{cv(mean)}$. To further assess the robustness and statistical confidence of the derived models, a bootstrapping analysis for 100 runs was performed. Bootstrapping involves the generation of many new datasets from the original dataset and is obtained by randomly choosing of samples from the original dataset. The statistical calculation is performed on each of these bootstrapping samplings. The difference between the parameters calculated from the original dataset and the average of the parameters calculated from the many bootstrapping samplings is a measure of the bias of the original calculations. Table 5 illustrates all corresponding results, where the r^2_{bs} values are 0.825 and 0.870 for the best CoMFA and CoMSIA models respectively, proving the robustness of present models. Besides, the $r^2_{cv(mean)}$ values for both the two models also present highly statistical significance.

Outliers, recognized as acting by a different mechanism of action from other molecules which may be well modeled by the QSAR, are compounds that do not fit the model, or are poorly predicted by it. When performed correctly, removal of significant outliers will allow for the development of stronger and more significant models. Thus, one should perform an outlier check in the developed models. However, in the present work, after establishment of the models, none of the outliers is found in both the training and test sets in terms of percentage relative error (RE, which can be defined as: $RE = (y_{obs} - y_{pred})/y_{obs}$, where y_{obs} and y_{pred} are the observed and predicted pIC_{50} values, respectively) [46] of prediction less than 30%, though the number of the dataset is unusually large in our study. Table 6 gives the representative predicted results for the optimal CoMFA and CoMSIA models, and Table S2 (Supplementary Information) exhibits the whole results of these models. Fig. 4A and B depict the plot of observed versus predicted pIC_{50} values for the training and the test set compounds based on the optimal CoMFA and CoMSIA models, respectively. The good agreement between experimental and predicted values for the 66 test set compounds proves the reliability of the constructed CoMFA and CoMSIA models.

Table 5
Results observed from various validation methods for the best CoMFA and CoMSIA models.

Parameters[a]	CoMFA[b]	CoMSIA[c]
q^2	0.503	0.616
SEP	0.657	0.579
PCs	5	6
SEE	0.459	0.396
r^2	0.757	0.820
F	119.038	144.672
Fraction		
S	0.597	0.217
E	0.403	0.341
H	–	–
D	–	0.442
A	–	–
r^2_{bs}	0.825	0.870
SD_{bs}	0.025	0.017
$r^2_{cv(mean)}$	0.500	0.609
r^2_{pred}	0.568	0.602

[a] q^2, cross-validation correlation coefficient derived from SAMPLS; SEP, standard error of prediction; PCs, principal components; SEE, non-cross-validated standard error of estimate; r^2, regression coefficient; F, Fisher's F-value; r^2_{bs}, r^2 obtained after bootstrapping; SD_{bs}, bootstrapping standard deviation; $r^2_{cv(mean)}$, mean r^2 of cross-validation in 10 groups; r^2_{pred}, predictive r^2 in the test set; S, steric field; E, electrostatic field; H, hydrophobic field; D, H-bond donor field; and A, H-bond acceptor field.
[b] The best CoMFA model from alignment I.
[c] The best CoMSIA model from alignment I with the combination of SED fields.

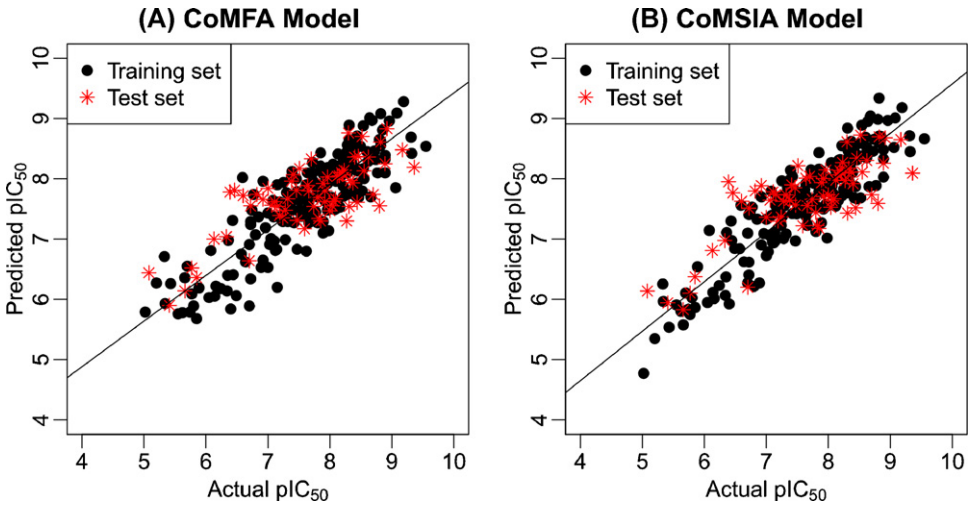


Fig. 4. The correlation plots of predicted versus actual pIC_{50} values using the training (black dot) and test (red asterisk) sets based on (A) CoMFA model and (B) CoMSIA model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 6
Representative experimental and predicted pIC_{50} values of the dataset.

