



## Development of in silico models for pyrazoles and pyrimidine derivatives as cyclin-dependent kinase 2 inhibitors

Fangfang Wang<sup>a</sup>, Zhi Ma<sup>a</sup>, Yan Li<sup>c</sup>, Shanna Zhu<sup>b</sup>, Zhengtao Xiao<sup>a</sup>, Hong Zhang<sup>a</sup>, Yonghua Wang<sup>a,d,\*</sup>

<sup>a</sup> Bioinformatics Center, Northwest A&F University, Yangling, Shaanxi, 712100, China

<sup>b</sup> College of Information Engineering, Northwest A&F University, Shaanxi, 712100, China

<sup>c</sup> School of Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

<sup>d</sup> College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, 712100, China

### ARTICLE INFO

#### Article history:

Accepted 14 June 2011

Available online 25 June 2011

#### Keywords:

CDK<sub>2</sub>

Inhibitors

CoMFA

CoMSIA

Molecular docking

Molecular dynamics

### ABSTRACT

CDK<sub>2</sub> (cyclin-dependent kinase 2) is an attractive target for therapeutic intervention in cancer. In this work, quantitative structure–activity relationship (QSAR), molecular docking, and molecular dynamics (MD) studies were performed on three sets of 155 CDK<sub>2</sub> inhibitors. The obtained models exhibit good predictive capability in both internal and external validations ( $q^2 = 0.73$ ,  $r_{\text{pred}}^2 = 0.94$  for 6, 6-dimethyl pyrrolo [3,4-c]pyrazoles analogs,  $q^2 = 0.62$ ,  $r_{\text{pred}}^2 = 0.63$  for imidazole pyrimidine amides analogs and  $q^2 = 0.56$ ,  $r_{\text{pred}}^2 = 0.58$  for 4-(pyrazol-4-yl)-pyrimidines analogs). Furthermore, a comparison between 3D-contour map, docking and MD simulation explore in detail the binding modes and the key structural features impacting the interaction of each series of inhibitors with the CDK<sub>2</sub> enzyme, which should be useful to aid the designing of new inhibitors with CDK<sub>2</sub> improved biological response.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

Uncontrolled cell proliferation is a hallmark of tumorigenesis, which may result from an increased expression of cell cycle up-regulators or from a reduced expression of down-regulators [1,2]. Cell proliferation is regulated by the cell cycle which is commonly viewed as an orderly progression through four distinct phases: G1, S, G2, and M. Transition between these four phases are mainly controlled by cyclin-dependent kinases (CDKs), a family of serine threonine protein kinases [3]. CDKs are inactive as monomers and their activation requires binding to the corresponding regulatory proteins (cyclins) and the phosphorylation by CDK-activating kinase (CAK) on a specific threonine residue. For example, cyclins A and B activate CDK<sub>1</sub>, cyclins A and E regulate the activity of CDK<sub>2</sub>, and the D-type cyclins are associated with CDK<sub>4</sub>. To date, 13 distinct CDKs and 25 different cyclin-box containing proteins have been identified by human genome sequencing [4]. Different CDK/cyclin combinations serve to regulate distinct check points in the cell division cycle.

CDK<sub>2</sub> is one of the cyclin-dependent Ser/Thr kinases which regulate the cell cycle progression at multiple levels. In complex with cyclin E, CDK<sub>2</sub> plays a paramount role during the G1/S transition

of the cell cycle, while in complex with cyclin A, it facilitates the progression of the S phase of the cell cycle [5]. It consists of an amino-terminal lobe rich in  $\beta$ -sheets and a larger, mostly  $\alpha$ -helical, carboxy-terminal lobe. The ATP-binding site is located in a deep cleft between the two lobes that contains the catalytic residues conserved among eukaryotic protein kinases. Monomeric CDK<sub>2</sub> is thought to be inactive because a region in the C-terminal lobe, named the T-loop (residues 146–166, contains the T160 phosphorylation site), blocks the entry of substrates to the catalytic cleft. When cyclin A binds to CDK<sub>2</sub>, the T-loop is displaced, allowing substrates access to the catalytic cleft. This rotation also exposes the T160 hydroxyl group to be phosphorylated by the CAK, allowing full activation of the CDK–cyclin complex. It has been revealed by genetic studies in mice that normal cells are not dependent on interphase CDKs for their growth, however, certain tumor cells, due to their pathogenic spectrum of mutation, may be sensitive to the inhibition of CDKs [6]. Additionally, evidence from in vitro studies suggests that the inhibition of CDK<sub>2</sub> selectively kills tumor cells with deregulated E2F-1 activity [7]. The CDK<sub>2</sub> activity is associated with apoptotic progression and also with the onset of apoptosis in the SK-HEP1 human hepatoma cell line [8] and HeLa cells [9]. Therefore, CDKs especially CDK<sub>2</sub> have been regarded as novel therapeutic targets for cancer chemotherapy [10] and may offer new opportunities for selective and tolerable therapy for human cancer.

CDK inhibitors are novel anti-cancer therapeutics that would inhibit CDKs by targeting the catalytic CDK subunit or by inter-

\* Corresponding author at: Bioinformatics Center, Northwest A&F University, Yangling, Shaanxi 712100, China. Tel.: +86 029 87092262; fax: +86 029 87092262.

E-mail address: [yh.wang@nwsuaf.edu.cn](mailto:yh.wang@nwsuaf.edu.cn) (Y. Wang).

fering the regulatory pathways that govern CDK activity [11]. Recently, several CDK inhibitors have entered clinical evaluation for treatment of cancer [12]. The first generation CDK inhibitors include flavopiridol, 7-hydroxystaurosporine (UCN-01), roscovitine (CYC202), BMS-387032 (SNS-032) [13], PD0332991 [14], and R547 [15]. Flavopiridol is the first CDK inhibitor entering clinical trials in oncology with more than 50 clinical trials undertaken or currently in progress. This flavonoid inhibitor has been reported to inhibit CDK<sub>1</sub>, CDK<sub>2</sub>, CDK<sub>4</sub> and CDK<sub>6</sub> selectively over other protein kinases involved in signal transduction and cell cycle regulation [16]. UCN-01 is a staurosporine analogue which is associated with the G1/S cell cycle arrest, dephosphorylation of CDK<sub>2</sub> [17]. CYC202 is a trisubstituted purine analogue selective for CDK<sub>2</sub>/cyclin E and possesses sub-micromolar activity on CDK<sub>7</sub>/cyclin H and CDK<sub>9</sub>/cyclin T, which has undergone Phase II trial [18]. Unlike flavopiridol which is a pan-CDK inhibitor, the second generation of CDK inhibitors primarily inhibits the G1/S transition by attacking CDK<sub>2</sub> kinase. For example, R-roscovitine, a small aminopurine molecule with an IC<sub>50</sub> of 100 nM, is orally bioavailable and has a single agent *in vitro* and *in vivo* activity against a broad range of tumor cell types [19]. BMS-387032 is an N-acyl-2-aminothiazole analogue which could inhibit the cell cycle and transcription [20]. In addition, AstraZeneca is a member of a novel class of CDK inhibitors with specificity for CDK<sub>2</sub> inhibition called imidazopyridines, showing enhancement of both gemcitabine- and cisplatin-induced apoptosis [21]. Another CDK<sub>2</sub> inhibitor E7070 is a synthetic chloro-indolyl sulfonamide derivative that has multiple cell cycle effects including the inhibition of CDK<sub>2</sub>. Meanwhile, studies have revealed that E7070 is tolerable in each schedule with the dose-limiting toxicities of neutropenia, thrombocytopenia, and anemia [22]. However, up to date, few CDK<sub>2</sub> inhibitors have entered clinical trials and displayed encouraging results.

The deficiency of these compounds led to significant efforts to identify novel inhibitors of CDK<sub>2</sub> which have more potent inhibitory activity and can ideally fit for combination therapies for cancer. However, *in vitro* assessment of the activity of CDK<sub>2</sub> inhibitors remains a labor-intensive and time-consuming operation. *In silico* molecular modeling approaches, as effective computational method, have been widely used to find out various interactive fields making impacts on activity, to predict the activities of the inhibitors and thus to help forecasting and designing of novel inhibitors [23–25]. In order to improve the medicinal properties and to eliminate the untoward effects, some compounds with good inhibitory activity have been synthesized [1,26]. Recently, several QSAR models have been built for these CDK<sub>2</sub> inhibitors, such as 3D-pharmacophore model [27], topochemical models [28], CoMFA and CoMSIA models [29], crystal structure model [30], docking [31] and SVM prediction [32]. For example, using a relatively small data set of 58 aminopyrazole derivatives, Mascarenhas and Ghoshal built a 3D-QSAR model using feature-shape pharmacophore of biologically active conformation [33]. In two other reports, the ligand–protein interaction energy between CDK<sub>2</sub> and 9H-purine derivatives was determined by using CoMFA and QM/MM hybrid methods [34,35]. Quite recently, combined 3D-QSAR with docking study on 4,5-dihydro-1H-pyrazolo [4,3-h] quinazoline derived CDK<sub>2</sub> inhibitors have identified the key structural requirements for the biological activity of the molecules [36].

More recently, three different series of CDK<sub>2</sub> inhibitors [37–43], with specific activities have been reported; however, they have not received much attention from a theoretical perspective so far. Thus, in this work, we applied ligand- and receptor-based QSAR technique to the three different classes of CDK<sub>2</sub> inhibitors. In addition, molecular docking and MD simulation were also performed to deeply explore the interaction mode between the inhibitor and the receptor. To our knowledge, this work provides the first QSAR

study for these CDK<sub>2</sub> inhibitors. The developed models can not only help in understanding the structure–activity relationship of these compounds but also serve as a useful guide for the design of novel CDK<sub>2</sub> inhibitors with better biological activities.

## 2. Methods and materials

### 2.1. Data sets and biological activity

The inhibitory activities of three different series of compounds towards CDK<sub>2</sub> enzyme reported in the literatures [37–43] were used as the dataset for molecular modeling in this study (Tables S1–S3). Discarding compounds with unspecified inhibitory activity, the data set comprising diverse DPP, IPA and PYP analogs possessed a wide spectrum of inhibitory activities against CDK<sub>2</sub> enzyme. The inhibitory activities (IC<sub>50</sub>) of DPP analogs have been previously investigated by *in vitro* kinase assay [37,38]. For IPA analogs, the Spatial Proximity Assay was used to illustrate the ability (IC<sub>50</sub>) of these inhibitors to modulate the phosphorylation of Retinoblastoma protein [39–42]. And for PYP group, the kinase phosphatase activity was monitored by using IMAF-FP assays and expressed by IC<sub>50</sub> values [43]. The IC<sub>50</sub> values were converted into the corresponding pIC<sub>50</sub> (–log IC<sub>50</sub>) values, which were further used as dependent variables in the QSAR analyses. Each group of the molecules was separated into two sets. The test set (Tables S1–S3 marked with <sup>a</sup>) was chosen by considering that the test compounds represent the structural diversity and a range of biological activities similar to that of the training set, which were used to determine the external predictivity of the derived models.

All molecular modeling and QSAR studies were performed using Sybyl package (Tripos Associates, St. Louis, MO). The Gasteiger–Hückel charges were added to DPP and PYP analogs, while Gasteiger–Marsili charges for IPA analogs were added. Energy minimization was used to compute the equilibrium configuration of the molecules, which employs the mathematical procedure of optimization to move atoms so as to obtain the stable conformational state of the molecular systems corresponding to the global and local minimum on their potential energy surface. In this work, the molecular mechanics method was applied to perform the energy minimization and conformational search, which were carried out in Sybyl by Tripos force field with the Powell conjugate gradient minimization algorithm. The energy gradient convergence criterion was set to 0.05 kcal/mol Å with the dielectric constant of 1.0 [44]. This molecular modeling process was carried out in vacuum. The minimized molecules were further employed to calculate the Dragon descriptors and construct the QSAR models.

### 2.2. Conformational sampling and alignment

Identification of the bioactive conformation and the molecular alignment are two important steps in a QSAR study [45]. Two different alignment rules were employed in this study, i.e., ligand-based alignment and receptor-based alignment. For the ligand-based alignment, the most active compound in each group was selected as a template (compounds 27, 122 and 126, respectively), and all the inhibitors for each group in the data set were then aligned to a common substructure (shown in Figs. 1–3) using the “align database” command in Sybyl software to produce the best possible models. The receptor-based alignment was based on the geometries obtained from docking.

