


Cite this: *RSC Adv.*, 2023, 13, 14808

# *In silico* studies of a novel scaffold of benzoxazole derivatives as anticancer agents by 3D-QSAR, molecular docking and molecular dynamics simulations†

Yuhan Jiang,<sup>‡a</sup> Wei Yang,<sup>‡b</sup> Fangfang Wang<sup>ID</sup> <sup>\*a</sup> and Bo Zhou<sup>c</sup>

The vascular endothelial growth factor receptor-2 kinases (VEGFR-2) expressed on tumor cells and vessels are attractive targets for cancer treatment. Potent inhibitors for the VEGFR-2 receptor are novel strategies to develop anti-cancer drugs. In this work, template ligand-based 3D-QSAR studies were performed on a series of benzoxazole derivatives toward different cell lines (HepG2, HCT-116 and MCF-7). Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) techniques were used to generate 3D-QSAR models. Good predictability was derived for the optimal CoMFA models (HepG2:  $R_{cv}^2 = 0.509$ ,  $R_{pred}^2 = 0.5128$ ; HCT-116:  $R_{cv}^2 = 0.574$ ,  $R_{pred}^2 = 0.5597$ ; MCF-7:  $R_{cv}^2 = 0.568$ ,  $R_{pred}^2 = 0.5057$ ) and CoMSIA models (HepG2:  $R_{cv}^2 = 0.711$ ,  $R_{pred}^2 = 0.6198$ ; HCT-116:  $R_{cv}^2 = 0.531$ ,  $R_{pred}^2 = 0.5804$ ; MCF-7:  $R_{cv}^2 = 0.669$ ,  $R_{pred}^2 = 0.6577$ ). In addition, the contour maps derived from CoMFA and CoMSIA models were also generated to illustrate the relationship between different fields and the inhibitory activities. Moreover, molecular docking and molecular dynamics (MD) simulations were also conducted to understand the binding modes and the potential interactions between the receptor and the inhibitors. Some key residues (Leu35, Val43, Lys63, Leu84, Gly117, Leu180 and Asp191) were pointed out for stabilizing the inhibitors in the binding pocket. The binding free energies for the inhibitors agreed well with the experimental inhibitory activity and indicated that steric, electrostatic and hydrogen bond interactions are the main driving force for inhibitor-receptor binding. Overall, a good consistency between theoretical 3D-SQAR and molecular docking and MD simulation studies would provide directions for the design of new candidates, avoiding time-consuming and costly synthesis and biological evaluations. On the whole, the results derived from this study could expand the understanding of benzoxazole derivatives as anticancer agents and would be of great help in lead optimization for early drug discovery of highly potent anticancer activity targeting VEGFR-2.

Received 26th February 2023  
Accepted 1st May 2023

DOI: 10.1039/d3ra01316b

rsc.li/rsc-advances

## 1. Introduction

In vertebrates, many physiological processes including embryogenesis, organ development, estrus, and wound healing require vascular growth and remodeling.<sup>1–3</sup> Furthermore, several diseases such as tumor growth, metastasis, psoriasis, rheumatoid arthritis, macular degeneration and retinopathy are closely related to angiogenesis.<sup>4–7</sup> Activation of tumor angiogenesis originates from the imbalance between pro-

angiogenesis and anti-angiogenesis factors.<sup>8,9</sup> Extensive studies have been carried out and have discovered several factors: vascular endothelial growth factor (VEGF), angiopoietin (Ang), and ephrins. VEGF as a specific endothelial cell growth factor has been shown to be essential for angiogenesis cascade.<sup>10</sup> Anti-VEGF monoclonal antibodies have been employed to inhibit the VEGF pathway, which result in inhibiting a variety of human tumors.<sup>11</sup>

Among angiogenic factors, vascular endothelial growth factor-A (VEGF-A) is involved in the basic signaling of angiogenesis, especially signals for endothelial cell growth and survival *in vivo*.<sup>12</sup> And it activates two tyrosine kinase receptors, VEGFR-1 (Flt-1) and VEGFR-2 (KDR in humans/Flk-1 in mice).<sup>13</sup> VEGFR-1 and VEGFR-2 possess different signaling pathways and functions. VEGFR-1 is mainly responsible for the regulation of monocyte and macrophage migration, VEGFR-2 is the major signal transducer for the differentiation of endothelial cells from precursor mesodermal cells and the growth of endothelial

<sup>a</sup>School of Life Science, Linyi University, Linyi, 276000, China. E-mail: yu100288@163.com

<sup>b</sup>National Clinical Research Center for Infectious Diseases, Shenzhen Third People's Hospital, Shenzhen, 518112, China

<sup>c</sup>State Key Laboratory of Functions and Applications of Medicinal Plants, College of Basic Medical, Guizhou Medical University, Guizhou, 550004, China

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3ra01316b>

‡ These authors contributed equally to this work.



cells in early embryogenesis.<sup>14</sup> Since VEGFR-2 plays a substantial role in tumor neoangiogenesis, thus it has become an appropriate target for suppression of solid tumor growth.<sup>15</sup>

VEGFR-2 including 1356 amino acids in humans can be divided into 4 domains: the extracellular ligand-binding domain with seven immunoglobulin (Ig)-like domains, transmembrane domain, tyrosine kinase domain with an about 70 amino acids insert and carboxyl terminal domain.<sup>16</sup> Studies have proven that VEGFR-2 can bind all VEGF-A isoforms, further induce a cascade of different signaling pathways. The dimerization of the receptor and the autophosphorylation of the intracellular tyrosine kinase domain result in the simultaneous activation of PLC- $\gamma$ -Raf kinase-MEK-MAP kinase and PI3K-AKT pathways, leading to following cellular proliferation and endothelial-cell survival. Therefore, specific inhibitors binding with the different domains of VEGFR-2 would be employed for the treatment of cancers.<sup>17,18</sup>

In the early years, a rat mAb named DC101 was developed against VEGFR-2, which mainly interferes the binding of VEGF to VEGFR-2, and blocks tumor growth in mice.<sup>19</sup> However, this ligand cannot bind the human VEGFR-2, thus limiting its further development. For this reason, more efforts are devoted to finding molecules for clinical use,<sup>20</sup> for example, ramucirumab (IMC-1121B) is the only tested in human subjects, which is a novel human IgG1 mAb that would inhibit VEGFR-2 (selectively bind to the extracellular domain) with IC<sub>50</sub> of 0.8–1.0 nM.<sup>21,22</sup> Additionally, orally active and small inhibitors of VEGFR-2 have been studied: phthalazines,<sup>23</sup> indolinones,<sup>24</sup> quinazolines,<sup>25</sup> isothiazoles,<sup>25</sup> thienopyridines,<sup>26</sup> thiazoles,<sup>27</sup> pyridines,<sup>28</sup> indoles,<sup>29</sup> imidazopyridines,<sup>30</sup> benzimidazoles,<sup>31</sup> pyrimidines,<sup>32</sup> biheteroaryls,<sup>33</sup> sorafenib,<sup>34</sup> sunitinib,<sup>34</sup> pazopanib<sup>34</sup> and apatinib.<sup>35</sup> Some VEGFR-2 inhibitors undergoing clinical evaluation are also developed, for example, PTK-787,<sup>36</sup> SU-11248,<sup>37</sup> ZD-6474,<sup>25</sup> CP-547632,<sup>38</sup> and CEP-7055.<sup>39</sup> Recent paper has found that a series of benzoxazole derivatives as VEGFR-2 inhibitors exert anticancer activity against HepG2, HCT-116, and MCF-7 cells. However, the structure–activity relationship for these inhibitors have not been performed on these inhibitors.

In the present work, a molecular modeling study combining three dimensional structure–activity relationship (3D-QSAR), molecular docking, molecular dynamics (MD), and molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) were employed to discover the detailed binding mode of these inhibitors and identify the key amino acid residues responsible for inhibition. Finally, the best 3D-QSAR models were thoroughly analyzed. The interpretations derived from molecular docking, MD simulation were compared for model validation. Overall, this study would offer useful references for the discovery and design of novel VEGFR-2 inhibitors, and serve as the approachable end resultant hits that can be applied further for clinical trials evaluation in cancer.

## 2. Methods and materials

### 2.1. Preparation of data set

A series of substituted benzoxazole derivatives as VEGFR-2 inhibitors expressed inhibitory activity (IC<sub>50</sub>) against the

HepG2, MCF-7, and HCT-116 cell lines were employed for the current study.<sup>40</sup> The inhibitory activities were initially converted into corresponding pIC<sub>50</sub> (–log IC<sub>50</sub>) values, which have been used as dependent variables in the 3D-QSAR analyses, as shown in Table 1. Additionally, to derive the 3D-QSAR models, the whole data set was divided into the training set for the construction of CoMFA and CoMSIA models and the test set for validating the reliability of the model.

According to related ref. 41–43, the following criteria was considered when selecting the test set (labeled as a in Table 1): the inhibitory values in the test set should be distributed in various orders of magnitude in proportion to the whole data set.

### 2.2. Molecular modeling and alignment

All structures of inhibitors were constructed using Sybyl-X 1.1 (Tripos Associates, St. Louis, MO). Then, the structures were optimized using the default Powell method with Gasteiger–Hückel charge, Tripos force field,<sup>44</sup> the maximum number of iterations of 1000, and the termination of 0.005 kJ mol<sup>–1</sup>. Furthermore, the other parameters were set as default values as in Sybyl software.

Molecular alignment is a crucial procedure in building accurate and reliable 3D-QSAR models.<sup>45</sup> Some alignments approaches have been developed, *i.e.*, rigid body-based, pharmacophore-based, and docking-based alignment to understand the effect of different substitutions on the inhibitory activity. In the present work, template ligand-based alignment was applied to derive the optimum 3D-QSAR models. In template ligand-based alignment, compound 11, exhibiting the highest activity in the whole data set, was selected as the template compound. The remaining compounds were aligned by the maximum common substructures (shown in blue in Fig. 1A). The alignment results of all inhibitors targeted on HepG2, HCT-116 and MCF-7 are shown in Fig. 1B–D.

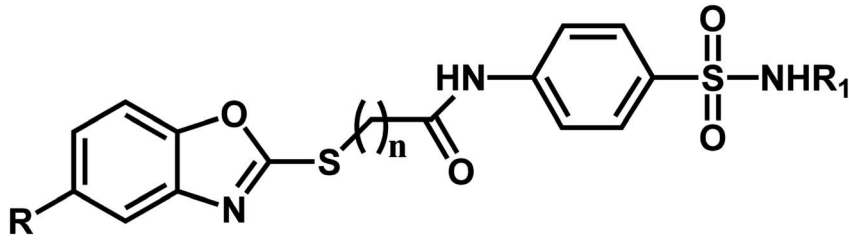
### 2.3 3D-QSAR studies

In this work, CoMFA and CoMSIA methods were employed to explore the structure–activity relationships of the active inhibitors, which can connect different force fields of compounds with related activity in a quantificational way.<sup>46</sup>

In CoMFA analysis,<sup>47</sup> the steric and electrostatic fields were computed by Lennard Jones and coulombic potential functions, respectively. The aligned compounds were placed in the 3D cubic lattice with a grid spacing of 2 Å. A sp<sup>3</sup> carbon atom with van der Waals radius of 1.52 Å and a charge of +1.0 were used to derive CoMFA descriptors. In addition, the CoMFA field was scaled using the CoMFA-STD method with a default energy of 30 kcal mol<sup>–1</sup>.