No.	pIC_{50} exp.	CoMFA		CoMSIA	
		pIC_{50} pred.	RE[b]	pIC_{50} pred	RE
1	5.34	5.93	0.11	5.97	0.12
12	5.62	5.78	0.03	5.80	0.03
13[a]	5.66	6.14	0.08	5.83	0.03
32	8.42	7.71	−0.08	7.68	−0.09
40	7.92	7.15	−0.10	7.50	−0.05
44	6.96	7.19	0.03	7.09	0.02
102[a]	7.85	7.30	−0.07	7.18	−0.09
105	7.08	7.64	0.08	7.80	0.10
110	7.92	7.71	−0.03	7.92	0.00
144[a]	8.23	8.14	−0.01	8.27	0.00
145	7.72	8.27	0.07	8.15	0.06
162	8.11	8.13	0.00	8.16	0.01
165	7.46	8.00	0.07	7.43	0.00
178	7.89	8.14	0.03	8.13	0.03
182	8.82	9.08	0.03	9.34	0.06

[a] Test set.
[b] RE, relative error.

In summary, based on the 3D-QSAR investigations of 263 PKCθ inhibitors with diverse structures by CoMFA and CoMSIA methods, it is found that the database alignment produced a model with better statistics than those with field fit and docking-based superimpositions. Though several other efforts have been made, including turning the grid spacing, region focusing and investigation of different charges, no significant enhancement of the statistical performance is achieved, indicating the default parameters using the present study are best. It is true that the QSAR model should present the high performance, however, to different research systems, such fully ideal goal is not easy to achieve. But, the developed model, to some degree, must satisfy certain statistical indexes (e.g., q^2, r^2, r^2_{pred}, etc.). According to the references [46,47], we can conclude that the current models are acceptable.

3.5. 3D-QSAR contour maps

Contour plots are beneficial to identify important regions where changes in the steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor fields affect the biological activity, and may also be helpful for identifying the possible interaction sites for ligand with target. Thus, presently after consideration of both internal and external predictive powers, the best CoMFA and CoMSIA models were selected for each conformation to construct the StDev*Coeff contour maps to view the field effects on the target features. To aid in visualization, the most active compound 174 is shown with the contour maps for the optimal CoMFA and CoMSIA models (Figs. 5 and 6).

Compared with CoMFA maps, the CoMSIA approach provides more contiguous contour diagrams, which allows physicochemical properties relevant for binding to be mapped back onto the molecular structures. Fundamentally, in CoMFA, steric and electrostatic interaction energies are calculated for each molecule at the intersections of a grid embedding that molecule. However, Gaussian functions are used in CoMSIA for describing the steric, electrostatic, hydrophobic, and H-bond-donor and -acceptor similarities. This approach avoids the steep potentials next to the molecular surfaces and, thus, determines the similarity indices close to the atoms. Hence, CoMSIA contour plots denote areas within the ligands that favor or disfavor certain properties, whereas CoMFA contours highlight those areas where the ligands would interact with a possible environment. Yet, the combination of the two different approaches enables one to check the convergence of the results, or to obtain conclusions that can complement each other. In such cases, exploring the results of both approaches leads to an optimal interpretation at the 3D level of the QSAR.

For steric field, the green and yellow contours describe regions of space around the molecules where an increase in steric bulk enhances and diminishes the activity, respectively. CoMFA sterically favorable (green) contours are observed (Fig. 5A) appearing adjacent to the 4-methyl group on the indole ring of compound 174 as well as to the linker between the benzofuran and piperazine groups. Hence, an addition of bulky groups at these positions near the green contours might aid in an improved potency of the inhibitors. For example, compared to compound 187, the 5-indolyl analog 188 have three-fold reduced potency (compared with their IC_{50}) due to the lack of methyl group at the C-4 position on the indole ring. Its also the same reason for compound 31 showing larger activity than compound 28 due to the absence of methyl group at the 4-position of indole ring. Therefore, most of analogs have a methyl substituent in C-4 position of the indole ring (see Table S1). Also, it can be seen that compound 174 exhibits the largest potency, while compound 170 gives relative low activity probably due to the absence of bulk group linked to the benzofuran compared to compound 174. The same case occurs between analogs 171 and 173. These results derived by the CoMFA contour maps are all in agreement with those of CoMSIA ones (Fig. 6A).

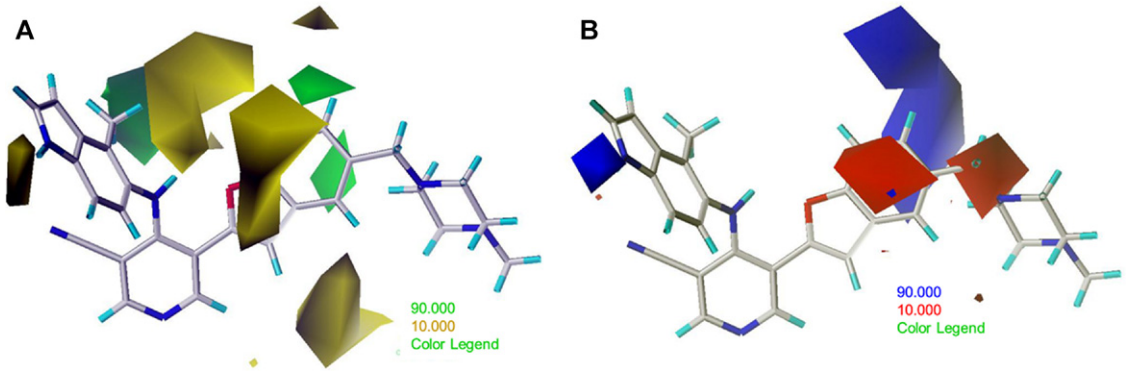


Fig. 5. CoMFA StDev*Coeff contour plots. (A) Steric contour map combined with compound 174. Green contours indicate regions where bulky groups increase activity; yellow contours indicate regions where bulky groups decrease activity. (B) Electrostatic contour map in combination with compound 174. Blue contours indicate regions where positive charges increase activity; red contours indicate regions where negative charges increase activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

