

Structural requirements of pyrimidine, thienopyridine and ureido thiophene carboxamide-based inhibitors of the checkpoint kinase 1: QSAR, docking, molecular dynamics analysis

Fangfang Wang · Zhi Ma · Yan Li · Jinan Wang · Yonghua Wang

Received: 1 June 2011 / Accepted: 23 November 2011 / Published online: 15 January 2012
© Springer-Verlag 2012

Abstract Our focus of current research is directed toward clarification of novel inhibitors (pyrazolo[1,5-a] pyrimidine (PP), thienopyridines (TP) and 2-ureido thiophene carboxamide (UTC) derivatives) targeting Checkpoint kinase 1 (CHK₁), which is an oncology target of significant current interest. Our computational approaches include: (i) QSAR analysis was carried out on the computed steric/electrostatic/hydrophobic/hydrogen bond donor/hydrogen bond acceptor interactions with the pseudoreceptor surface, which yielded predictive models capable of explaining much of the variance of inhibitors. The resultant optimum QSAR/CoMFA models exhibited ($N_{\text{training}}=51$, $N_{\text{test}}=16$, $R_{\text{cv}}^2=0.47$, $R_{\text{pred}}^2=0.7$) for PP, ($N_{\text{training}}=45$, $N_{\text{test}}=9$, $R_{\text{cv}}^2=0.52$, $R_{\text{pred}}^2=0.75$) for TP and ($N_{\text{training}}=58$, $N_{\text{test}}=15$, $R_{\text{cv}}^2=0.67$, $R_{\text{pred}}^2=0.88$) for UTC. (ii) Molecular docking and molecular dynamics simulations experiments of the inhibitors into the binding site of CHK₁ aided the interpretation of the QSAR models and demonstrated the binding modes in the aspects of inhibitor's conformation, subsite interaction, and hydrogen bonding interactions, which indicated

that a set of critical residues (Cys87, Glu91, Glu85, Ser147, Asp148, Glu17, Leu84 and Asn135) played a key role in the drug-target interactions. The obtained results in the present work will be fruitful for the design of new potent and selective inhibitors of CHK₁.

Keywords Checkpoint kinase 1 · Molecular docking · Molecular dynamics · Quantitative structure activity relationship

Introduction

Cancer is a top killer of human beings exhibiting genomic instability and heightened drug sensitivity due to underlying defects in DNA repair or cell cycle regulation [1, 2]. However, the clinical cancer therapies show lack of selectivity toward cancer cells and resistant to DNA-damaging agents [3]. Recently, checkpoint pathways that regulate the cell cycle in response to DNA damage have gained great interests [4].

Checkpoint kinase 1 (CHK₁), a serine/threonine protein kinase, plays a critical role in DNA damage-induced checkpoints responsible for the maintenance of mammalian genomic integrity and repair of damaged DNA [5]. CHK₁ is a 54 kDa protein comprised of a highly conserved N-terminal kinase domain (residues 1–265), a putative flexible linker region, an SQ/TQ domain, and a C-terminal domain with ill-defined function [6]. The SQ/TQ domain has several conserved Ser-Gln (SQ) or Thr-Gln (TQ) motifs, in which the serine or threonine residues are preferred phosphorylation sites by ATR in vitro [7]. Through phosphorylation in the SQ/TQ domain by ataxia telangiectasia mutated (ATM) and Rad3-related (ATR) kinases, CHK₁ arrests cells at various DNA-damaging checkpoints (G1, S and G2) to initiate the

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1321-z) contains supplementary material, which is available to authorized users.

F. Wang · Z. Ma · J. Wang · Y. Wang
Bioinformatics Center, Northwest A&F University,
Yangling, Shaanxi 712100, China

Y. Li
School of Chemical Engineering,
Dalian University of Technology,
Dalian, Liaoning 116024, China

Y. Wang (✉)
College of Life Sciences, Northwest A&F University,
Yangling, Shaanxi 712100, China
e-mail: yh_wang@nwsuaf.edu.cn

DNA repair process. The inhibition of CHK₁ abrogates the S and G2 checkpoints and disrupts the DNA repair process, resulting in premature chromosome condensation and leading to cell death, thereby they preferentially sensitize tumor cells, especially p53-null cells [8].

Extensive studies with various cancer cell lines have reported that the enhanced cell death is achieved under genotoxic stresses by eliminating CHK₁ via siRNA [9], anti-sense [10], and small molecule inhibitors [11]. Inhibition of CHK₁ represents a targeted approach to enhance the cytotoxicity of DNA-damaging agents toward tumor cells while having a lesser effect on normal cells, resulting in an attractive target in the oncology field [12]. Thus, the development of novel CHK₁ inhibitors is of great interest, not only from a drug development viewpoint, but also to provide molecular tools with which the biological functions of CHK₁ can be further elucidated.

More recently, several natural products have been proved to be potent CHK₁ inhibitors, which included various scaffolds, e.g., Granulatimide and isogranulatimide (IC₅₀ values of 0.25 and 0.1 nM, respectively) [13, 14], 7-hydroxystaurosporine (IC₅₀=10 nM) [15, 16], AZD7762 (IC₅₀ of 5 nM) [17], PF477736 (K_i value of 0.49 nM) [18] and SCH900776 (IC₅₀=3 nM) [19]. Unfortunately, so far none of them have progressed to being a clinical drug. One of the main bottlenecks is the difficulty in attaining selectivity, which stems from the diverse nature of the kinase substrates. The occurrence of severe side effects of these inhibitors (e.g., strong single agent activity, small volume of distribution, low systemic clearance and a prolonged half-life of elimination) urgently calls for the search of novel, potent, more selective, and less toxic CHK₁ inhibitors [20].

In vitro assessment of the activity of CHK₁ inhibitors remains a labor-intensive and time-consuming operation. As an important technology for drug design, in silico modeling methods, such as QSAR (quantitative structure-activity relationship) approaches, have been found to be valuable in further optimization and development of novel inhibitors [21–23]. Recently, 3D-QSAR studies were performed on a series of 1,4-dihydroindeno[1,2-c]pyrazoles inhibitors, using molecular docking and CoMFA, and the results suggested that Cys87 and Glu85 were mainly involved in hydrogen bonding interactions [24]. In another report, the QSAR models were developed (for 5,10-dihydro-dibenzo[b,e]1,4 diazepin-11-ones which the structures are different with those employed in our study) based on the molecular descriptors from both the ligand and the ligand-receptor complex, which indicated that the strong hydrophobic interaction between the ligand and the receptor, low RuleOf5 value and QPlogKhsa, more Amine or Amide group and less carbonyl group are favorable to activity improvement [25].

More recently, three different series of CHK₁ inhibitors (194 compounds) with diverse structures and biological

activities have been reported [26–29], however, to our knowledge there has been no attempt to quantitatively study the structure activity relationship of these derivatives and no comprehensive features for the ligand-receptor interactions have been demonstrated, previously. Thus in this work, we applied the ligand- and receptor-based QSAR techniques to these inhibitors. In addition, molecular docking and molecular dynamics (MD) simulation were also utilized to orient the compounds in the binding site of CHK₁ kinase, further to elucidate the probable binding modes of these inhibitors at the allosteric site of the enzyme. The QSAR models as well as the information gathered from molecular docking and 3D contour maps are beneficial to identify the structural features of CHK₁ and also helpful in the design of novel CHK₁ inhibitors with improved activities.

Methods and materials

Biological data and molecular structures

The structures and biological activities of CHK₁ inhibitors (194) were taken from the literatures [26–29] and listed in Tables S1–S3. These molecules comprised three different classes of diverse pyrazolo [1,5-a] pyrimidine, thienopyridines, and 2-ureido thiophene carboxamide derivatives. The IC₅₀ values were converted into the corresponding pIC₅₀ (–logIC₅₀) values, which were further used as the dependent variables in the QSAR analysis. For carrying out CoMFA/CoMSIA QSAR studies, each group of the molecules was divided into the training set for model generation and the test set for model validation. The test set compounds were selected by considering both the distribution of biological data and the structural diversity.

All molecular modeling and QSAR studies were performed using Sybyl package (Tripos Associates, St.Louis, MO). 3D structures of all molecules were constructed by using the Sketch molecule module. Partial atomic charges required for electrostatic interactions were computed by Gasteiger-Marsili (for PP analogs) and Gasteigere-Huckel (for TP and UTC analogs) methods. Each structure was fully geometry-optimized using the Tripos force field [30] with a distance-dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.05 kcal mol⁻¹.