In addition to steric and electrostatic fields, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields were also calculated in the CoMSIA method.<sup>48</sup> CoMSIA descriptors were also derived with the same lattice box as used in CoMFA. The probe atom possessed a hydrophobicity property of +1, and hydrogen bond donor and acceptor properties of +1. The attenuation factor and minimum column filtering was set to 0.3



Table 1 Molecular structures of benzoxazole derivatives and the binding affinity pIC<sub>50</sub> values



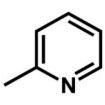
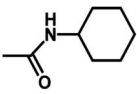
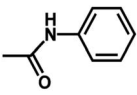
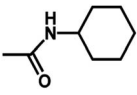
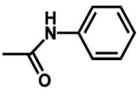
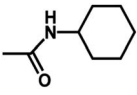
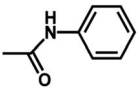
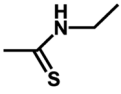
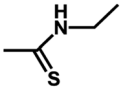
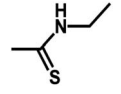
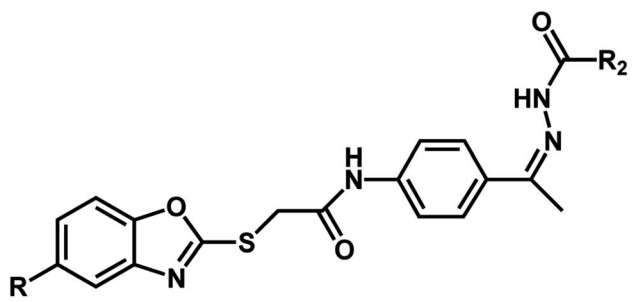

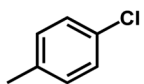


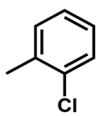
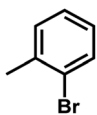
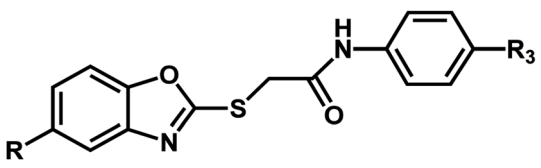
Compound	<i>n</i>	R	R <sub>1</sub>	pIC <sub>50</sub> (HepG2)	pIC <sub>50</sub> (HCT-116)	pIC <sub>50</sub> (MCF-7)
1	1	H	H	4.0508	4.3152	4.2278
2	1	H		4.6048 <sup>a</sup>	4.7650 <sup>a</sup>	4.7846 <sup>a</sup>
3	1	H		4.8142	4.7857	4.7375
4	1	CH <sub>3</sub>	H	4.3496 <sup>a</sup>	4.5955	4.5709
5	2	H	H	4.1712	4.2019 <sup>a</sup>	4.1352 <sup>a</sup>
6	2	CH <sub>3</sub>	H	4.7442	4.5782 <sup>a</sup>	4.4594 <sup>a</sup>
7	1	H		4.8811	4.7368	4.7111
8	1	H		4.8502	4.6828	4.5508
9	2	H		5.2269 <sup>a</sup>	5.1463 <sup>a</sup>	5.0491 <sup>a</sup>
10	2	H		5.058	5.0214	5.0022
11	2	CH <sub>3</sub>		5.384	5.1593	5.0620
12	2	CH <sub>3</sub>		5.1818	5.0410	4.9952
13	1	CH <sub>3</sub>		4.6702	4.5319	4.4342
14	2	H		5.0915	5.1018	4.9097
15	2	CH <sub>3</sub>		5.0022	4.9038	4.8041

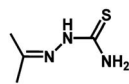
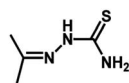
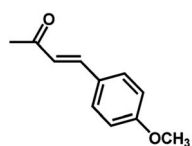
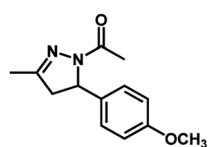


Table 1 (Contd.)



Compound	<i>n</i>	R	R <sub>2</sub>	pIC <sub>50</sub> (HepG2)	pIC <sub>50</sub> (HCT-116)	pIC <sub>50</sub> (MCF-7)
18	1	H		4.3669	4.5018	4.3408
19	1	H		4.3775	4.5158	4.4148
20	1	H		4.1661	4.1880 <sup>a</sup>	4.1236 <sup>a</sup>
21	1	H		4.1311 <sup>a</sup>	4.3363	4.2530
24	1	H		—	4.1230	4.1065
25	1	H		—	4.1066	4.0898



Compound	<i>n</i>	R	R <sub>3</sub>	pIC <sub>50</sub> (HepG2)	pIC <sub>50</sub> (HCT-116)	pIC <sub>50</sub> (MCF-7)
16	1	H		4.4444	4.3458 <sup>a</sup>	4.2608 <sup>a</sup>
17	1	CH <sub>3</sub>		4.9401 <sup>a</sup>	4.9851	4.8604
22	1	H		4.3143	4.3429	4.3746
23	1	H		4.4407	4.3536	4.3882

<sup>a</sup> Represents the test set.

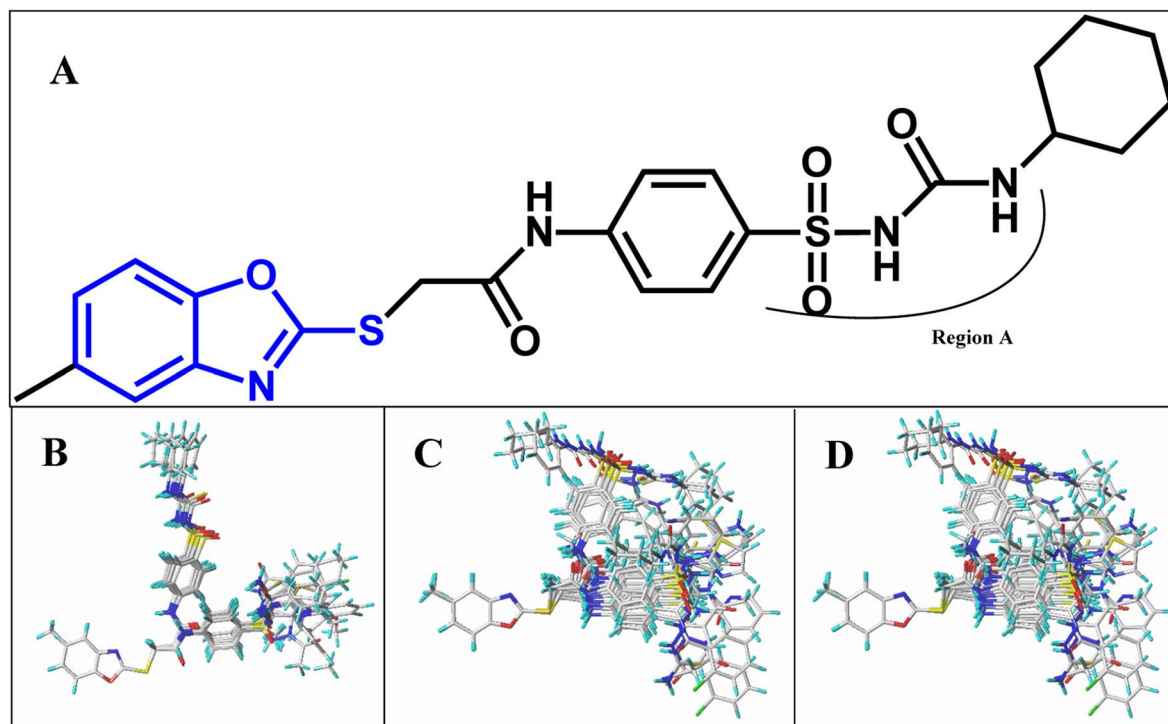


Fig. 1 (A) Cpd11 is used as a template for alignment. The common substructure is shown in blue. (B) Present the alignment for HepG2 compounds. (C) Present the alignment for HCT-116 compounds. (D) Present the alignment for MCF-7 compounds.

and 2.0 kcal mol<sup>-1</sup>, respectively. More importantly, the distance dependence between the probe atom and each inhibitor atom was measured by a Gaussian function.<sup>49</sup> Compared with CoMFA method, CoMSIA possesses a good performance in visualization and explanation of the contained correlations in terms of field contributions.<sup>50</sup>

#### 2.4 Partial least-squares (PLS) analysis

PLS regression analysis was applied to build statistically significant 3D-QSAR models, which establish correlations between the CoMFA/CoMSIA descriptors and inhibitory activities, as independent and dependent variables, respectively. Initially, the cross-validation using leave-one-out (LOO) method was performed to evaluate the reliability of the constructed models, resulting in the optimum number of components (ONC), the cross-validated correlation coefficient ( $R_{cv}^2$ ) and the standard predicted errors (SEP). The  $R_{cv}^2$  is computed with equation:

$$R_{cv}^2 = 1 - \frac{\sum (Y_{pred} - Y_{exp})^2}{\sum (Y_{exp} - Y_{mean})^2} \quad (1)$$

where  $Y_{pred}$ ,  $Y_{exp}$  and  $Y_{mean}$  represent the predicted activity, the experimental activity and the mean activity, respectively.

Then, non-cross-validation analysis was performed using the derived ONC to generate the final PLS regression models with the conventional correlation coefficient ( $R_{ncv}^2$ ), standard error of estimation (SEE) and the  $F$  values calculated.

#### 2.5 3D-QSAR model validation

The robustness and statistical significance of the 3D-QSAR models was evaluated with the test set inhibitors.<sup>51</sup> The inhibitory activities of the test set were predicted based on the derived models and the predictive correlation coefficient ( $R_{pred}^2$ ) is computed using the following equations:

$$R_{pred}^2 = 1 - \frac{PRESS}{SD} \quad (2)$$

$$PRESS = \sum (Y_{exp} - Y_{pred})^2 \quad (3)$$

$$SD = \sum (Y_{exp} - Y_{mean})^2 \quad (4)$$

where the PRESS is the sum of squared deviation between the experimental activity and the predicted activity of each inhibitor in the test set; SD is the sum of the squared deviations between the inhibitory activities of the test set and the mean activities of the training set.

#### 2.6 Molecular docking

The crystal structure of VEGFR-2 in complex with a novel 4-amino-furo[2,3-*d*]pyrimidine (PDB ID: 1YWN) obtained from the RCSB Protein Data Bank (<https://doi.org/10.2210/pdb1YWN/pdb>) was employed for molecular docking with AutoDock software (version 4).<sup>52</sup> Prior to molecular docking, all the hetero-atoms including co-crystallized ligand as well as water molecules and ions were firstly removed from the protein. Polar





hydrogen atoms and Gasteiger charges were added to the receptor. In addition, the inhibitors employed in the present work were prepared by merging non-polar hydrogens, adding Gasteiger charges, defining the rotatable bonds, and assigning AutoDock type.<sup>53</sup>

The 3D grid box with  $60 \times 60 \times 60$  Å grid points with grid spacing of 0.375 Å was generated using the AutoGrid algorithm to evaluate inhibitor-receptor interactions energy.<sup>54</sup> Additionally, Lamarckian algorithm (LGA) approach was also used to search for the globally optimized conformations. Each molecular docking procedure was derived from 100 different runs that were set to terminate after a maximum of 250 000 energy evaluations. Finally, the binding modes with lower binding free energies were chosen for further analysis.