In addition, it can be noted that a group of CoMFA sterically disfavored yellow contours is present on the opposite position of the indole ring, seemingly like a wall blocking the larger groups. For instance, the relative low potency of compounds like 191 and 195 compared to analogs 188 and 187 might be attributed to the presence of additional substituents at positions 1 on the indole ring, which collide with the yellow forbidden region. There is a large yellow region locating below the benzofuran group, indicating that too large substituents may decrease the activity. It could be noted that compounds 29 and 30 has reduced PKCθ inhibitory activity compared to inhibitors 18 and 17. This may because, for the former compounds there are larger substituents at 2-position of R′ group which locates in the yellow-colored region. In addition, there are two yellow-colored contours on the top of benzofuran ring indicating a large substituent in this position will decrease the inhibitory potency of PKCθ. In Fig. 6A, a CoMSIA sterically disfavored yellow contour also appears close to the CoMFA one which presents the same tendency as described in CoMFA. With respect to the C4-position on the indole ring, a steric map shows a green region around this position, but at the back of it there are also yellow fragmental fields, indicating careful substituent group selection for this region. This could explain why compound 23, the 4-ethyl analog, exhibits less activity than the 4-methyl analog 22.

In the CoMFA (Fig. 5B) and CoMSIA (Fig. 6B) electrostatic contour maps, the red and blue contours depict the favorable electronegative and electropositive charges regions, respectively. A close inspection of the electrostatic contour plots reveals that for the examined molecules electropositive groups are more preferred surrounding N-1 atom of the indole ring. For example, compound 14 having relatively electropositive group (5-indolyl) is more potent than compound 16 (with 5-indanyl instead). Additionally, it can be noted that a big blue-colored map is surrounding the H atom of the benzene ring suggest that positively charged substituents in these areas would improve the inhibitory activity. At last, the

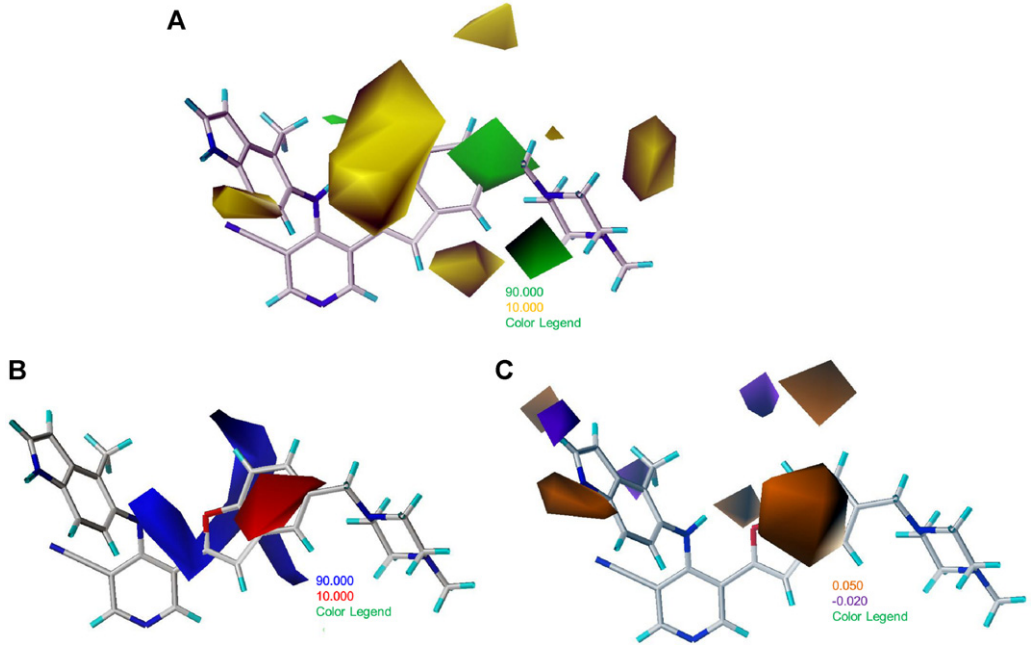


Fig. 6. CoMSIA StDev*Coeff contour plots. (A) Steric contour map combined with compound 174. Green contours indicate regions where bulky groups increase activity; yellow contours indicate regions where bulky groups decrease activity. (B) Electrostatic contour map in combination with compound 174. Blue contours indicate regions where positive charges increase activity; red contours indicate regions where negative charges increase activity. (C) H-bond donor contour map in combination with compound 174. Orange contours indicate regions where H-bond donor substituents enhance activity; purple contours indicate regions where H-bond donor substituents decrease activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

large red isopleths are present at the junction of benzofuran ring and piperazine ring. It indicates that an increase in negative charge in this region will increase the activity. For example, the unsubstituted phenyl analog 135 haves reduced activity compared to the 4,5-dimethoxyphenyl analog 128 touching the red contours (Figs. 5B and 6B).