Conformational sampling and alignment

Molecular alignment is considered as a sensitive parameter in QSAR analysis, and the quality of the model is directly dependent on the alignment rule. Here, the molecular alignment was performed using “database alignment” in Sybyl with the common substructures shown in the upper left corner of Figs. 1, 2 and 3. Two alignment rules were employed

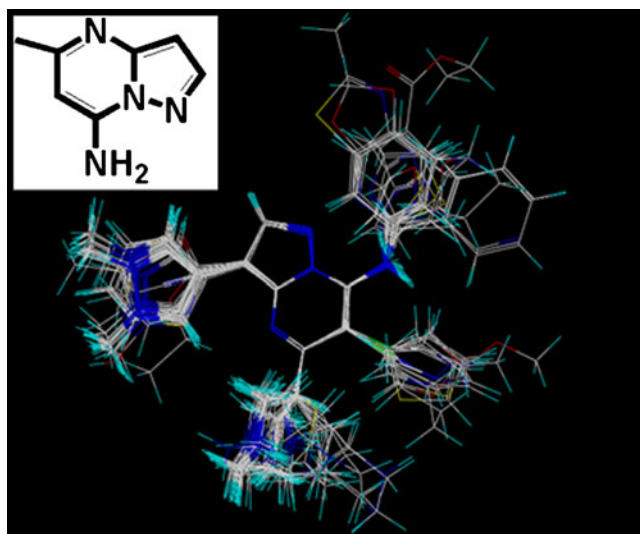


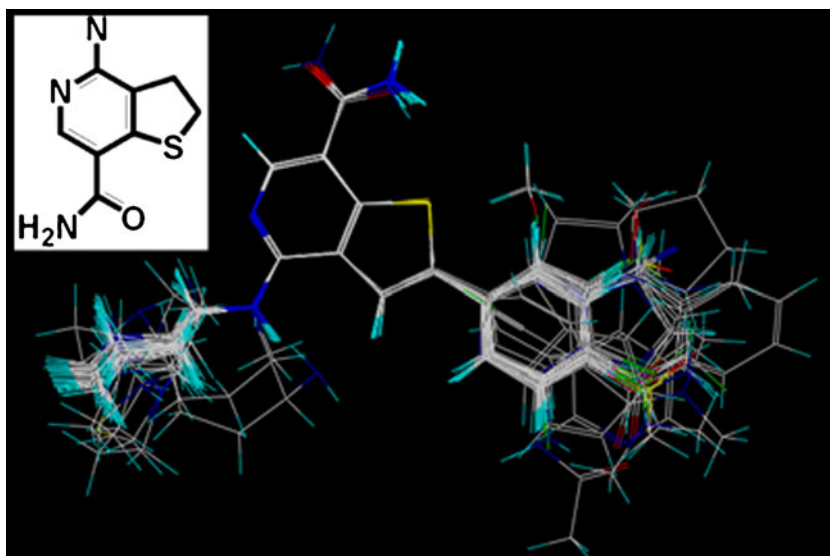
Fig. 1 Superimposition of PP compounds in the training and test sets with common substructure shown in the upper left corner

to produce valid and reliable CoMFA/CoMSIA models: 1) the ligand-based alignment, where the most active compound in each group (compounds 38, 112 and 185) was selected as a template to fit the remaining molecules (Figs. 1, 2 and 3). 2) The receptor-based alignment was based on the geometries obtained from docking, where all the molecules were docked into the active site of the receptors, then the top scored conformations were aligned together for CoMFA/CoMSIA analysis (Figs. S1-S3).

CoMFA and CoMSIA set up

In order to explore the steric (S)/electrostatic (E)/hydrophobic (H)/hydrogen bond donor (D)/hydrogen bond acceptor (A) fields contributions to the binding affinities of the inhibitors,

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



and to build predictive QSAR models, CoMFA/CoMSIA studies were performed based on the molecular alignment as described.

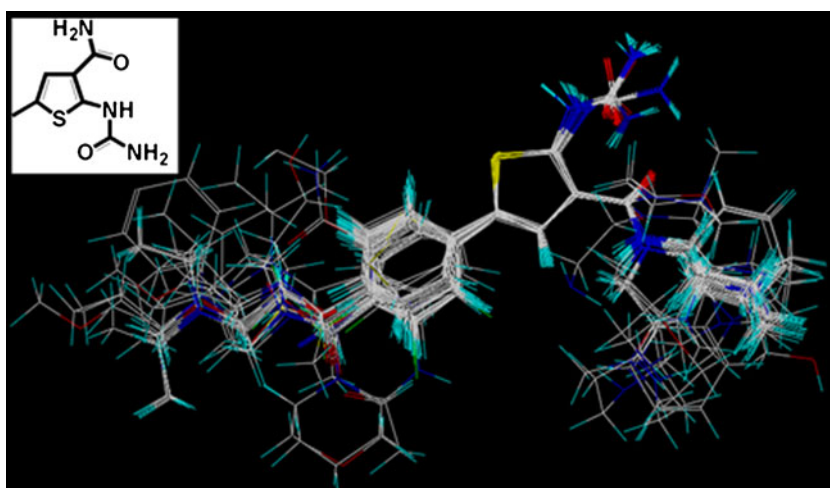
The CoMFA descriptors, S (Lennard-Jones 6–12 potential) and E (Coulombic potential) field energies were calculated using Sybyl. A 3D cubic lattice with grid spacing of 2 Å in x, y and z directions, were generated automatically to encompass the aligned molecules, the S and E fields were scaled by the CoMFA-STD method with default energy of 30 kcal mol⁻¹, and generated at each grid point with Tripos force field using a sp³ carbon atom probe carrying a +1 net charge with a van der Waals radius of 1.52 Å.

For CoMSIA analysis, five descriptors, namely S, E, and H parameters and the D and A fields were calculated using a sp³ carbon probe with a +1.0 charge atom and a radius of 1.0 Å placed at regular grid spacing of 2 Å, coupling with a similar lattice box as used in CoMFA calculations. S indices were related to the third power of the atomic radii, E descriptors were derived from atomic partial charges, H fields were derived from atom-based parameters [31], and D and A indices were obtained by a rule-based method based on experimental results [32].

Regression analysis and model validation

Partial least squares (PLS) method [33] was used to linearly correlate the CoMFA/CoMSIA fields to biological activity values. The CoMFA/CoMSIA descriptors were used as independent variables and pIC₅₀ values were used as dependent variables in PLS analyses to derive QSAR models. The performance of models was evaluated using the leave-one-out (LOO) cross-validation method which one compound was removed from the data set and its activity was predicted using the model derived from the rest of the data set. A

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



cross-validated correlation coefficient R_{cv}^2 was generated and provided as a statistical index of the predictive power. PLS was conjunct with the cross-validation option to determine the optimum number of components (ONC). Then, the ONC was used in the final non-cross-validated analysis; the Pearson coefficient (R_{ncv}^2), standard error of estimates (SEE) and F values were calculated as a measure of the quality of the models. Finally, the CoMFA/CoMSIA results were graphically represented by field contour maps using the field type ‘StDev*Coeff’. In addition to LOO method to validate QSAR models, the established test set was used for further evaluation, with the predictive correlation coefficient (R_{pred}^2) was produced.

Molecular docking

To investigate the protein-ligand interactions, the Surflex package was employed to generate an ensemble of docking conformations. The Surflex uses an empirical scoring function and a patented search engine to dock ligands into the receptor’s binding site [34]. Crystal structures of CHK₁ (3OT3, 3PA5 and 2WMT) were retrieved from the RCSB Protein Data Bank (<http://www.rcsb.org>), all the three crystal structures are in ligand binding state (active state), which possess 22 K (pyrazolo[1,5-a]pyrimidine analog), C73 (thienopyridines) and ZYT (imidazol) compounds in the binding site, respectively. Moreover, the structures of these molecules are similar with the ligands (PP, TP and UTC analogs) employed in the present work. The relative ligands were extracted and polar hydrogen atoms were added, and the water molecules in the crystal structures were not removed before docking since the water molecules were found conserved in different crystal structures. Protomol, an idealized representation of a ligand that makes every potential interaction with the receptor, was used to guide molecular docking. Establishment of protomol supplies three manners: (1) automatic probing of cavity in the receptor protein; (2)

ligand location in the same coordinate space in the receptor; (3) specification of interacting residues. During the protomol generating process, two parameters are critical for forming appropriate binding pocket, in which one is the protomol_bloat determining how far from a potential ligand the site should extend, and the other is the protomol_threshold determining how deep into the protein the atomic probes used to define the protomol can penetrate. In the present work, for 3OT3, 3PA5 and 2WMT docking analyses, the protomol_bloat values were set to 1, 1 and 2 and the protomol_threshold were set to 0.5, 0.5 and 0.01, respectively. 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 [35]. During the docking process, the protein was regarded as rigid and the ligand was flexible. According to the total score function, the docking poses were saved for each compound, and the top ranked conformations were extracted and aligned together for further QSAR analysis.

Calculation and selection of dragon descriptors

Molecular descriptors have been demonstrated to play important roles in developing QSAR models and have been successfully applied in many in silico modeling processes [36–38]. In this work, Dragon Professional, version 5.0 [39] was employed to calculate all molecular descriptors (shown in Table S10). And 1356 descriptors were calculated for each compound. The stepwise linear regression method as the variable selection in R software-version 2.13.0 (www.r-project.org) was employed to reduce the descriptor space. The results indicated that no single descriptor was capable of predicting the activity. Using the linear regression for variable selection, the extracted descriptors (shown in Table 1) are ATS_{3e} , D_u and $MATS_{8m}$ for PP analogs, D_p , $RDF060m$, G_{3m} and $Mor24_v$ for TP dataset, R_{4u} and $nRNHR$ for UTC analogs. The obtained

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

Symbols	Descriptor family	Definition
ATS _{3e}	2D autocorrelations	Broto-Moreau autocorrelation of a topological structure-lag 3/weighted by atomic Sanderson electro
D _u	WHIM descriptors	D total accessibility index/unweighted
MATS _{8m}	2D autocorrelations	Moran autocorrelation-lag 8/weighted by atomic masses
D _p	WHIM descriptors	D total accessibility index/weighted by atomic polarizabilities
RDF060m	RDF descriptors	Weighted by atomic masses
G _{3m}	WHIM descriptors	3st component symmetry directional WHM index/weighted by atomic masses
Mor24 _v	3D-MoRSE descriptors	3D-MoRSE-signal 24/weighted by atomic van der Waals volumes
nRNHR	Functional group counts	Number of secondary amines (aliphatic)
R _{4u}	GETAWAY descriptors	R autocorrelation of lag 4/unweighted

descriptors were further introduced into building QSAR models in order to improve the robustness and generalization ability of the models.