## 2.7. Molecular dynamics simulation

The docked complexes VEGFR-2/Cpd11 and VEGFR-2/Cpd15 were used as initial structures for the next MD simulations by using the AMBER v 20 (PMEMD) with CUDA acceleration by HPC cluster facility of Warshel Institute for Computational Biology.<sup>55</sup> The parameters for inhibitors Cpd11 and Cpd15 were generated by Antechamber of Ambertools 21.<sup>55,56</sup> The RESP charges<sup>57</sup> and the bond constraints were extracted from Gaussian 09 by calculating the DFT function of B3LYP/6-31G\*\*,<sup>58,59</sup> and Amber GAFF,<sup>60</sup> respectively. In addition, the AMBER 14SB force field<sup>61</sup> was used for VEGFR-2 receptor. The topology file of the inhibitor-receptor complexes was retrieved and the structures were simulated in a periodic solvent TIP3P box with thickness of 16 Å in layer for each system.<sup>62</sup> The total charge of the systems was then neutralized by adding counter ions ( $\text{Na}^+$  and  $\text{Cl}^-$ ). Finally, the system of VEGFR-2/Cpd11 and VEGFR-2/Cpd15 is  $82.09 \times 86.83 \times 86.55$  Å<sup>3</sup> (50 222 atoms) and  $82.09 \times 86.83 \times 86.54$  Å<sup>3</sup> (50 218 atoms), respectively.

The solvated systems were then minimized by 10 000 steps to remove bad contacts between water, ions and the complex systems. Then, the temperature of the systems were heated from 0 to 300 K with 50 ps in the NVT ensemble. Furthermore, the equilibrium simulations at 300 K and 1 atm were incorporated. A Langevin thermostat was used for temperature control and the SHAKE algorithm<sup>63</sup> was employed for hydrogen bonds, the detailed setting was as follows: 50 ps of density equilibration with weak restraints ( $5 \text{ kcal mol}^{-1}$ ) on the complex followed by 500 ps of constant pressure (1 atm) equilibration at 300 K at a time step of 2 fs, the non-bonded cut-off was set to 20.0 Å. Finally, a 500 ns production run with 3 replicates was conducted by setting different velocities of random seeds to the equilibrium structure. And the trajectories were sampled every 10 ps.

## 2.8 Binding free energy calculation

MM/PB & GBSA<sup>64,65</sup> with MMPBSA.py.MPI<sup>66</sup> provided in the Amber program was employed to calculate the Gibbs binding free energies ( $\Delta G_{\text{binding}}$ ) between receptors and inhibitors (Cpd11, Cpd15), which was computed according to the following equations:

$$\Delta G_{\text{binding}} = \Delta G_{\text{com}} - (\Delta G_{\text{rec}} + \Delta G_{\text{lig}}) \quad (5)$$

$$\Delta G_{\text{com/rec/lig}} = \Delta H - T\Delta S \quad (6)$$

$$\Delta H = \Delta E_{\text{gas}} + \Delta G_{\text{sol}} \quad (7)$$

$$\Delta E_{\text{gas}} = \Delta E_{\text{int}} + \Delta E_{\text{vdw}} + \Delta E_{\text{ele}} \quad (8)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{NP}} \quad (9)$$

$$\Delta G_{\text{NP}} = \gamma \text{SASA} + \beta \quad (10)$$

where  $\Delta G_{\text{com}}$ ,  $\Delta G_{\text{rec}}$  and  $\Delta G_{\text{lig}}$  is the free energy of the complex, the receptor and the ligand, respectively. Each of the term ( $\Delta G_{\text{com/rec/lig}}$ ) represents the discrepancy between enthalpy contribution ( $\Delta H$ ) and the conformational entropy ( $T\Delta S$ ).  $T$  and  $\Delta S$  is the temperature of the simulated environment and the entropy of the molecule, respectively.<sup>64</sup> The enthalpy ( $\Delta H$ ) includes the internal energy from gas phase ( $\Delta E_{\text{gas}}$ ) and the solvation free energy ( $\Delta G_{\text{sol}}$ ).  $\Delta E_{\text{gas}}$  is the interaction energy between the receptor and inhibitor in the gas phase, which consists of  $\Delta E_{\text{int}}$  (internal energy),  $\Delta E_{\text{vdw}}$  (van der Waals interactions), and  $\Delta E_{\text{ele}}$  (electrostatic energies). Since complex MD simulations are only performed here, the  $\Delta E_{\text{int}}$  to the binding free energy is zero.  $\Delta G_{\text{sol}}$  is the solvation energy is the sum of  $\Delta G_{\text{NP}}$  (non-polar energy) and  $\Delta G_{\text{PB/GB}}$  (electrostatic energy).  $\Delta G_{\text{PB}}$  is calculated from Poisson-Boltzmann function with the default cavity radii from AMBER pomtop files. While the  $\Delta G_{\text{GB}}$  is an alternative part for  $\Delta G_{\text{GB}}$  which uses Hawkins, Cramer, and Truhlar pairwise generalized Born model<sup>67,68</sup> with the parameters described by Tsui and Case.<sup>69</sup> The SASA (solvent accessible surface area) is calculated by the LCPO approach.<sup>70</sup> The  $\gamma$  ( $0.00542 \text{ kcal mol}^{-1} \times \text{\AA}^2$ ) and  $\beta$  ( $0.92 \text{ kcal mol}^{-1}$ ) are taken from a linear regression of the solvation free energy.<sup>71,72</sup>

## 3. Results and discussion

### 3.1. 3D-QSAR analysis

In the present work, template ligand-based alignment was carried out in both CoMFA and CoMSIA analyses. All combinations of different fields (Tables S1–S3†) were tried to produce the optimal model, which would be evaluated by some significant parameters, including the cross-validated correlation coefficient ( $R_{\text{cv}}^2$ ), the conventional correlation coefficient ( $R_{\text{ncv}}^2$ ), the standard error of estimation (SEE), standard error of prediction (SEP) and  $F$  values. Furthermore, the optimal CoMFA and CoMSIA models are summarized in Table 2.

**3.1.1. 3D-QSAR results for HepG2.** The statistical results (Table 2) obtained after PLS regression analysis for CoMFA and CoMSIA models (HepG2) were recorded to evaluate the reliability of the developed models.

In case of CoMFA<sub>HepG2</sub>, the cross-validated correlation coefficients  $R_{\text{cv}}^2$  and the regression coefficient  $R_{\text{ncv}}^2$  for the training set are 0.509 and 0.904 with optimal principal component  $N_{\text{c}}$  of 3, standard error of estimation SEE of 0.134, and  $F$  value of 44.077. Furthermore, the relative contributions of steric and electrostatic field of the CoMFA model are 72.9% and 27.1%, respectively, indicating that the steric field has higher impact on the inhibitory activity to HepG2 cell line than



Table 2 Summary of the results of CoMFA and CoMSIA analyses<sup>a</sup>

Parameters	HepG2		HCT-116		MCF-7	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
$R_{cv}^2$	0.509	0.711	0.574	0.531	0.568	0.669
$R_{ncv}^2$	0.904	0.865	0.746	0.984	0.755	0.981
SEE	0.134	0.153	0.171	0.047	0.157	0.048
$F$	44.077	48.191	49.903	218.135	52.355	179.713
$R_{pred}^2$	0.5128	0.6198	0.5597	0.5804	0.5057	0.6577
SEP	0.302	0.224	0.222	0.256	0.209	0.255
Nc	3	2	1	4	1	4
<b>Field contribution</b>						
S	0.729	0.262	0.623	—	0.623	—
E	0.271	—	0.377	0.375	0.377	0.372
H	—	—	—	—	—	—
D	—	0.738	—	0.625	—	0.628
A	—	—	—	—	—	—

<sup>a</sup>  $R_{cv}^2$  = cross-validated correlation coefficient using the leave-one-out methods;  $R_{ncv}^2$  = conventional 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.

electrostatic field. Overall, the parameters derived from the CoMFA model suggest the reliability of this model. Besides, the test set inhibitors were also employed to validate the CoMFA model, and the predicted correlation coefficient  $R_{pred}^2$  of 0.5128 suggests that the model is strong enough to fit new observations. The plot of experimental against calculated values of inhibitory activity by the CoMFA model is listed in Fig. 2A. As shown in this figure, most points are equally dispersed among the line  $y = x$ , further demonstrating the good quality of the CoMFA model.

The best CoMSIA<sub>HepG2</sub> model was obtained using a combination of steric and hydrogen bond donor fields. The CoMSIA model has a cross-validated correlation coefficient  $R_{cv}^2$  of 0.711 with 2 optimal components. PLS analysis with non-cross-validation method gives  $R_{ncv}^2$  of 0.865,  $F$  value of 48.191, and a SEE value of 0.153. The CoMSIA model also has a predicted correlation coefficient  $R_{pred}^2$  of 0.6198 which is higher than 0.5, indicating a good external predictive ability of this model. Additionally, the contributions of the steric and hydrogen bond donor fields are 26.2% and 73.8%, respectively, suggesting that the contribution of the hydrogen bond donor is predominant in the CoMSIA model. Fig. 2B shows predicted and experimental activities of training and test set compounds using the CoMSIA model.

**3.1.2. 3D-QSAR results for HCT-116.** In the CoMFA<sub>HCT-116</sub> model, the cross-validated correlation coefficients  $R_{cv}^2$  and optimal number of components Nc are 0.574 and 1, respectively, which indicate that the model has good internal predictive ability. The non-cross-validated correlation  $R_{ncv}^2$ , the standard error of estimation SEE, and  $F$  value of the model are 0.746, 0.171, and 49.903, respectively, indicating that the model has good fitting. The contribution rates of the steric and electrostatic fields of the model is 62.3% and 37.7%, respectively, which shows that the steric has more significant impact on the

inhibitory activity. The external validation results shows that the predicted correlation coefficient  $R_{pred}^2$  is 0.5597, and the plot of the predicted *versus* the experimental values is shown in Fig. 2C, which demonstrate that the model has good external predictive ability.

The best CoMSIA<sub>HCT-116</sub> model (combination of electrostatic and hydrogen bond donor fields) gives an  $R_{cv}^2$  of 0.531,  $R_{ncv}^2$  of 0.984, SEE of 0.047, Nc of 4,  $F$  value of 218.135, indicating that the model had good fitting ability and predictive ability. The proportions of electrostatic and hydrogen bond donor contributions account for 37.5% and 62.5%, respectively, indicating that the hydrogen bond donor distribution of the group has more effect on the activity of HCT-116. The activity of the test set compounds were also predicted to verify the accuracy of the model. Through PLS analysis, the external predicted correlation coefficient  $R_{pred}^2$  is 0.5804, which suggests the derived model possesses good external prediction ability of the 3D-QSAR model. Fig. 2D indicates a strong linear association between the predicted  $pIC_{50}$  values and those observed.

**3.1.3. 3D-QSAR results for MCF-7.** The results of CoMFA<sub>MCF-7</sub> studies are listed in Table 2. The optimum number of components Nc for CoMFA model is one, which were calculated by selecting the highest  $R_{cv}^2$  value. The generated CoMFA model illustrates an  $R_{cv}^2$  value of 0.568 with a Nc value of 1,  $R_{ncv}^2$  of 0.755, SEE of 0.157, and  $F$  value of 52.355. The contribution of the fields are 62.3% of the steric field and 37.7% of the electrostatic field. It is found that the steric field plays more important role in the CoMFA model. In addition, the CoMFA model has a statistically significant effect on the capability in predicting the activity. Fig. 2E indicates that the actual and predicted activities of the training and test set compounds possess strong linear correlations. The  $R_{pred}^2$  value of the CoMFA model is 0.5057, further indicating the external predictive capability.

