

# Structure-based CoMFA and CoMSIA study of indolinone inhibitors of PDK1

Kirandeep Kaur · Tanaji T. Talele

Received: 30 April 2008 / Accepted: 8 August 2008 / Published online: 9 September 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Phosphoinositide-dependent protein kinase-1 (PDK1) is a Ser/Thr kinase which phosphorylates and activates members of the AGC kinase group known to control processes such as tumor cell growth, protection from apoptosis, and tumor angiogenesis. In this paper, CoMFA and CoMSIA studies were carried out on a training set of 56 conformationally rigid indolinone inhibitors of PDK1. Predictive 3D QSAR models, established using atom fit alignment rule based on crystallographic-bound conformation, had cross-validated ( $r_{cv}^2$ ) values of 0.738 and 0.816 and non-cross-validated ( $r_{ncv}^2$ ) values of 0.912 and 0.949 for CoMFA and CoMSIA models, respectively. The predictive ability of the CoMFA and CoMSIA models was determined using a test set of 14 compounds, which gave predictive correlation coefficients ( $r_{pred}^2$ ) of 0.865 and 0.837, respectively. Structure-based interpretation of the CoMFA and CoMSIA field properties provided further insights for the rational design of new PDK1 inhibitors.

**Keywords** 3D QSAR · CoMFA · CoMSIA · Indolinone · PDK1

## Introduction

The signal transduction pathways, which mediate the cell's ability to respond to chemical and physical stimuli, can

roughly be divided into two major types of protein kinases, the tyrosine protein kinases and the serine–threonine protein kinases. Many of the serine–threonine protein kinases belong to one of two major families, the AGC family and MAPK family. The AGC protein kinase family includes such members as the cAMP dependent protein kinase (PKA), cGMP dependent protein kinase and protein kinase C (PKC), phosphoinositide-dependent protein kinase-1 (PDK1), and protein kinase B (PKB or Akt), which are all known to be activated by small second-messenger molecules generated either by G-protein coupled receptors (GPCRs) or by receptor tyrosine kinases (RTK's) [1].

A number of well-characterized insulin and growth factor peptides bind and activate RTK's [2], an event that subsequently leads to the activation of phosphatidylinositol 3-kinase (PI3K) [3]. In turn, PI3K phosphorylates phosphatidylinositol 4-phosphate [PI(4)P] and PI(4,5)P<sub>2</sub> to generate the second messengers PI(3,4)P<sub>2</sub> and PI(3,4,5)P<sub>3</sub>, for promoting cell survival through the PDK1/PKB-mediated pathway. Phosphatase and tensin homolog (PTEN) down regulates this pathway by acting as the phosphatidylinositol 3-phosphatase that inactivates the second messengers PI(3,4)P<sub>2</sub> and PI(3,4,5)P<sub>3</sub> [4].

PDK1 activates PKB by phosphorylation of its T-loop (Thr308 in PKB $\alpha$ ) [5, 6]. This phosphorylation creates a docking motif that specifically interacts with a pocket on the N-terminal lobe of the PDK1 kinase domain [termed PDK1 interacting fragment (PIF) pocket] [7, 8], bringing PDK1 together with its substrate and enabling PDK1 to phosphorylate PKB in its T-loop, thereby activating it. A significant number of human cancers possess elevated PI(3,4,5)P<sub>3</sub> levels due to mutations in a number of genes that regulate the production and degradation of PI(3,4,5)P<sub>3</sub>. One of the most frequently found mutations occurs in the PI(3,4,5)P<sub>3</sub> 3-phosphatase, resulting in constitutive

**Electronic supplementary material** The online version of this article (doi:10.1007/s10822-008-9235-2) contains supplementary material, which is available to authorized users.

K. Kaur · T. T. Talele (✉)  
Department of Pharmaceutical Sciences, College of Pharmacy  
and Allied Health Professions, St. John's University,  
Jamaica, NY 11439, USA  
e-mail: talelet@stjohns.edu

activation of PKB, thought to be a major contributor to the proliferation and the survival of such tumor cells [9]. Thus PDK1 and PKB are pivotable enzymes in the cell survival pathway, utilizing a combination of membrane targeting and a series of serine–threonine phosphorylation reactions to transducer cell growth and survival messages between cell compartments.

PDK1 was first purified from tissue extracts as an enzyme that could phosphorylate the T-loop of PKB $\alpha$  (Thr308) in the presence of PI(3,4,5)P<sub>3</sub>. It is a 556 amino acid long enzyme, possessing a kinase domain at its N-terminus (residues 70–359) and a pleckstrin homology (PH) domain at its C-terminus (residues 459–550), and interacting with PI(3,4,5)P<sub>3</sub> with high affinity. According to the high-resolution (2.0 Å) crystal structure of the PDK1 kinase domain<sup>8</sup>, PDK1 assumes the classical bilobal kinase fold in a manner analogous to PKA. As in other protein kinases [10], the PDK1 structure shows the  $\alpha$ C-helix (residues 124–136) to be a key element in the kinase core. One turn of the PDK1  $\alpha$ C-helix (residues 129–131), links together the N-terminal lobe and the active site. While Arg129 points towards the T-loop and forms two hydrogen bonds with the phosphorylated Ser241, Arg131 forms two hydrogen bonds with the sulfate in the phosphate pocket. Glu130 coordinates Lys111 which in turn, forms a hydrogen bond with the  $\alpha$ -phosphate of the bound ATP. Tyr126 is an additional residue that contributes with a third hydrogen bond between the  $\alpha$ C-helix and the phosphorylated Ser241. Val124 and Val127 are residues on the  $\alpha$ C-helix that are involved in formation of the PIF-pocket. The  $\alpha$ C-helix provides a structural link between the hydrophobic and phosphopeptide pockets and the phosphoserine in the T-loop. Phosphorylation at the T-loop of PDK1 plays a key role in stabilizing and positioning the  $\alpha$ C-helix, thereby modulating the overall conformation of the hydrophobic and phosphate pockets [11, 12].

The best characterized cellular substrate for PDK1 is Akt (PKB) [13]. All the three isoforms of Akt (Akt1, Akt2, and Akt3) are activated by PDK1 through phosphorylation of a regulatory threonine residue within the activation loop (e.g., Thr308 of Akt1), thereby, initiating a conformational change that activates the protein [14]. PDK1 has also been implicated in the activation of other potential cancer targets such as protein kinase C [14, 15]. Hence, inhibitors of PDK1 have the potential for acting as anticancer agents since they would be expected to suppress the activation of PKB to limit cell growth and to induce apoptosis in cancer cells possessing elevated levels of PI(3,4,5)P<sub>3</sub>.

To the best of our knowledge, no QSAR investigation for PDK1 inhibitors has been reported to this date. In the present study, the mechanism of intermolecular interaction between the most potent compound **30** and PDK1 was studied with the help of comparative molecular field

analysis (CoMFA) [16, 17] and comparative molecular similarity indices analysis (CoMSIA) [18] methodologies using the partial least squares (PLS) method [19, 20]. CoMFA analysis involves the alignment of molecules in a structurally and pharmacologically reasonable manner on the basis of the assumption that each compound acts via a common macromolecular target binding site. In this method, it is possible to predict the biological activity of molecules and represent the relationships between molecular properties (steric and electrostatic) and biological activity in the form of contour maps.