In the CoMSIA H-bond donor field (Fig. 6C), areas colored in orange where H-bond acceptors on the receptor are predicted to favor the binding, and areas colored in purple where H-bond acceptors on the receptor are predicted to reduce the binding. As shown in Fig. 6C, a large orange region near the N-1 of the indole ring indicates the favor of the position for a H-bond donor substituent, which is consistent with the corresponding experimental results [17]. The 1-Me-5-indolyl analog 191 has greatly decreased activity illustrating the importance of the proton at N-1 of the indole ring. In addition, the 1,4-diMe analog 134 and 1,4-diMe-5-indolyl analog 195 also present lower activity than their counterparts due to the replacement of H of -NH with a methyl group on the indole ring. In addition, the large yellow areas (Figs. 5A and 6A) and the blue contour (Fig. 5B) around the N-1 of indole at the same position suggest that a small, electropositive group could benefit from a hydrogen bond donor, which might increase the activity. There is another orange-colored contour at the distal of the benzofuran ring, illustrating the tendency that at this location an introduction of H-bond donors will increase the PKCθ inhibition. It is consistent with the fact that the 4-aminopiperidine-substituted compound 66 and the 4-alcohol-substituted analog 69 result in a significant increase in inhibitory activity of PKCθ over that unsubstituted molecule 54, which indicates that the distal position of the benzofuran plays a crucial role in PKCθ activity, especially, a H-bond donor substituent (compounds 53–70 in Table S1). Thus, the combination of the large H-bond donor contributions (Table 5 and Fig. 3) with the CoMFA and CoMSIA contour results illustrates that, in the position of N-1 on the indole ring and the distal N of the benzofuran ring of compound 174, the H-bond donors play a crucial role in PKCθ inhibition, which is in good agreement with our previous report [27].

3.6. Molecular docking

To validate the ability of the present docking protocol, reproducing the PKCθ–staurosporine binding structure [48] was performed, prior to docking of all inhibitors to the active center of PKCθ receptor. As the crystallographic conformation of the PKCθ–staurosporine complex is available at the protein data bank (PDB entry code: 1XJD), we directly extracted the staurosporine crystal structure from the complex and took it as the starting conformation to perform the docking simulation. As a result, the root-mean-square deviation between the conformations of the experimental and redocked PKCθ–staurosporine is 0.64 Å (Fig. 7), suggesting that Surflex-dock, in this work, is able to generate the binding model for PKCθ inhibitors and that the parameter set for the Surflex-dock simulation are appropriate for performing the present research.

The docking results could illustrate us the interaction models between inhibitors and PKCθ. In the present work, the most potent molecule 174 is taken as the representative to elucidate this point in detail. As illustrated in Fig. 8, the compound 174 is located at the binding site by interaction with amino acid residues within the receptor through H-bond and hydrophobic interactions. The ligand 174 is represented as carbon-chain in light blue and PKCθ as carbon-chain in bottle-green. The key amino acid residues (3.5 Å around the ligand atoms) are shown in line mode. It can be noted in this figure, a large hydrophobic cavity is formed among Leu461, Tyr460, Glu459, Met458 and Ala407, which accommodate a hydrophobic substituent. Herein, the pyridine ring is located in this hinge region. And another hydrophobic center is formed

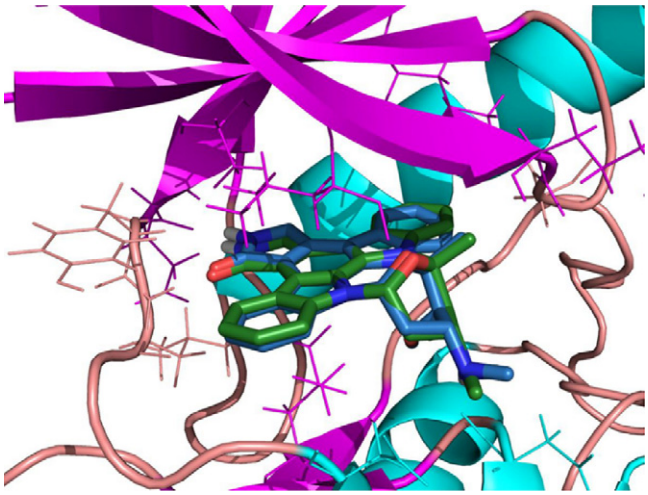


Fig. 7. Conformational comparison of from docking (represented as carbon-chain in light-blue) and the original ligand (represented as carbon-chain in bottle-green) in the crystal structure (PDB ID: 1XJD).