The optimal CoMSIA<sub>MCF-7</sub> model was constructed by using different combinations of steric, electrostatic, hydrophobic, hydrogen bond donor, and acceptor fields. Finally, CoMSIA-ED model gives significant results with  $R_{cv}^2$  of 0.669, Nc of 4,  $R_{ncv}^2$  of 0.981, SEE of 0.048, and  $F$  of 179.713. The contributions of electrostatic and hydrogen bond donor fields are 37.2% and 62.8%, respectively, which suggest that the hydrogen bond donor descriptor has more impact on compounds' inhibitory activities than electrostatic field. The above statistical values suggest that the developed CoMSIA model is satisfactory. Additionally, to further validate the model's predictive ability, the activities of the test set compounds were predicted (shown in Table 2). The CoMSIA model exhibits excellent result in terms of predictive correlation coefficient  $R_{pred}^2$  of 0.6577. Fig. 2F depicts the relationship between the observed and predicted activities for the CoMSIA model, suggesting that the predicted values are in agreement with the experimental values in the allowable error range.

### 3.2. Contour maps

The results of 3D-QSAR models were graphically represented by contour maps using the field type 'StDev\*Coeff'. In order to



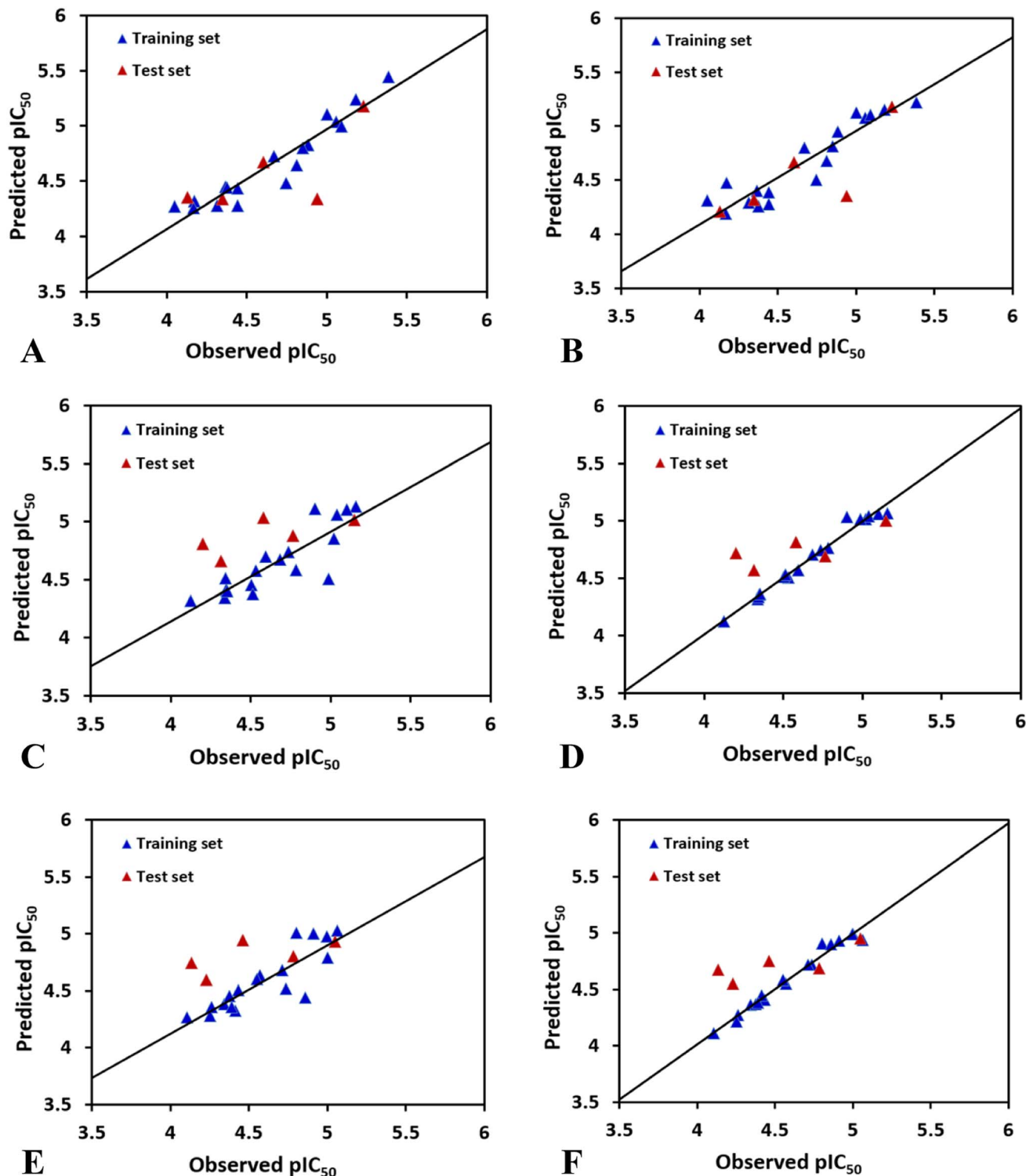


Fig. 2 Graphs of the predicted versus the experimental  $pIC_{50}$  values of the optimal models for (A) CoMFA model for HepG2. (B) CoMSIA model for HepG2. (C) CoMFA model for HCT-116. (D) CoMSIA model for HCT-116. (E) CoMFA model for MCF-7. (F) CoMSIA model for MCF-7.

facilitate the contour map analysis, the most potent **Cpd11** was chosen as the reference in the 3D coefficient contour maps. Furthermore, a default value of 80% contribution for favored regions was defined to visualize the contour maps, while the disfavored regions was 20%.

**3.2.1. Contour maps for HepG2.** Fig. 3 displays the steric and electrostatic contour maps of the CoMFA model. The steric field is presented in green and yellow contour maps, and the electrostatic field is show in blue and red contour maps.



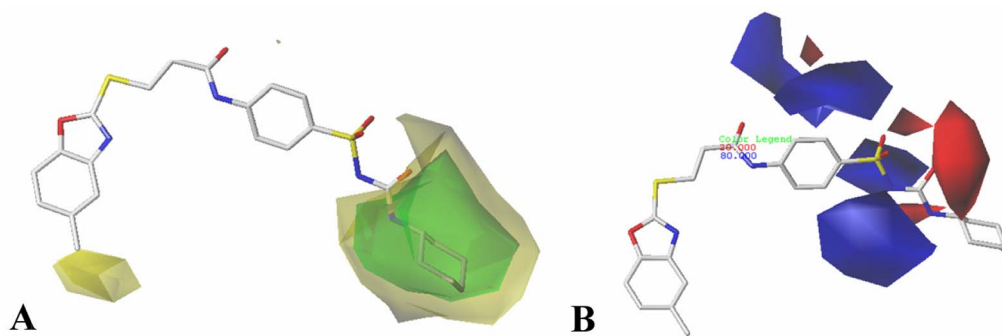


Fig. 3 CoMFA StDev\*Coeff contour plots for HepG2 compounds in combination of **Cpd11**. (A) The steric contour map, where the green and yellow contours represent 80% and 20% level contributions, respectively. (B) The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively.

For the CoMFA steric field (Fig. 3A), a green contour map is observed around  $R_1$  substituent, suggesting that bulky substitution in this area is favorable for the inhibitory activity. This explains the lower activity of **Cpd06** ( $pIC_{50} = 4.7442$ ) with  $-H$  group when compared to **Cpd11** ( $pIC_{50} = 5.384$ ) with  $-NH-C_6H_{11}$  group. In addition, there is also a larger yellow contour map is observed around the above green map, indicating that the size of the substituent at this place needs to be carefully selected. A yellow contour map covering the  $R$  substituent indicates that a small group in this region generally gets better inhibitory activity. This is in good agreement with the experimental data, for example, **Cpd14** ( $R = H$ ) > **Cpd15** ( $R = CH_3$ ).

Fig. 3B shows the CoMFA electrostatic field contour map. There is a red contour map located in the  $R_1$  position, indicating that compounds with electronegative groups are beneficial for inhibitory activity, as observed from **Cpd02** ( c1cc[nH]c1 ), **Cpd03** ( c1ccc[nH]1 ) > **Cpd01** ( $H$ ). Several blue contour maps are located

around  $-SO_2NH-$  group, suggesting that the introduction of electropositive moieties into this position would be of benefit to activity. For the most active **Cpd11**, the substituents here can be modified to improve its activity.

The contour maps of CoMSIA steric field are shown in Fig. 4A, which are found to be quite similar to the contour maps derived from the CoMFA model, thus are not discussed here.

In hydrogen bond donor contour map of CoMSIA (Fig. 4B), only purple contours and one cyan contour are discovered, illustrating that the hydrogen bond acceptor property is crucial for the inhibitory activity. In addition, the purple contour maps are mainly distributed at the nitrogen atom of the common skeleton and Region A, indicating that the atoms at these positions would act as hydrogen bond acceptors to interact favorably with the receptor. A cyan contour map is involved in the  $-NH$  group of Region A, indicating that a hydrogen bond interaction with the receptor (acting as a hydrogen bond donor) favors the activity, as is the case with **Cpd16** and **Cpd22**. The activity of **Cpd16** ( $pIC_{50} = 4.4444$ ) is higher than that of **Cpd22** ( $pIC_{50} = 4.3143$ ).

**3.2.2. Contour maps for HCT-116.** As shown in Fig. 5A, sterically beneficial and detrimental interactions are displayed as green and yellow contours, respectively. A small green polyhedron on  $R$  substitution indicates that bulky groups in this position could improve the activity. Different activities of **Cpd01** and **Cpd04** (**Cpd01**  $pIC_{50} = 4.3152$ , **Cpd04**  $pIC_{50} = 4.5955$ ) are probably caused by the presence of larger group ( $CH_3$ ) in **Cpd04** than **Cpd01** ( $H$ ). Similarly, the activity of **Cpd06** ( $pIC_{50} = 4.5782$ ) is higher than that of **Cpd05** ( $pIC_{50} = 4.2019$ ) because the  $-CH_3$  at this position in **Cpd06** is larger than that of the  $-H$  in **Cpd05**. A large green contour appeared near  $R_1$ , which indicates that adding substituents at this area might increase the activity, for example, **Cpd03** ( c1ccc[nH]1 ) > **Cpd02** ( c1cc[nH]c1 ), **Cpd19** ( c1ccc(Cl)cc1 ) > **Cpd18** ( c1ccccc1 ).

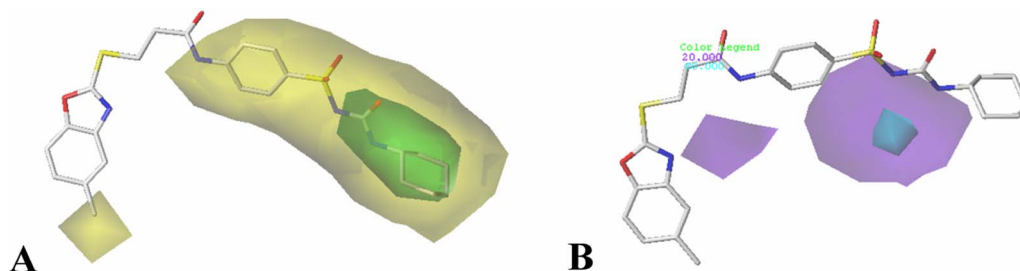


Fig. 4 CoMSIA StDev\*Coeff contour plots for HepG2 compounds in combination of **Cpd11**. (A) The steric contour map, where the green and yellow contours represent 80% and 20% level contributions, respectively. (B) The hydrogen bond donor contour map, where the cyan and purple contours represent 80% and 20% level contributions, respectively.