The more recently reported CoMSIA approach calculates similarity indices in the space surrounding each of the aligned molecules in the dataset. CoMSIA is believed to be less affected by changes in molecular alignment and provides smooth and interpretable contour maps as a result of employing Gaussian type distance dependence with the molecular similarity indices it uses. Furthermore, in addition to steric and electrostatic fields of CoMFA, CoMSIA defines explicit hydrophobic and hydrogen bond donor and acceptor fields. Such 3D QSAR models would be of great help in a drug development program since the activity of new analogs could be quantitatively predicted before attempting their synthesis and testing.

## Materials and methods

### Dataset for analysis

All indolinone derivatives and their in vitro biological activity (IC<sub>50</sub>) were taken from the literature [21, 22]. The IC<sub>50</sub> values in  $\mu$ M units were converted to the corresponding pIC<sub>50</sub> values using the formula (pIC<sub>50</sub> =  $-\log$  IC<sub>50</sub>) (Table 1). The pIC<sub>50</sub> values of the training set described in this manuscript span approximately 3.93 log units. A total of 70 indolinone compounds were used as a data set, of which a set of 14 compounds were chosen as the test set, while the remaining 56 compounds were treated as a training set. The structures of all of the compounds are shown in Table 1. The selected test set represented a range of inhibitory activity similar to that of a training set and was used to evaluate the predictive power of the CoMFA and CoMSIA models.

### Molecular modeling and alignment

All molecular modeling and calculation studies were performed using the SYBYL version 7.2 [23] on a Dell Precision 470n workstation with the RHEL 4.0 operating system. The position of each atom is important to CoMFA and CoMSIA because the descriptors are calculated based on the 3D space grid. Hence, the method to determine the

**Table 1** Structures of the training and test set of compounds

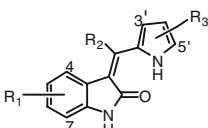
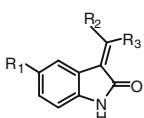
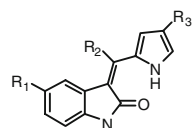
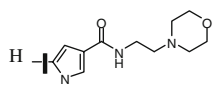
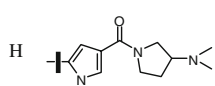
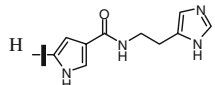
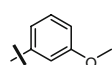
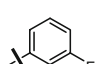
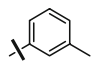
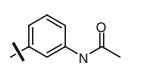
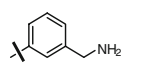

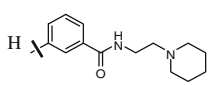






















<div></div> 1-31, 57-63				<div></div> 32-39, 64-65				<div></div> 40-56, 66-70			
Compd	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pIC <sub>50</sub>	Compd	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pIC <sub>50</sub>		
Training Set											
1	H	H	H	-0.25							
2	4-Me	H	H	0.29							
3	5-Me	H	H	-0.43	37	NHCONH <sub>2</sub>	H	<div></div>	1.54		
4	7-Me	H	H	-1.41							
5	4-OH	H	H	0.00	38	NHCONH <sub>2</sub>	H	<div></div>	1.51		
6	5-OH	H	H	0.47							
7	6-OH	H	H	0.05	39	NHCONH <sub>2</sub>	H	<div></div>	1.85		
8	5-OH	Me	H	1.10							
9	5-OH	H	3'-Me	0.55	40	NHCONH <sub>2</sub>	H	<div></div>	1.72		
10	5-OH	H	4'-Me	0.55							
11	5-OH	H	5'-Et	-0.36	41	NHCONH <sub>2</sub>	H	<div></div>	1.29		
12	5-OH	H	3',5'-Me	-0.04							
13	5-SO <sub>2</sub> NH <sub>2</sub>	H	H	0.54	42	NHCONH <sub>2</sub>	H	<div></div>	1.43		
14	5-COOMe	H	H	0.17							
15	5-COOH	H	H	0.59	43	NHCONH <sub>2</sub>	H	<div></div>	2.10		
16	5-CONH <sub>2</sub>	H	H	0.92							
17	5-NH <sub>2</sub>	H	H	0.24	44	NHCONH <sub>2</sub>	H	<div></div>	2.52		
18	5-CN	H	H	0.01	45	NHCONH <sub>2</sub>	CH <sub>3</sub>	<div></div>	2.00		
19	5-CH <sub>2</sub> NH <sub>2</sub>	H	H	0.00							
20	5-NHSO <sub>2</sub> Me	H	H	0.28	46	NHCONH <sub>2</sub>	H	<div></div>	2.40		
21	5-NHCOMe	H	H	1.26							
22	5-SO <sub>2</sub> NH <sub>2</sub>	Me	H	1.18	47	NHCONH <sub>2</sub>	H	<div></div>	1.68		
23	5-SO <sub>2</sub> NH <sub>2</sub>	Et	H	1.85							

Table 1 continued

24	5-SO <sub>2</sub> NH <sub>2</sub>	Ph	H	1.54	48	NHCONH <sub>2</sub>	H		2.10
25 <sup>a</sup>	5-SO <sub>2</sub> NH <sub>2</sub>	COOEt	H	1.47					
26	5-SO <sub>2</sub> NH <sub>2</sub>	CONH <sub>2</sub>	H	-0.59	49	NHCONH <sub>2</sub>	H		1.35
27	5-SO <sub>2</sub> NH <sub>2</sub>	CONHEt	H	-0.78					
28	5-SO <sub>2</sub> NH <sub>2</sub>	CONEt <sub>2</sub>	H	-0.73	50	NHCONH <sub>2</sub>	CH <sub>3</sub>		2.30
29	5-NHCONH <sub>2</sub>	Me	H	2.30					
30	5-NHCONH <sub>2</sub>	Et	H	2.52	51	NHCONH <sub>2</sub>	CH <sub>3</sub>		2.00
31	5-NHCONH <sub>2</sub>	4-pyridine	H	2.00					
32	NHCONH <sub>2</sub>	H		2.05	52	NHCONH <sub>2</sub>	CH <sub>3</sub>		2.22
33	NHCONH <sub>2</sub>	H		2.05	53	NHCONH <sub>2</sub>	CH <sub>3</sub>		2.40
34	NHCONH <sub>2</sub>	H		1.41	54	NHCONH <sub>2</sub>	CH <sub>3</sub>		2.30
35	NHCONH <sub>2</sub>	H		1.25	55	NHCONH <sub>2</sub>	CH <sub>3</sub>		2.00
36	NHCONH <sub>2</sub>	H		1.96	56	NHCONH <sub>2</sub>	H		2.22
Test Set									
57	7-OH	H	H	-1.41	66	NHCONH <sub>2</sub>	H		2.40
58	5-OH	H	5'-Me	0.17					
59	5-OMe	H	H	0.24	67	NHCONH <sub>2</sub>	H		2.10
60	5-tetrazole	H	H	0.70					
61	5-NHCONH <sub>2</sub>	H	H	1.75	68	NHCONH <sub>2</sub>	H		2.30
62	5-SO <sub>2</sub> NH <sub>2</sub>	COOH	H	-0.62					
63	5-NHCONH <sub>2</sub>	Ph-3-NH <sub>2</sub>	H	2.05	69	NHCONH <sub>2</sub>	H		2.52
64	NHCONH <sub>2</sub>	H		1.57					
65	NHCONH <sub>2</sub>	H		1.77	70	5-NHCONH <sub>2</sub>	H		2.30