between Leu386, Val394, Gly387, Gly389 and Lys388, which may form hydrophobic interactions with the indole ring of the compound 174. Thus a large substituent in this position may increase the inhibitory activity of PKCθ. This observation is supported by our CoMFA and CoMSIA results, which presents a bulk favored green map in this position. The third hydrophobic interaction is observed at the bottom of the benzofuran ring of compound 174 including the Gly464, Asp465, Leu511, Tyr468, Asn509, Ala521 and Asp522. Among them, Tyr468 can form pi–pi interaction with the piperazine ring. Besides the important role of hydrophobic interactions, several H-bond interactions formed between the ligand and PKCθ also attract our attention. For example, it can be noted that the pyridine nitrogen and the backbone NH of the hinge region residue Leu461 [11,14,48] with the bond length of 2.2 Å. In the previous report [48], it has been shown that this amino acid residue plays an important

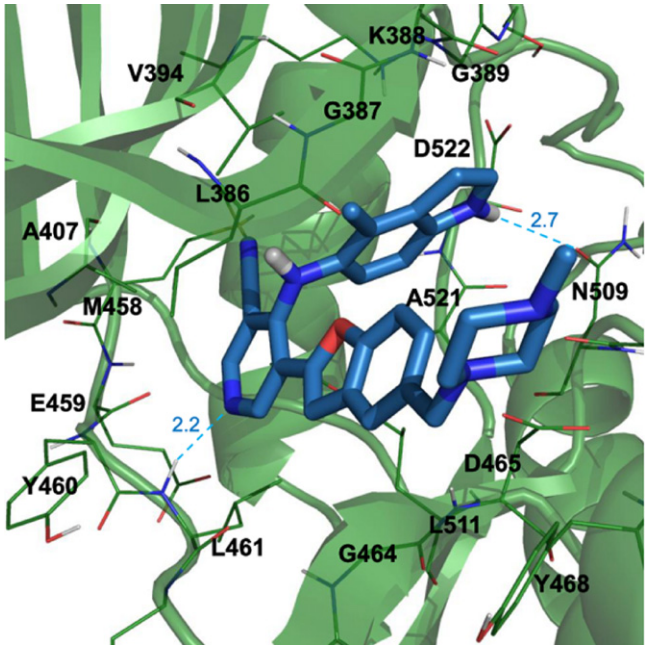


Fig. 8. Molecular models of compound 174 in the binding site of PKCθ. The dashed lines show the formation and distance of the hydrogen bonds. Key amino acid residues are represented as carbon-chain in bottle-green and compound 174 as carbon-chain in light-blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

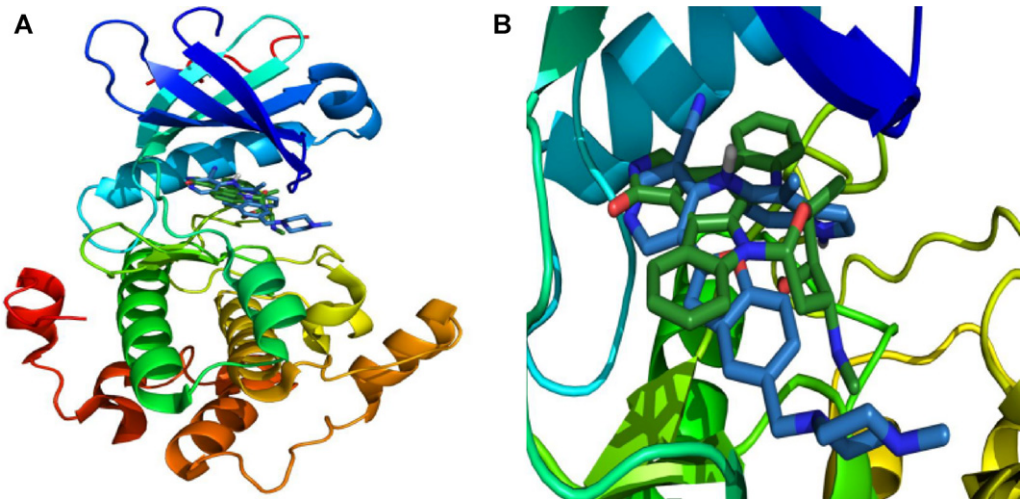


Fig. 9. The overlay of the staurosporine and most active inhibitor 174 in the binding pocket. Staurosporine is colored as carbon-chain in light-blue and inhibitor 174 as carbon-chain in bottle-green. (A) Distant view of the alignment. (B) Nearby view in the active center of PKCθ receptor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

role in ligand bonding. In addition, the indole-NH as a hydrogen bond donor interacts with Asn509 with the bond length of 2.7 Å, which is consistent with the fact our 3D-QSAR modeling suggests that the in the N-1 position of the indole ring one should introduce a H-bond donor to improve the PKCθ inhibition since there is a electropositive favored contour around the group (Fig. 5B). It is also noticed that a bulk disfavored yellow map is locating at the region as shown in Figs. 5A and 6A. Therefore, the substituent with small and electropositive group in this region is required to increase the inhibitory activity of PKCθ.

For validating the reliability of the docking simulation, the conformation of the most potent compound 174 has been superimposed with PKCθ–staurosporine complex [48] in the binding pocket. As shown in Fig. 9, the inhibitor of staurosporine and compound 174 overlay well and they both exhibit similar poses. Therefore, it can be concluded that the identified conformation of compound 174 is rational. Consequently, it is the combination of relatively high prediction accuracy and good consistence with the docking simulations that indicates the developed optimal 3D-QSAR models should be reasonable.

