



Cite this: *Mol. BioSyst.*, 2015,  
11, 2885

## *In silico* research to assist the investigation of carboxamide derivatives as potent TRPV1 antagonists†

Jinghui Wang,<sup>a</sup> Yan Li,<sup>\*a</sup> Yinfeng Yang,<sup>a</sup> Jian Du,<sup>a</sup> Shuwei Zhang<sup>a</sup> and Ling Yang<sup>b</sup>

The transient receptor potential vanilloid type 1 (TRPV1), a non-selective cation channel, is known for its essential role in the pathogenesis of various pain conditions such as nerve damage induced hyperalgesia, diabetic neuropathy and cancer pain. Therefore, TRPV1 is considered as a promising target for the development of new anti-inflammatory and analgesic drugs. In the present study, a theoretical study on the functionalities of the molecular interactions between 236 active ligands and TRPV1 was carried out, using three-dimensional quantitative structure–activity relationships (3D-QSAR), molecular docking and molecular dynamics (MD) simulation approaches. The ligand-based CoMSIA model (obtained by use of a random division method for splitting the training and test sets) exhibits optimum predictivity ( $Q^2 = 0.522$ ,  $R_{ncv}^2 = 0.935$ ,  $R_{pred}^2 = 0.839$ ). The results show that the models are useful tools for the prediction of the test sets as well as newly designed structures against TRPV1 activity. In addition, to verify the rationality of the random division method for splitting the dataset, we also used a self-organizing map (SOM) division approach for establishing the QSAR models. Interestingly, the obtained optimal CoMSIA model based on the SOM division exhibits almost the same proper statistical results ( $Q^2 = 0.521$ ,  $R_{ncv}^2 = 0.929$ ,  $R_{pred}^2 = 0.829$ ) as the random division-derived model, proving the reasonability of both division methods for building the models. The contour plots of molecular fields along with docking and MD simulation have identified several key structural requirements responsible for the activity. The present work provides extremely useful guidelines for future structural modifications of this class of compounds towards the development of superior TRPV1 antagonists. The new computational insights presented in this study are expected to be valuable for the guideline and development of new potent TRPV1 antagonists.

Received 28th May 2015,  
Accepted 30th July 2015

DOI: 10.1039/c5mb00356c

[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

## 1 Introduction

The unmistakable pungency of capsaicin, a component of chilli peppers and other natural products, that acts as a ‘vanilloid receptor’ has led to many basic research studies into the transient receptor potential vanilloid type 1 (TRPV1) function and an appreciation that this ion channel acts as a key signaling complex in the pain pathway.<sup>1</sup> TRPV1 is primarily found on small diameter primary afferents, particularly unmyelinated C-fibres and thinly myelinated A- $\delta$  fibres but is also expressed, albeit at apparently lower levels, in upstream parts of the pain pathway such as the

spinal cord and brain.<sup>1</sup> The pain research witnessed the breakthrough work that led to the cloning of the first vanilloid (capsaicin) receptor, TRPV1.<sup>2</sup> The cDNA for the vanilloid receptor was identified using the expression cloning method and its characterization as a nonselective cation channel was reported,<sup>3,4</sup> which sparked immense interest in academia and drug-discovery companies alike. The vanilloid receptor was re-named as ‘transient receptor potential vanilloid type 1’ (TRPV1) as a part of the re-classification of a large group of cation channels that contain a ‘TRP’ motif in the C terminus.<sup>3</sup>

TRPV1 is pursued as a target for the development of a new class of antiinflammatory and analgesic drugs. As a member of the six mammalian TRP channels of the vanilloid subfamily (TRPV), TRPV1 plays an important role in sensory and pain perception as well as calcium homeostasis.<sup>5</sup> It is a nonselective cation channel expressed on small diameter (C-fiber) nociceptor sensory neurons responsible for sensing noxious heat and transducing inflammatory pain signals.<sup>6</sup> An opening of TRPV1 channels expressed on sensory neurons caused by agonists [e.g., capsaicin, resiniferatoxin (RTX), or anandamide], heat or acidic pH, always leads to a

<sup>a</sup> Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Faculty of Chemical, Environmental and Biological Science and Technology, Dalian University of Technology, Dalian, Liaoning, 116024, PR China. E-mail: [jhwang\\_dlut@163.com](mailto:jhwang_dlut@163.com), [yinfengyang@yeah.net](mailto:yinfengyang@yeah.net), [dujian@dlut.edu.cn](mailto:dujian@dlut.edu.cn), [zswei@dlut.edu.cn](mailto:zswei@dlut.edu.cn), [yanli@dlut.edu.cn](mailto:yanli@dlut.edu.cn); Tel: +86-0411-84986062