<sup>a</sup> An outlier in the CoMFA and CoMSIA models

conformation of each molecule and the way to align molecules together are critical to obtaining a reasonable model. An attractive quality of these inhibitors is their relative conformational rigidity, a feature that makes them more amenable to meaningful CoMFA and CoMSIA analyses than flexible molecules. Furthermore, the crystallographic bound inhibitor, compound **29**, available in the Protein Data Bank (PDB ID: 2PE1) [22], was used as a starting geometry to construct the remaining compounds included in the study. Following the addition of hydrogen atoms “Clean Up” procedure was used to relieve any strain from the ring atoms and the bond lengths from the newly built structural moieties without the need to subject them to any energy minimization. Partial atomic charges were calculated using the Gasteiger-Marsili method [24]. The 3D QSAR studies are meaningful when carried out on a experimentally bound (bioactive) conformation instead of on the energy minimized free-state conformation. The reference atoms C-1, C-2, C-3, and C-4 in a crystallographically-bound inhibitor, compound **29**, were used for atom-fit alignment (Fig. 1a). The resulting alignment model (Fig. 1b) was then subjected to CoMFA and CoMSIA studies.

#### CoMFA and CoMSIA descriptors

CoMFA steric and electrostatic interaction fields were calculated at each lattice intersection on a regularly spaced grid of 2 Å. The grid pattern was generated automatically by the Sybyl/CoMFA routine and extended in the *x*, *y* and *z* directions. An  $sp^3$  carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 was used to generate steric (Lennard-Jones 6–12 potential) and electrostatic (Coulombic potential) field energies with a distance-dependent dielectric at each lattice point. Values of steric and electrostatic energy were truncated at a default value of 30 kcal/mol. The CoMFA steric and electrostatic fields thus generated were scaled by the CoMFA-STD method in SYBYL.

CoMSIA descriptors were derived with the same lattice box as that used for the CoMFA calculations, with a grid

spacing of 2 Å and employing a  $C^{1+}$  probe atom with a radius of 1.0 Å as implemented in SYBYL. CoMSIA similarity indices (AF) for a molecule *j* with atoms *i* at a grid point *q* were calculated using Eq. 1:

$$A_{F,k}^q(j) = - \sum \omega_{\text{probe},k} \omega_{ik} e_{iq}^{-\alpha r} \quad (1)$$

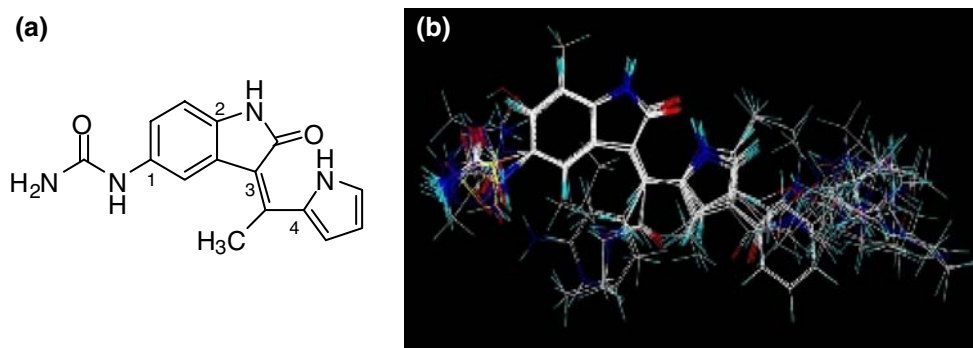
Being an extension to the CoMFA approach which has two fields, the CoMSIA method incorporates five physicochemical properties (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) denoted as *k* in Eq. 1, which were evaluated using the probe atom. A Gaussian type distance-dependence was used between the grid point *q* and each atom *i* in the molecule. A default value of 0.3 was used as the attenuation factor  $\alpha$ . 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 [25], and the hydrogen-bond donor and acceptor atoms within a putative protein environment were derived from experimental values [26, 27].

A partial least-squares (PLS) [19, 20] approach, which is an extension of multiple regression analysis, was used to derive the 3D QSAR models using the standard implementation in the SYBYL package, in which the CoMFA and CoMSIA descriptors were used as independent variables and the  $pIC_{50}$  values were used as dependent variables. CoMFA and CoMSIA columns with a variance 2.0 kcal/mol were filtered by using column filtering to improve the signal-to-noise ratio. The cross-validation with the leave-one-out (LOO) [16, 28] option was carried out to obtain optimal number of components (ONC). The cross-validated correlation coefficient,  $r_{cv}^2$ , was calculated using Eq. 2:

$$r_{cv}^2 = 1 - \frac{\sum (Y_{\text{predicted}} - Y_{\text{observed}})^2}{\sum (Y_{\text{observed}} - Y_{\text{mean}})^2} \quad (2)$$

where  $Y_{\text{predicted}}$ ,  $Y_{\text{observed}}$  and  $Y_{\text{mean}}$  are the predicted, observed, and mean values of the target property ( $pIC_{50}$ ), respectively.  $(Y_{\text{predicted}} - Y_{\text{observed}})^2$  is the predictive

**Fig. 1** (a) Structure of compound **30**, used as a template for atom-based alignment. The atoms for alignment are numbered 1–4. The atom numbering does not follow IUPAC rules. (b) Atom-based alignment of all the compounds



residual sum of squares (PRESS). The optimal number of components obtained from the cross-validated PLS analysis was used with the same column filtering set to derive the final QSAR model leading to the highest correlation coefficient ( $r^2$ ) and the lowest standard error using the compounds in the training set (without cross-validation).

The boot strapping analysis [29] for 100 runs and the number of cross-validations (e.g., two and five) were carried out and confirmed by the average value for 50 runs from each cross-validation. To test the utility of the model as a predictive tool, an external set of compounds (the test set) with known activities, but not used in model generation, were used. The predictive  $r^2$ , calculated by using Eq. 3, was based on molecules from the test set and it was used to evaluate the predictive power of the CoMFA and CoMSIA models as given below:

$$\text{Predictive } r^2 = 1 - (\text{“press”}/\text{SD}) \quad (3)$$

Here SD is the sum of the squared deviations between the actual activities of the compounds in the test set and the mean activity of the compounds in the training set. “Press” is the sum of the squared deviations between predicted and actual activities for every compound in the test set. The activity of the test set was predicted by the CoMFA and CoMSIA models using the predict property command. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pair-wise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values.