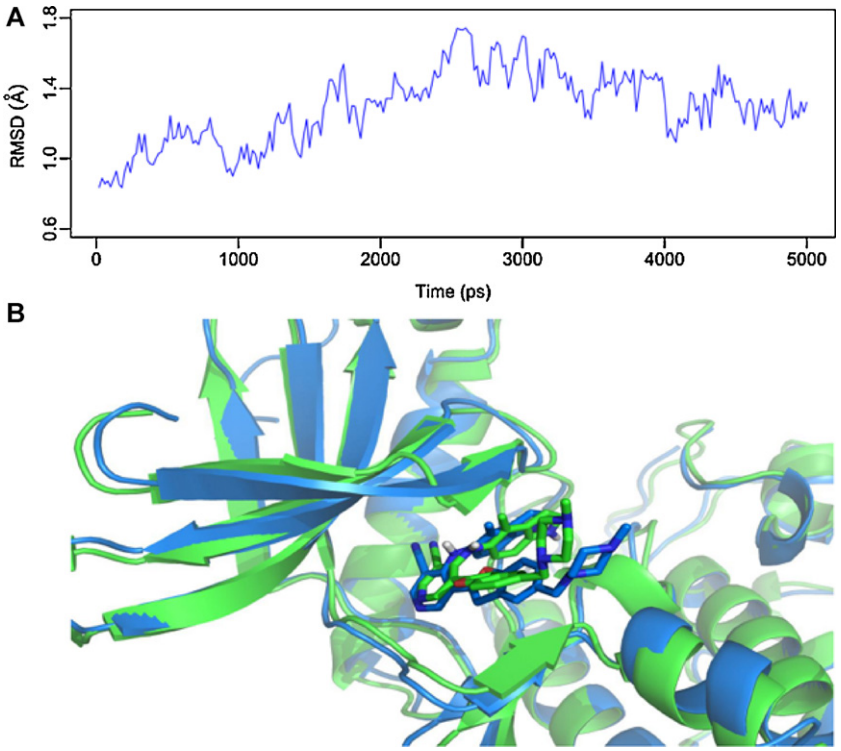


Fig. 10. MD simulation result. (A) Plot of the RMSD of the docked complex versus the MD simulation time in the MD-simulated structure. (B) View of the aligned backbone atoms of the average structure of the MD simulation (green) and the initial structure (blue) for compound 174-PKCθ complex. Compound 174 is represented as carbon-chain in blue and green for the initial complex and the average complex, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3.7. Molecular dynamics simulation

For validating the reliability of the docking model, the docked complex PKCθ-inhibitor 174 has been performed by a 5 ns simulation, and the root mean square deviation (RMSD) of backbone is presented in Fig. 10A. After 2 ns, the RMSD tends to be stable (at about 1.4 Å), indicating stabilization of the complex. Thus, we extracted the average structure in the last 3 ns illustrated in Fig. 10B, where the original structure is colored in blue and the average structure in green. It is noted that there is no significant difference (RMSD of superimposition is about 1.3 Å) between the two structures (average structure from MD calculations and the docked structure of the complex), suggesting the reliability of the current docked model is.

4. Conclusions

Presently, two predictive CoMFA and CoMSIA 3D-QSAR models based on up to now the largest dataset of 263 structurally diverse of PKCθ inhibitors with wide range of pIC_{50} were developed successfully indicated by significant statistical parameters in terms of internal and external validations. The consistency among 3D-QSAR, molecular docking and molecular dynamics calculations confirms that the current *in silico* modeling is reliable. The resultant contour maps produced by the best 3D-QSAR models illustrate useful information about the intermolecular interactions of the inhibitors with their surrounding environment. Taking the most potent compound 174 as a reference, the 3D-QSAR studies reveal that (1) the introduction of a small H-bond donor group in N-1 of the indole ring increases the potency; (2) in the C-4 position of the indole head of the inhibitor, a substituent with proper volume (like methyl) is beneficial to PKCθ inhibition; (3) an electropositive substituent at the distal position of the benzofuran may be required (e.g. $-NH_2$). Also, Leu461 and Asn509 have been identified as the key amino acid residues to form H-bond interaction with the ligand compound. The adopted models and included above information will be of help for designing and optimization of novel molecules with improved spectrum of PKCθ inhibition.

Acknowledgments

Thanks for the financial support given by the National Natural Science Foundation of China (Grant No. 10801025 and No. 30973590), the National High Technology Research and Development Program ("863") of China (No. 2009AA02Z205), and the help of Hai-Bin Luo from School of Pharmaceutical Sciences, Sun Yat-Sen University. G.H. Li appreciates the supports from the National Natural Science Foundation of China (31070641), the National Key Basic Research Development Program (2012CB721000) and "Hundreds Talents Program" of the Chinese Academy of Sciences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.12.010.