Molecular dynamics simulations

The MD simulations were performed using the GROMACS package 4.0.7 [40] using the GROMOS96 force field [41]. The molecular topology files were generated by the program PRODRG2.5 [42]. The whole complex was solvated in SPCE water molecules, in addition, counter ions (five Na⁺, four Na⁺ and three Na⁺ ions) were added to neutralize the charge of each system, i.e., 3OT3.pdb and 3PA5.pdb and 2WMT.pdb, respectively. The size of the rectangular cell was 7.21×6.16×8.48 nm, 7.24×5.95×8.49 nm and 7.087×6.126×8.164 nm, respectively, containing 11,058, 10,956, 9039 water molecules in each complexes. Additionally, the minimum distance between the protein and the box walls was set to more than 8 Å so that the protein does not directly interact with its own periodic image given the cutoff in each complex. The solvated systems contain approximately 35,877 atoms for 3OT3, 35,543 atoms for 3PA5 and 29,736 atoms for 2WMT including the protein-ligand complexes and waters, respectively. And the full system has been optimized without constraints using the steepest descent integrator for 5000 steps.

Simulations were run in the NPT ensemble and at a temperature of 300 K. The temperature was kept constant by the Berendsen thermostat [43], the coupling time was 2 ps. The pressure was maintained by coupling to a reference pressure of 1 bar using the Parrinello-Rahman scheme [44]. Electrostatic interactions were calculated using the particle mesh Ewald method [45]. The LINCS algorithm was used to restrain all bond lengths [46]. Cutoff distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. The system was equilibrated via a 200 ps MD simulation. Finally, a 5 ns simulation was performed with a time step of 2 fs.

Results and discussion

QSAR model

To judge whether a QSAR model is reliable for prediction of unknown molecules, several statistical parameters including R_{cv}^2 , R_{ncv}^2 , SEE and F values should be evaluated. Table 2 summarizes the statistical results of the optimum models (ligand-based models).

PP

During the molecular modeling process, 51 compounds out of the total 67 CHK₁ inhibitors were selected as training set and the remaining 16 compounds were used as test set. The statistical parameters of the optimal models are summarized in Table 2.

The best CoMFA model gives a good R_{cv}^2 of 0.467 with six components, indicating a proper internal predictive capacity of the model. A high R_{ncv}^2 of 0.823 with a low SEE of 0.503 for the final model shows its self-consistency. In terms of the relative contributions, S and E account for 51% and 29.7%, respectively, indicating that the S field contributes more to the binding affinities. Additionally, another three descriptors ATS_{3e}, D_u and MATS_{8m} also make 10.3%, 29.7% and 0.5%, respective contributions to this model, which effectively increase the robustness of the model. ATS_{3e} is 2D autocorrelation descriptors which is the Broto-Moreau autocorrelation of a topological structure-lag 3, and is weighted by atomic Sanderson electro-negativities. The D_u descriptor is the WHIM descriptors which belong to the 3D descriptor. The other descriptor is MATS_{8m}, Moran autocorrelation-lag8/weighted by atomic masses, which is a 2D-autocorrelation descriptor, the positive regression coefficient suggests in favor of increased autocorrelation contents of eight-member structural graphs weighted by atomic masses for the activity [47].

To understand and explain the chemical meaning of the descriptors, the dataset was used. As can be appreciated the

Table 2 The best results of the CoMFA analyses for the training and test set compounds

Parameters	PP analogs CoMFA	TP analogs CoMFA	UTC analogs CoMFA
R_{cv}^2	0.467	0.522	0.673
R_{ncv}^2	0.823	0.824	0.873
SEE	0.503	0.371	0.42
F	34.966	29.745	71.607
R_{pred}^2	0.701	0.745	0.877
SEP	0.874	0.612	0.674
N_c	6	6	5
Field contribution			
S	0.51	0.382	0.497
E	0.297	0.215	0.238
H	–	–	–
D	–	–	–
A	–	–	–
ATS _{3e}	0.103	–	–
Du	0.085	–	–
MATS _{8m}	0.005	–	–
D _p	–	0.098	–
RDF _{060m}	–	0.135	–
G _{3m}	–	0.048	–
Mor24v	–	0.121	–
<i>n</i> RNHR	–	–	0.16
R ₄ U	–	–	0.105

R_{cv}^2 = Cross-validated correlation coefficient after the leave-one-out procedure; 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; N_c = Optimal number of principal components; S = Steric; E = Electrostatic; H = Hydrophobic; D = Hydrogen-bond-donor; A = Hydrogen-bond-acceptor

greater values of ATS_{3e} are for compounds 28, 29 and 30, with R₃ substituted with hydrogen atoms. It is explained by the worse interaction with the surface of the binding site for these compounds. Once the R₃ position of compounds 38 and 65 are functionalized with different -Cl or -Br side chains, the particular conformation adopted by its R₃ substituents significantly enhance the electrostatic interactions with the receptor residues, possessing lower values of ATS_{3e}; this phenomenon is consistent with the contour maps with red contours at this position.

The CoMSIA model (SE model) shows poor internal predictions relative to the CoMFA model, which ends in statistical results of $R_{cv}^2=0.39$, $R_{ncv}^2=0.581$, F=16.275 with four optimal components using the same Dragon descriptors as the CoMFA model.

The test set molecules are further used to validate the efficacy of the obtained models. The R_{pred}^2 is 0.701 for the

optimal CoMFA model and 0.667 for CoMSIA model. For the CoMFA model, molecules 16, 57 and 58 are regarded as outliers since their residuals between the experimental value and the predicted value are up to 2.0 log unit. All these results suggest that the CoMFA model is superior to the CoMSIA one.

The correlation between the predicted activities and the actual activities is displayed in Fig. 4a. Clearly, all points are uniformly distributed around the regression line, suggesting the reliable predictive ability of the model.

TP

The total of 54 CHK₁ inhibitors was divided into a training set composed of 45 compounds and a test set consisting of 9 chemicals. The statistical results obtained from standard CoMFA model constructed with S and E fields are summarized in Table 2. Statistical data shows R_{cv}^2 0.522, R_{ncv}^2 0.824, SEE 0.371, F 29.745 for the optimum CoMFA model. The S field descriptor explains 38.2% of the variance, while the E descriptor explains 21.5%, indicating that the S field makes more contributions to the inhibitory activity. In developing this model, another four descriptors D_p, RDF060m, G_{3m} and Mor24_v are also employed, contributing 9.8%, 13.5%, 4.8% and 12.1% to this model, respectively. RDF060m is radial distribution descriptors centered on different interatomic distances interpreted as the probability distribution of finding an atom in a spherical volume of certain radius. RDF can be interpreted as the probability distribution to find an atom in a spherical volume of radius r. Based on this definition it is plausible to presume that RDF060m encodes a partial contribution of the Casari's bioactivity at regions where atoms are distanced $r=6.0$ Å from the geometrical center of each molecule. Mor24_v is a 3D-MoRSE descriptor which is weighted by atomic van der Waals volumes. This descriptor extracts information from the 3D atomic coordinates by using the same transform as in electron diffraction studies [47]. The descriptor G_{3m} belongs to the WHIM descriptors, which is based on the statistical indices calculated from projection of atoms along principal axes. The algorithm consists of performing a principal-components analysis on the centered Cartesian coordinates of a molecule by using a weighted covariance matrix obtained from different weighting schemes for the atoms. D_p descriptor is the WHIM descriptors belongs to the 3D descriptors, it is the D total accessibility index which is weighted by atomic polarizabilities.

The Dragon descriptor RDF060m explained in the model encoded specific structure information. For example, for compounds 112 and 121, the lower value of RDF060m corresponds to compound 121 where the substituent in position R₃ is hydrogen atom which is located far from the receptor residues. In the case of compound 112 when the R₃ is positioned with larger group could significantly interact with the receptor, thus enhancing its inhibitory

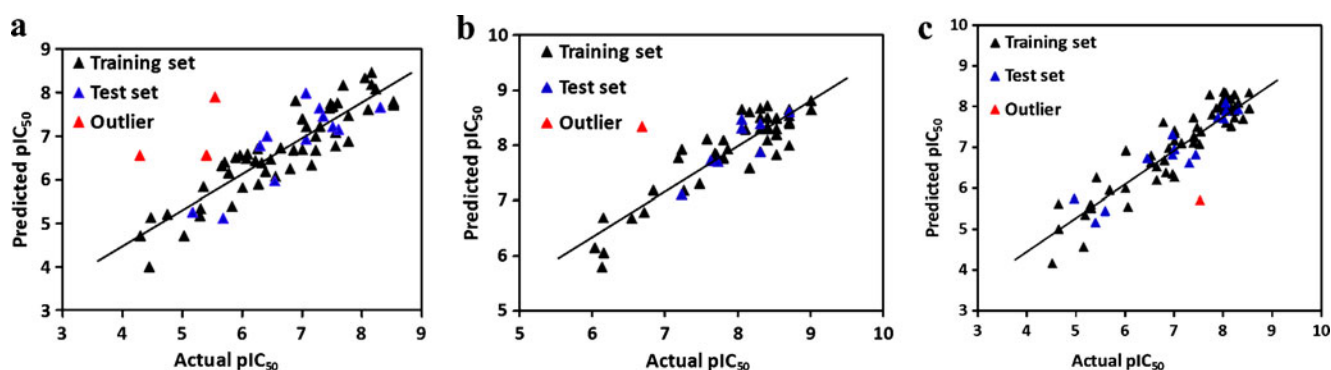



Fig. 4 Graphs of the predicted pIC_{50} versus the experimental pIC_{50} values of the optimal models. (a) CoMFA model of PP analogs. (b) CoMFA model of TP analogs. (c) CoMFA model of UTC analogs

activity, which is also supported by the steric contour maps (blue).

For CoMSIA analysis, the model was developed by using S and E fields, giving a result of $R_{cv}^2=0.587$ using nine variables. Inclusion of a third field such as H, D or A to this model also exhibits good statistical parameters with R_{cv}^2 of 0.508, 0.505 and 0.535, respectively. The SEHD, SEDA, SHDA and EHDA combinations also produce significant models (R_{cv}^2 of 0.521, 0.544, 0.548 and 0.544, respectively). So a conclusion can be drawn that the D field plays a key role in building the models, which is also explained by the greatest field contributions of D field in each model. The external predictive R_{pred}^2 is 0.745 for the CoMFA model and 0.436 for the CoMSIA model, respectively. Overall, the performance of the CoMFA model is superior to that of the CoMSIA one.