### 2.3. QSAR analysis

In order to derive the CoMFA and CoMSIA descriptor fields, a 3D cubic lattice with a grid spacing of 2 Å and extending

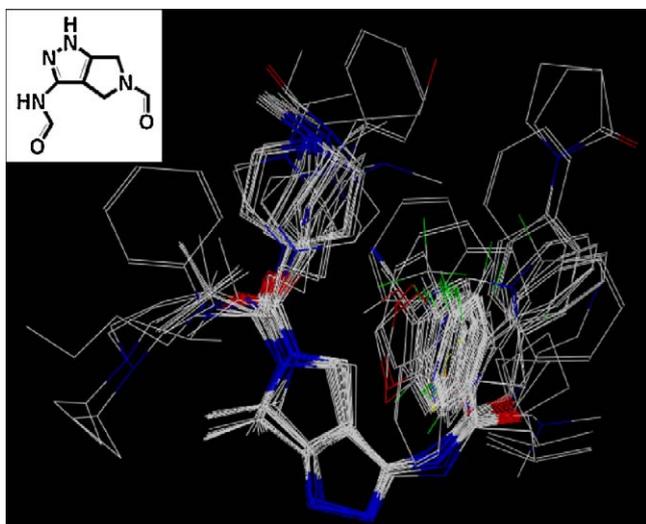


Fig. 1. Superimposition of DPP compounds in the training and test sets with common substructure shown in the upper left corner.

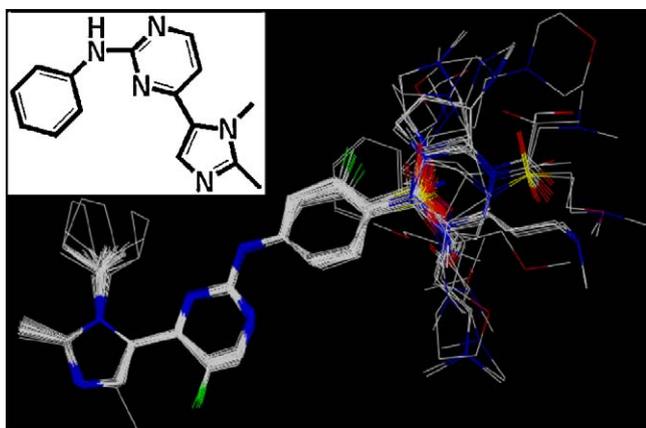


Fig. 2. Superimposition of IPA compounds in the training and test sets with common substructure shown in the upper left corner.

4 Å units beyond the aligned molecules in *x*, *y* and *z* directions were created. In CoMFA, steric interactions were calculated using a Lennard–Jones potential, while electrostatic interactions were calculated using a Coulomb potential. In CoMSIA, five different similarity fields: steric (*S*), electrostatic (*E*), hydrophobic (*H*), H-

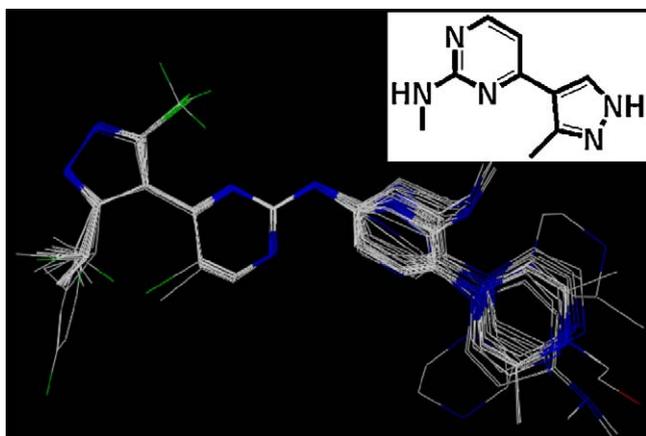


Fig. 3. Superimposition of PYP compounds in the training and test sets with common substructure shown in the upper left corner.

bond donor (*D*), and H-bond acceptor (*A*) were calculated. The Dragon descriptors as independent variables were also employed to develop quantitative models.

For the CoMFA analysis, *S* and *E* fields were scaled by the CoMFA-STD method with default energy of 30 kcal/mol. CoMFA descriptors were calculated using a  $sp^3$  carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0.

The CoMSIA models were also derived with the same lattice box that was used in CoMFA. However, it had several advantages over CoMFA such as greater robustness regarding both region shifts and small shifts within the alignments [44]. The CoMSIA descriptors were calculated using a probe atom with radius 1.0 Å, +1.0 charge, +1 hydrophobicity, +1 H-bond donor, and +1 H-bond acceptor properties. A Gaussian method was applied to evaluate the distance between the probe atom and each molecule atom. In general, similarity indices,  $A_{F,K}$  between the compounds of interest were computed by placing a probe atom at the intersections of the lattice points and using Equation (1):

$$A_{F,K}^q(j) = - \sum \omega_{\text{probe},k} \omega_{ik} e^{-\alpha r_{iq}^2} \quad (1)$$

where  $q$  represents a grid point,  $i$  is the summation index over all atoms of the molecule  $j$  under computation,  $k$  represents five physicochemical properties (*S*, *E*, *H*, *D* and *A*),  $\omega_{ik}$  is the actual value of physicochemical property  $k$  of atom  $i$ , and  $\omega_{\text{probe},k}$  is the value of the probe atom.  $r_{iq}$  is the mutual distance between the probe atom at grid point  $q$  and atom  $i$  of the test molecule.  $\alpha$  is the attenuation factor, an optimal value normally between 0.2 and 0.4. Larger values result in a steeper Gaussian function, and a strong attenuation of the distance-dependent effects of molecular similarity. In this work,  $\alpha$  was set to 0.3. With this selection, at a given lattice point the property value of an atom of the molecule under investigation (e.g., the partial atomic charge) is experienced in 1 Å distance by 74.1%, in 2 Å by 30.1% and in 3 Å by 6.7% of its total value, respectively.

Partial Least Squares (PLS) method [46] was used to linearly correlate the CoMFA and CoMSIA fields to biological activity values. The predictive value of the QSAR models was evaluated first by leave-one-out (LOO) method in which one compound was removed from the data set and its activity was then predicted by the model derived from the rest of the data set. PLS was conjunct with the cross-validation option to determine the optimum number of components ( $N_c$ ) and the cross-validated correlated correlation coefficient  $R_{cv}^2$ . The  $N_c$  was then used to derive the final QSAR model using all of the training set compounds with non-cross-validation and to obtain the conventional correlation coefficient ( $R_{ncv}^2$ ). Additionally, the CoMFA and CoMSIA results were graphically represented by field contour maps using the field type 'StDev\*Coeff'. To validate the CoMFA- and CoMSIA-derived models, the predictive ability for the test set of compounds (expressed as  $r_{\text{pred}}^2$ ) was determined by using the following equation:

$$r_{\text{pred}}^2 = \left( \frac{SD - \text{PRESS}}{SD} \right) \quad (2)$$

where PRESS represents the sum of the squared deviations between the predicted and the actual  $\text{pIC}_{50}$  values for the test set compounds, SD is the sum of the squared deviations between the biological activities of the test set molecules and the mean activity of the training set compounds.

#### 2.4. Molecular docking

To predict the appropriate binding orientation of CDK<sub>2</sub> inhibitors, the Surflex module in Sybyl package was employed to generate an ensemble of docking conformations. Surflex is

a methodology that combines Hammerhead's empirical scoring function [47] with a molecular similarity method (morphological similarity) [48] to generate putative poses of ligand fragments. Crystal structures of CDK<sub>2</sub> (2WPA, 2W17 and 2C68) were retrieved from RCSB Protein Data Bank (<http://www.rcsb.org>). The PDB files occasionally have missing residues, for which coordinates were not determined. For these three structures, 2WPA misses residues 299–304, which are located at the C-terminal end of the protein and do not affect the ligand binding pocket; 2C68 misses residues 38–43 and 297–298, which are distant from the active/binding site and are not involved in the ligand–receptor interactions. Therefore, all these missing residues are not further rebuilt in this work. As for 2W17 protein, the missing residues of 13, 39–43 and 151–162 were reconstructed using the fragment library of the Biopolymer module in Sybyl package (Tripos Associates, St. Louis, MO), since they participate in establishment of the active center in the loop area. All these structures were further optimized by MD approach in the solvent system to obtain appropriate geometry. The docking method employed an idealized active site ligand which was called a protomol [49] as a target to generate putative poses of molecules or molecular fragments to guide the molecular docking. During the docking procedure, relative ligands were extracted and polar hydrogen atoms were added to the template protein. Then the protomol was produced using ligand-based method: ligand location in the same coordinate space in the receptor. The definition of the protomol is an important step, and the docking performance depends on (1) the size, location and shape of the area considered to form the binding site and, (2) how far the site (protomol\_bloat) should extend from a potential ligand, and (3) how deep the atomic probes that are used to define the protomol can penetrate (protomol\_threshold) into the protein. Finally, each of the inhibitors was docked into the receptor 10 times and various scores were calculated for each conformation of the inhibitors to evaluate the docking analysis, i.e., D\_score, G\_score, Chemscore, PMF\_score and the total score [50]. The docking poses were saved for each compound, and were ranked according to the score function. The highest-ranking poses (conformations) for each compound, according to the docking score, were extracted and aligned together for further QSAR analysis.

### 2.5. Calculation and selection of Dragon descriptors

Dragon descriptors have been successfully used for quantitatively representing the structural and physicochemical features of a molecule in the development of various QSAR models for special pharmacological and biological properties [51,52]. In general, molecular descriptors can be collected from different sources such as substituent constants, physicochemical properties, quantum chemical calculations and the theoretical structural parameters derived from one or two dimensional molecular structures [53]. In this work, Dragon Professional, version 5.0 software [54] was employed to calculate the molecular descriptors from the optimized 3D structure of the molecules. It calculates 1644 molecular descriptors that are divided into 20 groups (shown in Table S4). To reduce the descriptor space and find more informative and appropriate descriptors, a stepwise linear regression method as the variable selection in R software: version 2.13.0 ([www.r-project.org](http://www.r-project.org)) was used [55]. Generally, the purpose of linear regression is to quantify the relationship between several independent (descriptors) variables and a dependent (bioactivity values) variable. The connectivity indices descriptor X<sub>5</sub>Av was well correlated with activity ( $R^2 = 0.25$ ) in the IPA dataset, whereas for the PYP model, RDF090v, H-046 and Hypertens-50 showed important influences with  $R^2 = 0.66$ . The purpose of employing the Dragon descriptors is to increase the robustness and generalization of the system. However, for DPP analogs, the original derived CoMFA or CoMSIA models

were statistically significant, thus the Dragon descriptors were not introduced in building models.

### 2.6. MD simulations

The MD simulations were performed using the GROMACS package 4.0.7 [56], the force field for the simulations described here was GROMOS96 [57]. The molecular topology files were generated by the program PRODRG2.5 [58]. Prior to MD simulation, each of the complexes was minimized without constraints using the steepest descent approach for 5000 steps to remove unfavorable atomic contacts. All complexes were solvated in cubic periodic box of SPCE water molecules. The box dimension was  $9.06 \times 9.06 \times 9.06$  nm,  $8.74 \times 8.74 \times 8.74$  nm and  $8.46 \times 8.46 \times 8.46$  nm, respectively for 2WPA, 2W17 and 2C68, and contained 21,593, 19,049 and 16,987 water molecules in each complex, respectively. Additionally, the minimum distance between the protein and the box walls was set to more than  $8 \text{ \AA}$  so that the protein does not directly interact with its own periodic image given the cut-off in each complex. Counter ions (three Cl<sup>-</sup>, five Cl<sup>-</sup> and seven Cl<sup>-</sup> ions, respectively) were added to neutralize the charge of each system, i.e., 2WPA.pdb and 2W17.pdb and 2C68.pdb, respectively. At the end, the total number of the atoms was 67,921 for 2WPA, 60,061 for 2W17 and 53,987 for 2C68 including the protein complexes and waters, respectively.

During the MD simulation, the temperature was kept constant at 300 K by the Berendsen thermostat [59], the coupling time was 2 ps while the pressure was maintained at 1 bar using the Parrinello–Rahman scheme [60], the coupling time was 2 ps. The electrostatic interactions were calculated using the particle mesh Ewald method [61]. The LINCS algorithm [62] was used to constrain bond lengths and the coupling time for MD integration is 2 ps. The values of the isothermal compressibility were set to  $4.5 \times 10^{-5} \text{ bar}^{-1}$  for water simulations; the cut-off distances for the calculation of Coulomb and van der Waals interactions were 1.0 nm and 1.4 nm, respectively. For the MD simulation, the time steps were 1 fs during the equilibration and 2 fs during the production dynamics. Snapshots were taken from the trajectory every 5 ps and subsequent result analysis was performed. All the simulations lasted for 5 ns to guarantee the stability of the whole system.