## Results and discussion

### CoMFA and CoMSIA statistical results

The major objective when working with CoMFA and CoMSIA analysis is to find the best predictive model within the system. The structures of all of the compounds used in this study are shown in Table 1. PLS analysis results based on atom-fit alignment on receptor-bound inhibitor are listed in Table 2, which shows that all the statistical indexes are reasonably high. For the 56 compounds in the training set, the model yielded ( $r_{cv}^2 = 0.672$ ,  $SEP = 0.666$ ,  $r_{ncv}^2 = 0.887$ , and  $r_{pred}^2 = 0.851$ ) and ( $r_{cv}^2 = 0.777$ ,  $SEP = 0.543$ ,  $r_{ncv}^2 = 0.920$ , and  $r_{pred}^2 = 0.816$ ) values for the CoMFA and CoMSIA models, respectively (Table 2). The next step was to examine the training set for outliers so as to achieve a superior predictive model. There are several reasons for the existence of outliers, including unique structural differences and a higher residual between the observed and the predicted biological activity of the inhibitor. According to this criterion, compound 25 was found to be an outlier since it had residual values of 1.68

**Table 2** Summary of CoMFA and CoMSIA results

PLS statistics	56-compound model		55-compound model	
	CoMFA	CoMSIA	CoMFA	CoMSIA
$r_{ncv}^2$ <sup>a</sup>	0.887	0.920	0.912	0.949
SEE <sup>b</sup>	0.361	0.301	0.321	0.245
$F_{test}$ <sup>c</sup>	78.66	146.43	101.86	181.17
$r_{cv}^2$ <sup>d</sup>	0.672	0.777	0.738	0.816
SEP <sup>e</sup>	0.666	0.543	0.631	0.509
$r_{pred}^2$ <sup>f</sup>	0.851	0.816	0.865	0.837
PLS components <sup>g</sup>	5	4	5	5
<i>Contribution</i>				
Steric	0.603	0.108	0.585	0.118
Electrostatic	0.397	0.195	0.415	0.230
Hydrophobic	–	0.364	–	0.322
H-bond donor	–	0.334	–	0.330
$r_{boot}^2$ <sup>h</sup>	–	–	0.952	0.966
SEE <sub>boot</sub> <sup>i</sup>	–	–	0.235	0.197
$r_{LHO}^2$ <sup>j</sup>	–	–	0.484	0.444
SD <sub>LHO</sub> <sup>k</sup>	–	–	0.683	0.675
$r_{5cv}^2$ <sup>l</sup>	–	–	0.627	0.558
SD <sub>5cv</sub> <sup>m</sup>	–	–	0.622	0.633

<sup>a</sup> Correlation coefficient

<sup>b</sup> Standard error of estimate

<sup>c</sup> Ratio of  $r^2$  explained to unexplained =  $r^2/(1 - r^2)$

<sup>d</sup> Cross-validated correlation coefficient after leave-one-out procedure

<sup>e</sup> Standard error of prediction

<sup>f</sup> Predicted correlation coefficient for test set of compounds

<sup>g</sup> Optimal number of principal components

<sup>h</sup> Average of correlation coefficient for 100 samplings using bootstrapped procedure

<sup>i</sup> Average standard error of estimate for 100 samplings using bootstrapped procedure

<sup>j</sup> Average cross-validated correlation coefficient for 50 runs using leave-half-out (LHO) group

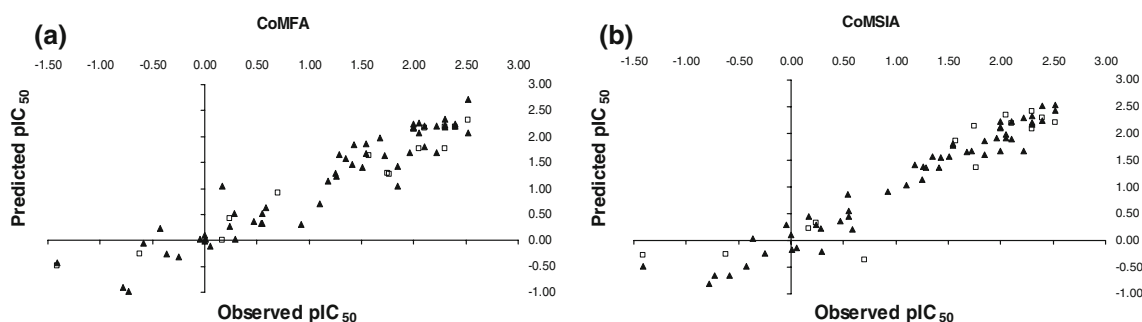
<sup>k</sup> Standard deviation of average cross-validated correlation coefficient for 50 runs

<sup>l</sup> Average cross-validated correlation coefficient for 50 runs using five cross-validation group

<sup>m</sup> Standard deviation of average cross-validated correlation coefficient for 50 runs

and 1.25 for the CoMFA and CoMSIA models, respectively (Supplementary information, Table S1). The corresponding pIC<sub>50</sub> values are somewhat reflective of the differences in substituent at the R<sub>2</sub>-position, which seems to play an important role in the PDK1 inhibitory activity. Compounds 22–24, containing hydrophobic groups at the R<sub>2</sub>-position, demonstrated high inhibitory activity. In contrast, compounds 26–28, which lack such functional groups demonstrated a poor activity. In compounds 26, 27, and 28 the R<sub>2</sub>-position is occupied by a primary,





**Fig. 2** Plot of observed versus predicted  $\text{pIC}_{50}$  values for the training set ( $\blacktriangle$ ) and the test set ( $\square$ ) compounds based on (a) CoMFA model and (b) CoMSIA model

secondary, and tertiary amide group, respectively, and in compound **25** by an ester group. Compound **25** is unique in the sense that the carbonyl group is attached to an electronegative oxygen atom. On the other hand, in compounds **26**, **27**, and **28** the carbonyl group is attached to a less electronegative nitrogen atom. Thus, after dropping compound **25** from the training set, the PLS analyses were re-performed on the remaining 55 compounds, and led to the  $r_{\text{cv}}^2$  value of 0.738, with the optimal number of principal components (ONC) being 5 for CoMFA and 0.816 for CoMSIA again with (ONC) 5 (Table 2). Since statistically significant CoMFA and CoMSIA models were derived from the compounds remaining in the training set, these compounds served as the basis for further assessment and discussion. The observed and predicted  $\text{pIC}_{50}$  values derived from the training set by the best CoMFA model are presented in (Supplementary Information, Table S1) and Fig. 2a.

All possible combinations of fields (Supplementary Information, Table S2) were evaluated to determine which of the five fields are actually needed for the generation of a predictive model. The hydrophobic and hydrogen bond donor fields yielded the best individual field models with  $r_{\text{cv}}^2$  of 0.834 and 0.762, respectively. In the combined models, the best  $r_{\text{cv}}^2$ ,  $r_{\text{ncv}}^2$ , and least standard error of estimation were obtained by the combination of steric, electrostatic, hydrophobic, and hydrogen bond donor fields ( $r_{\text{cv}}^2 = 0.816$ ,  $r_{\text{ncv}}^2 = 0.949$  and  $\text{SEE} = 0.245$ ). Thus, the model with steric, electrostatic, hydrophobic, and hydrogen bond donor fields appeared to be the superior among all the derived models. The detailed observed and predicted  $\text{pIC}_{50}$  values based on the selected CoMSIA model for the training set are shown in (Supplementary Information, Table S1) and Fig. 2b.

Bootstrapping analysis for 100 runs was performed (Table 2) to assess the robustness and statistical confidence of the derived models. Bootstrapping involves the generation of multiple new datasets from the original datasets, after randomly choosing samples from the original dataset. The bootstrapped  $r^2$  of 0.952 for CoMFA and 0.966 for

CoMSIA suggest that a good internal consistency exists within the underlying dataset.