Compound 79 in the optimum CoMFA model is treated as an outlier. Inclusion of this compound, the model shows poor predictive ability with R_{pred}^2 of 0.192. Omission of compound 79 results in an increased R_{pred}^2 value of 0.745. Its outlier status is due to the low inhibitory activity ($pIC_{50}=6.6757$) which beyond the predictive power of this model and the residual between the experimental value and predicted value is nearly to 2.0 log unit. In addition, it could stem from its structural uniqueness (the only one compound

has  substitution at R_2), when compared to its counterparts, compounds 68–78 and 80. The correlation between the predicted activities and the actual activities is displayed in Fig. 4b, 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 and accuracy of the model.

UTC

The statistical results for the final CoMFA model are summarized in Table 2. The best CoMFA model (model CoMFA-SE)

has an R_{cv}^2 value of 0.673 using four components; the model also explains R_{ncv}^2 of 0.873, SEE of 0.42 and F value of 71.607, which indicates that the QSAR model are reliable and able to predict binding affinities of new derivatives accurately. CoMFA-SE model indicates contributions of S and E fields are 49.7% and 23.8%, respectively. In this work, $nRNHR$ and R_4U descriptors are also applied to constitute the models, which possess 16% and 10.5% contributions to the inhibitory activity. R_4U , a GETAWAY descriptor, represents R autocorrelation of lag 4 (unweighted), demonstrates the positive effect of the molecule geometry and size and shape properties. $nRNHR$ is a Functional group counts descriptor, which considers the number of secondary amines (aliphatic) of the molecules in building models, which can form H-bonds with the receptor. For example, compounds 131 and 132 ($nRNHR$ value=1) have higher $nRNHR$ values than compounds 128 and 129 ($nRNHR$ value=0), being in agree with the inhibitory potency, which is due to that the secondary amines in compounds 131 and 132 serve as H-bond donor to interact with the protein, thus enhancing its inhibitory activity.

For the CoMSIA analysis, the CoMSIA-SE model exhibits a proper internal predictivity demonstrated by the statistical results of $R_{cv}^2=0.72$, $R_{ncv}^2=0.833$, $SEE=0.478$ and $F=65.967$ based on the use of four components. Additionally, incorporation of other fields (H, D and A) also shows comparable internal predictivity (Table S8).

The external predictive R_{pred}^2 is 0.701 for the CoMFA model; examination of the residuals between the actual and the predicted values suggests that compound 152 might be an outlier. Omission of compound 152 results in an increased R_{pred}^2 value of 0.877. The outlier status is due to the higher residue between the observed and the predicted biological activity which further confirms the robustness and statistical confidence of the derived model. Figure 4c depicts the correlation between the experimental and the predictive activities of the CoMFA model. Clearly, all points are rather uniformly distributed around the regression line, indicating no existence of systematic errors in the model, which further proves the satisfactory predictive ability of the derived model.

All the models built for the three classes of compounds have shown that CoMFA models possess greater robustness compared to CoMSIA ones. The above QSAR models culminating from the training set yield a regression equation with a high degree of statistical significance and perform well in predicting the biological activities of compounds in the test set, which indicate the models are predictive and unbiased.

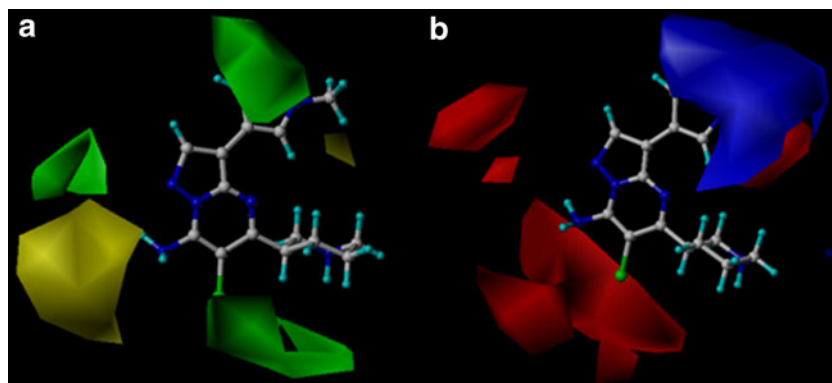
3D-QSAR contour maps

The $\text{stdev} \times \text{coeff}$ contour maps were constructed to view the field effects on the target features. The CoMFA contour maps denote the region in the space where the aligned molecules would favorably or unfavorably interact with the receptor while the CoMSIA contribution maps denote those areas within the specified region where the presence of a group with a particular physicochemical property will be favored or disfavored for biological activity.

PP

Figure 5 shows the CoMFA S and E contour maps. Compound 38 is served as reference molecule. In the CoMFA S field (shown in Fig. 5a), a large green contour situates near R_1 substituent indicates that bulky groups at this position are beneficial to activity. This discovery is well illustrated by the example that compound 4 (nitrile) is more potent than compound 1 (hydrogen). Another large green contour shown around R_3 substituent represents that bulky group would enhance the activity. The fact that compound 45 ($-\text{CH}=\text{CH}-\text{CH}_2\text{OMe}$) is more potent than compound 46 ($-\text{CN}$) is a good example. A large yellow contour is located around R_2 substituent, indicates the areas where steric bulk would decrease the inhibitory activity, which can be shown by the order of compounds 19 ($-\text{Me}$) > 20 ($-\text{Et}$) > 24 ($-\text{C}_6\text{H}_5$). Additionally, a large green contour is observed beside the yellow contour implies that both of the bulky favored and unfavored contours emerge at the same region,

Fig. 5 CoMFA $\text{StDev} \times \text{Coeff}$ contour plots for PP analogs. Compound 38 is superposed as the reference molecule in the map. **(a)** Steric field: favored-green (80%) and disfavored-yellow (20%); **(b)** Electrostatic field: favored-blue (85%) and disfavored-red (15%)



indicating that a balance of these properties among the groups present at this region is required for optimum binding.

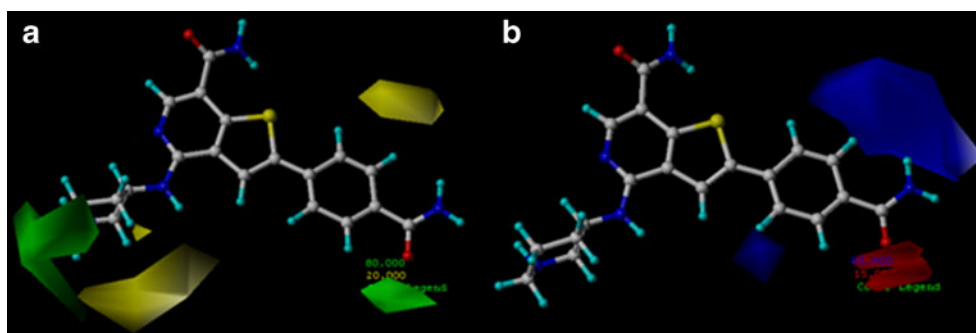
The E field contour is presented in Fig. 5b. A large and two small red isopleths around the R_3 substituent represent an area where an electronegative group is favored. Hydroxyethyl group (compound 44) at this position, bearing negative GH charges, increases the activity; while ethyl (compound 41) and propynyl group (compound 42) having positive charges decrease the activity. A red contour in the vicinity of the side chain of R_2 reveals that the electron-withdrawing substituent is essential for the inhibitory activity. This may be the reason why compound 27, possessing electron-withdrawing group (phenyl ring) has higher potencies than compound 22 with electron-donating group ($-\text{C}(\text{CH}_3)_3$). However, for the most potent compound 38, with an electropositive group ($-\text{H}$) attaching this contour is unfavorable for the inhibitory activity, thus modifications on R_2 substituent may be requirement for increasing the activity. Another red contour near R_1 substituent is conflict with the most potent compound 38 ($-\text{CH}_3$), suggests that modification may be made in this region. Furthermore, a large blue contour is observed outside the R_1 substituent and around the red contour, which strongly delimits the sideward relocatability, also indicates that electropositive groups extended to this contour would benefit for the inhibitory activity.

TP

The S and E field distributions of CoMFA are depicted in Fig. 6. Compound 112 is displayed in the map in aid of visualization. Figure 6a depicts the S contour map of the CoMFA model. Sterically favored regions (colored green) appear around R_3 substituent, which explains the experimental results that compounds 112, 120 and 121 have the activity order of 112 ($-\text{C}_6\text{H}_5$) > 120 ($-\text{Br}$) > 121 ($-\text{H}$).