## 3. Results and discussion

### 3.1. QSAR model

In an attempt to build robust QSAR models, various parameters were considered during the QSAR studies, i.e., the cross-validated correlation coefficient ( $R_{cv}^2$ ), non-cross-validated correlation coefficient ( $R_{ncv}^2$ ), standard error of estimate (SEE) and *F*-statistic values (*F*), etc. *S* and *E* fields were generated for CoMFA models, while *H*, *D* and *A*, three other fields in addition to the *S* and *E* were generated for CoMSIA ones. Since it has been argued that the five different descriptor fields may not be totally independent of each other and such dependencies of individual fields usually decrease the statistical significance of the models [63,64], in the present work, all 31 possible combinations of the descriptors for each group were used to generate various QSAR models, with a full summation as represented in Tables S5–S10. It is clearly observed that no accurate models were obtained for IPA and PYP analogs without employing any Dragon descriptors. It is because the inhibitory activities of diverse set of compounds cannot be related to the single structural features of the molecules, and thus a combination of different structural features should be searched for building desirable models. Therefore, various 2D- and 3D-molecular descriptors belonging

**Table 1**  
Symbols of the descriptors used in the models and their definitions.

Symbols	Descriptor family	Definition
X <sub>5</sub> Av	Connectivity indices	Average valence connectivity index chi-5
RDF090v	RDF descriptors	Weighted by atomic van der Waals volumes
H-046	Atom-centered fragments	H attached to CO (sp <sup>3</sup> ) no X attached to next C
Hypertens-50	Molecular properties	Ghose–Viswanadhan–Wendoloski antihypertensive-like index at 50%

to a wide variety of structural features were considered for the IPA (X<sub>5</sub>Av) and PYP (RDF090v, H-046 and Hypertens-50) systems. The molecular descriptor types and definition of the symbols are shown in Table 1. Evaluated by statistical properties, the ligand-based approach showed superiority to the structure-based one and the obtained models are summarized in Table 2.

### 3.1.1. DPPs

For this class of inhibitors, the dataset consisting of 54 molecules is divided into a training set of 42 and a test set of 12 molecules considering the structural diversity and activity range.

The best CoMSIA-HD model gives an  $R_{cv}^2$  of 0.73 (>0.5) with an optimized component of seven, which suggest that this model should be considerably reliable to predict the IC<sub>50</sub> values. The non-cross-validated PLS analysis produces an  $R_{ncv}^2$  of 0.95,  $F$  value of 94.87 and a low SEE of 0.18. Contributions of  $H$  and  $D$  fields are 70.1% and 29.9%, respectively, which indicates that the  $H$  field makes more contributions to the ligand binding affinity.

The CoMFA model employing the  $S$  and  $E$  fields with an  $R_{cv}^2$  of 0.62 for four components is obtained (shown in Table S5), the PLS analysis results in an  $R_{ncv}^2$  of 0.896,  $F=79.36$  and a SEE value of 0.253. The  $S$  field descriptors explain 81.9% of the variance, while the  $E$  field descriptors explain 18.1%.

These models are further validated using an external test set compounds. Large  $R_{pred}^2$  values, 0.94 and 0.85, are generated for

**Table 2**  
The optimal QSAR models for the three types of CDK<sub>2</sub> inhibitors.

Parameters	DPPs	IPAs	PYPs
	CoMSIA	CoMSIA	CoMFA
$R_{cv}^2$	0.727	0.621	0.557
$R_{ncv}^2$	0.951	0.930	0.689
SEE	0.180	0.182	0.396
$F$	94.865	62.021	24.398
$R_{pred}^2$	0.944	0.626	0.575
SEP	0.427	0.422	0.473
Nc	7	10	2
Field contribution			
$S$	–	0.411	0.086
$E$	–	–	0.007
$H$	0.701	–	–
$D$	0.299	0.553	–
$A$	–	–	–
X <sub>5</sub> Av	–	0.036	–
RDF090v	–	–	0.459
H-046	–	–	0.220
Hypertens-50	–	–	0.228

$R_{cv}^2$  = Cross-validated correlation coefficient using the leave-one-out method;  
 $R_{ncv}^2$  = Non-cross-validated correlation coefficient; SEE = Standard error of estimate;  $F$  = Ratio of  $R_{ncv}^2$  explained to unexplained =  $R_{ncv}^2 / (1 - R_{ncv}^2)$ ;  
 $R_{pred}^2$  = Predicted correlation coefficient for the test set of compounds;  
 SEP = Standard error of prediction; Nc = Optimal number of principal components;  $S$  = Steric,  $E$  = Electrostatic,  $H$  = Hydrophobic,  $D$  = H-bond donor,  $A$  = H-bond acceptor.

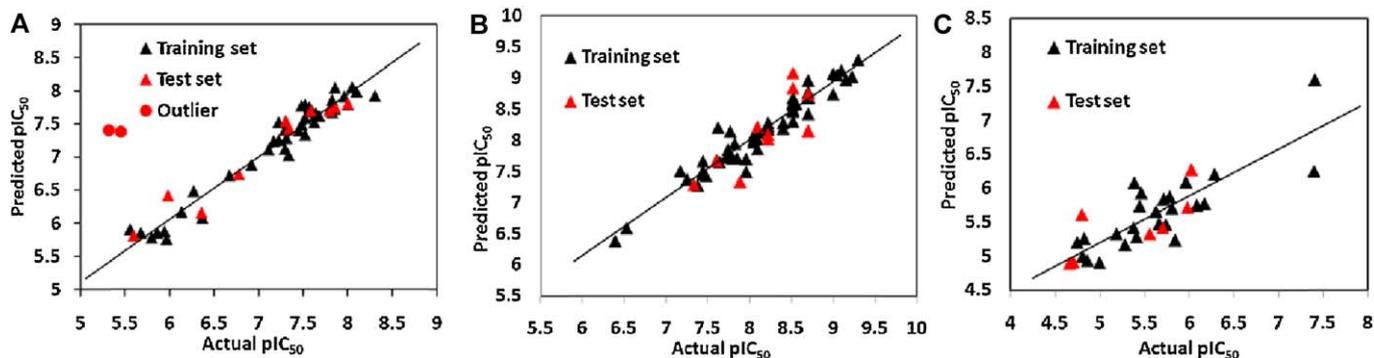
CoMSIA and CoMFA models, respectively, suggesting that these models all have good predictions. The correlation between the predicted and the actual activities is displayed in Fig. 4A. From this figure, we can see that all points are uniformly distributed along the regression line which suggests that no systematic error exists in the method. Additionally, two compounds 12 and 16 are detected as outliers, due to their low inhibitory activity and large residual between the experimental value and the predicted value that is nearly up to 2.0 log unit.

### 3.1.2. IPAs

During the modeling process, the data set is divided into a training set of 58 and a test set of 11 molecules. The statistical parameters obtained from the optimum CoMSIA-SD analysis are listed in Table 2. The PLS analysis gives a relatively high  $R_{cv}^2$  value of 0.62 at 10 components, suggesting that this model is a useful tool for predicting IPA inhibitory activities. The conventional regression coefficient  $R_{ncv}^2$  of 0.93 and SEE of 0.182 show its self-consistency. Statistical parameter for external validation,  $R_{pred}^2 = 0.63$ , indicates that the predicted activity coincides with the observed activity values for the test compounds and this model is reasonable with good external predictability. The respective contributions of  $S$  and  $D$  fields are 41.1% and 55.3%, respectively, indicating that the  $D$  field contributes more to the inhibitory activity. Additionally, one 2D-topological index X<sub>5</sub>Av [65] also makes 4% contributions to the model, effectively increasing the fitting degree of the model. In term of the X<sub>5</sub>Av, it is the average valence connectivity index of order 5, which encodes the presence of double and triple bonds as well as heteroatoms in the molecule such as N and O. In general, the double and triple bonds coupling with heteroatoms are key for the binding affinity of IPA analogue. Take compound 122 as an example, the heteroatom nitrogen atom at rings A and B (shown in Fig. 5B) serves as H-bond acceptor to interact with the protein, thus enhancing the inhibitory activity of the molecule. Additionally, molecular docking reveals that amino acids composed of benzene ring, such as Phe80 and Phe82, might produce  $\pi$ – $\pi$  interactions with ring B, which explains why the presence of this descriptor in the derived model is so important. The correlation between the predicted and the actual activities is displayed in Fig. 4B.

### 3.1.3. PYPs

In terms of PYPs derivatives, a total of 32 molecules are divided into a training set of 25 molecules and a test set of 7 molecules. The resulted optimum CoMFA model derived from the pools of different types of descriptors is listed in Table 2. The CoMFA model has an  $R_{cv}^2$  value of 0.56 and an  $R_{ncv}^2$  of 0.69. The model is developed with two components. It has an  $F$  value of 24.398 and an SEE value of 0.396, suggesting a good internal predictive ability of the model. The  $S$  and  $E$  contributions are found to be 8.6% and 0.7%, respectively, indicating the steric interactions could be important for the CDK<sub>2</sub> antagonistic activity. Three descriptors RDF090v, H-046 and Hypertens-50 show high contribution rate (22.8%, 22.0% and 45.9%) to the CoMFA model. For this reason, we believe that these descriptors are useful for increasing the reliability of the model. Generally, RDF090v is the radial distribution function at 9.0 Å interatomic distances weighted by atomic van der Waals volumes contributed positively. The RDF also provides valuable information about bond distances, ring types, planar and non-planar systems, atom types and other important structural motifs [66]. The H-046 is a short-range atom-centered descriptor that describes each atom by its own atom type and the bond types and atom types of its first neighbors [56]. A great number of hydrogen atoms present in the molecular structure attached to carbon sp<sup>3</sup> (H-046) may lead to an increase of the size of the molecule and the increase of the number of electrostatic interactions in specific reactive sites of the



**Fig. 4.** Graphs of the predicted versus the experimental  $pIC_{50}$  values of the optimal models. (A) CoMSIA model of DPPs analogs. (B) CoMSIA model of IPAs analogs. (C) CoMFA model of PYPs analogs.

inhibitor. Additionally, Hypertens-50 (molecular properties) is the Ghose–Viswanadhan–Wendoloski 50%-antihypertensive drug-like index.

Additionally, the CoMSIA-SE model also shows reasonable statistical features of  $R_{cv}^2$  0.556 with two components,  $R_{ncv}^2$  0.682,  $F$  value 23.56 and SEE value 0.4. Incorporation of other fields ( $H$ ,  $D$  and  $A$ ) also show comparable internal predictivity (shown in Table S9).

To test the predictive ability of the model, a test set of seven molecules excluded from the model derivation is employed. The  $R_{pred}^2$  is 0.58 (in Table 2) and 0.57 (in Table S9) for CoMFA and CoMSIA models, respectively. The graph of the actual versus the predicted  $pIC_{50}$  values of the training and test sets is illustrated in Fig. 4C, where the plots represent a uniform distribution around the regression line with respective slope and intercept very close to one and zero, indicating the satisfactory predictive capability of the model.

### 3.2. 3D-QSAR contour maps

One of the attractive features of the CoMFA and CoMSIA modeling is the visualization of the information content of the derived QSAR models. Presently, the contour maps were generated to rationalize the regions in 3D space around the molecules where changes in the fields were predicted to increase or decrease the activity.

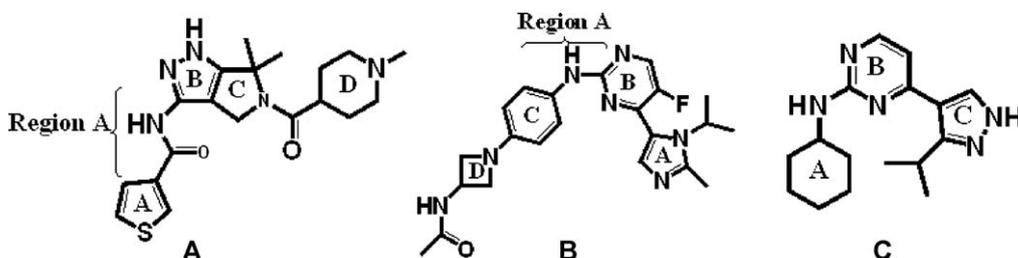
#### 3.2.1. DPPs

The CoMSIA hydrophobic and hydrogen bond (H-bond) donor field contour maps with compound 27 as a reference are shown in Fig. 6. In the hydrophobic contour map, the yellow contours indicate regions where hydrophobic groups would be favorable, while the white contours represent the regions where the hydrophobic groups would decrease the inhibitory activity. The H-bond donor field is represented by cyan and purple contours, in which the cyan

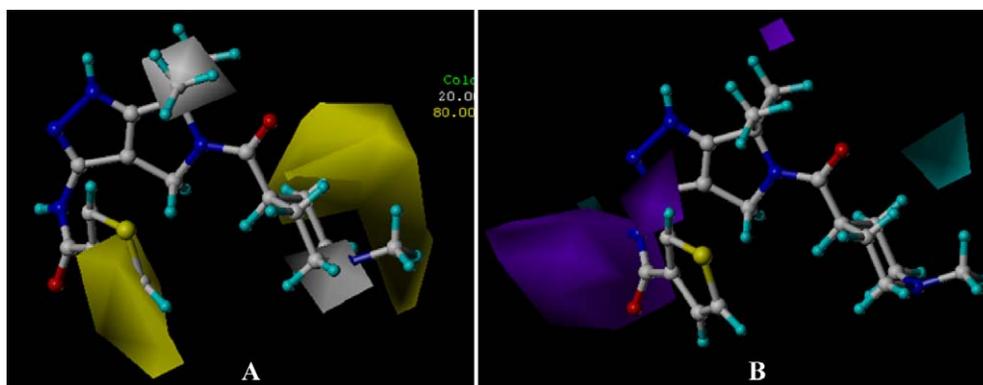
contours indicate areas where H-bond donor substitutions are preferred and purple regions represent areas where H-bond donor substitutions are disfavored.