References

[1] Z. Sun, C.W. Arendt, W. Ellmeier, E.M. Schaeffer, M.J. Sunshine, L. Gandhi, et al., PKCθ is required for TCR-induced NF-κB activation in mature but not immature T lymphocytes, Nature 404 (2000) 402–407.
[2] S. Osada, K. Mizuno, T. Saido, K. Suzuki, T. Kuroki, S. Ohno, A new member of the protein kinase C family, nPKC theta, predominantly expressed in skeletal muscle, Mol. Cell. Biol. 12 (1992) 3930–3938.
[3] G. Baier, D. Telford, L. Giampa, K. Coggeshall, G. Baier-Bitterlich, N. Isakov, et al., Molecular cloning and characterization of PKC theta, a novel member of the protein kinase C (PKC) gene family expressed predominantly in hematopoietic cells, J. Biol. Chem. 268 (1993) 4997–5004.
[4] J. Chang, Y. Xu, M. Raychowdhury, J. Ware, Molecular cloning and expression of a cDNA encoding a novel isoenzyme of protein kinase C (nPKC). A new member of the nPKC family expressed in skeletal muscle, megakaryoblastic cells, and platelets, J. Biol. Chem. 268 (1993) 14208–14214.
[5] K. Hayashi, A. Altman, Protein kinase C theta (PKCθ): a key player in T cell life and death, Pharmacol. Res. 55 (2007) 537–544.
[6] A. Healy, E. Izmailova, M. Fitzgerald, R. Walker, M. Hattersley, M. Silva, et al., PKCθ-deficient mice are protected from Th1-dependent antigen-induced arthritis, J. Immunol. 177 (2006) 1886–1893.
[7] S. Salek-Ardakani, T. So, B. Halteman, A. Altman, M. Croft, Protein kinase Cθ controls Th1 cells in experimental autoimmune encephalomyelitis, J. Immunol. 175 (2005) 7635–7641.
[8] B. Marsland, T. Soos, G. Späth, D. Littman, M. Kopf, Protein kinase Cθ is critical for the development of in vivo T helper (Th)2 cell but not Th1 cell responses, J. Exp. Med. 200 (2004) 181–189.
[9] N. Berg-Brown, M. Gronski, R. Jones, A. Elford, E. Deenick, B. Odermatt, et al., PKCθ signals activation versus tolerance in vivo, J. Exp. Med. 199 (2004) 743–752.
[10] D.H. Boschelli, Small molecule inhibitors of PKCθ as potential antiinflammatory therapeutics, Curr. Top. Med. Chem. 9 (2009) 640–654.
[11] D.C. Cole, M. Asselin, A. Brennan, R. Czerwinski, J.W. Ellingboe, L. Fitz, et al., Identification, characterization and initial hit-to-lead optimization of a series of 4-arylamino-3-pyridinecarbonitrile as protein kinase C theta (PKCθ) inhibitors, J. Med. Chem. 51 (2008) 5958–5963.
[12] D.H. Boschelli, B.Q. Wu, A.C.B. Sosa, J. Chen, M. Asselin, D.C. Cole, et al., Synthesis and PKCθ inhibitory activity of a series of 4-(indol-5-ylamino)thieno 2,3-b pyridine-5-carbonitriles, Bioorg. Med. Chem. Lett. 18 (2008) 2850–2853.
[13] L.N. Tumey, D.H. Boschelli, J. Lee, D. Chaudhary, 2-Alkenylthieno-2,3-b-pyridine-5-carbonitriles: potent and selective inhibitors of PKCθ, Bioorg. Med. Chem. Lett. 18 (2008) 4420–4423.
[14] L.N. Tumey, N. Bhagirath, A. Brennan, N. Brooijmans, J. Lee, X.K. Yang, et al., 5-Vinyl-3-pyridinecarbonitrile inhibitors of PKCθ: optimization of enzymatic and functional activity, Bioorg. Med. Chem. 17 (2009) 7933–7948.
[15] B.Q. Wu, D.H. Boschelli, J.L. Lee, X.K. Yang, D. Chaudhary, Second generation 4-(4-methyl-1H-indol-5-ylamino)-2-phenylthieno-2,3-b pyridine-5-carbonitrile PKCθ inhibitors, Bioorg. Med. Chem. Lett. 19 (2009) 766–769.
[16] R.G. Dushin, T. Nittoli, C. Ingalls, D.H. Boschelli, D.C. Cole, A. Wissner, et al., Synthesis and PKCθ inhibitory activity of a series of 4-indolylamino-5-phenyl-3-pyridinecarbonitriles, Bioorg. Med. Chem. Lett. 19 (2009) 2461–2463.
[17] D.H. Boschelli, D. Wang, A.S. Prashad, J. Subrath, B. Wu, C. Niu, et al., Optimization of 5-phenyl-3-pyridinecarbonitriles as PKCθ inhibitors, Bioorg. Med. Chem. Lett. 19 (2009) 3623–3626.
[18] J. Subrath, D. Wang, B.Q. Wu, C.S. Niu, D.H. Boschelli, J. Lee, et al., C-5 substituted heteroaryl 3-pyridinecarbonitriles as PKCθ inhibitors: part I, Bioorg. Med. Chem. Lett. 19 (2009) 5423–5425.
[19] A.S. Prashad, D. Wang, J. Subrath, B.Q. Wu, M. Lin, M.Y. Zhang, et al., C-5 substituted heteroaryl-3-pyridinecarbonitriles as PKCθ inhibitors: part II, Bioorg. Med. Chem. Lett. 19 (2009) 5799–5802.
[20] C.S. Niu, D.H. Boschelli, L.N. Tumey, N. Bhagirath, J. Subrath, J. Shim, et al., First generation 5-vinyl-3-pyridinecarbonitrile PKCθ inhibitors, Bioorg. Med. Chem. Lett. 19 (2009) 5829–5832.
[21] J. Shim, C. Eid, J. Lee, E. Liu, D. Chaudhary, D.H. Boschelli, Synthesis and PKCθ inhibitory activity of a series of 5-vinyl phenyl sulfonamide-3-pyridinecarbonitriles, Bioorg. Med. Chem. Lett. 19 (2009) 6575–6577.
[22] D.H. Boschelli, J. Subrath, C.S. Niu, B.Q. Wu, Y. Wang, J. Lee, et al., Optimization of 5-vinylaryl-3-pyridinecarbonitriles as PKCθ inhibitors, Bioorg. Med. Chem. Lett. 20 (2010) 1965–1968.
[23] D. Agrafiotis, D. Bandyopadhyay, J. Wegner, H. van Vlijmen, Recent advances in chemoinformatics, J. Chem. Inf. Model. 47 (2007) 1279–1293.
[24] K. Roy, S. Paul, Docking and 3D-QSAR studies of acetohydroxy acid synthase inhibitor sulfonyl urea derivatives, J. Mol. Model. 16 (2010) 951–964.
[25] Y. Wang, Y. Li, S. Yang, L. Yang, Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach, J. Chem. Inf. Model. 45 (2005) 750–757.
[26] K. Roy, S. Paul, Docking and 3D-QSAR studies of protoporphyrinogen oxidase inhibitor 3H-pyrazolo[3,4-d][1,2,3]triazin-4-one derivatives, J. Mol. Model. 16 (2010) 137–153.
[27] M. Hao, Y. Li, Y. Wang, S. Zhang, Prediction of PKCθ inhibitory activity using the random forest algorithm, Int. J. Mol. Sci. 11 (2010) 3413–3433.
[28] D. Richard, E. David, D. Jeffrey, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[29] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4146.
[30] S.S. Chaudhaery, K.K. Roy, A.K. Saxena, Consensus superiority of the pharmacophore-based alignment, over maximum common substructure (MCS): 3D-QSAR studies on carbamates as acetylcholinesterase inhibitors, J. Chem. Inf. Model. 49 (2009) 1590–1601.
[31] A.N. Jain, Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine, J. Med. Chem. 46 (2003) 499–511.
[32] G. Klebe, T. Mietzner, F. Weber, Methodological developments and strategies for a fast flexible superposition of drug-size molecules, J. Comput. Aided Mol. Des. 13 (1999) 35–49.