Similarly, a broad region of green contour is displayed near the R_2 substituent, indicates that a bulky substituent at this

Fig. 6 CoMFA StDev*Coeff contour plots for TP analogs. Compound 112 is superposed as the reference molecule in the map. (a) Steric field: favored-green (80%) and disfavored-yellow (20%); (b) Electrostatic field: favored-blue (85%) and disfavored-red (15%)



site would be favorable. Exactly, the inhibitory activity of compounds 69 (substituted by hexatomic ring), 70 (five-membered ring), and 71 (tetratomic ring) is placed in the order as $71 < 70 < 69$. A yellow contour below the R_1 substituent shows that appropriately minor substituents at the position would improve the biological activities. For example, compound 68 with the minor substituent ($-H$) exhibits a considerable gain in binding affinity, while compounds 72

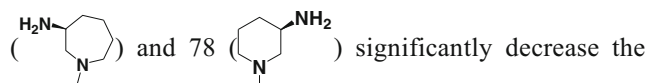
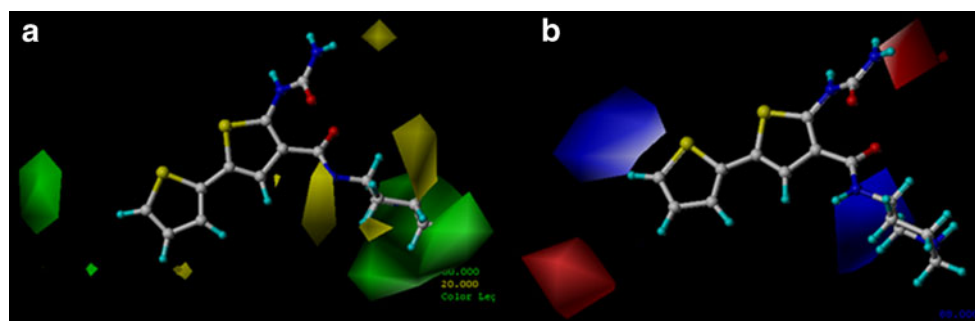


Fig. 7 CoMFA StDev*Coeff contour plots for UTC analogs. Compound 185 is superposed as the reference molecule in the map. (a) Steric field: favored-green (80%) and disfavored-yellow (20%); (b) Electrostatic field: favored-blue (88%) and disfavored-red (12%)



PP

As a reference for other compounds, ligand 38 was used as a model drug (Fig. 8) to show the docking modes of this class of CHK₁ inhibitors. The ligand-enzyme interaction analysis shows that Leu137, Glu91, Cys87, Asp148, Ser88, Gly90 and Tyr86 are the important residues present in the active site. Compound 38 binds to the kinase through key H-bonds interactions: (1) Between the -NH of ring A (Fig. 9a) and the -COO of Glu91 (-O...HN, 1.90 Å, 160.9°); (2) Between the backbone carbonyl oxygen of Cys87 and the amino of ring B (Fig. 9a) (-O...HN, 2.46 Å, 129.8°; -O...HN, 3.47 Å, 150.2°). (3) Between the nitrogen atom of ring C (Fig. 9a) and the -NH group of Cys87 residue (-N...HN, 2.37 Å, 157.6°). (4) The interaction between the nitrogen atom of ring D (Fig. 9a) and the side chain of Asp148 is bridged by a structural molecule W502. (5) Ring B and ring C all form arene-cation interaction with Leu137 which further enhance the binding activity.

Interestingly, the docking result is well consistent with the CoMFA contour map analysis, which further validate the

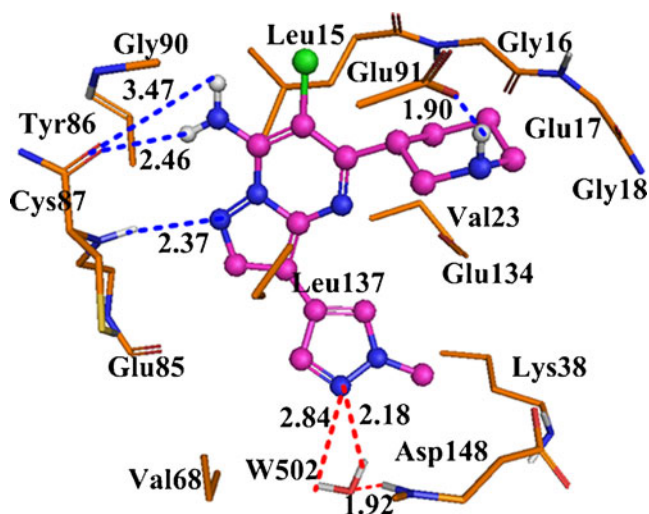


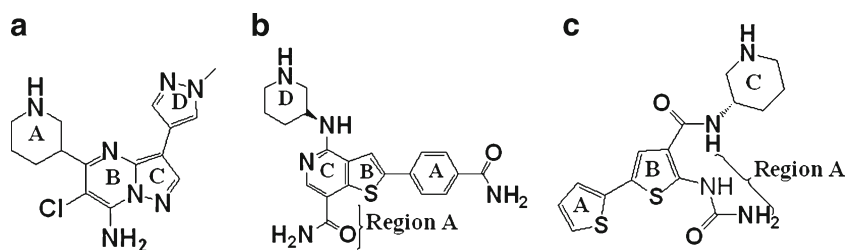
Fig. 8 Docked conformations derived for molecule 38 of PP analogs (shown in ball and stick). H-bonds formed by residues and molecule directly and mediated by water indirectly are shown as dotted lines with blue and red color, respectively. W502 represents water molecule. The nonpolar hydrogens are removed for clarity

QSAR model overall. The R₂ substituent is fitted nicely into a large hydrophobic pocket composed by Leu84, Ala36, Val23 and Val168, suggesting that a relatively bulky substituent is needed, which is evidenced by the presence of a green contour around this area (shown in Fig. 5a). Additionally, it is obvious that a large empty cavity is present on R₃ substituent. For compound 38, the R₃ substituent almost extends outside this binding site, indicating that in this area the steric interaction may be favorable, which is in accord with the contour map with a green contour in this region. The yellow contour observed near R₂ substituent indicates that steric potentials are not desirable, which is confirmed by the docking results that bulky substitutions at this position will make clash with residues Ser88, Gly90 and Tyr86. A red contour near R₁ substituent is surrounded by the positively charged Lys38 and the neutral amino acid residues Ala36 and Leu84, which may explain the increased activity expected from the introduction of electronegative groups in this region. Due to the H-bond formed by the -NH of ring A and backbone -COO group of Glu91, where the -NH as an H-bond acceptor and Glu91 as an H-bond donor, so negative charged substituents in this region are favorable for the inhibitory activity, which is evidenced by the red contour near this area. Interestingly, the -Cl at R₃ substituent almost has no interaction with the receptor, which incites us to investigate the reasons responsible for it. We compare compound 38 with compound 37 (possesses a -H at this position) by calculating the ClogP values, the partition coefficient of compound 38 is 1.25 while compound 37 is 0.53, the difference may account for the phenomenon that the -Cl at this position may effectively increase the liposoluble of the compound, which is favorable for its membrane transporting.

TP

The binding mode of compound 112 is shown in Fig. 10. The specific cleft in which the ligand binds (within 4.5 Å) contains polar (Glu17, Glu85, Tyr86, Cys87, Ser88, Gly89, Gly90, Glu91, Ser147 and Asp148) and non polar (Ala36, Leu84, Val23) amino acids. Compound 112 is placed inside the active site and demonstrates the following interactions:

Fig. 9 The structures of the most active molecules used in the contour analyses. (a) Compound 38 for the group of PP analogs. (b) Compound 112 for the group of TP derivatives. (c) Compound 185 for the group of UTC analogs



The oxygen atom of region A (Fig. 9b) forms H-bond with -NH group of Cys87 ($-O\cdots HN$, 2.07 Å, 149°). The -NH₂ group of region A forms two H-bonds with the backbone -CO of Glu85 ($-O\cdots HN$, 1.89 Å, 140.5°; $-O\cdots HN$, 3.5 Å, 147.4°). The -NH group of ring D (Fig. 9b) acts as a H-bond donor by binding to the -CO group of Glu17 ($-O\cdots HN$, 2.93 Å, 80°). The -NH group of region A is also involved in hydrogen bonding interaction with the backbone of Ser147 through a structural water molecule W401. Additionally, the -NH group between ring C (Fig. 9b) and ring D interacts with Glu91 through the bridge water molecule W360 and signifies the importance of the water molecules in ligand binding.

It is remarkable to note that on R₂ substituent there is sufficient room to introduce a large group into the hydrophilic pocket composed by Asp148, Glu17 and Glu91, which is similar to the S contour map (shown in Fig. 6a). Docked conformation of compound 112 within the binding pocket indicates that the electropositive residue Glu91 and the neutral amino acids Gly89, Gly90 appear near the phenyl group of R₃ substituent, where electronegative substituent would favor the interaction between ligands and CHK₁,

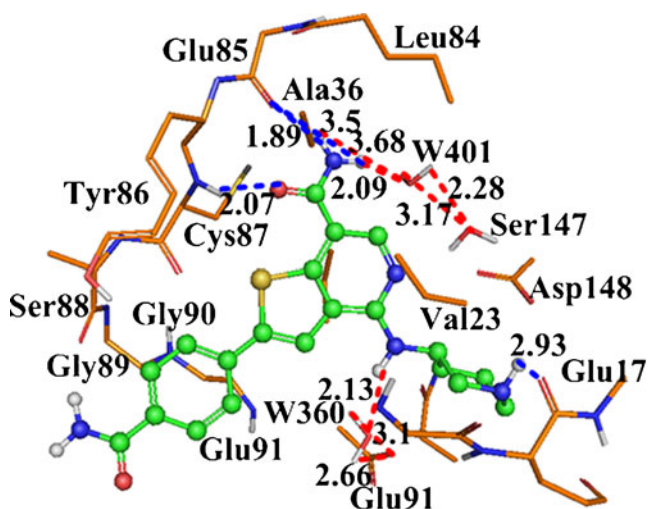


Fig. 10 Docked conformations derived for molecule 112 of TP (shown in ball and stick) CHK₁ kinase. H-bonds formed by residues and molecule directly and mediated by water indirectly are shown as dotted lines with blue and red color, respectively. W360, W401 and W419 represent water molecules. The nonpolar hydrogens are removed for clarity

which is evident from the presence of a blue contour at this position (Fig. 6b).