In Fig. 6A, a large yellow contour near the  $R_1$  substituent indicates that hydrophobic substituent at this site would be favorable. It is consistent with the fact that compounds 12, 15 bearing acetyl groups and compound 14 possessing a hydroxyethyl substituent have higher potency than compound 13 with a methyl group at this position. In addition, another yellow contour is found at  $R_2$  region demonstrating that a hydrophobic substituent may be favorable. This may be the reason why compound 36 with a relative hydrophobic substituent (fluorophenylamino) shows higher activity than compound 16 with a fluorophenyl at this position. A white contour around  $R_3$  and  $R_4$  areas indicates that a hydrophobic substituent would be disfavored. The fact is that all of the derivatives involved in the present study possess hydrophobic substituents at this site; therefore, it is maybe better for the inhibitory activity if the hydrophobic substituent is transformed to other hydrophilic groups. Compound 27 possesses a hydrophobic methyl group at the end of  $R_1$  substituent (white contour), which is unfavorable for the inhibitory activity. Thereby, we suggest replacing the hydrophobic methyl with hydrophilic group which may be beneficial to the activity.

In Fig. 6B, only a few purple and cyan contours distribute in the contour map indicate that the H-bond donor field is not very important for the inhibitory activity. In fact, the H-bond donor makes smaller contribution (29.9%) than the hydrophobic one (70.1%) to the  $CDK_2$  inhibitory activity. One cyan contour mapped up the  $R_1$  substituent suggests that H-bond donor groups favor the activity. Compared compound 46 ( $-NH$ ) with compound 45 ( $-CH(CH_3)_2$ ), it could be easily found that their activity discrepancies could be explained by this cyan contour. This can also be explained by compound 47 with  $-CH_3$  in this region exhibited lower activity than compound 49 possessing  $-NH_2$  group. A large purple polyhedral around Region A (shown in Fig. 5A) implies that H-bond



**Fig. 5.** The most active molecule in each series of  $CDK_2$  inhibitors which are used in contour analysis. (A) Compound 27 in the group of DPPs analogs. (B) Compound 122 in the class of PYPs derivatives. (C) Compound 126 in the series of IPAs analogs.

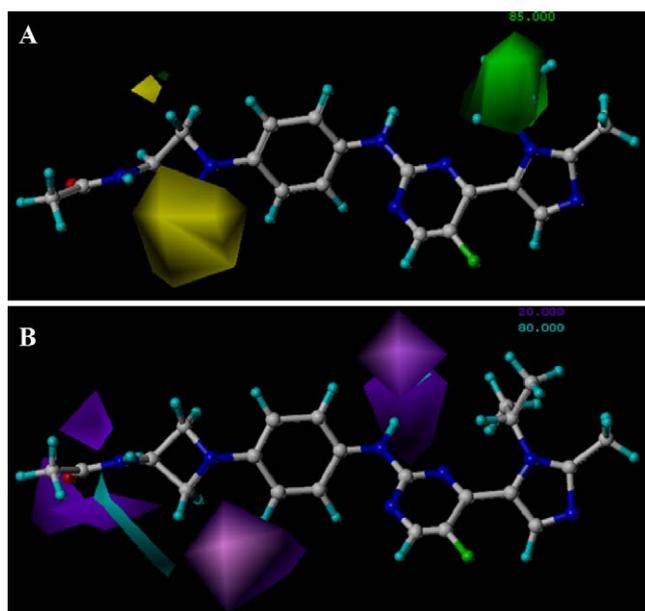


**Fig. 6.** CoMSIA StDev\*Coeff contour plots for DPPs analogs in combination of compound 27. (A) The hydrophobic contour map, where the yellow and white contours represent 80% and 20% level contributions, respectively. (B) The H-bond donor contour map, where the cyan and purple contours represent 80% and 20% level contributions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

donor groups could have a negative effect on the inhibitory activity. However, all the molecules in this work have  $-NH$  groups in this area suggest that there is room for modification at this position considering the effect of H-bond donor groups. In addition, another medium sized purple contour is observed around  $R_2$  substituent suggesting that any H-bond donor groups at this position are unfavorable for the inhibitory activity. This is consistent with the fact that there are no H-bond donor groups existing among all the molecules.

### 3.2.2. IPAs

The most potent compound 122 in this series is shown superimposed with the CoMSIA contour map in Fig. 7. The steric fields are represented by green- and yellow-colored contours (green, bulky substitution favored; yellow, bulky substitution disfavored). The H-bond donor fields are indicated by cyan- and purple-colored contours (cyan, H-bond donor favored; purple, disfavored).



**Fig. 7.** CoMSIA StDev\*Coeff contour plots for IPAs analogs in combination of compound 122. (A) The steric contour map, where the green and yellow contours represent 85% and 15% level contributions, respectively. (B) The H-bond donor contour map, where the cyan and purple contours represent 80% and 20% level contributions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

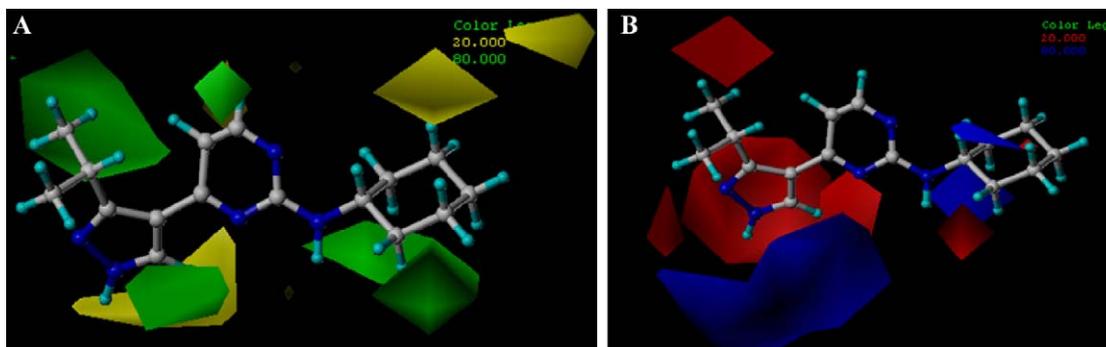
The steric favorable green contour at  $R_5$  substituent indicates that bulky groups may promote the  $CDK_2$  inhibitor activity. Compounds 74 and 75, where the substituents are  ${}^cPrCH_2$  and  ${}^cPhetyl$ , respectively, fit into the green contour and their activities are thus higher while compound 73 with ethyl substituent shows less activity. Additionally, the plot shows that the  $R_1$  substituent of molecule 122 places between one big yellow polyhedron and a small yellow contour. Thus a bulky substituent occupying this region would depress the biological activity. Compound 101, with  $-SO_2(CH_2)_3N(CH_2CH_2)_2O$  substituent exhibits lower activity than compound 100, with  $-SO_2(CH_2)_3N(Me)_2$  at this position. This contour also could be used to explain why those molecules substituted by  $-SO_2NH_2$  are much less potent than molecules 103, 104, 105 and 106 ( $R_1 = SO_2NCH_3$ ,  $SO_2NCH_2CH_2OCH_3$ ,  $SO_2NCH_2CH_2CH_2OCH(CH_3)_2$  and  $SO_2NCH_2CH_2N(CH_3)_2$ , respectively).

The H-bond donor field contour map of CoMSIA is shown in Fig. 7B. The large cyan contour near the  $-NH$  of Region A (shown in Fig. 5B) demonstrates the extreme significance of the H-bond donor group, and if a H-bond acceptor occupies this contour, the inhibitory activity would be depressed, as seen in compound 123. Another cyan contour is located between rings B and C. In fact, all of the molecules bear  $-NH$  groups at this position which are benefit for the activity. In addition, three purple contours around the  $-CO$  group of Region A depict the unfavorable interactions with the H-bond donor groups. In compound 122, the  $-CO$  group serving as the H-bond acceptor is favorable for the activity and its replacement would lead to lower activity such as compounds 58, 59 and 60 with  $-NH$  extended to this contour.

### 3.2.3. PYPs

Fig. 8 shows the steric and electrostatic maps for the CoMFA model on compound 126, which is the most active compound in this series of inhibitors. The steric field contours are shown in green (more bulk favored) and yellow (less bulk favored), while the electrostatic field contours are shown in red (electronegative substituents favored) and blue (electropositive substituents favored), respectively.

The steric field is shown in Fig. 8A. Two yellow contours far away from the  $R_1$  substituent indicate bulky groups at this position would decrease the potency. This may explain why compounds 141, 143–155 with bulkier groups touch this contour showing significantly decreased activities. A large region of yellow contour at  $R_3$  substituent suggests that bulky groups would decrease the activity. The activity of compounds 141, 143–154 decreases after the  $-Cl$  is replaced with bulky groups, such as compound 155 ( $-CF_3$ ). A large green contour is found below the  $R_1$  substituent indicat-



**Fig. 8.** CoMFA StDev<sup>Coeff</sup> contour plots for PYPs analogs in combination of compound 126. (A) The steric contour map, where the green and yellow contours represent 80% and 20% level contributions, respectively. (B) The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

ing that bulky substituents at this position would be favorable for the CDK<sub>2</sub> inhibitory activity. This is in accordance with all the molecules employed in the present study bearing large groups in this area. In addition, another large green contour around R<sub>2</sub> substituent indicates areas where steric bulk would enhance the inhibitory activity. This could be explained by the order of the activity of some compounds, such as the activity of compounds 126–128 (isopropyl) > 124–125 (methyl).

In the contour map of the electrostatic field of CoMFA model, shown in Fig. 8B, a large and three small red contours around R<sub>2</sub> and R<sub>3</sub> substituents indicate that electronegative groups in this region benefit the potency. This could be explained by the fact that the activity of compound 50 (–Cl) is higher than compound 51 (–CH<sub>3</sub>). On the other hand, all molecules have electropositive substituents at R<sub>2</sub> region suggesting that there is room for modification in this area. The electrostatic contours also show blue polyhedra outside the R<sub>1</sub> substituent where electropositive groups are expected to increase the activity. Hence, compounds 143–155 having electronegative groups (–Cl) attached to this contour exhibit decreased activity.

### 3.3. Docking analysis and comparison with 3D-contour maps

In the present study, molecular docking was employed to understand the interactions between CDK<sub>2</sub> and its inhibitors. The advantage of using docking-based alignment is that molecular docking explores the binding pocket more extensively. In our work, all the molecules in the database were docked into the active site of CDK<sub>2</sub> crystal structures. The docking scores of the most potent ligand 27 for 2WPA, 122 for 2W17, and 126 for 2C68 are 7.02, 5.10 and 5.00, respectively.

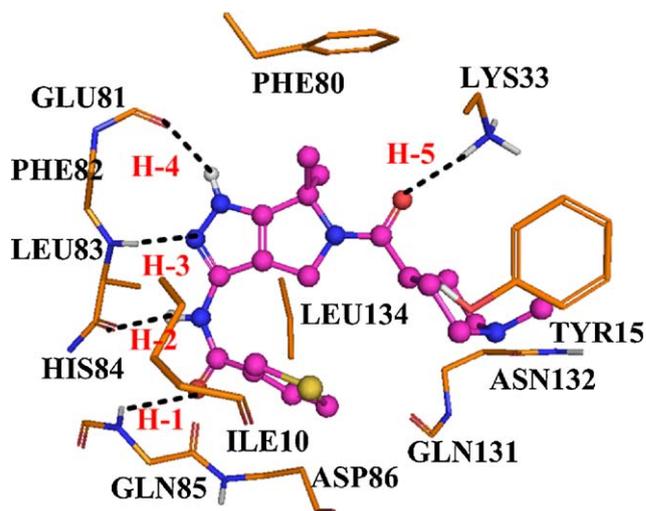
#### 3.3.1. DPPs

To elucidate the interaction mechanism, compound 27 is selected for more detailed analysis. The best possible interacting model of compound 27 with CDK<sub>2</sub> is depicted in Fig. 9. The analysis shows that Glu81, Phe82, Leu83, Gln85, Ile10, Asp86, Gln131, Asn132, Tyr15, Leu134, Lys33 and Phe80 are the most important residues present at the active site. The docked model reveals that the –NH group between rings A and B (shown in Fig. 5A) acts as H-bond donor to form H-bond with the oxygen atom of carbonyl of Leu83 (–O...HN, 1.88 Å, 137.9°) (H-2), so negative charged substituents in this area are favored for the inhibitor activity. The nitrogen atom of ring B forms H-bond with the –NH of Leu83 (–N...HN, 2.23 Å, 175.5°) (H-3), and the –NH group of ring B forms H-bond with the oxygen atom of Glu81 (–O...HN, 2.34 Å, 153.2°) (H-4). In addition, the oxygen atom of carbonyl between ring C and ring D forms H-bond with backbone of Lys33 (–O...HN, 2.13 Å, 172.6°) (H-5), and thus the positively charged substituents around