Furthermore, to investigate the stability of the CoMFA and CoMSIA models, a cross-validation analysis was applied to the set of compounds in the training set. The training set model was cross-validated using two (Leave-Half-Out) and five (Leave 20% out) cross-validation groups, 50 times each. The average and standard deviation values of  $r_{\text{cv}}^2$  are shown in Table 2. When two cross-validation groups were used, the average  $r_{\text{cv}}^2$  values for CoMFA and CoMSIA were 0.484 (standard deviation = 0.683) and 0.444 (standard deviation = 0.675), respectively. By using five cross-validation groups, the average  $r_{\text{cv}}^2$  and SD values were 0.627 and 0.622, respectively, for CoMFA. For CoMSIA, these values were 0.558 and 0.633, respectively. Thus, two and five cross-validation analyses of the training set composition for each run were consistent for the CoMFA and CoMSIA models.

#### Predictive power of the final CoMFA and CoMSIA models

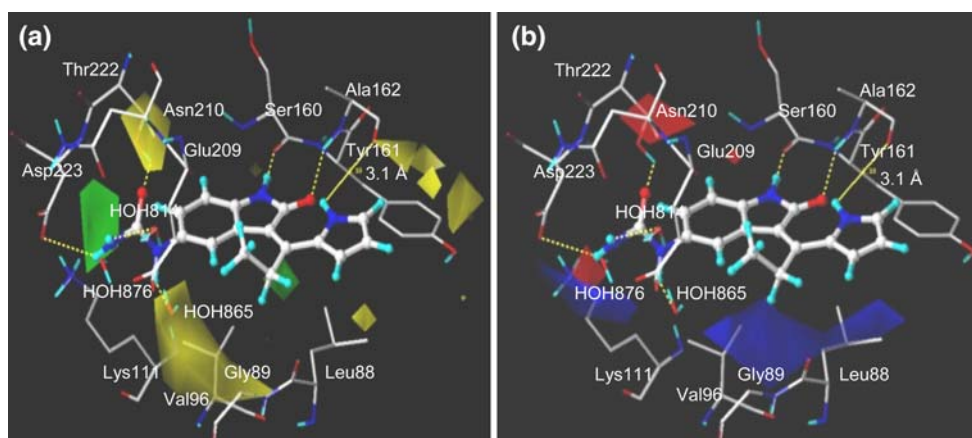
The predictive abilities of the CoMFA and CoMSIA models were determined from a set of 14 test compounds not included in the model generation study. The predicted  $r^2$  values from the CoMFA and CoMSIA models were found to be 0.865 and 0.837, respectively (Table 2). The detailed observed and predicted  $\text{pIC}_{50}$  values based on the selected CoMFA and CoMSIA models for the test set compounds are listed in the Supplementary Information included in Table S1 and graphically shown in Fig. 2a and b. It is interesting to note that the predicted  $\text{pIC}_{50}$  residuals for both models were less than one log unit. In the CoMFA model 79% of the compounds had predicted residuals of less than 0.5 log units, in the CoMSIA model 86% of the compounds had predicted residuals of less than 0.5 log units, and four compounds were found to have residuals  $\geq 0.5$  log units. The performance of both models on these four compounds is described below. The  $\text{pIC}_{50}$  value of compound **57**, bearing a 7-OH group at the  $\text{R}_1$  position,

was poorly predicted by both models, possibly because the training set did not include this type of substituent. The small residual value for compound **60**, bearing a bulky tetrazole group at the  $R_1$  position, in the CoMFA model could be due to its exact fit within the large green contour; in contrast, the large residual value for this compound based on the CoMSIA model could be due to its distant location from the small green contour seen in its steric/electrostatic contour map. The underestimation of the  $pIC_{50}$  value for compound **65** based on CoMFA and CoMSIA may be related to the absence of an alkyl substituent at the  $R_2$  position, which is found to occupy a sterically favorable green contour in both models. Likewise, underestimation of the  $pIC_{50}$  value for compound **70** by the CoMFA model may stem from the absence of a green contour around the  $R_3$  substituent present in the corresponding CoMSIA steric/electrostatic contour map.

The statistically significant predictions as described above support the validity of the derived models in predicting the inhibitory activity of the untested compounds.

### Structure-based interpretation of 3D QSAR models

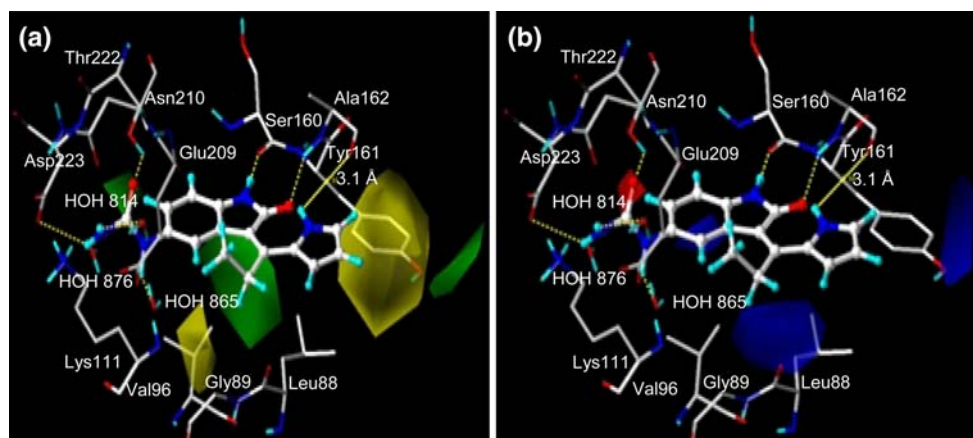
The major advantage of CoMFA and CoMSIA modeling is that the information content of the derived 3D QSAR models can be visualized as 3D coefficient contour plots. In addition, ligand-based methods such as CoMFA and CoMSIA are widely used not only because they are not very computationally intensive but also because they can lead to the rapid generation of QSARs from which the biological activity of newly designed compounds can be predicted. In contrast, an accurate prediction of activity of untested compounds based on the computation of binding free energies is both complicated and lengthy. In Figs. 3, 4,



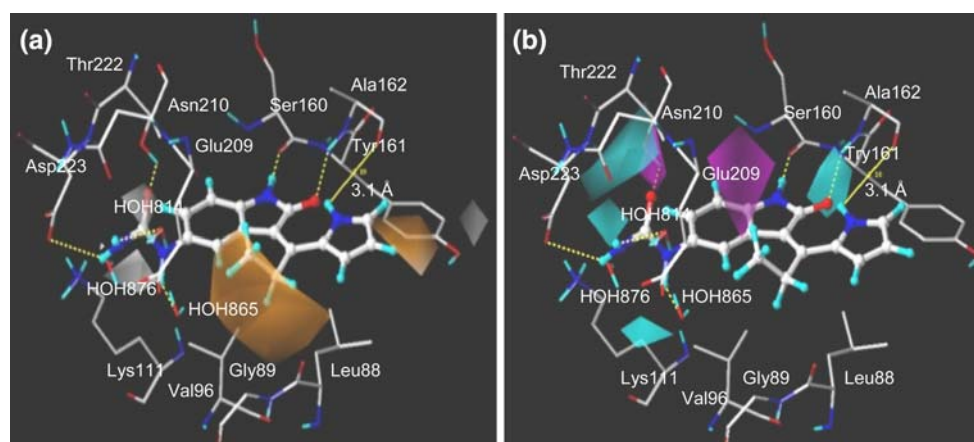
**Fig. 3** CoMFA contour maps displayed with most potent compound **30** and the key residues in the binding site of PDK1. Amino acid residues are represented as sticks and the inhibitor is shown in as a ball-and-stick model. The hydrogen bonds are shown as broken yellow lines. **(a)** CoMFA stdev \* coeff steric contour map; green contours indicate regions where bulky groups increase activity,

whereas yellow contours indicate regions where bulky groups decrease activity. **(b)** CoMFA stdev \* coeff electrostatic contour map; blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups increase activity