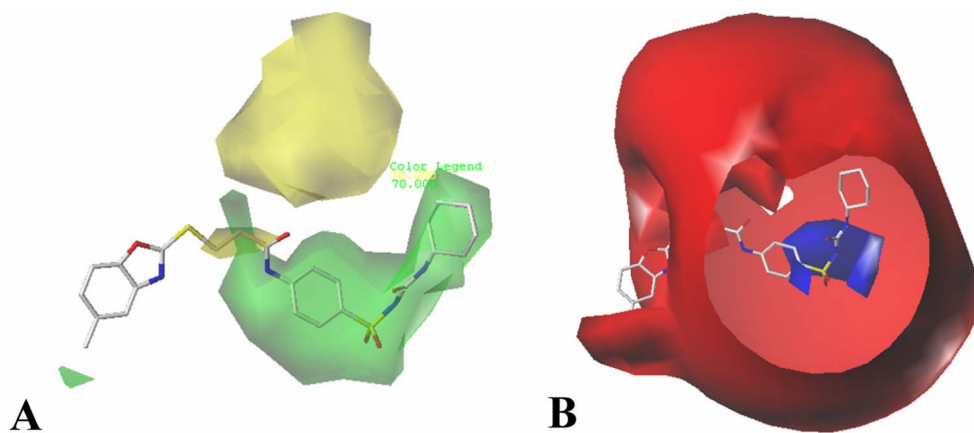


Fig. 5 CoMFA StDev\*Coeff contour plots for HCT-116 compounds in combination of **Cpd11**. (A) The steric contour map, where the green and yellow contours represent 70% and 30% level contributions, respectively. (B) The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively.

For CoMFA-electrostatic map in Fig. 5B, blue and red contours represent regions where electron positive and electron negative groups would increase the activity, respectively. There is red contour near  $R_1$  substituent, where electropositive occupation will enhance the activity. This can explain well why **Cpd18** ( $pIC_{50} = 4.5018$ ) has lower activity than **Cpd19** ( $pIC_{50} = 4.5158$ ). The red contour map near R position indicates that groups with negative charges may increase the activity. After analysis, it is found that all the compounds used in the present work are positively charged groups. Therefore, it can be modified to improve the activity. Furthermore, some blue contour maps are situated around Region A, thereby indicate the existence of electropositive groups in this region would improve the activity. Therefore, modifications can be made for **Cpd11** at this position to improve the activity.

The electrostatic field of CoMSIA model (Fig. 6A) is similar to that of CoMFA model, for example: a red contour map at  $R_1$  substituent, a blue contour at Region A. However, the difference is that the red contour map at R substituent disappears.

The contour maps of the CoMSIA hydrogen bond donor field are shown in Fig. 6B. There is a purple contour at the sulfur

atom of the common skeleton, which indicate hydrogen bond acceptor substituents here would increase the activity. It is also noted a cyan contour map appeared near the -NH of Region A, suggesting hydrogen bond donor in this position is important to the activity. This result can be explained by the **Cpd11** and **Cpd06** ( $pIC_{50} = 5.1593$  for **Cpd11**,  $pIC_{50} = 4.5782$  for **Cpd06**).

**3.2.3. Contour maps for MCF-7.** The steric contour map of CoMFA is displayed in Fig. 7A. A small green contour is near R position of the template molecule, indicating that introducing bulky atoms on this position can improve the activity. This can be illustrated by the fact that **Cpd04** with  $-CH_3$  at this position shows higher activity than the corresponding **Cpd01** with  $-H$  at this site. This phenomenon is also consistent with **Cpd06** ( $R = CH_3$ ,  $pIC_{50} = 4.4594$ ) and **Cpd05** ( $R = H$ ,  $pIC_{50} = 4.1352$ ). Another big green contour map near  $R_1$  substituent suggests that bulky groups would enhance the inhibitory potency. This finding is consistent with the fact that **Cpd19** bearing larger group (c1ccc(Cl)cc1) exhibits higher value than **Cpd18** (c1ccccc1).

The contour map of the electrostatic field in Fig. 7B shows a red contour map around R substituent, which indicate that

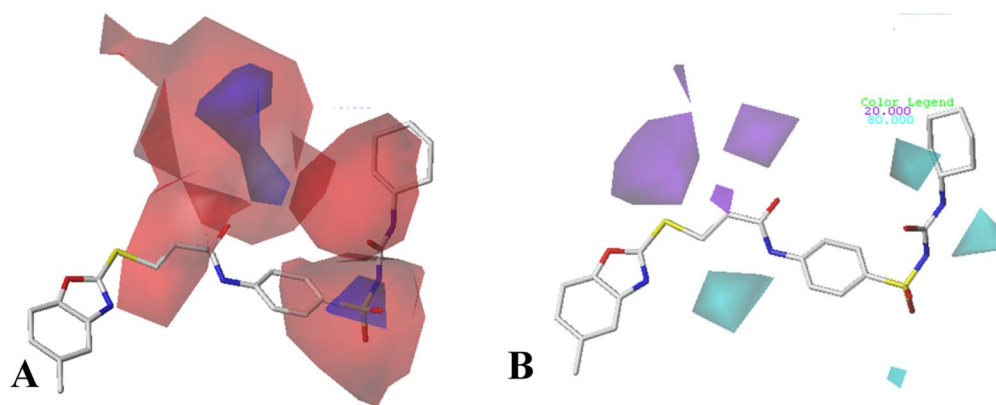


Fig. 6 CoMSIA StDev\*Coeff contour plots for HCT-116 compounds in combination of **Cpd11**. (A) The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively. (B) The hydrogen bond donor contour map, where the cyan and purple contours represent 80% and 20% level contributions, respectively.

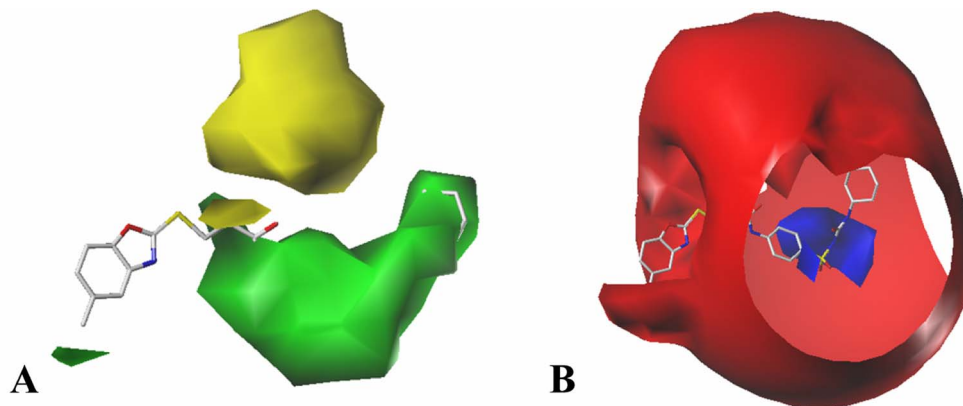


Fig. 7 CoMFA StDev\*Coeff contour plots for MCF-7 compounds in combination of **Cpd11**. (A) The steric contour map, where the green and yellow contours represent 70% and 30% level contributions, respectively. (B) The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively.

electronegative groups at this location are favorable for the activity. However, all compounds possess positively charged groups here, therefore, it can be electrically modified to improve the activity. On the other, another red contour map is observed near  $R_1$  group, it results that the introduction of electronegative group in this region favors the inhibitory activity. Therefore, the higher activity observed in **Cpd19** than **Cpd18** may be explained by the presence of electronegative group  $-Cl$  at this position. In addition, around Region A, several blue contour maps are observed, indicating that electropositive substituents would enhance the inhibitory activity, therefore, modifications can be made to this substitution.

The electrostatic CoMSIA contour (Fig. 8A) is somewhat similar as the CoMFA contour. Therefore, it is no reason to describe it here as it is already explained in detail during CoMFA contour analysis. Here, we mainly focus on the hydrogen bond donor field.

The CoMSIA hydrogen bond donor contour maps are shown in Fig. 8B. A purple contour is appeared at the sulfur atom of the common skeleton, signifying that the hydrogen bond acceptor substituent would increase the activity. Moreover, two cyan

contour maps near the  $-NH$  group of Region A indicate that hydrogen bond donor groups are preferred at this position. Thus, **Cpd09**, **Cpd10**, **Cpd11** with hydrogen bond donor groups are more active than **Cpd06**.

### 3.3. Molecular docking studies

To validate the accuracy of molecular docking, the co-crystallized ligand of the protein was extracted and re-docked into the same binding site of the receptor. The conformation with lowest energy and the co-crystallized ligand were superimposed, as shown in Fig. 9, and the RMSD value is 0.863 Å, which is within the reliable range of 2 Å,<sup>73</sup> illustrating that the docking procedure is rational. Although small rotation angle is presented, the binding site is same for the re-docked compound and the original structure. In addition, cross-docking with other VEGFR-2 structures 3VO3 was also done on **Cpd11**, the docked structures of 1YWN-Cpd11 and 3VO3-Cpd11 are aligned together to compare (Fig. S1†), the RMSD value is 0.324 Å and we find that the two docking structures are located at the same binding site, and the conformations are slightly different, further validating the reliability of molecular docking.

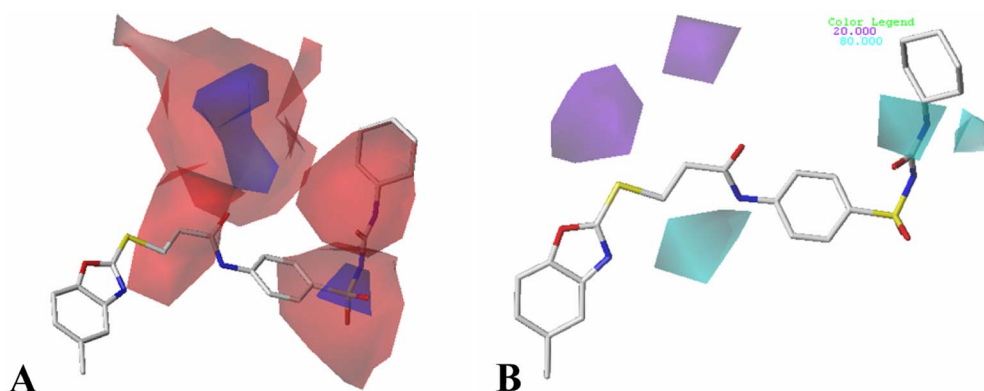


Fig. 8 CoMSIA StDev\*Coeff contour plots for MCF-7 compounds in combination of **Cpd11**. (A) The electrostatic contour map, where the blue and red contours represent 80% and 20% level contributions, respectively. (B) The hydrogen bond donor contour map, where the cyan and purple contours represent 80% and 20% level contributions, respectively.





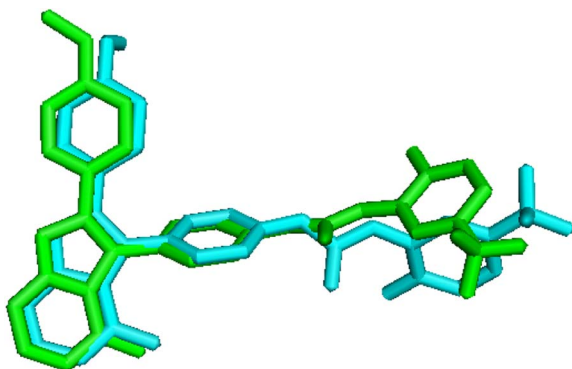


Fig. 9 Re-docking pose and the crystal ligand (cyan = docked, green = original).

Therefore, the relevant settings were employed for docking the most active **Cpd11** and the lowest active **Cpd15** into the binding pocket of VEGFR-2 to evaluate the detailed binding mode of this series of benzoxazole derivatives into the active site of VEGFR-2.