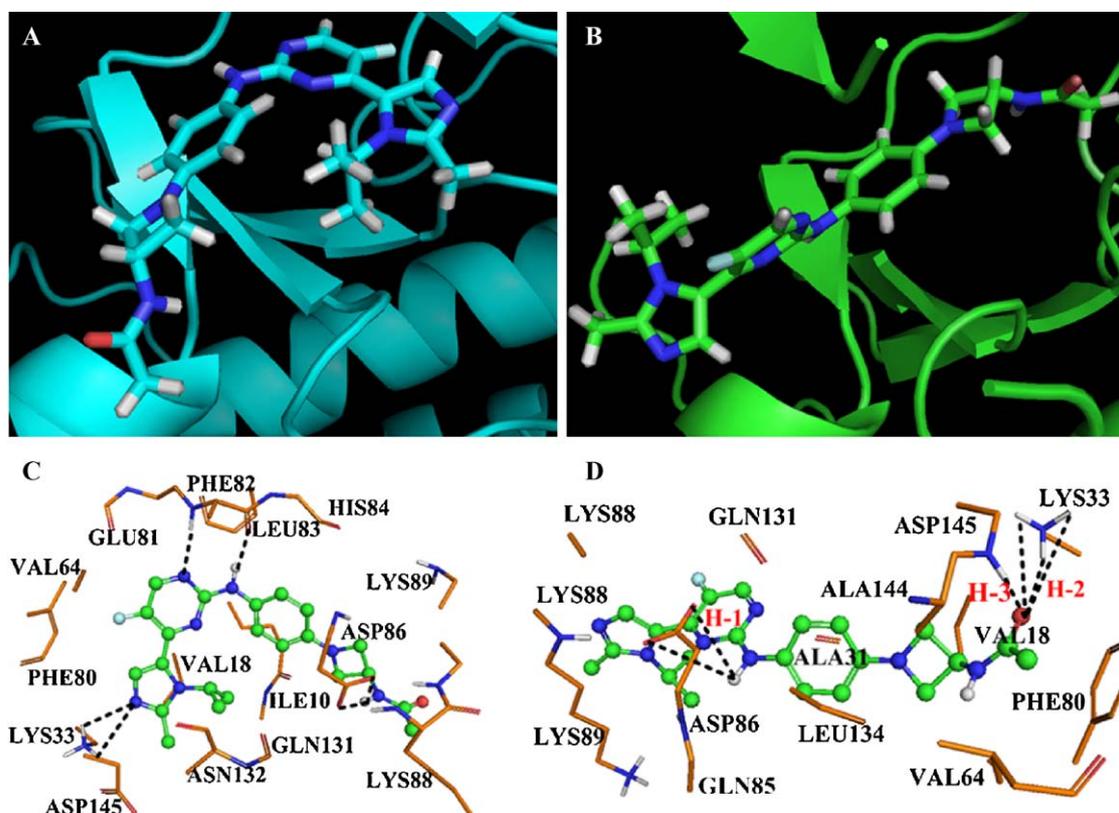
the carbonyl oxygen are favored for the activity. Moreover, the oxygen atom at the linker, between rings A and B, forms H-bond with the hydrogen atom of –NH group of Gln85 (–O...HN, 3.89 Å, 59.7°) (H-1), which is consistent with the H-bond donor contour map (Fig. 6B). The small purple contour around this position suggests that this is an H-bond acceptor favorable region. The hydrophobic favorable yellow contour at R<sub>2</sub> substituent is enclosed by Leu34 and Ile10, which explains the increased activity after introduction of hydrophobic groups to this area. As displayed in Fig. 9, the hydrophobic amino acid residues Gln131, Tyr15 and Asn132 located around R<sub>1</sub> substituent form a hydrophobic binding pocket, which exerts strong hydrophobic interactions with the hydrophobic substituents of the Ligand, which correlates well with the analysis of the hydrophobic contour map (shown in Fig. 6A).

#### 3.3.2. IPAs

Generally, during the docking procedure, two criteria are applied to select the best-docked poses: (1) ligand binding position; (2) scores comparison. The best ligand binding position is often determined by the RMSD between the docked and the crystallographic model. In order to find the correct orientation of the ligand in the binding site, compound 122 was docked into the active site of CDK<sub>2</sub> receptor 10 times, the highest scored pose (pose A with docking score of 8.86) is shown in Fig. 10A and C. The



**Fig. 9.** Docked conformation derived for compound 27 in complex with the active site of CDK<sub>2</sub> enzyme (2WPA). Compound 27 is displayed in ball and sticks, H-bonds are shown as dotted green lines, respectively, and the nonpolar hydrogens were removed for clarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



**Fig. 10.** (A–B) Docked conformation derived for compound 122 in complex with the active site of CDK<sub>2</sub> enzyme (2W17) (pose A and pose B). (C–D) The enlargement for the ligand in the binding site for pose A and pose B, the ligand is displayed in ball and sticks, H-bonds are shown as dotted black lines, and the nonpolar hydrogens were removed for clarity.

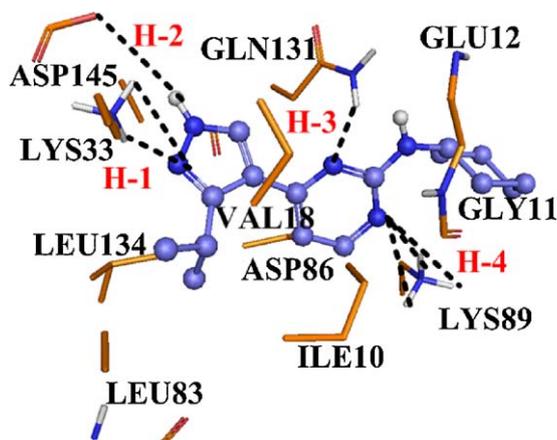
ligand core is anchored in the binding site via six H-bonds. The –NH group between rings B and C involves in hydrogen bonding interaction with the backbone of Leu83. The –NH group which is located outside ring D acts as an H-bond donor to form two H-bonds with the –COO of Asp86. Ring A forms two H-bonds with the –NH<sub>2</sub> of Lys33. Ring B also forms H-bond with Leu83. Surprisingly, the docking results are conflict with the contour maps shown in Fig. 7. For example, a large green contour map appears at the R<sub>5</sub> substituent, suggesting that in this area the steric interaction may be favorable; however, the docking model shows a hydrophilic pocket around R<sub>5</sub> substituent, suggesting that too large groups may introduce severe steric clash. A large yellow contour is mapped at the R<sub>1</sub> substituent; conversely, the docking model reveals that this substituent is almost located outside the binding pocket. In addition, rings A and B serve as H-bond acceptor to interact with the protein, which are not in agreement with the CoMSIA contour map (Fig. 7B) without purple contour near these positions. Thus the docked complex is submitted to MD simulation to further study their characteristics and stabilities. After MD simulation, the R<sub>5</sub> substituent (hydrophobic) is fitted into a hydrophilic pocket composed of Gln124, Asn125, Glu12 and Thr13, indicates that large ligand with surrounding residues causes the steric hindrance, which was unfavorable for the activity. Additionally, the distance between fluorine and the oxygen atom of Glu74 is 3.14 Å, which might result in inappropriate electrostatic repulsion. It is evident that the long tail of ring D turns ~90° relative to its initial structure (shown in Fig. 14A), and large fluctuations reminiscent of the unstabilization of the ligand, which further indicates that the docking protocol may be unreasonable.

According to 3D-contour indication and our chemical intuition, pose B (docking score of 5.10) was selected for further analysis. The possible interacting model and the main residues involved

in the interaction are depicted in Fig. 10B and D. The –NH group between rings B and C is H-bonded with Asp86 (–O…HN, 3.05 Å, 69.7°, –O…HN, 3.45 Å, 52.4°) (H-1). The oxygen atom which is located outside ring D acts as an acceptor to form H-bond with the –COO of Asp145 (–O…HN, 2.98 Å, 122.9°) (H-3), and it is also involved in hydrogen bonding with the backbone of Lys33 (–O…HN, 2.19 Å, 158.3°, –O…HN, 3.77 Å, 46.4°, –O…HN, 3.45 Å, 65.4°) (H-2). The docking model exhibits a comparatively large empty cavity around R<sub>5</sub> substituent, indicating that in this area the steric interaction may be favorable, which is consistent with the contour map (shown in Fig. 7A) with a green contour in this region. In addition, amino acids Leu134, Ala31, Val18 and Ala144 appear around R<sub>1</sub> substituent, which is in accordance with the steric contour map. Due to the H-bond formed by the oxygen atom outside ring D and the backbone of Lys33 and Asp145, the oxygen atom in the ligand serves as an H-bond acceptor and amino acids as H-bond donors, which is evident from the presence of two purple contours nearby. The –NH group appearing between rings B and C forms H-bond with Asp86, suggesting the importance of H-bond donor groups at this position, which is supported by the presence of a favorable cyan contour near ring D (CoMSIA H-bond donor).

Additionally, the RMS difference between the docked ligand and the X-ray structure of the ligand which the interaction mode is similar to compound 122 is 3.47 Å for pose A, 0.29 Å for pose B, respectively, suggesting the reasonability of pose B and the accuracy of the docking procedure.

These analyses suggest that pose B might be the reliable docking-predicted conformation, regardless of the relatively lower score. However, for pose B, some H-bond patterns also show non-typical geometric parameters, that is, the H-bond distance between Asp86 and ligand is 3.05 Å, but with an angle of only 69.7°. Thus, in order to confirm the binding modes of the ligand and to optimize



**Fig. 11.** Docked conformation derived for compound 126 in complex with the active site of CDK<sub>2</sub> enzyme (2C68), compound 126 is displayed in ball and sticks, H-bonds are shown as dotted green lines, respectively, and the nonpolar hydrogens were removed for clarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the structures of the final complex, MD simulation was performed. (The detailed analysis is shown in Section 3.5.2).

### 3.3.3. PYPs

The binding mode of compound 126 is shown in Fig. 11, involving some key residues such as Lys89, Asp86, Gln131, Asp145, Lys33, Leu83, Ile10, Gly11, Val18, Leu134 and Glu12 in the active site. It is found that the nitrogen atom of ring B forms three H-bonds with backbone of Lys89 ( $-N\cdots HN$ , 2.02 Å, 141.3°;  $-N\cdots HN$ , 3.16 Å, 64.6°;  $-N\cdots HN$ , 3.66 Å, 33.5°) (H-4). Ring B and the backbone  $-NH$  in Gln131 also form an H-bond ( $-N\cdots HN$ , 2.66 Å, 140.9°) (H-3). The nitrogen atom of ring C acts as H-bond acceptor to form two H-bonds with Lys33 ( $-N\cdots HN$ , 1.90 Å, 145.5°;  $-N\cdots HN$ , 2.97 Å, 69.7°) (H-1). In addition, ring C is seen involved in hydrogen bonding interaction with the carboxyl oxygen atom of Asp145 ( $-O\cdots HN$ , 3.26 Å, 118.1°) (H-2).

Our docking model also shows that the R<sub>1</sub> substituent extends outside the cavity, where a steric interaction may be favorable, which is validated by a sterically favorable green contour at this position in the CoMFA model (Fig. 8). Moreover, the electropositive favorable blue contour near the R<sub>1</sub> substituent is surrounded by some neutral amino acids Gly11, Ile10 and electronegative amino acids Glu12, Asp86 which can explain the increased inhibitory activity expected from the introduction of electropositive groups in this region. At R<sub>3</sub> substituent, introduction of too large groups would potentially lead to steric clash with residues Lys33 and Asp145, which is evident from the presence of a yellow contour at this position. It also shows that the R<sub>2</sub> substituent is fitted nicely into a large pocket; therefore, bulky groups are needed, correlating well with the CoMFA model with a green contour at this position. Additionally, an electronegative favored red contour is surrounded by some amino acids Ala31 (not shown in Fig. 11), Asp86 and Leu134, which explains the increased activity after introducing electronegative groups to this area.

The results of the docking studies and QSAR can complement and validate each other, indicating that the QSAR models developed are reasonable and can offer constructive suggestions to further modification of CDK<sub>2</sub> inhibitors.

### 3.4. Comparison of binding modes for each class

The docking results of the three different classes of CDK<sub>2</sub> inhibitors are compared to explore their similarities and differences

and to get better understanding of the variations in the biological activities. The docking study has revealed that the hydrogen bonding is an important interaction between the inhibitors and the receptor.

Interestingly, only one common amino acid residue Lys33 is discovered among these three series of CDK<sub>2</sub> inhibitors. Asp145 and Lys33 residues are found to possess hydrogen bonding interactions with both the inhibitors (compounds 122 and 126). Therefore, it can be concluded that Lys33 and Asp145 are vital amino acids for the ligand–receptor interactions in CDK<sub>2</sub> kinase.