**Fig. 4** CoMSIA contour maps displayed with most potent compound **30** and the key residues in the binding site of PDK1. Amino acid residues are represented as sticks and the inhibitor is shown in as a ball-and-stick model. **(a)** CoMSIA stdev \* coeff steric contour map; the color scheme is the same as in Fig. 3a. **(b)** CoMSIA stdev \* coeff electrostatic contour map; the color scheme is the same as in Fig. 3b







**Fig. 5** CoMSIA contour maps displayed with most potent compound **30** and the key residues in the binding site of PDK1. Amino acid residues are represented as sticks and the inhibitor is shown as a ball-and-stick model. The hydrogen bonds are shown as broken yellow lines. (a) CoMSIA stdev \* coeff hydrophobic contour maps; orange

and **5** the contour plots were generated as the scalar products of coefficients and standard deviation (stdev \* coeff), associated with each CoMFA or CoMSIA column. The contour plots are extremely useful to identify the possible interaction sites and, in addition, are helpful to identify important regions where any change in the steric, electrostatic, hydrophobic, and hydrogen bond donor fields may affect the biological activity.

#### CoMFA contour map

A CoMFA-generated contour map of steric contribution is shown in Fig. 3a. The green contours represent the regions of high steric tolerance (80% contribution), while yellow contours represent regions of unfavorable steric effects (20% contribution). The large green contour localized at the urea group at 5-position of the indolinone ring shows that the steric interaction is favored in this region. This is consistent with the reported experimental results since compounds **29–56** have a urea group as a substituent and all have an activity with  $\text{pIC}_{50} > 1.25$ . As discussed for CoMSIA fields, this urea group enters in several hydrogen bonding interactions. The large green contour was found unoccupied by any of the active site residues, a consequence of which, any bulky substitution is favored. In the least active compounds, i.e., compounds **26, 27** (both with sulfonamide group), **59** (methoxy), and **60** (tetrazole), their  $\text{R}_1$  substituent is positioned away from this green contour. In the case of the least active compounds, i.e., compounds **26, 27, 28**, and **62**, the  $-\text{NH}_2$  group of their sulfonamide moiety was seen approaching the medium yellow contour near the  $\text{R}_1$ -position of the indolinone ring. Since this yellow contour was occupied by the  $-\text{CH}_3$  group in the side chain of Thr222, there may be potential for its steric clash

and white contours indicate favorable and unfavorable hydrophobic groups, respectively, for PDK1 inhibitory activity. (b) CoMSIA stdev \* coeff hydrogen bond donor contour maps; cyan and magenta contours indicate favorable and unfavorable hydrogen bond donor groups, respectively, for PDK1 inhibitory activity

with the side chain of Thr222. Another green contour near the  $\text{R}_2$  substituent indicates the presence of a bulky group at this position is necessary to attain PDK1 inhibitory activity. This is exemplified by compounds **29, 30**, and **61** with  $-\text{CH}_3$  ( $\text{pIC}_{50} = 2.30$ ),  $-\text{C}_2\text{H}_5$  ( $\text{pIC}_{50} = 2.52$ ), and  $-\text{H}$  ( $\text{pIC}_{50} = 1.75$ ) as  $\text{R}_2$  substituents, respectively. It is obvious that as the bulkiness of the alkyl group at the  $\text{R}_2$  position increases, the PDK1 inhibitory activity also increases. Also, finding the favorable green contour surrounded by an unfavorable large yellow contour signifies that any group bulkier than an ethyl group would decrease the PDK1 inhibitory activity. This possibility is exemplified by the least active compounds, **27** having  $-\text{CONH}\text{Et}$  ( $\text{pIC}_{50} = -0.78$ ) and **28** with  $-\text{CONEt}_2$  ( $\text{pIC}_{50} = -0.73$ ). Furthermore, in compounds **31** and **63**, a 4-pyridine ( $\text{pIC}_{50} = 2.00$ ) and Ph-3- $\text{NH}_2$  ( $\text{pIC}_{50} = 2.05$ ) substituent, respectively, were found to be in the vicinity of the large yellow contour. Structure-based analysis of these two compounds led to the conclusion that these bulky groups may be involved in steric hindrance with the side chains of Leu88 and Val96 residues. Since no difference in inhibitory activity was noted between compounds **30** and **44**, then the 3-aminomethylphenyl moiety at the 4'-position of the pyrrole ring of compound **44** and the  $-\text{C}_2\text{H}_5$  group at  $\text{R}_2$  position of compound **30** have no role in modulating the activity of the compounds associated with them. Compounds **22–24** and **26–28**, with a similar  $\text{R}_1$  substituent but a different  $\text{R}_2$  substituent, also showed varying degrees of activity. Less active compounds, as exemplified by compounds **27** and **28**, are found to orient their  $\text{R}_2$  substituent into the yellow contour. Compounds **13–21** and **61**, by virtue of having different  $\text{R}_1$  substituents, demonstrate different activities, with the highest activity being derived when the  $\text{R}_1$  substituent is a urea group as is the case of

compound **61**. The presence of five small yellow contours on the right side and at positions 4' and 5' of the pyrrole ring signifies that compounds having substituents that enter these contours will be less active than those with substituents that do not. Not surprisingly, most of the analogs in this series showed reasonable potency ( $\text{pIC}_{50} > 1.25$ ) by virtue of possessing a 4'-substituent on the pyrrole ring that did not enter the small yellow contours.

The electrostatic contour map of the CoMFA model is shown in Fig. 3b. To facilitate the visualization, the most potent compound **30** is overlaid on the map. Two red contour maps around the  $R_1$ -substituent indicate that an electronegative group needs to be present at this position for potent inhibitory activity. The most active compounds have either a urea group (compounds **29–56**, **61**, and **63**) or a sulfonamide group (compounds **22–24**) at this position. In contrast, the least active compounds, as exemplified by compounds **3** and **4**, lack an electronegative group at this position. Furthermore, the  $-\text{NH}_2$  group of the sulfonamide moiety in compounds **22** and **23** and the carbonyl oxygen atom in compounds **29** and **30** were mapped in the vicinity of the upper left large red contour. The sulfonamide analogs tend to be less active than the urea analogs on account of the nitrogen atom in the former to be less electronegative than the carbonyl oxygen present in the latter. A medium blue contour along with a red contour seen in close proximity to the  $R_1$  substituent are viewed as being an extension of the urea group at the terminal  $-\text{NH}_2$  group with the electropositive group such as alkylcarbonyl would lead to increase in PDK1 inhibitory activity. In addition, the  $-\text{NH}_2$  group of the sulfonamide moiety at the 5-position of the indolinone ring is found to approach the upper left side large red contour containing the  $-\text{CH}_3$  group of Thr222. Thus, the existence of a possible steric clash is implied by the co-localization of red and yellow contours, as exemplified by the least potent compounds **26**, **27**, **28**, and **62**. A large blue contour located near the  $R_2$  substituent suggests that the presence of an electropositive group such as an alkyl chains would increase the inhibitory activity towards PDK1, as is the case for the most potent compounds **29** ( $R_2 = -\text{CH}_3$ ) and **30** ( $R_2 = \text{C}_2\text{H}_5$ ). Conversely, the poor inhibitory activity of compound **62** ( $\text{pIC}_{50} = -0.62$ ), stemming from the presence of an electronegative  $-\text{COOH}$  group as the  $R_2$  substituent, is mapped near the large blue contour.