<sup>b</sup> Laboratory of Pharmaceutical Resource Discovery, Dalian Institute of Chemical Physics, Graduate School of the Chinese Academy of Sciences, Dalian, Liaoning, 116023, China. E-mail: [yling@dicp.ac.cn](mailto:yling@dicp.ac.cn)

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5mb00356c

depolarization of the cell membrane *via* calcium in flux, whereby painful stimuli are transmitted and proinflammatory neuropeptides (such as the substance P and calcitonin generated peptides) are released.<sup>7</sup> Therefore, TRPV1 is also assumed playing an essential role in the pathogenesis of various pain conditions and chronic inflammatory disorders. TRPV1 antagonists are predictable to prevent pain by blocking receptors where pain is generated instead of preventing the transmission of pain.<sup>8</sup> Since TRPV1 can be activated or sensitized by multiple mechanisms and exists both in brain and various non-neuronal tissues, it is considered the principal integrator of multiple pain-producing stimuli, described as an attractive therapeutic target for the treatment of various neuroinflammatory disorders in the transmission of pain signaling.<sup>9,10</sup>

The increase in intracellular  $\text{Ca}^{2+}$  upon TRPV1 activation causes an excitation of the primary sensory neurons and the consequent central perception of pain. TRPV1 antagonists inhibit this transmission of nociceptive signaling from the periphery to the CNS as well as block other pathological states associated with this receptor. In recent years, a number of TRPV1 antagonists have been developed as novel analgesic and antiinflammatory agents, particularly for the treatment of chronic pain and inflammatory hyperalgesia.<sup>11</sup> As the first generation of TRPV1 antagonist, capsaizepine, was less promising than expected due to its poor pharmacokinetics,<sup>12</sup> a lot of pharmaceutical companies like GlaxoSmithKline (GSK), Novartis, Astra Zeneca, Amgen, Johnson & Johnson, Neurogen/Merck, Renovis/Pfizer and so on have launched attempts to develop the second generation of TRPV1 antagonists, and some of these concerted efforts are beginning to bear fruit as clinical trials are in progress. A number of promising antagonists have been developed, including SB705498, SB366791, BCTC, NGD-8243, AMG-517, AMG-9810, A-425619, KJM429, JYL1421, JNJ17203212, NGX-4010, WL-1001, WL-1002, A-4975, and GRC-6127.<sup>13</sup> GlaxoSmithKline was among the first to complete its lead's phase 1 results, which ended up with SB-705498.<sup>14</sup> In preclinical studies, SB-705498 blocked the three major means of channel activation (capsaicin, protons and heat) with similar potencies of 3, 6, and 7 nM.<sup>14</sup> In a proof-of-concept study, a single oral dose of 400 mg SB-705498 substantially reduced the pain and flare evoked by cutaneous capsaicin challenge (0.075% capsaicin cream applied to the forearm) compared to placebo.<sup>15</sup> Subsequently, the effect of peros SB-705495 on dermal heat sensitivity following the capsaicin or UVB challenge was evaluated. Somewhat unexpectedly, SB-705498 caused a modest but significant (up to 1.3 °C) increase in the noxious heat perception threshold both in rodents and humans. In 2005, an active-controlled, placebo-controlled, randomized, single-blind, phase 2 trial was initiated in subjects with dental pain following the third molar tooth extraction.<sup>15</sup> A-425619 (Abbott) [1-isoquinolin-5-yl-3-(4-trifluoromethyl-benzyl)-urea] inhibited the mechanical allodynia induced by skin incision and reversed mechanical hyperalgesia in rat models of chronic and neuropathic pain.<sup>16–18</sup> BCTC [*N*-(4-tertiarybutylphenyl)-4-(3-cholorophyridin-2-yl)tetrahydropyrazine-1(2*H*)-carbox-amide] and Nrgn (Neurogen) significantly attenuated the thermal and mechanical hyperalgesia in the chronic constriction injury

(CCI) model of neuropathic pain and the CFA model of inflammatory pain in rats.<sup>18,19</sup> JNJ-17203212 relieved osteolytic sarcoma-related bone pain in mouse.<sup>18,20</sup> TRPV1 antagonists are expected to have therapeutic potential for various disorders, such as hyperalgesia, neuropathic pain, migraine, irritable bowel syndrome, osteoarthritis, and overactive bladder.<sup>21</sup> Therefore, it appears to be of interest to utilize both the ligand- and structure-based drug design techniques for probing the structural requirements of potential TRPV1 antagonists, which may be useful for design of novel potent antimalarial agents.