On the other hand, the three classes of inhibitors all form more than five H-bonds with CDK<sub>2</sub> kinase, indicating that they all possess potent inhibitory activities. However, PYPs form seven H-bonds which are more than DPPs and IPAs (Fig. 12A, B and C), indicating that PYPs analogs may be more potent than DPPs and IPAs analogs.

### 3.5. Molecular dynamics simulations

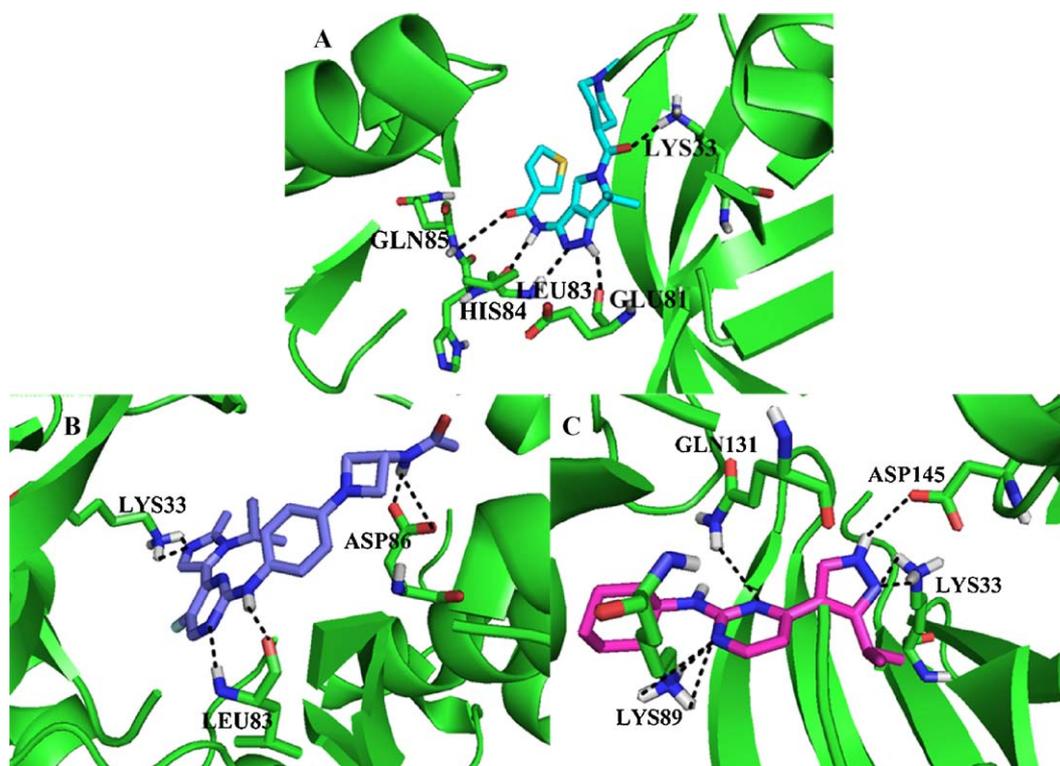
To further explore the conformational alterations that take place in the ligands and the receptor, MD simulations were performed on three docked complex structures, i.e., 2WPA + ligand 27, 2W17 + ligand 122, 2C68 + ligand 126 for a total of 5 ns.

#### 3.5.1. DPPs

The RMSDs of the trajectory with respect to their initial structure ranging from 1.5 Å to 2.5 Å are depicted in Fig. 13A (colored in red). After 2 ns, the RMSD of the complex reaches about 2 Å and retains this value throughout the simulation. This clearly suggests a stable conformation after 2 ns of simulation for docked complex structure. Fig. 13A also shows the RMSD of the ligand in the binding site versus the simulation time. As expected, the RMSD of the ligand stabilizes at 0.9 Å, while there is a 0.7 Å increase between 2.4 ns and 2.6 ns and the jump is then stabilized further between 2.6 ns and 2.7 ns. This result reveals that the ligand may switch between two kinds of conformations in the binding site, and finally maintains the initial structure. This can explain why the position in the ligand and the amino acids involved in the hydrogen bonding are conserved. However, the length and the number of the H-bonds have undergone changes. For example, the distance between the  $-NH$  group at Region A (shown in Fig. 5A) and Leu83, the nitrogen atom of ring B and Leu83, the  $-NH$  group of ring B and the backbone of Glu81 is shortened from 1.88 Å to 1.78 Å, 2.23 Å to 1.71 Å and 2.34 Å to 1.81 Å, respectively, after the MD simulation. For Lys33, before the MD simulation, only one H-bond was formed with the ligand, after 5 ns simulation, the ligand moves much closer to the amino acid residue Lys33 to form more H-bonds. These results suggest that the ligand moves deeper into the binding pocket during the MD simulation, which is consistent with the report which highlighted that an inhibitor with its ability to enhance the hydrogen bonding capacity to the residues located in the hinge region (residues 81–84) of CDK<sub>2</sub> would certainly possess a greater affinity towards the active site [67].

#### 3.5.2. IPAs

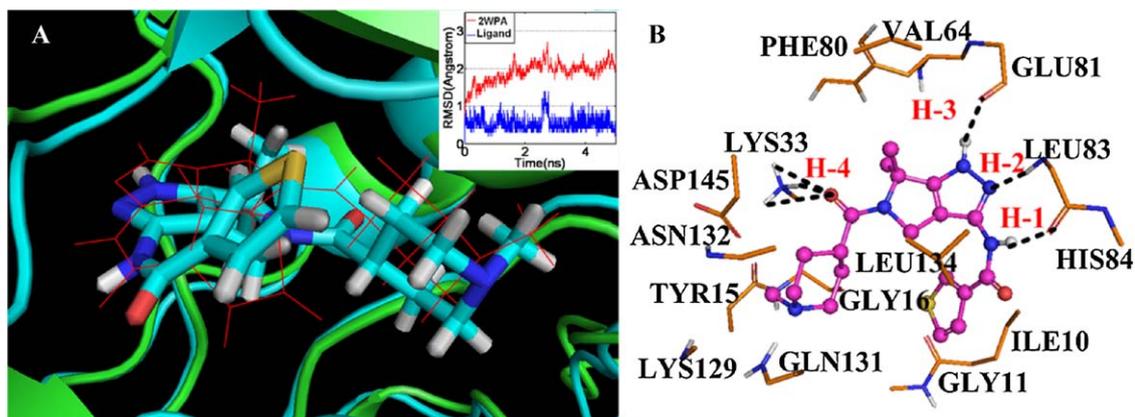
To take into account of the protein flexibility, the behavior of the docked complex (pose B) is studied in a dynamic context. Fig. 14D shows the RMSD of Ca backbone with respect to the initial structure (in red), which clearly shows that the RMSD reaches 2.2 Å, suggesting that a relatively stable conformation of the protein is achieved through the MD simulation. Fig. 14D also gives the RMSD of the ligand (in blue) in the protein binding site, where, interestingly, the RMSD reaches 2 Å during the first 2 ns, then levels off to 1 Å in the following 1 ns. It also increases to 2 Å again from 3 ns to 3.5 ns, after which the RMSD descends back to 1 Å again. This result reveals that compound 122 may switch between two kinds of conformations in the binding site of CDK<sub>2</sub>, which is confirmed by the well superposi-



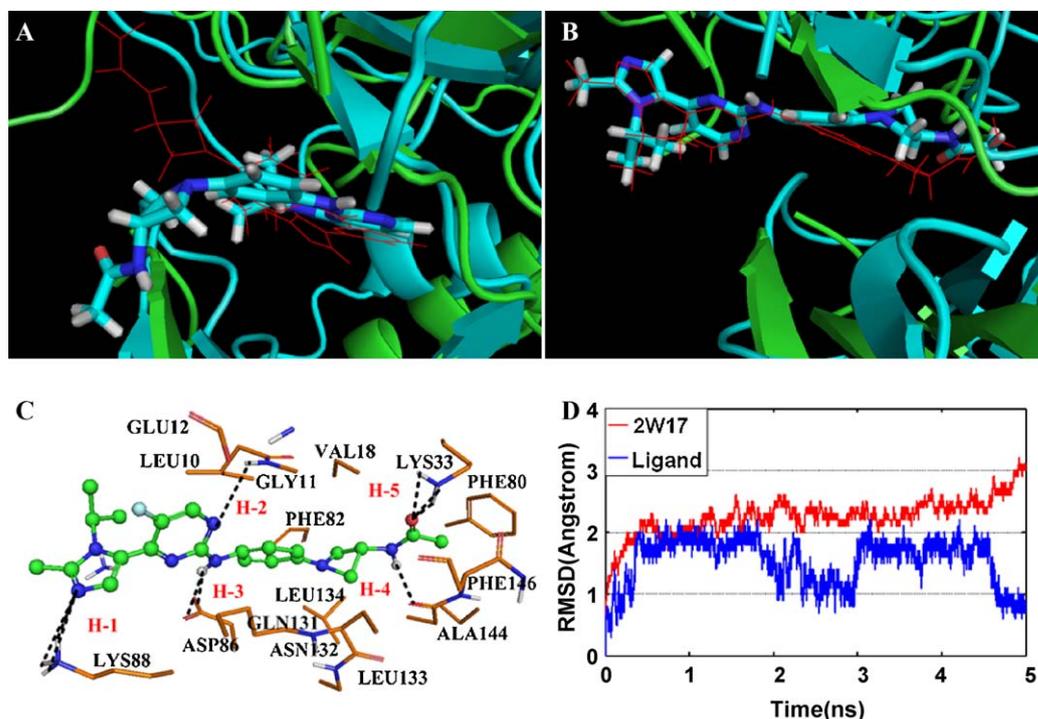
**Fig. 12.** Stereoview of the docked conformations of compounds 27, 122 and 126, respectively, in the active site of CDK<sub>2</sub> kinase. The hydrogen bonds are shown by broken lines. Compounds 27, 122 and 126, colored cyan, blue and purple, are presented in pictures A, B and C, respectively. The important amino acid residues, Lys33, Gln85, His84, Leu83, Glu81, Asp86, Lys89, Gln131 and Asp145 (stick rendering) are colored by atom type (C, green; N, blue; H, white; O, red).

tion of the coordinates of the protein after MD simulation onto the initial conformation (shown in Fig. 14B). Furthermore, the stability of the hydrogen bonding network predicted by molecular docking method is further examined by analysis of the predicted H-bonds during the simulation time. The main findings are summarized as follows: (1) the main position involving hydrogen bonding in the ligand is conserved after MD simulation; however, the H-bond length and angles have altered. (2) The H-bonds (the –NH between ring B and ring C...CO of Asp86 at H-3 region, the O outside ring D...NH<sub>3</sub> of Lys33 at H-5 area) are preserved after MD simulation. (3) After MD simulation, the ligand–receptor complex has more

possibilities for hydrogen bonding, ring A and Lys88 (H-1), ring B and Gly11 (H-2), the –NH outside ring D and Ala144 (H-4). (4) The loss of the H-bond between the oxygen atom outside ring D and the backbone of Asp145 is compensated by the newly formed H-bond between the –NH outside ring D and Ala144. (5) The H-bond between the ligand and Asp86 did not show typical parameters for H-bond angle (69.7°) during the docking process. However, after MD simulation, the angle changes to 93.3° which coincides with the typical geometrical parameters. Although the ligand has undergone several movements to yield different conformations during the MD simulation, the binding site is always stable, lending credit to the



**Fig. 13.** (A) The root-mean-square deviation (RMSDs) of docked complex (in red) and ligand (in blue) versus the MD simulation time in the MD-simulated structures shown in the upper right corner. Plot of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound 27–2WPA complex. Compound 27 is represented as line in red for the initial complex and stick in cyan for the lowest energy complex, respectively. (B) Plot of the MD-simulated structures of the binding site with compound 27. H-bonds are shown as dotted black lines; active site amino acid residues are represented as sticks; the inhibitor is shown as stick and ball model, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



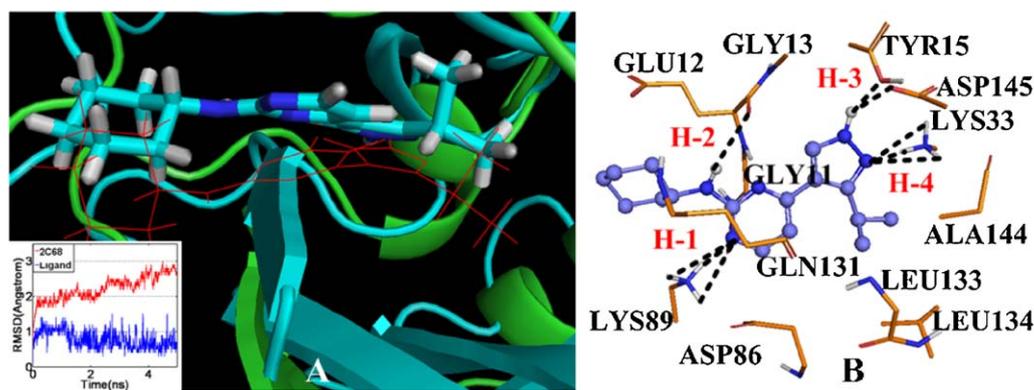
**Fig. 14.** (A) View of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound 122-2W17 complex (pose A). (B) Plot of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound 122-2W17 complex (pose B). (C) Plot of the MD-simulated structures of the binding site with compound 122. H-bonds are shown as dotted black lines; active site amino acid residues are represented as sticks; the inhibitor is shown as stick and ball model, respectively. (D) The root-mean-square deviation (RMSDs) of docked complex (in red) and ligand (in blue) versus the MD simulation time in the MD-simulated structures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

reliability of the active conformation (pose B) obtained by docking. The results also indicate that the MD simulation obligates the ligand to optimize its orientation and distance to the binding site for maximum interaction with receptor.