UTC

Compound 185 was selected as a template to show the docking mode of this class of inhibitors. The ligand-enzyme interaction shows that Cys87, Glu85, Leu84, Asn135, Ser147, Tyr86 and Ser88 are the most important residues present at the binding pocket. In Fig. 11, the binding mode is presented, which further reveals the main interactions within the enzyme active site. The observed interactions are: (1) The sulfur atom of ring B (Fig. 9c) forms H-bond with the backbone -NH of Cys87 ($-S\cdots HN$, 2.76 Å, 160°). (2) Three interactions of region A (Fig. 9c) (-NH group, -NH₂ group and the oxygen atom) with -CO of Glu85 ($-O\cdots HN$, 2.74 Å, 154.1°), the backbone of Glu85 ($-O\cdots HN$, 2.29 Å, 149.1°; $-O\cdots HN$, 3.94 Å, 142.4°) and -CO of Leu84 ($-O\cdots HN$, 3.76 Å, 67.8°), respectively. (3) The oxygen atom between ring B and ring C (Fig. 9c) interacts with Asn135 and Ser147 bridged by a water molecule W2021.

As displayed in Fig. 11, the hydrophilic amino acid residues Asp148, Asn135, Lys38 and Gly16 locate around R₂ substituent, which correlates well with the S contour map

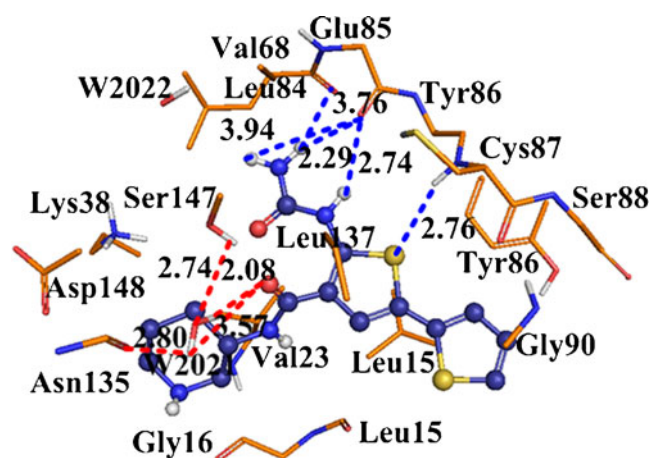


Fig. 11 Docked conformations derived for molecule 185 of UTC (shown in ball and stick) CHK₁ kinase. H-bonds formed by residues and molecule directly and mediated by water indirectly are shown as dotted lines with blue and red color, respectively. W2021 represents water molecule. The nonpolar hydrogens are removed for clarity

of the CoMFA model (shown in Fig. 7a). The R₃ substituent is surrounded by polar amino acids Gly90, Tyr86 and Ser88, suggesting that polar groups at this position are favorable for the binding affinity, this observation is corroborated by compound 112 possessing a polar group in this region, in addition, a yellow contour in this region suggests that bulky group would decrease the binding affinity; replacement with too large substituents would potentially lead to steric clash with the above described residues. Residues Asp148, Val23, Gly16 and Leu15 around R₂ substituent suggest that electropositive substituent would enhance the binding activities. The results match well with E contour map of the CoMFA model (Fig. 7b). The blue contour near the sulfur atom of R₃ substituent indicates an electropositive favorable region, which is confirmed by the docking results that electropositive substituents will have a favorable interaction with residues Leu15, Leu137 and Gly90.

Based on the above observations, we can speculate that the network of H-bonds formed between CHK₁ and the ligands are crucial for molecular recognition. For PP analogs, it is shown in Fig. 12a that Cys87 and Glu91 solidly form H-bonds with compound 38, while water molecule (W502) mediates interactions between compound 38 and residue Asp148. Additionally, two arene-cation interactions involving Leu137 are found to enhance the binding affinity. As to TP analogs, residues Glu17, Glu85, Cys87, Glu91 and Ser147 coupling with two water molecules (W401 and W360) form H-bonds with compound 112 (shown in Fig. 12b). Interestingly, residues Leu84, Glu85 and Cys87 are identified as the main contributors to the binding affinity of UTC analogs. And water molecule W2021 is found to mediate the hydrogen bonding interactions with residues Asn135 and Ser147 (shown in Fig. 12c).

By comparison, the main conclusions are summarized as follows: (1) One common amino acid residue Cys87 is found to possess hydrogen bonding interactions with the three series of inhibitors. (2) Glu91 is common for PP and TP analogs; For the TP and UTC derivatives, amino acid residues Glu85, Cys87 and Ser147 are involved in the

binding mode at the same time. (3) PP, TP and UTC inhibitors all form more than five H-bonds with CHK₁ kinase, indicating that they all exhibit potent inhibitory activity. (4) In the three systems, water mediated interactions are also significant for the binding affinity.

Consequently, the docking analysis reveals that the S, E, H and H-bonding interactions between ligands and key amino acid residues in the binding pocket of CHK₁ correlate well with the contour maps, indicating that the QSAR model is reasonable and can offer constructive suggestions for further modification of CHK₁ inhibitors.

Molecular dynamics simulations

To further evaluate the complementarity of the ligands and the receptor, MD simulation in explicit water is carried out. The derived best docked structure has been subjected to MD simulation. The three docked complex structures of 3OT3-38, 3PA5-112 and 2WMT-185 have performed 5 ns simulations.

PP

The RMSD of 3OT3 complex is shown in Fig. 13a (in blue). The RMSDs of the trajectory with respect to their initial structure range from 1.2 to 2 Å. Especially, between 2 and 2.5 ns, the value is totally off the baseline even up to 2.4 Å. However, after 0.5 ns, the complex seems to gradually stabilize and finally reach equilibrium. It is because that the new neighboring residues make the ligand incompatible with the surroundings, then it is pulled back to the active site again by that hydrogen bonding interactions. Additionally, the average conformation derived from the trajectory takes almost the original hydrogen binding mode. The RMSD of the ligand is also shown in Fig. 13a (shown in red), the RMSD of compound 38 increases to 1 Å after 0.2 ns and maintains this value until the end of the simulation, suggesting that the position and orientation of the ligand is stable.

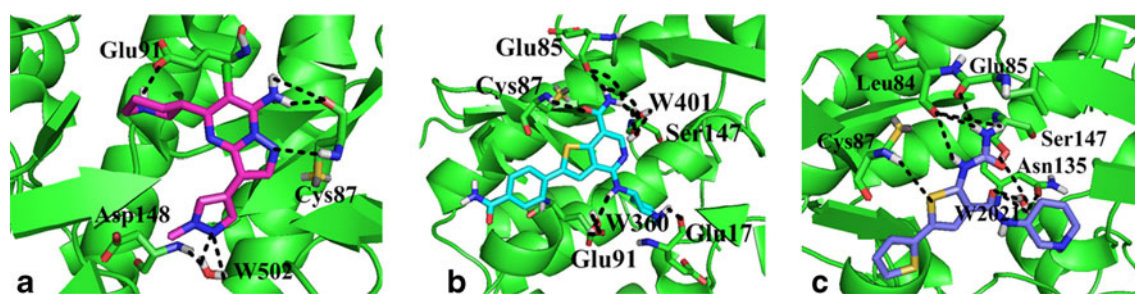


Fig. 12 Stereoview of the docked conformations of compounds 38, 112 and 185, respectively, in the active site of CHK₁ kinase. The H-bonds are shown by broken lines. Compounds 38, 112 and 185, colored purple, cyan and blue are presented in pictures a, b and c,

respectively. The important amino acid residues, Glu91, Cys87, Asp148, Glu85, Glu17, Leu84, Ser147, and Asn135 (stick rendering) are colored by atom type (C, yellow; N, blue; H, white; O, red)

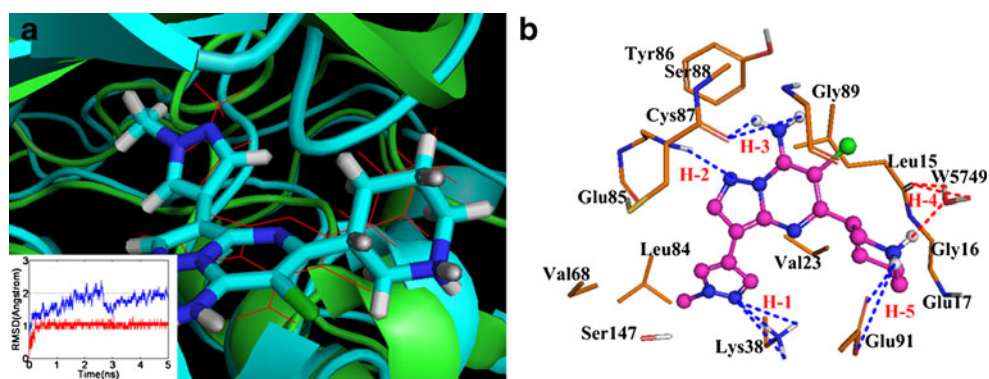


Fig. 13 (a) The root-mean-square deviation (RMSDs) of docked complex (in blue) and ligand (in red) versus the MD simulation time in the MD-simulated structures shown in the lower left corner. View of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound 38-3OT3 complex. Compound 38 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 38. H-bonds are shown as dotted blue and red lines; Active site amino acid residues are represented as sticks; the inhibitors are shown as stick and ball model

The binding mode of compound 38 with 3OT3 is shown in Fig. 13b; the ligand is anchored into the binding site via a network of H-bonds involving Glu91, Cys87, Lys38 and Leu15. Docking results of compound 38 into the receptor indicate that after MD simulation almost all H-bonds are preserved (between ring A and Glu91, ring B and Cys87, ring C and Cys87), three new H-bonds are also found to exist between the nitrogen atom of ring D and Lys38 of 3OT3 (H-1).

As can be seen from the docking results (shown in Fig. 8), one water molecule near the nitrogen atom of ring D bridges the ligand and receptor interaction, however, disappears after MD simulation, rather, a novel water molecule is involved in hydrogen-bonded network with the -NH of ring A.