[33] S. Wold, A. Ruhe, H. Wold, W. Dunn III, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, SIAM J. Sci. Stat. Comput. 5 (1984) 735–743.
[34] A.D. MacKerell, D. Bashford, R.L. Bellott, J.D. Dunbrack, M.J. Evanseck, Field, et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[35] A.D. Mackerell, M. Feig, C.L. Brooks, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, J. Comput. Chem. 25 (2004) 1400–1415.
[36] J. Saam, I. Ivanov, M. Walther, H.-G. Holzhütter, H. Kuhn, Molecular dioxygen enters the active site of 12/15-lipoxygenase via dynamic oxygen access channels, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 13319–13324.
[37] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[38] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al., Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[39] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[40] Y. Ren, H. Liu, X. Yao, M. Liu, Prediction of ozone tropospheric degradation rate constants by projection pursuit regression, Anal. Chim. Acta 589 (2007) 150–158.
[41] Talete srl, DRAGON for Windows (Software for Molecular Descriptor Calculations), Version 5.4. http://www.talete.mi.it/, 2006.
[42] K.C. Tsai, Y.C. Chen, N.W. Hsiao, C.L. Wang, C.L. Lin, Y.C. Lee, et al., A comparison of different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies, Eur. J. Med. Chem. 45 (2010) 1544–1551.
[43] K. Roy, G. Ghosh, QSTR with extended topochemical atom (ETA) indices. 12. QSAR for the toxicity of diverse aromatic compounds to *Tetrahymena pyriformis* using chemometric tools, Chemosphere 77 (2009) 999–1009.
[44] A. Nayyar, A. Malde, R. Jain, E. Coutinho, 3D-QSAR study of ring-substituted quinoline class of anti-tuberculosis agents, Bioorg. Med. Chem. 14 (2006) 847–856.
[45] A. Golbraikh, A. Tropsha, Beware of q^2!, J. Mol. Graph. Model. 20 (2002) 269–276.
[46] M. Shahlaei, A. Fassihi, L. Saghaie, Application of PC-ANN and PC-LS-SVM in QSAR of CCR1 antagonist compounds: A comparative study, Eur. J. Med. Chem. 45 (2010) 1572–1582.
[47] A. Tropsha, P. Gramatica, V. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci. 22 (2003) 69–77.
[48] Z. Xu, D. Chaudhary, S. Olland, S. Wolfrom, R. Czerwinski, K. Malakian, et al., Catalytic domain crystal structure of protein kinase C-θ (PKCθ), J. Biol. Chem. 279 (2004) 50401–50409.