### 3.5.3. PYPs

The RMSDs from the ligand–protein complex versus simulation time is depicted in Fig. 15A (colored in red), which reveals that the RMSD value for the complex has a rising in the first 2.5 ns and then maintains stable in the rest of the simulation time. Fig. 15A also shows the evolution of the ligand along the reaction coordinate (colored in blue). For ligand 126, it moves away from the initial conformation, occurring about 3 Å, which is confirmed by the RMSD

of the ligand inside the binding pocket. Interestingly, the RMSD of the ligand reaches about 1.4 Å during the first 1.5 ns, then descends to 1.2 Å and holds to an altered state till the end of the simulation. This result reveals that the structure of the ligand is stable after 5 ns simulation. The ligand binding induces perturbations in the average H-bonding interaction pattern which is found in the docking model. The analysis of the binding mode of compound 126 after the MD simulation suggests the presence of several H-bonds between the inhibitor and CDK<sub>2</sub> kinase with higher frequencies. The alteration of this ligand in the binding pocket impedes a conformational change and stabilizes the whole complex. At the end of the MD simulation, the position of the ligand in the binding site is changed; however, its orientation has been reserved. The MD sim-



**Fig. 15.** (A) The root-mean-square deviation (RMSDs) of docked complex (in red) and ligand (in blue) versus the MD simulation time in the MD-simulated structures shown in the lower left corner. Plot of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound 126-2C68 complex. Compound 126 is represented as line in red for the initial complex and stick in cyan for the lowest energy complex. (B) Plot of the MD-simulated structures of the binding site with compound 126. H-bonds are shown as dotted black lines; active site amino acid residues are represented as sticks; the inhibitor is shown as stick and ball model, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

ulation on the complex reveals that, except for Gly11, Glu12, Lys33, Asp86, Lys89, Gln131, Leu134 and Asp145, the rest of residues in the active site determined by docking have changed and some new residues such as Tyr15, Leu133 and Ala144 are located in proximity of the ligand and could be significant for the interaction. At the end of MD simulation, two novel H-bonds are formed between the docked compound and amino acids Glu12 and Tyr15, in addition, the previous H-bond between the ligand and Gln131 has vanished. The crystal structure has uncovered that Glu12, Tyr15 and Gln131 are equally distributed on both sides of the binding groove; the ligand interacts with residues Glu12 and Tyr15 after MD simulation, further inducing the alteration of the ligand binding position to be much closer to these residues. Additionally, these changes may be explained by the fluctuation of the T-loop residues (146–166), which becomes labile when bound to the ligand rather than in the unliganded form [67].

Here, the MD simulations present an attractive alternative for structural refinement of the final docked complexes. They incorporate the flexibility of both the ligand and the receptor, improving interactions and enhancing complementarities between them. In addition, the evolution of the whole system during the simulation provides a dynamic picture of the complex. The stable RMSD indicates the reliability of the docking procedure. In a word, the conformations obtained after molecular dynamics are more reasonable than the docked conformations. The reason is that the docking procedure, on one hand, ignores the flexibility of the protein, which is not permitted to adjust its conformation upon the ligand binding, and on the other hand, lacks a unique and widely applicable scoring function necessary to generate a reliable ranking of the final complexes. Whereas, the MD simulation treats both the ligand and the protein in a flexible way, allowing for an induced fit of the receptor-binding site around the ligand [68].

#### 4. Conclusions

In this study, the ligand- and receptor-based QSAR, molecular docking and molecular dynamics methods, for the first time, are employed to explore the structure–activity relationship of three different series of CDK<sub>2</sub> inhibitors. The QSAR models exhibit good  $R_{cv}^2$  values, suggesting their good predictive ability. The characteristics of the 3D-contour plots produced by the models provide useful information about the intermolecular interactions of the inhibitors with the surrounding environment. The docking results propose a general binding mode of these ligands to the receptor and determine the key residues in the binding pocket of the CDK<sub>2</sub> enzyme. To validate the binding mode and elucidate the effects of ligand binding on the receptor, MD simulation is performed. The main findings are summarized as follows:

- (1) For DPPs, the CoMSIA model yields relatively better prediction than the CoMFA one. The contour map reveals that hydrophobic and H-bond donor groups at R<sub>1</sub> region, hydrophobic and H-bond acceptor substituents at R<sub>2</sub> region and hydrophilic groups at R<sub>3</sub> and R<sub>4</sub> areas would be beneficial to the activity. The majority of interactions have been observed from the docking structure which is conserved throughout the MD simulation. Meanwhile, the simulation also shows minor changes, such as the length and number of H-bonds after MD simulation, and the findings also reveal that the ligand moves deeper into the binding pocket which is correctly elicited by the MD simulation.
- (2) For IPAs, the optimal CoMSIA model shows that bulky groups at R<sub>5</sub> substituent, small groups at R<sub>1</sub> substituent, H-bond donor groups in Region A all exhibit positive effect on the activity. Results indicate that Pose B is the correct conformation in the active site of CDK<sub>2</sub>, which is evidenced by the contour map and

MD simulation. It has also been observed that, the inhibitors form H-bonds with amino acid residues (Asp86, Lys33, Lys88, Gly11, Ala144 and Asp145) to bind properly with the enzyme.

- (3) For PYPs, the contour map shows that small and electropositive substituents at R<sub>1</sub> region, small and electronegative groups at R<sub>3</sub> substituent, bulky and electronegative groups at R<sub>2</sub> area have great impact on the overall inhibitory activities. The MD simulation suggests that other than Gly11, Glu12, Lys33, Asp86, Asp89, Gln131, Leu134 and Asp145, the rest of the residues in the active site determined by docking have changed and some novel residues such as Tyr15, Leu133, and Ala144 are involved in the ligand–receptor interactions. And the increased hydrogen bonding with active site residues results in more stable complexes during the MD simulation.

The present study provides the first example of identifying the correct binding mode of CDK<sub>2</sub> inhibitors using the QSAR, molecular docking, and molecular dynamics approaches. The results indicate that the derived models are reliable and robust, which might be useful in predicting the activity of new CDK<sub>2</sub> inhibitors and offer guidelines for further corresponding ligand design.

#### Acknowledgements

The research is supported by high-performance computing platform of Northwest A & F University, and the financial aid of the Fund of Northwest A & F University. The authors are grateful to Prof. L. Yang for access of Sybyl software.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.06.006.

#### References

- [1] S. Vadivelan, B.N. Sinha, S.J. Irudayam, S.A.R.P. Jagarlapudi, Virtual screening studies to design potent CDK2-cyclin A inhibitors, *J. Chem. Inf. Model.* 47 (2007) 1526–1535.
- [2] M. Barvian, D.H. Boschelli, J. Cossrow, E. Dobrusin, A. Fattaey, Pyrido[2,3-d]pyrimidin-7-one inhibitors of cyclin-dependent kinases, *J. Med. Chem.* 43 (2000) 4606–4616.
- [3] J.W. Harper, P.D. Adams, Cyclin-dependent kinases, *Chem. Rev.* 101 (2001) 2511–2526.
- [4] M. Knockaert, P. Greengard, L. Meijer, Pharmacological inhibitors of cyclin-dependent kinases, *Trends. Pharm. Sci.* 23 (2002) 417–425.
- [5] D. Morgan, Cyclin-dependent kinases: engines, clocks, and microprocessors, *Annu. Rev. Cell. Dev. Biol.* 13 (1997) 261–291.
- [6] M.G. Brasca, N. Amboldi, D. Ballinari, A. Cameron, E. Casale, G. Cervi, M. Colombo, F. Colotta, V. Croci, R. D'Alessio, F. Fiorentini, A. Isacchi, C. Mercurio, W. Moretti, A. Panzeri, W. Pastori, P. Pevarello, F. Quartieri, F. Roletto, G. Traquandi, P. Vianello, A. Vulpetti, M. Ciomei, Identification of N,1,4,4-tetramethyl-8-([4-(4-methylpiperazin-1-yl)phenyl]amino)-4,5-dihydro-1H-pyrazolo[4,3-h]quinazoline-3-carboxamide (PHA-848125), a potent, orally available cyclin dependent kinase inhibitor, *J. Med. Chem.* 52 (2009) 5152–5163.
- [7] Y.N.P. Chen, S.K. Sharma, T.M. Ramsey, L. Jiang, M.S. Martin, K. Baker, P.D. Adams, K.W.W.G.K.J.R. Bair, Selective killing of transformed cells by cyclin/cyclin-dependent kinase 2 antagonists, *Proc. Natl. Acad. Sci. U.S.A.* 96 (1999) 4325–4329.
- [8] Y.H. Jin, K.J. Yoo, Y.H. Lee, S.K. Lee, Caspase 3-mediated cleavage of 21WAF1/CIP1-associated with the cyclinA-cyclin-dependent kinase 2 complex is a prerequisite for apoptosis in SK-HEP-1 cell, *J. Biol. Chem.* 275 (2000) 30256–30263.
- [9] J.A. Park, K.W. Kim, S.I. Kim, S.K. Lee, Caspase-3 specifically cleaves p21WAF1/CIP1 in the earlier stage of apoptosis in SK-HEP-1 human hepatoma cells, *Eur. J. Biochem.* 257 (1998) 242–248.
- [10] Y.M. Ahn, L. Vogetti, C.J. Liu, H.K.R. Santhapuram, J.M. White, V. Vasandani, L.A. Mitscher, G.H. Lushington, P.R. Hanson, D.R. Powell, R.H. Himes, K.F. Roby, Q. Ye, G.I. Georg, Design, synthesis, and antiproliferative and CDK2-cyclin A inhibitory activity of novel flavopiridol analogues, *Bioorg. Med. Chem.* 15 (2007) 702–713.
- [11] A.M. Senderowicz, Cyclin-dependent kinases as targets for cancer therapy, in: G. Giaccone, R. Schilsky, P. Sondel (Eds.), *Cancer Chemotherapy and Biological Response Modifiers*, 20, Elsevier Science, 2002, pp. 169–188.