#### CoMSIA contour map

The color scheme for CoMSIA steric and electrostatic fields is the same as described for CoMFA model. In the CoMSIA hydrophobic field, the orange (hydrophobic favorable) and white (hydrophobic unfavorable or hydrophilic favorable) contours represent 80% and 20% level

contributions, respectively. In the CoMSIA hydrogen bond donor field, the cyan (hydrogen bond donor favorable) and magenta (hydrogen bond donor unfavorable) contours represent 80% and 10% level contributions, respectively.

The CoMSIA steric (Fig. 4a) and electrostatic (Fig. 4b) contour maps are rather similar to those obtained by the CoMFA model (Fig. 3a, b, respectively), except for a few unexplained steric and electrostatic contour maps that may be important to understand the interactions between the active site of PDK1 and an inhibitor described below. In Fig. 4a, the green contour in the vicinity of the large yellow contour was found to be associated with long side chains present at the 4-position of the pyrrole ring. The green contour seen at the  $R_2$  position of compound **30** is found to be larger than the corresponding contour present in the CoMFA model and to completely engulf the ethyl moiety. While in the CoMFA model several small yellow contours are seen around the 5-position of the pyrrole ring, in the CoMSIA model they are replaced by a single large yellow contour. Finding this yellow contour to be exactly mapped on the side chain of Tyr161 suggests that any larger substituent at the 5-position of the pyrrole ring will be unfavorable since it may enter into potential steric clash with the Tyr161 residue present at the active site of PDK1.

In Fig. 4b, the red contour has moved from the side chain of Thr222 to the carbonyl oxygen atom of the urea moiety should be more reasonable compare to the one present in CoMFA electrostatic contour map because the CoMSIA fields are better interpretable than the CoMFA fields. A blue contour exactly fitting within the centroid of the phenyl ring of the indolinone moiety indicates the existence of a favorable activity when this ring is electropositive. The additional blue contour, appearing on the right side of the image for the moderately active compounds **64** and **65**, is a region that is typically occupied by substituents at 4-position of the pyrrole ring that are preferentially electropositive.

The hydrophobic contour map of the CoMSIA model in the presence of most potent compound **30** is displayed in Fig. 5a. The presence of two white contour maps around the  $R_1$  substituent suggests that its occupancy by a hydrophilic group would favor an inhibitory activity. It is interesting to note that the side chains of polar amino acids such as Lys111, Glu209, Asn210, Thr222, and Asp223 and of water molecules 814, 865, and 876 were mapped within 3 Å radius of the urea group at  $R_1$  position. Hence, the presence of white contours is justified based on active site environment. Whereas in the most active compounds (i.e., compounds **29–56**),  $R_1$  substituent was a polar group, in the least active compounds (i.e., compounds **3** and **4**) the  $R_1$  substituent was nonpolar. The large orange contour at the  $R_2$ -position suggests that the presence of a hydrophobic group at this position would increase the PDK1 inhibitory

activity. The most potent compounds (i.e., compounds **29** and **30**) possess  $-\text{CH}_3$  and  $-\text{C}_2\text{H}_5$  groups at this position, and the less active ones either lack (compounds **3** and **4**) or have a significantly large hydrophobic group at this position (compounds **27** and **28**). This large orange contour was mapped within the hydrophobic pocket formed by Leu88, Gly89, and Val96. Another small orange contour at the 4'- and 5'-positions of the pyrrole ring is due to the presence of small hydrophobic substituents at these two positions and which would increase the inhibitory activity. This small orange contour is seen surrounded by two hydrophobic amino acid residues such as Tyr161 and Leu88. The small white contour seen just behind the orange contour is solvent-exposed; is occupied by the amino group of the aminomethyl moiety in the highly active compound **44** ( $\text{pIC}_{50} = 2.52$ ).

The hydrogen bond donor contour map of the CoMSIA model in the presence of the most potent compound **30** is depicted in Fig. 5b. Three cyan contours seen near the  $\text{R}_1$ -position of the indolinone ring indicate that the hydrogen bond donor group is favorable for the inhibitory activity. The most potent compounds exhibit a urea (compounds **29–56**, **61** and **63**), and a sulfonamide moiety (compounds **22–25**) as  $\text{R}_1$ -substituent. The urea and sulfonamide functions possess a hydrogen bond donor group which seems to play a key role in their predominantly positive effect on PDK1 inhibitory activity. A total of six hydrogen bonds are formed between the most potent compound **30** and PDK1. A carboxyl oxygen atom of Asp223 and the oxygen atom of water molecule 814 are involved in hydrogen bond interaction with the  $-\text{NH}_2$  of the urea group ( $\text{Asp223 } -\text{C}=\text{O}\cdots\text{H}_2\text{N}$ , 2.66 Å) and ( $814\text{H}_2\text{O}\cdots\text{H}_2\text{N}$ , 2.63 Å). The  $-\text{NH}$  of the urea group forms a hydrogen bond with the oxygen atom of water molecule 865 ( $865\text{H}_2\text{O}\cdots\text{HN}-$ , 2.62 Å). Another cyan contour is seen in the vicinity of the  $-\text{NH}$  of the indolinone moiety indicates the presence of hydrogen bond donor group at this position would increase the inhibitory activity. The indolinone  $-\text{NH}$  was seen hydrogen bonded to the backbone of the hinge residue Ser160 ( $\text{Ser160 } -\text{C}=\text{O}\cdots\text{HN}-$ , 2.07 Å). The same cyan contour is also close to the acidic  $-\text{NH}$  of the pyrrole ring. The acidic H-bond donor group at the 1-position of the pyrrole ring was found to be solvent exposed and could potentially form the hydrogen bond with the backbone of another hinge residue Ala162 ( $\text{Ala162 } -\text{C}=\text{O}\cdots\text{HN}-$ , 3.1 Å). A small magenta contour located on the carbonyl oxygen of the urea moiety suggests a possible positive effect on activity by the presence of a hydrogen bond acceptor in this region. The carbonyl oxygen of the urea moiety is engaged in hydrogen bond interaction with the side chain of Thr222 ( $\text{Thr222 } -\text{OH}\cdots\text{O}=\text{C}-$ , 2.20 Å). A large magenta contour distantly located from the indolinone ring is suggestive of a favorable interaction with the

receptor if it is occupied by hydrogen bond acceptor group. The carbonyl oxygen of the indolinone ring is seen involved in hydrogen bond interaction with the backbone of Ala162 ( $\text{Ala162 } -\text{NH}\cdots\text{O}=\text{C}$ , 2.32 Å). These results point to the importance of hydrogen bond donor and acceptor groups in these inhibitors and whose presence is inferred from the corresponding cyan and magenta contour maps. Compounds **3** and **4**, by possessing 5- $\text{CH}_3$  and 7- $\text{CH}_3$  groups on the indolinone ring, respectively, displayed the least activity. Compound **60** having a tetrazole substituent at the  $\text{R}_1$  position exhibited a moderate activity as it is involved only in two hydrogen bonds with the backbone of Ser160 and Ala162.