TP

In the simulation of the 3PA5-112 complex, the RMSD is stable after 3 ns (in blue). The RMSD of the ligand is also

calculated to obtain information on position fluctuations and movements of ligand atoms (in red). The RMSD of compound 112 decreases to 0.3 Å in 1.2 ns and then rises up again to 1.3 Å after about 2.2 ns. As a result (Fig. 14), the H-bonds involving the -O of region A and Cys87 (H-7), the -NH₂ at region A and Glu85 (H-5), the -NH of ring D and Glu17 (H-2) are persistent after the simulation, leading to define such an interaction as the major contact between the protein and inhibitors. Moreover, new H-bonds are formed between the -S of ring B and Cys87 (H-8), the -NH₂ of region A and Leu15 (H-6), the -N of ring C and Glu17 (H-3), enhancing the stability of compound 112 in the binding site. It is also interesting to note the impact of water molecules on the interaction, the original water molecule presenting in the docking structure moves away during MD simulation and the interactions mediated by the two water molecules also vanished. However, another two new H-bonds are formed between the -NH of ring D and Glu91 (H-1), mediated by one

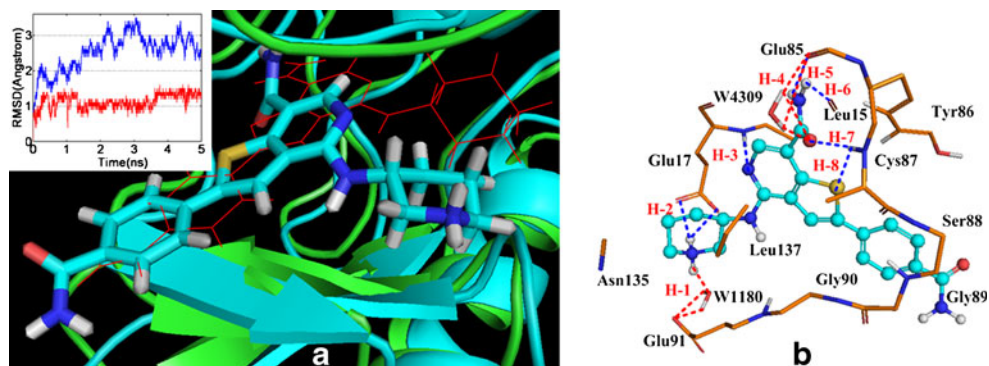


Fig. 14 (a) The root-mean-square deviation (RMSDs) of docked complex (in blue) and ligand (in red) versus the MD simulation time in the MD-simulated structures shown in the upper left corner. View of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound 112-3PA5 complex. Compound 112 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 112. H-bonds are shown as dotted blue and red lines; Active site amino acid residues are represented as sticks; the inhibitors are shown as stick and ball model

water molecule (W1180). Additionally, the H-bond between ring D and Glu17 with a distance of 2.93 Å is not strong enough for the docking model, after MD simulation, the increased H-bond interactions as the driving force to push ring D away from the original site (shown in Fig. 14a).

UTC

Based on the docking results, MD simulation is performed on the 2WMT-185 complex by using the Gromacs program. The stability of the complex is examined by a 5 ns MD simulation and a following RMSD calculation. Results show that the RMSDs of the trajectory with respect to the initial structure are depicted in Fig. 15a (in blue). The RMSD of ligand is obtained to get information on position fluctuations (in red). After an initial minor increase in the magnitude of ligand atoms fluctuation, the ligand reaches an equilibrium state characterized by the RMSD profile. The superimposition of the averaged structure of the CHK₁ with the initial model shows a plane rotated 30° compared with the initial docking structure, which can be explained by the significant conformational flexibility of ring C. The orientation of ring C in the binding pocket changes after MD simulation, due to the altered H-bond interactions. The results of the MD simulation suggest that the H-bonds (between the -S of ring B and the -NH of Cys87, the -NH₂ at region A and Leu84) have vanished, indicating that these hydrogen bonding interactions are weak. The H-bonds (-NH at region A and the oxygen between ring B and ring C) are located in the same way as obtained after molecular docking study. The vanished H-bonds are substituted by new ones which locate at the -NH of ring C with Asp148 (H-3), with Glu134 (H-1) and with Asn135 (H-2), the -NH between ring B and ring C with Gly16 (H-8). These H-bonds further increase the ligand binding affinity. Additionally, the original structural molecule W2021 has disappeared, which

account for the rotation of ring C. Conversely, two water molecules enter the binding pocket after MD simulation, mediate the H-bond interactions between the oxygen atom at region A and Asp148 (H-5), Glu22 (H-6), respectively. There is an improvement of the docking results for the complex, as water molecules occupy the position which further complements the disabled H-bond interaction.

Molecular docking offers reasonable binding structures for investigated ligands, while the MD simulation accounts for even the smallest variances. The obtained results of molecular docking and MD simulations confirm the existence of a suitable ligand binding site located inside the receptor. The MD simulation incorporates the flexibility of both ligand and receptor, improving interactions and enhancing complementarity between them. The stable RMSDs for the three classes indicate that the reliability of the docking procedure. Some new interactions are produced after MD simulation: 1) for PP analogs, there are three new H-bonds formed between the nitrogen atom of ring D and Lys38 of 3OT3, which indicates that Lys38 is important for the binding interactions; 2) for TP analogs, the altered orientation of ring D gives opportunity to produce more potent and stable interactions (H-1, H-3, H-6 and H-8); 3) for UTC analogs, the flexibility of ring C derived from the disappeared water molecule, makes some H-bonds vanished, conversely, the entered water molecules further produce stronger interactions (H-5 and H-6). In a word, the conformations obtained after MD simulation are more reasonable than the docked conformations.

Conclusions

For the first time we have described in this paper the employment of the QSAR method on three different series of CHK₁ inhibitors to investigate the structural relationship with the

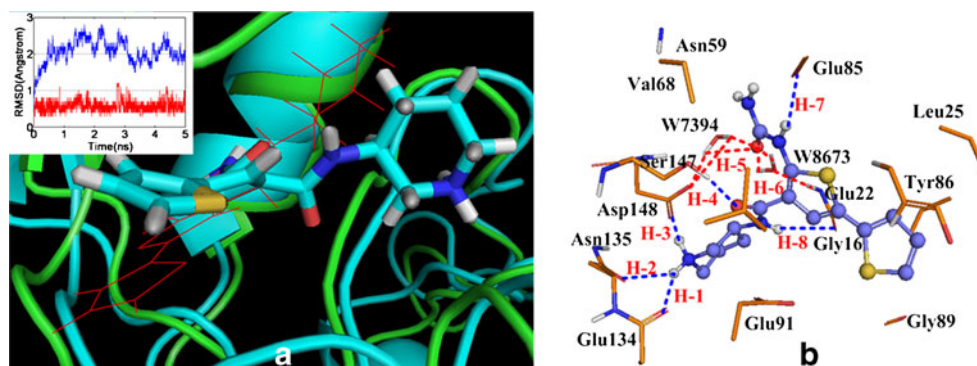


Fig. 15 (a) The root-mean-square deviation (RMSDs) of docked complex (in blue) and ligand (in red) versus the MD simulation time in the MD-simulated structures shown in the upper left corner. View of superimposed backbone atoms of the lowest energy structure of the MD simulation (cyan) and the initial structure (green) for compound

185-2WMT complex. Compound 185 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 185. H-bonds are shown as dotted blue and red lines; Active site amino

inhibitory activity. We also compared two alignment schemes employed in CoMFA and CoMSIA, namely, ligand-based and receptor-based alignment, with respect to the predictive ability and the robustness of the models. As a result, the models using the ligand-based alignment are superior to that based on the receptor alignment. Contour maps obtained with these models provide useful information about the intermolecular interactions of inhibitors with the surrounding environment. In addition, the docking and MD analysis reveal the important interactions between the receptor active site residues and the compound's functional groups, which are also found that one common residue Cys87, in the kinase active site played a significant role in recognition of the inhibitors by presenting hydrogen bonding interactions in the three classes of CHK₁ inhibitors.

- 1) For **PP analogs**, the contour maps reveal that at R₁ substituent, a bulky group can enhance the kinase activity; bulky, electronegative and hydrophobic substituents at R₂ substituent also can improve the inhibitory activity; minor and electronegative groups appearing around R₃ substituent would be beneficial for the activity. Moreover, the majority of interactions have been observed from docking structure which is conserved throughout the MD simulation, etc., the R₁ substituent mainly acts as H-bond acceptor and the R₂ substituent as H-bond donor to interact with the receptor. And Leu137, Glu91, Cys87, Asp148, Ser88, Gly90 and Tyr86 are the key residues for the ligand-receptor interactions.
- 2) For **TP analogs**, the hybrid effects of steric and electrostatic interactions are found to be crucial to the inhibitory activity in the CoMFA model. Small substituent at R₁ area, bulky and hydrophilic groups at R₂ substituent, bulky and electronegative groups in R₃ region are favorable for the inhibitory activity. The docking and MD simulation study further reveal that the key residues are Glu17, Glu85, Glu91, Cys87 and Ser147. The -NH group between ring C and ring D, the -NH at ring D, the -NH₂ of region A form the main H-bonds with the receptor.
- 3) For **UTC analogs**, according to the statistical analysis of the standard deviation residues, we assume the steric and electrostatic model derived from the CoMFA as the best predictive model. Substitution of small and polar groups at R₂ substituent, bulky, electropositive and hydrophilic groups at R₃ substituent exhibit positive effect on the activity. Key residues impacting the ligand-receptor interactions are Cys87, Glu85, Leu84, Asn135, Ser147, Tyr86 and Ser88. Additionally, the oxygen atom at R₂ substituent, the S atom of ring B, the -NH and -NH₂ groups of region A form H-bonds with the receptor. And the increased hydrogen bonding

with active-site residues result in more stable complexes during the MD simulation.

The present study provides an example of identifying the correct binding mode of CHK₁ inhibitors using the QSAR, molecular docking, and molecular dynamics approaches. Substantial ability of the model obtained to predict the external test set molecules supports that the deduced model can be used for predicting the related activity of the inhibitors and designing of novel potent CHK₁ inhibitors.