- [12] P.M. Fischer, A. Gianella-Borradori, CDK inhibitors in clinical development for the treatment of cancer, *Expert Opin. Investig. Drugs* 12 (2003) 955–970.
- [13] R.N. Misra, H. Xiao, K.S. Kim, S. Lu, W. Han, S.A. Barbosa, J.T. Hunt, D.B. Rawlins, W. Shan, S.J. Ahmed, L. Qian, B. Chen, R. Zhao, M.S. Bednarz, K.A. Kellar, J.G. Mulheron, R. Batorsky, U. Roongta, A. Kamath, P. Marathe, S.A. Ranadive, J.S. Sack, J.S. Tokarski, N.P. Pavletich, F.Y.F. Lee, K.R. Webster, S.D. Kimball, N-(Cycloalkylamino) acyl-2-aminothiazole inhibitors of cyclin-dependent kinase 2 N-[5-[[[5-(1,1-dimethylethyl)-2-oxazolyl]methyl]thio]-2-thiazolyl]-4-piperidinecarboxamide (BMS-387032), a highly efficacious and selective antitumor agent, *J. Med. Chem.* 47 (2004) 1719–1728.
- [14] P.L. Toogood, P.J. Harvey, J.T. Repine, D.J. Sheehan, S.N. VanderWel, H. Zhou, P.R. Keller, D.J. McNamara, D. Sherry, T. Zhu, J. Brodfuehrer, C. Choi, M.R. Barvian, D.W. Fry, Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6, *J. Med. Chem.* 48 (2005) 2388–2406.
- [15] X. Chu, W. DePinto, D. Bartkovitz, S. So, B.T. Vu, K. Packman, C. Lukacs, Q. Ding, N. Jiang, K. Wang, P. Goelzer, X. Yin, M.A. Smith, B.X. Higgins, Y. Chen, Q. Xiang, J. Moliterni, G. Kaplan, B. Graves, A. Lovey, N. Fotouhi, Discovery of [4-amino-2-(1-methanesulfonylpiperidin-4-ylamino)pyrimidin-5-yl]-(2,3-difluoro-6-methoxyphenyl) methanone (R547). A potent and selective cyclin-dependent kinase inhibitor with significant *in vivo* antitumor activity, *J. Med. Chem.* 49 (2006) 6549–6560.
- [16] P.M. Fischer, A. Gianella-Borradori, Recent progress in the discovery and development of cyclin-dependent kinase inhibitors, *Expert Opin. Investig. Drugs* 14 (2005) 457–477.
- [17] T. Akiyama, T. Yoshida, T. Tsujita, M. Shimizu, T. Mizukami, M. Okabe, S. Akinaga, G1 phase accumulation induced by UCN-01 is associated with dephosphorylation of Rb and CDK2 proteins as well as the induction of CDK inhibitor p21/CIP/WAF1/Sdi1 in p53-mutated human epidermoid carcinoma A431 cells, *Cancer Res.* 57 (1997) 1495–1501.
- [18] C. Benson, J. White, J.D. Bono, A. O'Donnell, F. Raynaud, C. Cruickshank, H. McGrath, M. Walton, P. Workman, S. Kaye, J. Cassidy, A. Gianella-Borradori, I. Judson, C. Twelves, A phase I trial of the selective oral cyclin-dependent kinase inhibitor seliciclib (CYC202; R-Roscovitin), administered twice daily for 7 days every 21 days, *Br. J. Cancer* 96 (2007) 29–37.
- [19] S.J. McClue, D. Blake, A.R. Clarke, A. Cowan, L. Cummings, P.M. Fischer, M. MacKenzie, J. Melville, K. Stewart, S.D. Wang, N. Zhelev, D. Zhelev, D.P. Lane, *In vitro* and *in vivo* antitumor properties of the cyclin dependent kinase inhibitor CYC202 (R-roscovitin), *Int. J. Cancer* 102 (2002) 463–468.
- [20] S.F. Jones, H.A. Burris, M. Kies, N. Willcutt, P. Degen, A phase I study to determine the safety and pharmacokinetics (PK) of BMS-387032 given intravenously every three weeks in patients with metastatic refractory solid tumors, *Am. Soc. Clin. Oncol.* 22 (2003) 199, Abstr 798.
- [21] D. Cai, K. Byth, G. Shapiro, A novel CDK inhibitor induces cell cycle blockade E2F-1 dependent apoptosis, and cytotoxic synergy with DNA damaging agents, *Proc AACR* 45 (2004) 190.
- [22] C. Terret, S. Zanetta, H. Roche, J.H. Schellens, M.N. Faber, J. Waners, A phase I clinical and pharmacokinetic study of E7070, a novel sulfonamide given as a 5-day continuous infusion repeated every 3 weeks in patients with solid tumors, *Eur. J. Cancer* 39 (2003) 1097–1104.
- [23] Y. Li, Y.H. Wang, L. Yang, S.W. Zhang, C.H. Liu, Structural determinants of flavones interacting with the C-terminal nucleotide-binding domain as p-glycoprotein inhibitors, *J. Mol. Des.* 5 (2006) 1–12.
- [24] Y. Li, Y.H. Wang, L. Yang, S.W. Zhang, C.H. Liu, S.L. Yang, Comparison of steroid substrates and inhibitors of p-glycoprotein by 3D-QSAR analysis, *J. Mol. Struct.* 733 (2005) 111–118.
- [25] X. Wang, W. Yang, X. Xu, H. Zhang, Y. Li, Y.H. Wang, Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics, *Curr. Med. Chem.* 17 (2010) 2788–2803.
- [26] A. Echalié, K. Bettayeb, Y. Ferandin, O. Lozach, Meriolins (3-(pyrimidin-4-yl)-7-azaindoles): synthesis, kinase inhibitory activity cellular effects, and structure of a CDK2/Cyclin A/Meriolin complex, *J. Med. Chem.* 51 (2008) 737–751.
- [27] E.A. Hecker, C. Duraiswami, T.A. Andrea, D.J. Diller, Use of catalyst pharmacophore models for screening of large combinatorial libraries, *J. Chem. Inf. Comput. Sci.* 42 (2002) 1204–1211.
- [28] H. Dureja, A.K. Madan, Topochemical for prediction of cyclin-dependent kinase 2 inhibitory activity of indole-2-ones, *J. Mol. Model.* 11 (2005) 525–531.
- [29] T. Naumann, H. Matter, Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes, *J. Med. Chem.* 45 (2002) 2366–2378.
- [30] D.A. Nugiel, A. Vidwans, A.M. Etzkorn, K.A. Rossi, P.A. Benfield, C.R. Burton, S. Cox, D. Doleniak, S.P. Seitz, Synthesis and evaluation of indenopyrazoles as cyclin-dependent kinase inhibitors 2. Probing the indeno ring substituent pattern, *J. Med. Chem.* 45 (2002) 5224–5232.
- [31] A. Vulpetti, P. Crivori, A. Cameron, J. Bertrand, M.G. Brasca, R.D. Alessio, P. Pevarello, Structure-based approaches to improve selectivity: CDK2-GSK3 $\beta$  binding site analysis, *J. Chem. Inf. Model.* 45 (2005) 1282–1290.
- [32] R.N. Jorissen, M.K. Gilson, Virtual screening of molecular databases using a support vector machine, *J. Chem. Inf. Model.* 45 (2005) 549–561.
- [33] N.M. Mascarenhas, N. Ghoshal, An efficient tool for identifying inhibitors based on 3D-QSAR and docking using feature-shape pharmacophore of biologically active conformation – a case study with CDK2/CyclinA, *Eur. J. Med. Chem.* 43 (2008) 2807–2818.
- [34] J. Alzate-Morales, J. Caballero, Computational study of the interactions between Guanine derivatives and cyclin-dependent kinase 2 (CDK2) by CoMFA and QM/MM, *J. Chem. Inf. Model.* 50 (2010) 110–122.
- [35] J.H. Alzate-Morales, J. Caballero, A. Vergara-Jaque, F.D. Gonza lezNilo, Insights into the structural basis of N2 and O6 substituted guanine derivatives as cyclin-dependent kinase 2 (CDK2) inhibitors: prediction of the binding modes and potency of the inhibitors by docking and ONIOM calculations, *J. Chem. Inf. Model.* 49 (2009) 886–899.
- [36] Y. Ai, S.T. Wang, P.H. Sun, F.J. Song, Molecular modeling studies of 4,5-dihydro-1H-pyrazolo[4,3-h]quinazoline derivatives as potent CDK2/Cyclin A inhibitors using 3D-QSAR and docking, *Int. J. Mol. Sci.* 11 (2010) 3705–3724.
- [37] M.G. Brasc, C. Albanese, R. Alzani, R. Amici, N. Avanzi, D. Ballinari, Optimization of 6,6-dimethyl pyrrolo[3,4-c]pyrazoles: identification of PHA-793887, a potent CDK inhibitor suitable for intravenous dosing, *Bioorg. Med. Chem.* 18 (2010) 1844–1853.
- [38] P. Pevarello, D. Fancelli, A. Vulpetti, R. Amici, M. Villa, 3-Amino-1,4,5,6-tetrahydropyrrolo [3, 4-c] pyrazoles: a new class of CDK2 inhibitors, *Bioorg. Med. Chem. Lett.* 16 (2006) 1084–1090.
- [39] C.D. Jones, D.M. Andrews, A.J. Barker, K. Blades, P. Daunt, S. East, The discovery of AZD5597, a potent imidazole pyrimidine amide CDK inhibitor suitable for intravenous dosing, *Bioorg. Med. Chem. Lett.* 18 (2008) 6369–6373.
- [40] C.D. Jones, D.M. Andrews, A.J. Barker, K. Blades, K.F. Byth, M.R.V. Finlay, Imidazole pyrimidine amides as potent, orally bioavailable cyclin-dependent kinase inhibitors, *Bioorg. Med. Chem. Lett.* 18 (2008) 6486–6489.
- [41] M. Anderson, D.M. Andrews, A.J. Barker, C.A. Brassington, J. Breed, K.F. Byth, Imidazoles: SAR and development of a potent class of cyclin-dependent kinase inhibitors, *Bioorg. Med. Chem. Lett.* 18 (2008) 5487–5492.
- [42] M.R.V. Finla, D.G. Acton, D.M. Andrews, A.J. Barker, M. Dennis, Imidazole piperazines: SAR and development of a potent class of cyclin-dependent kinase inhibitors with a novel binding mode, *Bioorg. Med. Chem. Lett.* 18 (2008) 4442–4446.
- [43] Y.S. Cho, M. Borland, C.C. Brain, T. Chen, H. Cheng, R. Chopra, K. Chung, 4-(Pyrazol-4-yl)-pyrimidines as selective inhibitors of cyclin-dependent kinase 4/6, *J. Med. Chem.* 53 (2010) 7938–7957.
- [44] S. Pirhadi, J.B. Ghasemi, 3D-QSAR analysis of human immunodeficiency virus entry-1 inhibitors by CoMFA and CoMSIA, *Eur. J. Med. Chem.* 45 (2010) 4897–4903.
- [45] R. Thaimattam, P.R. Daga, R. Banerjee, J. Iqbal, 3D-QSAR studies on c-Src kinase inhibitors and docking analyses of a potent dual kinase inhibitor of c-Src and c-Abl kinases, *Bioorg. Med. Chem.* 13 (2005) 4704–4712.
- [46] R.D. Cramer, J.D. Bunce, D.E. Patterson, Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies, *Quant. Struct. Act. Relat.* 7 (1988) 18–25.
- [47] A.N. Jain, Scoring noncovalent protein ligand interactions: a continuous differentiable function tuned to compute binding affinities, *J. Comput. Aided Mol. Des.* 10 (1996) 427–440.
- [48] A.N. Jain, Morphological similarity: a 3D molecular similarity method correlated with protein ligand recognition, *J. Comput. Aided Mol. Des.* 14 (2000) 199–213.
- [49] J. Ruppert, W. Welch, A.N. Jain, Automatic identification and representation of protein binding sites for molecular docking, *Protein Sci.* 6 (1997) 524–533.
- [50] R. Miglani, I.A. Cliffe, S.R. Voleti, Assessment of the putative binding conformation of a pyrazolopyridine class of inhibitors of MAPKAPK2 using computational studies, *Eur. J. Med. Chem.* 1 (2009) 98–105.
- [51] Z.Z. Wang, Y. Li, C.Z. Ai, Y.H. Wang, *In silico* prediction of estrogen receptor subtype binding affinity and selectivity using statistical methods and molecular docking with 2-arylnaphthalenes and 2-arylquinolines, *Int. Mol. Sci.* 11 (2010) 3434–3458.
- [52] G. Wang, Y. Li, X. Liu, Y.H. Wang, Understanding the aquatic toxicity of pesticide: structure-activity relationship and molecular descriptors to distinguish the ratings of toxicity, *QSAR Comb. Sci.* 28 (2009) 1418–1431.
- [53] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH Verlag, Weinheim, 2000.
- [54] Dragon, 5.3, Milano Chemometrics and QSAR Research Groups Inc., 2002.
- [55] Y.H. Wang, Y. Li, J. Ding, Z. Jiang, Y. Chang, Estimation of bioconcentration factors using molecular electro-topological state and flexibility, *SAR QSAR Environ. Res.* 19 (2008) 375–395.
- [56] E. Lindahl, B. Hess, D.A. van der Spoel, Package for molecular simulation and trajectory analysis, *J. Mol. Mod.* 7 (2001) 306–317.
- [57] D.M.F. Van Aalten, R. Bywater, J.B.C. Findlay, M. Hendlich, R.W.W. Hooft, G. Vriend, PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules, *J. Comput. Aided Mol. Des.* 10 (1996) 255–262.
- [58] <http://davap1.bioch.dundee.ac.uk/programs/prodrq/prodrq.html>.
- [59] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Dinola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Comput. Chem.* 81 (1984) 3684–3690.
- [60] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (1981) 7182–7190.
- [61] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [62] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINC: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [63] M. Bohm, J. Stürzbecher, G. Klebe, Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa, *J. Med. Chem.* 42 (1999) 458–477.

- [64] G. Bringmann, C.J. Rummey, 3D-QSAR investigations on antimalarial naphthylisoquinoline alkaloids, by comparative molecular similarity indices analysis (CoMSIA), based on different alignment approaches, *J. Chem. Inf. Comput. Sci.* 43 (2003) 304–316.
- [65] N. Trinajstić, *Chemical Graph Theory*, 8, CRC Press, Boca Raton (FL), 1992, pp. 161–198.
- [66] A.K. Gupta, R.A. Gupta, L.K. Soni, S.G. Kaskhedikar, Exploration of physicochemical properties and molecular modelling studies of 2-sulfonyl-phenyl-3-phenyl-indole analogs as cyclooxygenase-2 inhibitors, *Eur. J. Med. Chem.* 43 (2008) 1297–1303.
- [67] N.M. Mascarenhas, D. Bhattacharyya, N. Ghoshal, Why pyridine containing pyrido [2,3-d]pyrimidin-7-ones selectively inhibit CDK4 than CDK2: insights from molecular dynamics simulation, *J. Mol. Graph. Model.* 28 (2010) 695–706.
- [68] H. Alonso, A.A. Bliznyuk, J.E. Gready, Combining docking and molecular dynamic simulations in drug design, *Med. Res. Rev.* 26 (2006) 531–568.