## Conclusions

Use of 3D QSAR analysis has led to the development of satisfactory CoMFA and CoMSIA models for predicting the biological activity of new compounds. Understanding the intermolecular interactions of indolinone derivatives with PDK1 was achieved by integrating statistically significant and predictable CoMFA and CoMSIA models. The  $r_{\text{cv}}^2$  and  $r_{\text{ncv}}^2$  values of the CoMFA and CoMSIA models were good enough to suggest that all the reported inhibitors bind to PDK1 in an almost a similar fashion. The contour maps generated from the developed CoMFA and CoMSIA models provided useful concepts about the types of the protein environment surrounding the inhibitors. The CoMFA model point to the importance of steric descriptor field as it explained nearly 59% of the variance in PDK1 inhibitory activity. On the other hand, the CoMSIA model suggested significant contribution (>65%) by hydrophobic and hydrogen bond donor fields. All of the results obtained from structure-based 3D QSAR analysis are consistent with the experimental activity data, and provide useful insights on the future design of novel PDK1 inhibitors.

**Acknowledgements** Support to TT in the form of start-up funds and resources from the College of Pharmacy and Allied Health Professions of St. John's University are gratefully acknowledged. We thank Maulik Patel, Pallav Patel, Shridhar Kulkarni, and Dr. Cesar Lau-Cam for critical readings of the manuscript and for helpful discussions.

## References

1. Parker PJ, Parkinson SJ (2001) *Biochem Soc Trans* 29:860. doi: [10.1042/BST0290860](https://doi.org/10.1042/BST0290860)
2. Schlessinger J (2000) *Cell* 103:211. doi: [10.1016/S0092-8674\(00\)00114-8](https://doi.org/10.1016/S0092-8674(00)00114-8)
3. Cantley L (2002) *Science* 296:1655. doi: [10.1126/science.296.5573.1655](https://doi.org/10.1126/science.296.5573.1655)
4. Maehama T, Dixon JE (1998) *J Biol Chem* 273:13375. doi: [10.1074/jbc.273.22.13375](https://doi.org/10.1074/jbc.273.22.13375)

5. Brazil DP, Hemmings BA (2001) Trends Biochem Sci 26:657. doi:[10.1016/S0968-0004\(01\)01958-2](https://doi.org/10.1016/S0968-0004(01)01958-2)
6. Scheid MP, Woodgett JR (2001) Nat Rev Mol Cell Biol 2:760. doi:[10.1038/35096067](https://doi.org/10.1038/35096067)
7. Biondi RM, Cheung PCF, Casamayor A, Deak M, Currie RA, Alessi DR (2000) EMBO J 19:979. doi:[10.1093/emboj/19.5.979](https://doi.org/10.1093/emboj/19.5.979)
8. Biondi RM, Komander D, Thomas CC, Lizcano JM, Deak M, Alessi DR et al (2002) EMBO J 21:4219. doi:[10.1093/emboj/cdf437](https://doi.org/10.1093/emboj/cdf437)
9. Leslie NR, Downes CP (2002) Cell Signal 14:285. doi:[10.1016/S0898-6568\(01\)00234-0](https://doi.org/10.1016/S0898-6568(01)00234-0)
10. Huse M, Kuriyan J (2002) Cell 109:275. doi:[10.1016/S0092-8674\(02\)00741-9](https://doi.org/10.1016/S0092-8674(02)00741-9)
11. Yang J, Cron P, Good VM, Thompson V, Hemmings BA, Barford D (2002) Nat Struct Biol 9:940. doi:[10.1038/nsb870](https://doi.org/10.1038/nsb870)
12. Yang J, Cron P, Thompson V, Good VM, Hess D, Hemmings BA et al (2002) Mol Cell 9:1227. doi:[10.1016/S1097-2765\(02\)00550-6](https://doi.org/10.1016/S1097-2765(02)00550-6)
13. Alessi DR, James SR, Dowens CP, Homes AB, Gaffney PR, Reese CB et al (1997) Curr Biol 7:261. doi:[10.1016/S0960-9822\(06\)00122-9](https://doi.org/10.1016/S0960-9822(06)00122-9)
14. Mora A, Komander D, Van Aalten DM, Alessi DR (2004) Semin Cell Dev Biol 15:161. doi:[10.1016/j.semcdb.2003.12.022](https://doi.org/10.1016/j.semcdb.2003.12.022)
15. Storz P, Toker A (2002) Front Biosci 7:886. doi:[10.2741/storz](https://doi.org/10.2741/storz)
16. Cramer RDIII, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959. doi:[10.1021/ja00226a005](https://doi.org/10.1021/ja00226a005)
17. Cramer RDIII, Bunce JD, Patterson DE (1988) Quant Struct Act Relat 7:18. doi:[10.1002/qsar.19880070105](https://doi.org/10.1002/qsar.19880070105)
18. Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130. doi:[10.1021/jm00050a010](https://doi.org/10.1021/jm00050a010)
19. Wold S, Albano C, Dunn WJ, Edlund U, Esbenson K, Geladi P et al (1984) In: Kowalski B (ed) Chemometrics. Reidel, Dordrecht, The Netherlands
20. Staahle L, Wold S (1987) J. Chemom 1:185
21. Islam I, Bryant J, Chou YL, Kochanny MJ, Lee W, Phillips GB et al (2007) Bioorg Med Chem Lett 17:3814. doi:[10.1016/j.bmcl.2007.04.071](https://doi.org/10.1016/j.bmcl.2007.04.071)
22. Islam I, Brown G, Bryant J, Hrvatin P, Kochanny MJ, Phillips GB et al (2007) Bioorg Med Chem Lett 17:3819. doi:[10.1016/j.bmcl.2007.05.060](https://doi.org/10.1016/j.bmcl.2007.05.060)
23. SYBYL 7.2, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO, 63144
24. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219. doi:[10.1016/0040-4020\(80\)80168-2](https://doi.org/10.1016/0040-4020(80)80168-2)
25. Viswanadhan VN, Ghose AK, Revenkar GR, Robins RK (1989) J Chem Inf Comput Sci 29:163. doi:[10.1021/ci00063a006](https://doi.org/10.1021/ci00063a006)
26. Klebe G (1994) J Mol Biol 237:212. doi:[10.1006/jmbi.1994.1223](https://doi.org/10.1006/jmbi.1994.1223)
27. Klebe G, Mietzner T, Weber F (1999) J Comput Aided Mol Des 13:35. doi:[10.1023/A:1008026702439](https://doi.org/10.1023/A:1008026702439)
28. Wold S (1978) Technometrics 4:397. doi:[10.2307/1267639](https://doi.org/10.2307/1267639)
29. Vong R, Geladi P, Wold S, Esbensen K (1988) J Chemom 2:281