With the rapid growth of molecular biology, chemical biology, and computer technology, computer simulation technology plays an important role in the development of new drugs.<sup>22–24</sup> Structure- and ligand-based drug design approaches have become fundamental components of modern drug discovery.<sup>25,26</sup> During the past few decades, QSAR methods including especially the three-dimensional quantitative structure activity relationship (3D-QSAR) approaches have been successfully employed to assist the design of new drug candidates, ranging from enzyme inhibitors to receptor ligands. In the present work, in order to quantitatively probe the structure–activity relationship, we have applied two methodologies, comparative molecular field analysis (CoMFA), and comparative molecular similarity indices analysis (CoMSIA), to build 3D-QSAR models based on 236 TRPV1 antagonists,<sup>27–35</sup> for exploring those key structural features affecting the antagonistic activities. In addition, the docking study and molecular dynamics (MD) simulation could provide deep insight into the QSAR in view of the structural properties of active sites of the human TRPV1 receptor, and thus could more effectively utter the design of novel potential antagonists. The established models not only predict the biological activity values of carboxamide derivatives rapidly and accurately, but also provide some valuable information in structural modifications for designing new possible lead antagonists with higher activity.

## 2 Methods and computational details

### 2.1 Selection of training and test sets using the random division approach

Random division is a widely used approach in building the QSAR model's robustness. It consists of rebuilding the models using randomized activities of the modeling set, and a subsequent assessment of the model's statistics.<sup>36</sup> The advantage of this approach is that the test set compounds are "unknown" to the models since they are excluded from the model development procedure, especially the variable selection.<sup>37</sup> However, the disadvantage of dividing the overall set randomly is that it does not provide any rationale for selecting test set chemicals. In view of this, in order to verify the rationality of the random splitting approach, the modeling set was subdivided into a training set (80% of the modeling set) and a test set (20%) using Kohonen's self-organizing map (SOM) for comparison. From the validation results, it was obvious that the obtained random and SOM models could be used subsequently for the discovery of novel alternative potent lead compounds of TRPV1 antagonists.

In the present work, to investigate the descriptor space, all molecular descriptors were calculated for each TRPV1 antagonist using the Dragon professional package (version 5.4). Based on the SOM visualization of the whole data, the construction of the training and test sets was made.

### 2.3 Data set

In this work, a chemical structure database of 236 carboxamides-based analogs was investigated as TRPV1 antagonists.<sup>27–35</sup> The biological activities of these TRPV1 antagonists are expressed in IC<sub>50</sub> (nM) values, which were converted into log IC<sub>50</sub> (pIC<sub>50</sub>) in order to make normal distribution of the biological data for use in 3D-QSAR studies. All molecules exhibit a wide range of biological activity and proper variation in structural features. The data set of 236 compounds as TRPV1 antagonists was divided into a training set and a test set by using random and SOM splitting methods.<sup>43</sup> The training set was used to construct 3D-QSAR models and the test set was used for the model validation. The selection of the test set was made based

on the fact that they can appropriately represent the structural diversity of the whole data set and cover the range of  $\text{pIC}_{50}$  values. The experimental values for the training and test sets are listed in the ESI.†

## 2.4 Superposition of ligands

Molecular alignment is considered to be one of the most critical steps for the correlation of structurally diversified compounds in 3D-QSAR studies. During the process, the most active compound **197** was used as a template for alignment, and the rest of the molecules were aligned to fit the template with the common substructure as shown in blue in Fig. 1A. Fig. 1B shows the atom-based superimposition of all the compounds in the dataset. In order to avoid bias towards a particular alignment method, receptor-based superimposition derived from the docking analysis was also adopted.

## 2.5 Preparation of the small molecules and receptor preparation

The quality of docking results depends on reasonably the starting structures of both the protein and the ligand. Molecular modeling and database alignment were performed by using the SYBYL 6.9 program package.<sup>44</sup> The energy minimization processes for geometry optimization of ligands were performed using the Tripos force field followed by 1000 iterations with Powell's conjugate gradient algorithm with a  $0.05 \text{ kcal (mol } \text{\AA}^{-1})^{-1}$  convergence criterion.<sup>45</sup> Gasteiger-Hückel charges were assigned to each ligand to calculate the partial atomic charges for 3D-QSAR studies. As the rat and human TRPV1s have more than 85% sequence identity, and only five residues in the ligand binding site are different, we constructed the tetramer homology model of human TRPV1 based on the rat TRPV1 (rTRPV1) homology model<sup>46</sup> and performed docking studies on the whole dataset of TRPV1 ligands. The obtained TRPV1 complex was used to generate the receptor site and the energetic grid for the following docking and calculations. The 3D structure of the docked complex with compound **197** was analyzed in detail and served as a starting structure for MD simulations.