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

References

1. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396:643–649
2. Kastan MB, Bartek J (2004) Cell-cycle checkpoints and cancer. *Nature* 432:316–323
3. Shinohara KI, Narita A, Oyoshi T, Bando T, Teraoka H, Sugiyama H (2004) Sequence-specific gene silencing in mammalian cells by alkylating pyrrole-imidazole polyamides. *J Am Chem Soc* 126:5113–5118
4. Kawabe T (2004) G2 checkpoint abrogators as anticancer drugs. *Mol Cancer Ther* 3:513–519
5. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 73:39–85
6. Chen P, Luo C, Deng YL, Ryan K, Register J, Margosiak S, Tempezyk-Russel A, Nguyen B, Myers P, Lundgren K (2000) The 1.7 Å crystal structure of human cell cycle checkpoint kinase Chk1: implications for Chk1 regulation. *Cell* 100:681–692
7. Abraham RT (2001) Cell cycle checkpoint signaling through the ATM and ATR kinase. *Genes Dev* 15:2177–2196
8. Takai H, Tominaga K, Motoyama N, Minamishima YA, Nagahama H, Tsukiyama T, Ikeda K, Nakayama K, Nakanishi M (2000) Aberrant cell cycle checkpoint function and early embryonic death in *chk1*^{-/-} mice. *Genes Dev* 14:1439–1447
9. Chen Z, Xiao Z, Chen J, Ng SC, Sowin T, Sham H, Rosenberg S, Fesik S, Zhang H (2003) Human Chk1 expression is dispensable for somatic cell death and critical for sustaining G2 DNA damage checkpoint. *Mol Cancer Ther* 2:543–548
10. Wang H, Wang X, Zhou XY, Chen DJ, Li GC, Iliakis G, Wang Y (2002) Ku affects the ataxia and Rad 3-related/CHK1-dependent S phase checkpoint response after camptothecin treatment. *Cancer Res* 62:2483–2487
11. Huang S, Garbaccio RM, Fraley ME, Steen J, Kreatsoulas C, Hartman G, Stirdivant S, Drakas K, Rickert E, Walsh K, Hamilton CA, Buser J, Hardwick X, Mao M, Abrams S, Beck B, Tao W, Lobell R, Sepp-Lorenzino L, Yan Y, Ikuta M, Murphy JZ, Sardana V, Munshi S, Kuo L, Reilly M, Mahan E (2006) Development of 6-substituted indolylquinolinones as potent Chk1 kinase inhibitors. *Bioorg Med Chem Lett* 16:5907–5912
12. Zhou BBS, Elledge SJ (2000) The DNA damage response: putting checkpoints in perspective. *Nature* 408:433–439
13. Berlinck RGS, Britton R, Piers E, Lim L, Roberge M, Moreira da Rocha R, Andersen RJ (1998) Granulatimide and isogranulatimide, aromatic alkaloids with G2 checkpoint inhibition activity

- isolated from the Brazilian Ascidian *Didemnum granulatum*: structure elucidation and synthesis. *J Org Chem* 63:9850–9856
14. Jiang X, Zhao B, Britton R, Lim LY, Leong D, Sanghera JS, Zhou BBS, Piers E, Andersen RJ, Roberge M (2004) Inhibition of Chk1 by the DNA damage checkpoint inhibitor iso-granulatimide. *Mol Cancer Ther* 3:1221–1227
 15. Sausville EA, Arbuck SG, Messmann R, Headless D, Bauer KS, Lush RM, Murgo A, Figg WD, Lahusen T, Jaken S, Jing XX, Roberge M, Fuse E, Kuwabara T, Senderowicz AM (2001) Phase I trial of 72-hour continuous infusion UCN-01 in patients with refractory neoplasms. *J Clin Oncol* 19:2319–2333
 16. Shao RG, Cao CX, Shimizu T, O'Connor PM, Kohn KW, Pommier Y (1997) Abrogation of an S-phase checkpoint and potentiation of camptothecin cytotoxicity by 7-hydroxystaurosporine (UCN-01) in human cancer cell lines, possibly influenced by p53 function. *Cancer Res* 57:4029–4035
 17. Zabludoff SD (2008) AZD7762, a novel checkpoint kinase inhibitor, drives checkpoint abrogation and potentiates DNA-targeted therapies. *Mol Cancer Ther* 7:2955–2966
 18. Blasina A, Hallin J, Chen E, Arango ME, Kraynov E, Register J, Grant S, Ninkovic S, Chen P, Nichols T, Connor PO, Anderes K (2008) Breaching the DNA damage checkpoint via PF00477736, a novel small-molecule inhibitor of checkpoint kinase 1. *Mol Cancer Ther* 7:2394–2404
 19. Parry DA, Shanahan F, Davis N, Wiswell D, Seghezzi W, Pierce R, Hsieh Y, Paruch K, Guzi T, Biopharma SP (2009) Targeting the replication checkpoint with a potent and selective CHK1 inhibitor. In: Proceedings of the 100th Annual Meeting of the American Association for Cancer Research; 2009 Apr 18–22; Denver, CO. Philadelphia (PA): AACR; 2009. Abstract nr {2490}
 20. Fuse E, Tani H, Kurata N, Kobayashi H, Shimada Y, Tamura T, Sasaki Y, Tanigawara Y, Lush RD, Headlee D, Figg WD, Arbuck SG, Senderowicz AM, Sausville EA, Akinaga S, Kuwabara T, Kobayashi S (1998) Unpredicted clinical pharmacology of UCN-01 caused by specific binding to human alpha1-acid glycoprotein. *Cancer Res* 58:3248–3253
 21. Liu JL, Wang FF, Ma Z, Wang X, Wang YH (2011) Structural determination of three different series of compounds as Hsp90 inhibitors using 3D-QSAR modeling, molecular docking and molecular dynamics methods. *Int J Mol Sci* 12:946–970
 22. Wang FF, Li Y, Ma Z, Wang X, Wang YH (2011) Structural determinants of benzodiazepinedione/peptide-based p53-HDM2 inhibitors using 3D-QSAR, docking and molecular dynamics. *J Mol Model*. doi:10.1007/s00894-011-1041-4
 23. Li Y, Wang YH, Yang L, Zhang SW, Liu CH, Yang SL (2005) Comparison of steroid substrates and inhibitors of P-glycoprotein by 3D-QSAR analysis. *J Mol Struct* 733:111–118
 24. Wang RW, Zhou L, Zuo ZL, Ma X, Yang M (2010) 3D-QSAR studies of checkpoint kinase 1 inhibitors based on molecular docking and CoMFA molecular simulation. *Mol Simul* 36:87–110
 25. Du U, Xi LL, Lei BL, Lu J, Li JZ, Liu HX, Yao XJ (2010) Structure-based quantitative structure-activity relationship studies of checkpoint kinase 1 inhibitors. *J Comput Chem* 31:2784–2793
 26. Dwyer MP, Paruch K, Labroli M, Alvarez C, Keertikar KM, Poker C (2011) Discovery of pyrazolo[1,5-a]pyrimidine-based CHK1 inhibitors: a template-based approach-part 1. *Bioorg Med Chem Lett* 21:467–470
 27. Labroli M, Paruch K, Labroli M, Alvarez C, Keertikar KM, Poker C (2011) Discovery of pyrazolo[1,5-a]pyrimidine-based CHK1 inhibitors: a template-based approach-part 2. *Bioorg Med Chem Lett* 21:471–474
 28. Zhao LY, Zhang YX, Dai CY, Guzi T, Wiswell D, Seghezzi W (2010) Design, synthesis and SAR of thienopyridines as potent CHK1 inhibitors. *Bioorg Med Chem Lett* 20:7216–7221
 29. Janetka JW, Almeida L, Ashwell S, Brassil PJ, Daly K, Deng C (2008) Discovery of a novel class of 2-ureido thiophene carboxamide checkpoint kinase inhibitors. *Bioorg Med Chem Lett* 18:4242–4248
 30. Clark M, Cramer RDV (1989) Validation of the general-purpose tripos 5.2 force field. *J Comput Chem* 10:982–1012
 31. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J Chem Inf Comput Sci* 29:163–172
 32. Klebe G (1994) The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J Mol Biol* 237:212–235
 33. Cramer RD, Bunce JD, Patterson DE (1988) Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Struct Act Relat* 7:18–25
 34. Jain AN (2003) Surfex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* 46:499–511
 35. Rajni M, Ian AC, Sreedhara RV (2009) Assessment of the putative binding conformation of a pyrazolopyridine class of inhibitors of MAPKAPK2 using computational studies. *Eur J Med Chem* 1:98–105
 36. Li Y, Wang YH, Ding J, Wang Y, Chang YQ, Zhang SW (2008) In silico prediction of androgenic and nonandrogenic compounds using random forest. *QSAR Comb Sci* 27:1183–1192
 37. Wang YH, Li Y, Yang SL, Yang L (2005) Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach. *J Chem Inf Comput Sci* 45:750–757
 38. Wang YH, Li Y, Yang SL, Yang L (2005) An insilico approach for screening flavonoids as P-glycoprotein inhibitors based on Bayesian-regularized neural network. *J Comput Aided Mol Des* 19:137–147
 39. Dragon, 5.3, Milano Chemometrics and QSAR Research Groups Inc. 2002
 40. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Med* 7:306–317
 41. Van Aalten DMF, Bywater R, Findlay JBC, Hendlich M, Hooft RWW, Vriend G (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des* 10:255–262
 42. <http://davapc1.bioch.dundee.ac.uk/programs/prodrg/prodrg.html>
 43. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Comput Chem* 5:3684–3690
 44. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52:7182–7190
 45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh ewald method. *Chem Phys* 103:8577–8593
 46. Hess B, Bekker H, Berendsen HJC, Fraaije J (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472
 47. Dragon molecular descriptor theory. In: Talet srl, DRAGON for Windows, v. 5.4, 2006