Fig. 10A shows the binding environment of **Cpd11** in VEGFR-2, which is surrounded by Leu35, Gly36, Val43, Ala61, Val62, Lys63, Glu80, Ile83, Leu84, Ile87, Val93, Val94, Phe113, Cys114, Lys115, Phe116, Gly117, Asn118, Leu164, His171, Ile189, Cys190, Asp191, and Phe192. As seen in Fig. 10B, the –NH group in Region A that acts as a hydrogen bond donor, forms a hydrogen bond with the backbone of Asp191 ( $\text{O}\cdots\text{HN}$ , 2.19 Å, 143.1°) (H-1). Furthermore, the backbones of Lys63, which act as hydrogen bond donors and form hydrogen bonds with the –SO group at Region A ( $\text{O}\cdots\text{HN}$ , 2.89 Å, 70.8°) (H-2), ( $\text{O}\cdots\text{HN}$ , 2.29 Å, 105.3°) (H-3). In addition, the cyan and purple contour maps from the above CoMSIA models (Fig. 4B, 6B and 8B) are fallen in this region, thus, the obtained results from molecular docking and QSAR models are harmonious. It also reveals that the groups at Region A are significant for the activity of VEGFR-2 inhibitors, and agree with the higher inhibitory activity of some inhibitors.

It can be seen from Fig. 10A that the  $\text{R}_1$  substituent is anchored in a bulky and electropositive pocket made by Ile83,

Leu84, Ile87, Leu164 and His171, suggesting that large and electronegative groups are favorable in the site. This is in consistent with the positioned green and red contour maps (Fig. 3, 5 and 7). Additionally, the substituent at Region A is adjacent to amino acid residues Glu80, Val94, and Asp191, which indicates that electropositive groups are favored in this direction. This result is in concordance with the CoMFA results for electrostatic interactions shown in Fig. 3B, 5B and 7B.

The interactions between **Cpd15** and the active site of VEGFR-2 are depicted in Fig. 11. The **Cpd15** is surrounded by amino acid residues Met1, Cys12, Val43, Ala61, Val62, Lys63, Ser79, Glu80, Ile83, Leu84, Ile87, Val93, Val94, Val109, Ile110, Val111, Leu164, Cys169, Ile170, His171, Arg172, Ile189, Cys190, Asp191 and Phe192. Furthermore, the ligand is docked into the binding pocket *via* two hydrogen bonds and one arene–cation interaction. The hydrogen bond distances are observed to be 2.13 Å ( $\text{Met1-NH}\cdots\text{O}=\text{O}$ ) (H-1), and 2.64 Å ( $\text{Met1-NH}\cdots\text{O}=\text{O}$ ) (H-2). The arene–cation interaction is observed between Lys63 and the benzene ring and five-membered heterocyclic ring at the common skeleton. Analyses show that **Cpd15** possess several interactional amino acids as **Cpd11**, but different residues are also exist, especially residues Lys63, Glu80, and Asp191, which can interact with **Cpd11** *via* hydrogen bond interactions, affecting the inhibitory activity of this series of inhibitors from the field distributions in the CoMSIA models. This is the reason why **Cpd11** has higher activity than **Cpd15**.

On the other hand, to explore the detailed binding modes between the receptor VEGFR-2 and the inhibitors in the aqueous solution, MD simulations were carried out to describe the most active and least active inhibitors during the dynamic binding process.

### 3.4. MD simulations

Two docked complexes (VEGFR-2/**Cpd11** and VEGFR-2/**Cpd15**) were performed with MD simulations to examine the reliability of the molecular docking process, and in turn to gain insights into the stability and dynamics of the systems.

**3.4.1 Plasticity of the MD systems.** The resilience of the receptor and the inhibitors were investigated by calculating the

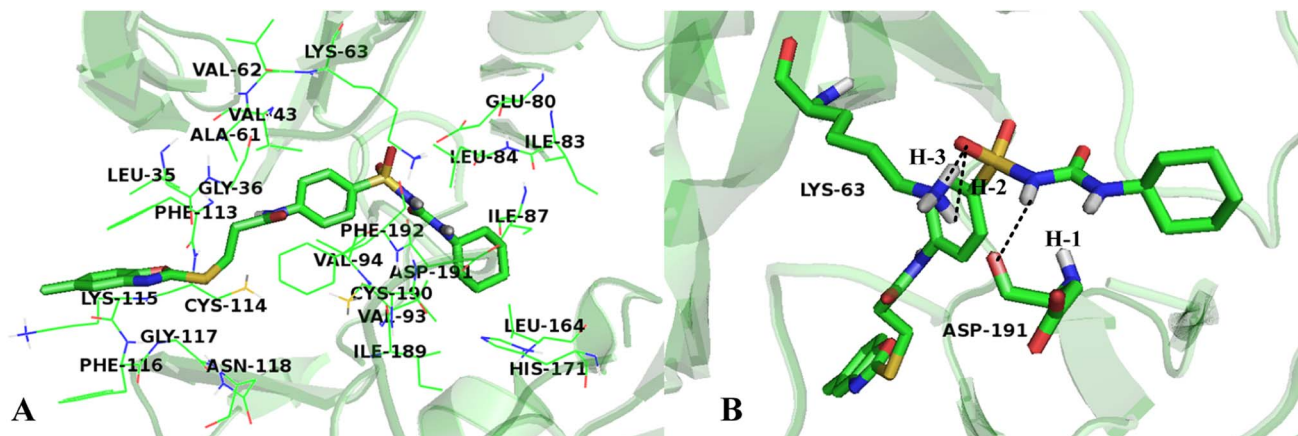


Fig. 10 (A) The residues in VEGFR-2 active site around **Cpd11**. (B) The enlargement for **Cpd11** in the binding site, which is displayed in stick, hydrogen bonds are shown as dotted black lines, and the nonpolar hydrogens were removed for clarity.

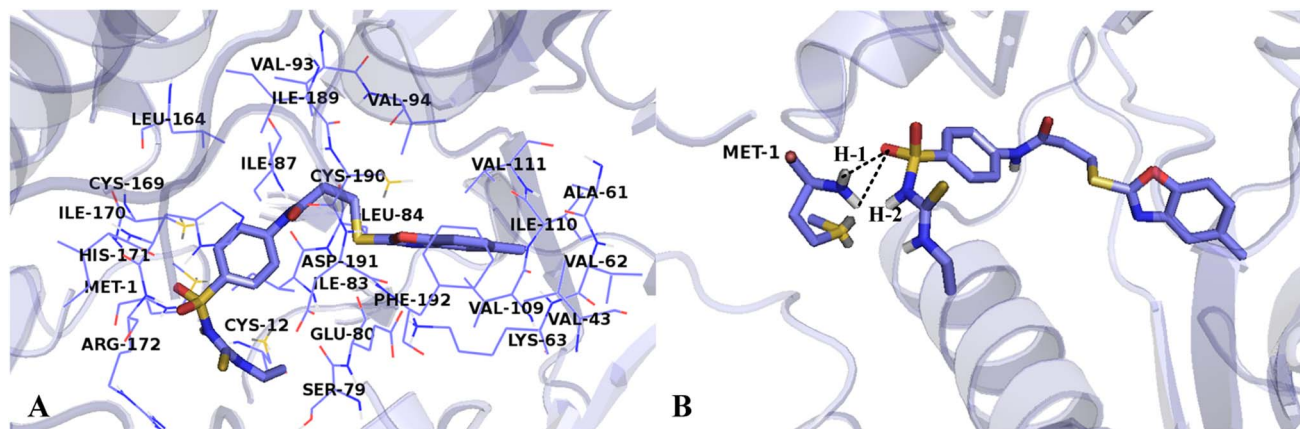


Fig. 11 (A) The residues in VEGFR-2 active site around Cpd15. (B) The enlargement for Cpd15 in the binding site, which is displayed in stick, hydrogen bonds are shown as dotted black lines, and the nonpolar hydrogens were removed for clarity.

Root Mean Square Deviation (RMSD) of the backbone atoms ( $C\alpha$ , N, O, C) and the heavy atoms of the docked ligands from the production phase trajectories by using Gromacs v 5.11.<sup>74</sup> The RMSD of each inhibitor and the VEGFR-2 receptor is depicted in Fig. 12. The curve shows that the VEGFR-2 protein in the Cpd15 is more volatile throughout the MD simulation than that in the Cpd11. Additionally, the RMSD of Cpd11 is slightly lower than Cpd15. In run2 of the VEGFR-2/Cpd15 system, the RMSD value drops sharply, and then go up sharply in run3, indicating that the Cpd11 binds more stable to VEGFR-2 than Cpd15. In addition, the change of the secondary structure of VEGFR-2/Cpd11 and VEGFR-2/Cpd15 with the simulation time is shown in Fig. 13, different colors are used to represent the secondary structure.<sup>75</sup> From this figure, we can see that the secondary structures has been stable for the two systems.

Additionally, the Root Mean Square Fluctuation (RMSF) of the backbone atoms that were summarized to exhibit the contribution on the residue-level. The trajectories were aligned by the  $C\alpha$  atoms of the starting structures of MD simulations beforehand to eliminate the rotation and transitions of the whole systems. The detailed analysis of RMSF of VEGFR-2 residues for the two MD systems are shown in Fig. 11. Firstly,

the N-terminal of the protein is also seen large fluctuation for both systems at residues ( $7.28 \pm 3.22$  Å), this might because the truncation effect of recombinant protein itself. Secondly, the second largest fluctuation area locates in the flexible loop, Thr154 to Ser166 of domain II ( $6.87 \pm 2.18$  Å for VEGFR-2/Cpd11,  $5.24 \pm 1.25$  Å for VEGFR-2/Cpd15). It is because of the missing 12 residues in the original crystal structure. Thirdly, the ligand binding domain (LBD) of Cpd11 containing system fluctuated lower ( $0.89 \pm 0.14$  Å) than that in Cpd15 system ( $1.23 \pm 0.34$  Å), which also suggesting the binding of Cpd15 is not as stable as that of Cpd11 (Fig. 14).

**3.4.2 MM/PB & GBSA analysis.** Moreover, to gain insights into the interactions between inhibitors and receptor VEGFR-2, the binding free energy of VEGFR-2-Cpd11/15 was elaborately calculated by MM/PBSA & GBSA. The snapshots extracted from the most stable trajectories were used in these calculations.

The results of binding free energy using MM-GBSA and MM-PBSA method are shown in Table S4.<sup>†</sup> The calculated binding free energy for Cpd11 and Cpd15 is  $-27.16$  kcal mol<sup>-1</sup> and  $-1.99$  kcal mol<sup>-1</sup>, respectively for GBSA, and  $-7.87$  kcal mol<sup>-1</sup>,  $14.08$  kcal mol<sup>-1</sup> for PBSA (Table S1<sup>†</sup>). These total free energies indicate that Cpd11 is quite a better ligand for VEGFR-2

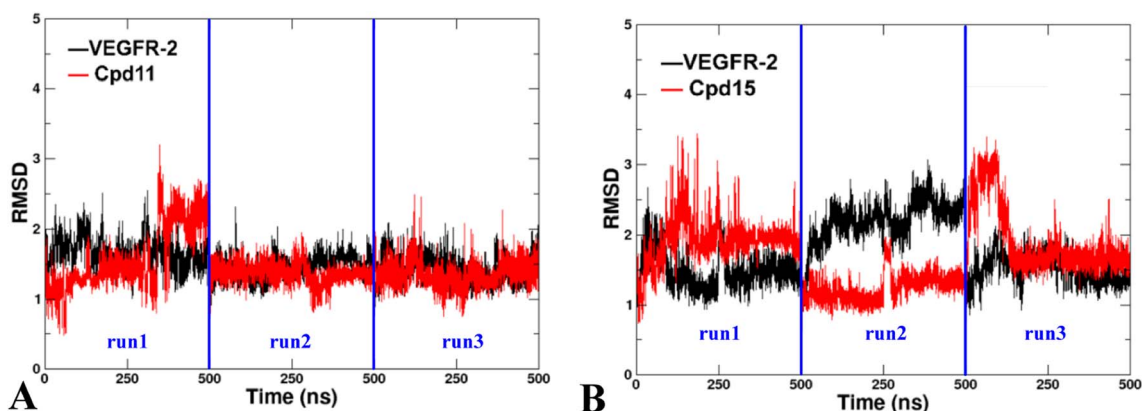


Fig. 12 The RMSD of the backbone atoms relative to the docking structures as function of time. (A) The three replicates of VEGFR-2/Cpd11. (B) The three replicates of VEGFR-2/Cpd15.