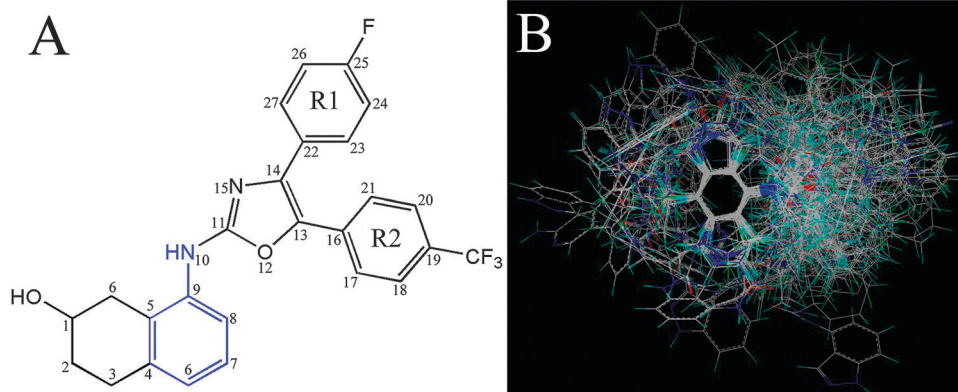


Fig. 1 (A) Compound **197** was used as a template for alignments with a common substructure shown in blue. (B) The atom-based alignment of all the compounds as shown in panels. Molecules are colored in white for common C, blue for N, red for O, yellow for S, cyan for H atoms, respectively.

## 2.6 CoMFA and CoMSIA descriptors

For the generation of CoMFA descriptor fields, a 3D cubic lattice with a grid spacing of 2.0 Å in all directions was created to encompass the aligned molecules. The steric (Lennard-Jones potential) and electrostatic (Coulomb potential) fields were calculated at each grid point using an sp<sup>3</sup> carbon probe atom with +1 charge, 0.2 nm grid spacing, and 0.152 nm van der Waals radius. To minimize the domination by large steric and electrostatic energies, an energy cutoff of 30 kcal mol<sup>-1</sup> for both the steric and electrostatic contributions was set as threshold.

In CoMSIA studies, the same lattice box was used as in CoMFA. Molecular similarity is expressed in terms of five different properties: steric, electrostatic, hydrophobic, H-bond donor and acceptor fields. The properties were studied using a probe atom with a radius of 2.0 Å, +1.0 charge and hydrophobic and hydrogen bond properties of +1. The steric indices were related to the third power of the atomic radii; electrostatic descriptors were derived from atomic partial charges; hydrophobic fields were derived using atom-based parameters; and H-bond donor and acceptor indices were obtained using a rule-based method based on experimental results.<sup>47</sup>

Partial least squares (PLS) have been universally applied as 3D-QSAR methods for analyzing the significant statistical relationship of the antagonistic testing set in CoMFA and CoMSIA models.<sup>48</sup> The cross-validation analysis was performed using the leave-one-out (LOO) method in which one molecule was removed from the data set and its activity was predicted using the model derived from the rest of the dataset. The cross-validated correlation coefficient ( $Q^2$ ) that resulted in the optimum number of components and the lowest standard error of prediction was considered for further analysis.

To assess the predictive power of the derived 3D-QSAR models, the activities of the test set molecules were predicted. The predictive correlation coefficient ( $R_{\text{pred}}^2$ ), based on the molecules of the test set, was calculated according to the equation shown below.

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

SD is the sum of the squared deviation between the biological activity of the test set molecules and the mean activity of the training set molecules, and PRESS is the predictive residual sum of square and is calculated by taking difference in predicted and actual activities of the test set molecules.

## 2.7 Molecular docking calculations

The ligand–receptor interaction is considered as an essential parameter for modulating the activity for series of inhibitors.<sup>49</sup> Knowledge of the preferred orientation may be used to predict the strength of association or binding affinity between the ligand and the receptor and quantified by a scoring function.<sup>50</sup> The advanced molecular docking program GOLD (version 5.1), with a powerful genetic algorithm (GA) method for conformational search and docking programs,<sup>51</sup> was employed to generate an ensemble of docked conformations. The GOLD score fitness function was implemented as a measure for evaluation of the

TRPV1 binding affinity, and takes into account the hydrogen-bonding energy, van der Waals energy and ligand torsion strain.<sup>52</sup> The binding site was created within 5 Å. In addition, we considered the knowledge of crucial residues determined by the mutational studies and experimental analysis obtained from the literature. At the end of the computation, each conformer of all 236 antagonists was docked into the binding pocket, and 10 conformations were obtained through GOLD for each ligand.