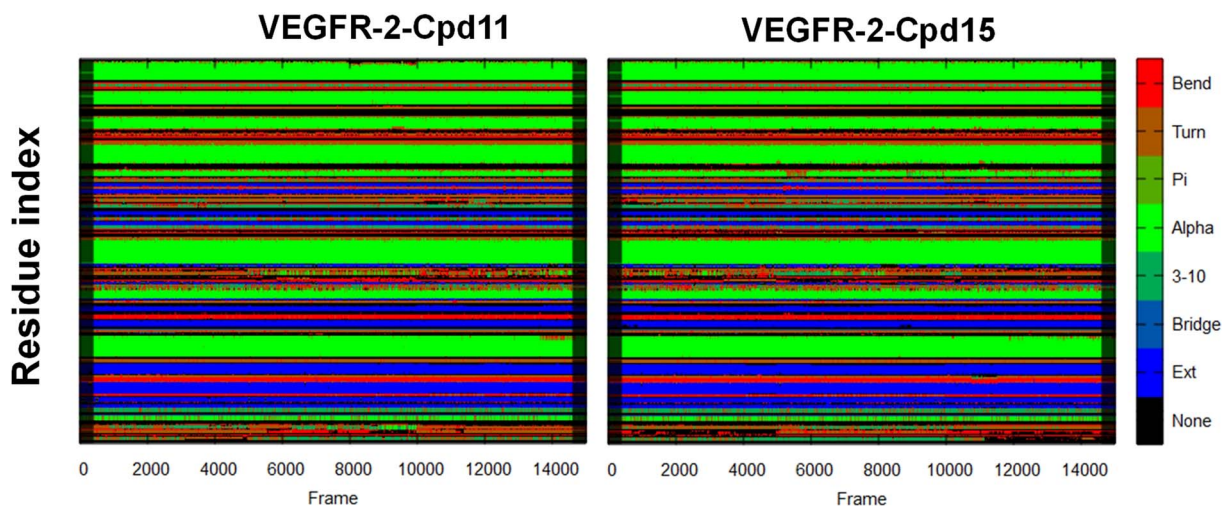


Fig. 13 Secondary structures as a function of time for VEGFR-2-Cpd11 and VEGFR-2-Cpd15.

compared to **Cpd15**. Indeed, the ranking of the binding free energies of the two compounds are consistent with the predicted ranking that was calculated from the developed 3D-QSAR models, which further validated the reliability of the constructed models. It can be seen from the binding free energy components that the van der Waals and electrostatic interactions in the gas phase are the pivotal contributions for the inhibitor binding. On the other hand, the solvation energy ( $\Delta G_{\text{solv}}$ ) calculated from both GB and PB contributes negatively to the total binding affinity (131.84 kcal mol<sup>-1</sup> and 54.82 kcal mol<sup>-1</sup> for **Cpd11**, and 151.13 kcal mol<sup>-1</sup> and 70.89 kcal mol<sup>-1</sup> for **Cpd15** respectively), illustrating that the solvation energies are unfavorable for stabilizing the binding of the two compounds. It might be because of the large volume of the binding pocket that exposes to the massive solvent. Interestingly, the entropy contribution of the two systems are similar (29.23 kcal mol<sup>-1</sup> for **Cpd11**, and 29.16 kcal mol<sup>-1</sup> for **Cpd15**), owing to the similar size of the two compounds.

**3.4.3 Binding free energy decomposition.** To get the detailed effects of individual residues on binding affinity of the simulated systems, the binding free energy (GBSA) was decomposed to the contributions of residues (Fig. 12). For

**Cpd11**, major favorable energy contributions are originated predominantly locating at Leu35, Val43, Lys63, Leu84, Gly117, Leu180 and Asp191, while Met1, Leu84 and Arg172 as favorable contributions can be detected in the binding of **Cpd15**. Notably, the residue Glu80 has negative effect on the binding in **Cpd15**, suggesting that the structure of **Cpd15** can still be altered at the R<sub>1</sub> part to improve the activity (Fig. 15).

**3.4.4 Comparison with previous work.** For this series of VEGFR-2 inhibitors, to our best knowledge, no corresponding 3D-QSAR models have been conducted on them. Only El-Adl *et al.* employed molecular docking for this series of compounds to assess the binding pattern and affinity toward the VEGFR-2 active site, but the analysis is simple. Furthermore, no MD simulations have been used to study this series of VEGFR-2 inhibitors, which would provide the inhibitor-receptor interactions more precisely. Overall, the flow of computational calculation employed in the manuscript are carried out for the first time.

**3.4.5 Design of new compounds.** Based on the above 3D-QSAR analyses, the binding conformations of molecular docking and the binding free energies, new VEGFR-2 inhibitors were designed, shown in Table 3. And the inhibitory activities were predicted by the developed CoMSIA (HepG2), CoMFA (HCT-116), and CoMSIA (MCF-7) models. The structure-activity relationship indicates that Region A of **Cpd11** is mainly responsible for the hydrogen bond interactions with the receptor, so this group is kept. Therefore, the alteration is mainly focused on the R and R<sub>1</sub> substituents. Based on the contour maps, small substituent for HepG2, bulky and electronegative groups for HCT-116 and MCF-7 at R substituent are favorable for the activity, thus -Cl, -OCH<sub>3</sub> and -NH<sub>2</sub> groups are introduced to this region. In addition, large and electronegative groups at R<sub>1</sub> are favored for enhancing the activity, therefore, groups -C<sub>6</sub>H<sub>4</sub>-Cl and -C<sub>6</sub>H<sub>4</sub>-NH<sub>2</sub> are introduced. Additionally, we

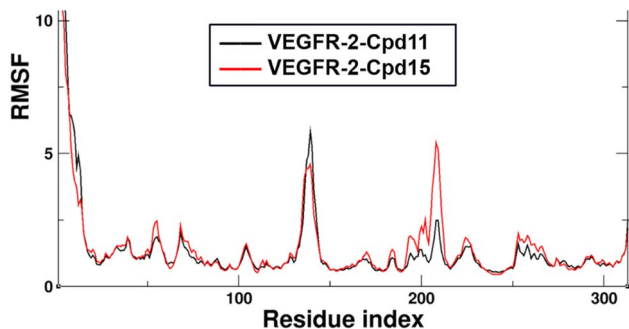


Fig. 14 The RMSF plot of the protein two MD systems (residue number are ranging from 1 to 310). Replicate trajectories of the same system are combined.

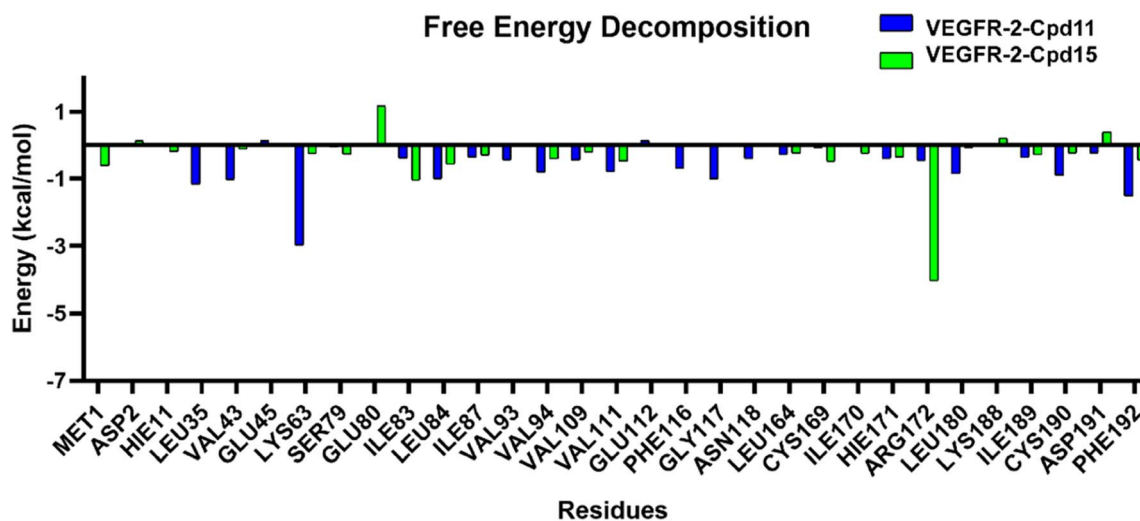


Fig. 15 Per residue average energy contributions to binding free energy.

Table 3 The structures and predicted activities of novel designed compounds

No	R	R <sub>1</sub>	Predicted pIC <sub>50</sub> (HepG2)	Predicted pIC <sub>50</sub> (HCT-116)	Predicted pIC <sub>50</sub> (MCF-7)	Docking scores
D1	-Cl		5.4010	5.1634	5.0912	-10.5 kcal mol <sup>-1</sup>
D2	-Cl		5.3980	5.2023	5.1011	-9.8 kcal mol <sup>-1</sup>
D3	-OCH <sub>3</sub>		5.3975	5.1874	5.0897	-9.3 kcal mol <sup>-1</sup>
D4	-NH <sub>2</sub>		5.4062	5.2053	5.0989	-11.7 kcal mol <sup>-1</sup>

model, thus these 3D-QSAR models were chosen to predict the inhibitory activities. Then, the designed compounds are docked into the active site to the receptor. We find that D4 exhibits higher predicted inhibitory activities and docking scores. D4 is introduced one bulk and electronegative to R<sub>1</sub> substituent, one moderate electronegative -NH<sub>2</sub> to R substituent, which improve the binding activity, especially the R substituent form electrostatic interactions with amino acid residues Lys33 and Lys115. In conclusion, based on the analysis discussed above, some novel VEGFR-2 inhibitors, especially D4 shows higher predicted binding activity to the receptor VEGFR-2, further suggesting that the above findings would provide useful guidance to rational design novel and potent VEGFR-2 inhibitors.

## 4. Conclusion

In the present work, the 3D-QSAR study combined with molecular docking and MD simulations were conducted on a series of VEGFR-2 inhibitors against HepG2, HCT-116, and

MCF-7 cell lines. The CoMFA and CoMSIA models possess excellent internal and external validation. The steric, electrostatic and hydrogen bond donor fields of these models for different cell lines have been verified that alterations for these characteristics can affect the inhibitory activity. The developed models were analyzed to show the effects associated with position of each group in the ligand, resulting in better insights for designing novel VEGFR-2 inhibitors. Molecular docking was used to provide initial conformation for MD simulation, which help in the investigation on the behavior of the inhibitors and point possible effects that are responsible for the inhibitory activity. Based on the derived information, a bulky group with higher electron-withdrawing ability at the R<sub>1</sub> position, minor group at R position, and electron-donating ability and hydrogen bond donor group at Region A would be favorable for anticancer activity against HepG-2. In addition, using a bulky group and electronegative group in R<sub>1</sub>, bulky and electronegative group in R, and electropositive and hydrogen bond donor group at Region A would lead to an increase in inhibition activity for HCT-116. Furthermore, bulky and electronegative groups at R<sub>1</sub> and R, electropositive and hydrogen bond donor groups at



Region A would be favorable for the activity against MCF-7 cell line. The powerful approaches, including 3D-QSAR studies, molecular docking and MD simulations led to the design of inhibitors with enhanced VEGFR-2 inhibitory activities. The newly designed inhibitors have improved VEGFR-2 inhibitory activities. Overall, the findings could represent good drug candidates for the treatment of cancer.

## Conflicts of interest

The authors declare that there are no conflicts of interest.

## Acknowledgements

The study was supported by the National Natural Science Foundation of China (No. 32001699).

## References

- 1 W. Risau, *Nature*, 1997, **386**, 671–674.
- 2 J. Folkman, *Biology of endothelial cells*, 1984, pp. 412–428.
- 3 W. Risau, *FASEB J.*, 1995, **9**, 926–933.
- 4 R. Kuiper, J. Schellens, G. Blijham, J. Beijnen and E. Voest, *Pharmacol. Res.*, 1998, **37**, 1–16.
- 5 R. Kumar and I. J. Fidler, *In Vivo*, 1998, **12**, 27–34.
- 6 Z. Szekanecz, G. Szegedi and A. E. Koch, *J. Invest. Med.*, 1998, **46**, 27–41.
- 7 M. J. Tolentino and A. P. Adamis, *Int. Ophthalmol.*, 1998, **38**, 77–94.
- 8 G. Bergers and L. E. Benjamin, *Nat. Rev. Cancer*, 2003, **3**, 401–410.
- 9 R. K. Jain, *Nat. Med.*, 2003, **9**, 685–693.
- 10 G. Neufeld, T. Cohen, S. Gengrinovitch and Z. Poltorak, *FASEB J.*, 1999, **13**, 9–22.
- 11 P. Borgström, K. J. Hillan, P. Sriramaraio and N. Ferrara, *Cancer Res.*, 1996, **56**, 4032–4039.
- 12 N. Ferrara and T. Davis-Smyth, *Endocr. Rev.*, 1997, **18**, 4–26.
- 13 M. Shibuya and L. Claesson-Welsh, *Exp. Cell Res.*, 2006, **312**, 549–560.
- 14 M. K. Sobhy, S. Mowafy, D. S. Lasheen, N. A. Farag and K. A. Abouzid, *Bioorg. Chem.*, 2019, **89**, 102988.
- 15 M. Shibuya, *Cancer Sci.*, 2003, **94**, 751–756.
- 16 M. A. McTigue, J. A. Wickersham, C. Pinko, R. E. Showalter, C. V. Parast, A. Tempczyk-Russell, M. R. Gehring, B. Mroczkowski, C.-C. Kan and J. E. Villafranca, *Structure*, 1999, **7**, 319–330.
- 17 N. Ferrara, K. J. Hillan, H.-P. Gerber and W. Novotny, *Nat. Rev. Drug Discovery*, 2004, **3**, 391–400.
- 18 P. Hamerlik, J. D. Lathia, R. Rasmussen, Q. Wu, J. Bartkova, M. Lee, P. Moudry, J. Bartek Jr, W. Fischer and J. Lukas, *J. Exp. Med.*, 2012, **209**, 507–520.
- 19 L. Witte, D. J. Hicklin, Z. Zhu, B. Pytowski, H. Kotanides, P. Rockwell and P. Böhlen, *Cancer Metastasis Rev.*, 1998, **17**, 155–161.
- 20 J. Huang, Y. Tan, Q. Tang, X. Liu, X. Guan, Z. Feng and J. Zhu, *Cytotechnology*, 2010, **62**, 61–71.
- 21 J. L. Spratlin, K. E. Mulder and J. R. Mackey, *Future Oncol.*, 2010, **6**, 1085–1094.
- 22 D. Lu, X. Jimenez, H. Zhang, P. Bohlen, L. Witte and Z. Zhu, *Int. J. Cancer*, 2002, **97**, 393–399.
- 23 J. M. Wood, G. Bold, E. Buchdunger, R. Cozens, S. Ferrari, J. Frei, F. Hofmann, J. Mestan, H. Mett and T. O'Reilly, *Cancer Res.*, 2000, **60**, 2178–2189.
- 24 L. Sun, C. Liang, S. Shirazian, Y. Zhou, T. Miller, J. Cui, J. Y. Fukuda, J.-Y. Chu, A. Nematalla and X. Wang, *J. Med. Chem.*, 2003, **46**, 1116–1119.
- 25 M. F. McCarty, J. Wey, O. Stoeltzing, W. Liu, F. Fan, C. Bucana, P. F. Mansfield, A. J. Ryan and L. M. Ellis, *Mol. Cancer Ther.*, 2004, **3**, 1041–1048.
- 26 M. J. Munchhof, J. S. Beebe, J. M. Casavant, B. A. Cooper, J. L. Doty, R. C. Higdon, S. M. Hillerman, C. I. Soderstrom, E. A. Knauth and M. A. Marx, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 21–24.
- 27 M. T. Bilodeau, A. E. Balitza, T. J. Koester, P. J. Manley, L. D. Rodman, C. Buser-Doepner, K. E. Coll, C. Fernandes, J. B. Gibbs and D. C. Heimbrook, *J. Med. Chem.*, 2004, **47**, 6363–6372.
- 28 M. T. Bilodeau, L. D. Rodman, G. B. McGaughey, K. E. Coll, T. J. Koester, W. F. Hoffman, R. W. Hungate, R. L. Kendall, R. C. McFall and K. W. Rickert, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 2941–2945.
- 29 M. E. Fraley, K. L. Arrington, C. A. Buser, P. A. Ciecko, K. E. Coll, C. Fernandes, G. D. Hartman, W. F. Hoffman, J. J. Lynch and R. C. McFall, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 351–355.
- 30 Z. Wu, M. E. Fraley, M. T. Bilodeau, M. L. Kaufman, E. S. Tasber, A. E. Balitza, G. D. Hartman, K. E. Coll, K. Rickert and J. Shipman, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 909–912.
- 31 M. T. Bilodeau, A. M. Cunningham, T. J. Koester, P. A. Ciecko, K. E. Coll, W. R. Huckle, R. W. Hungate, R. L. Kendall, R. C. McFall and X. Mao, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 2485–2488.
- 32 P. J. Manley, A. E. Balitza, M. T. Bilodeau, K. E. Coll, G. D. Hartman, R. C. McFall, K. W. Rickert, L. D. Rodman and K. A. Thomas, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 1673–1677.
- 33 G.-H. Kuo, A. Wang, S. Emanuel, A. DeAngelis, R. Zhang, P. J. Connolly, W. V. Murray, R. H. Gruninger, J. Sechler and A. Fuentes-Pesquera, *J. Med. Chem.*, 2005, **48**, 1886–1900.
- 34 L. Claesson-Welsh and M. Welsh, *J. Intern. Med.*, 2013, **273**, 114–127.
- 35 S. Tian, H. Quan, C. Xie, H. Guo, F. Lü, Y. Xu, J. Li and L. Lou, *Cancer Sci.*, 2011, **102**, 1374–1380.
- 36 G. Bold, K.-H. Altmann, J. Frei, M. Lang, P. W. Manley, P. Traxler, B. Wietfeld, J. Brüggén, E. Buchdunger and R. Cozens, *J. Med. Chem.*, 2000, **43**, 2310–2323.
- 37 T. J. Abrams, L. J. Murray, E. Pesenti, V. W. Holway, T. Colombo, L. B. Lee, J. M. Cherrington and N. K. Pryer, *Mol. Cancer Ther.*, 2003, **2**, 1011–1021.

- 38 J. S. Beebe, J. P. Jani, E. Knauth, P. Goodwin, C. Higdon, A. M. Rossi, E. Emerson, M. Finkelstein, E. Floyd and S. Harriman, *Cancer Res.*, 2003, **63**, 7301–7309.
- 39 D. E. Gingrich, D. R. Reddy, M. A. Iqbal, J. Singh, L. D. Aimone, T. S. Angeles, M. Albom, S. Yang, M. A. Ator and S. L. Meyer, *J. Med. Chem.*, 2003, **46**, 5375–5388.
- 40 A. G. A. El-Helby, H. Sakr, I. H. Eissa, H. Abulkhair, A. A. Al-Karmalawy and K. El-Adl, *Arch. Pharm.*, 2019, **352**, 1900113.
- 41 C. Roy, *Mol. Inf.*, 2012, **31**, 817–835.
- 42 P. Gramatica, *Int. J. Quant. Struct.-Prop. Relat.*, 2020, **5**, 61–97.
- 43 M. F. Andrada, E. G. Vega-Hissi, M. R. Estrada and J. C. Garro Martinez, *Sar & Qsar in Environmental Research*, 2017, pp. 1011–1023.
- 44 M. Clark, R. D. Cramer III and N. Van Opdenbosch, *J. Comput. Chem.*, 1989, **10**, 982–1012.
- 45 U. Chaube and H. Bhatt, *Mol. Diversity*, 2017, **21**, 741–759.
- 46 K. Gerhard, *J. Comput.-Aided Mol. Des.*, 1999, **13**, 1–10.
- 47 R. D. Cramer, D. E. Patterson and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959–5967.
- 48 G. Klebe, U. Abraham and T. Mietzner, *J. Med. Chem.*, 1994, **37**, 4130–4146.
- 49 A. Xiao, Z. Zhang, L. An and Y. Xiang, *J. Mol. Model.*, 2008, **14**, 149–159.
- 50 J. Liu, F. Wang, Z. Ma, X. Wang and Y. Wang, *Int. J. Mol. Sci.*, 2011, **12**, 946–970.
- 51 A. Tropsha, P. Gramatica and V. K. Gombar, *QSAR Comb. Sci.*, 2003, **22**, 69–77.
- 52 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**(16), 2785–2791.
- 53 M. F. Sanner, *Int. J. Mol. Sci.*, 1999, **17**, 57–61.
- 54 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639–1662.
- 55 I. Y. B.-S. D. A. Case, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco., 2018.
- 56 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *Int. J. Mol. Sci.*, 2006, **25**, 247–260.
- 57 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
- 58 P. Stephens, F. Devlin, C. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 59 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian, Inc., Wallingford, CT, USA, 2009.
- 60 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 61 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 62 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 63 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 64 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham 3rd, *Acc. Chem. Res.*, 2000, **33**, 889–897.
- 65 J. Wang, P. Morin, W. Wang and P. A. Kollman, *J. Am. Chem. Soc.*, 2001, **123**, 5221–5230.
- 66 B. R. Miller III, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
- 67 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.*, 1996, **100**, 19824–19839.
- 68 T. Hou, K. Chen, W. A. McLaughlin, B. Lu and W. Wang, *PLoS Comput. Biol.*, 2006, **2**, e1.
- 69 A. Onufriev, D. Bashford and D. A. Case, *J. Phys. Chem. B*, 2000, **104**, 3712–3720.
- 70 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
- 71 R. B. Hermann, *J. Phys. Chem.*, 1972, **76**, 2754–2759.
- 72 D. Sitkoff, K. A. Sharp and B. Honig, *J. Phys. Chem.*, 1994, **98**, 1978–1988.
- 73 K. Onodera, K. Satou and H. Hirota, *J. Chem. Inf. Model.*, 2007, **47**, 1609–1618.
- 74 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.
- 75 L. R. Lima, R. S. Bastos, E. F. Ferreira, R. P. Leão, P. H. Araújo, S. S. d. R. Pita, H. F. De Freitas, J. M. Espejo-Román, E. L. Dos Santos and R. d. S. Ramos, *Int. J. Mol. Sci.*, 2022, **23**, 9927.