## 2.8 Molecular dynamics simulations

To confirm the docking results, the docked complex of TRPV1 with compound **197** was adopted in the MD simulations using the GROMACS software.<sup>53</sup> To relax the TRPV1 models, the system was embedded into a Dioleoylphosphatidylcholine (DOPC) lipid bilayer on a water box (TIP3). The lipid bilayer system was built applying the Charmm input generator, graphical interface.<sup>54,55</sup> The full system was subjected to energy minimization using the steepest descent algorithm for 6990 steps, and then the system was equilibrated by a 2000 ps MD simulation at 300 K. The simulation used a normal pressure and temperature (NPT) ensemble at 300 K with periodic boundary conditions, the temperature was maintained constant using the Berendsen thermostat, and the values of the isothermal compressibility were set to be  $4.5 \times 10^{-5}$  bar<sup>-1</sup> while the pressure was maintained at 1 bar using the Parrinello–Rahman scheme.<sup>56</sup> The particle mesh Ewald (PME) technique was used to evaluate the electrostatic interactions, and the linear constraint solver (LINCS)<sup>57</sup> algorithm was applied to fix all bonds involving hydrogen. The cut-off distances for the calculation of Coulomb and van der Waals interactions were 1.0 and 1.4 nm, respectively. Finally, a 10 ns simulation was run with a time step of 2 fs.

# 3 Results

## 3.1 Reliability of the 3D-QSAR model based on random selection

It is well known that the CoMFA and CoMSIA models are alignment sensitive, with the quality and the predictive ability of the models directly dependent on the alignment rules, and differences in the statistical values are observed with different alignments.<sup>50</sup> As a comparison, atom-based superimposition CoMSIA models obtain better statistical results than those obtained from the receptor-based superimposition. Therefore, we focus on the atom-based conformational models in the following discussion. For the TRPV1 3D-QSAR studies, using random selection for splitting the dataset, good correlations are observed in the obtained CoMFA and CoMSIA models demonstrated by the high values of  $Q^2$  and other statistical results. As a result, the optimal CoMSIA model with proper predictive performance based on the training (177 molecules) and test set (59 molecules) was obtained. The statistically significant results for the CoMSIA model are summarized in Table 1.

The analysis of the resulted 3D-QSAR models shows that the best CoMFA model was established by a combinational use of steric and electrostatic fields, which yields a cross-validated



Table 1 Summary of CoMFA and CoMSIA results

								Contribution (%)				
PLS statistics	N	Q <sup>2</sup>	SEP	R <sub>ncv</sub> <sup>2</sup>	SEE	F	R <sub>pred</sub> <sup>2</sup>	S	E	H	D	A
CoMFA												
S	10	0.261	0.764	0.893	0.284	138.184	0.610	100	—	—	—	—
E	3	0.194	0.763	0.532	0.581	65.654	0.441	—	100	—	—	—
SE	10	0.387	0.679	0.945	0.204	283.530	0.766	50.4	49.6	—	—	—
CoMSIA												
S	6	0.112	0.808	0.562	0.568	36.313	0.535	100	—	—	—	—
E	4	0.267	0.730	0.656	0.500	82.116	0.581	—	100	—	—	—
H	6	0.186	0.774	0.742	0.435	81.606	0.520	—	—	100	—	—
D	3	0.032	0.836	0.208	0.757	15.138	0.265	—	—	—	100	—
A	2	0.042	0.830	0.252	0.733	29.272	0.291	—	—	—	—	100
SE	6	0.366	0.683	0.817	0.367	126.716	0.723	34.1	65.9	—	—	—
SH	10	0.237	0.758	0.879	0.302	120.224	0.595	35.7	—	64.3	—	—
SD	3	0.064	0.822	0.357	0.682	32.035	0.497	46.8	—	—	53.2	—
SA	7	0.176	0.781	0.713	0.461	59.856	0.555	52.2	—	—	—	47.8
EH	9	0.477	0.626	0.929	0.230	243.214	0.790	—	52.5	47.5	—	—
ED	6	0.286	0.734	0.731	0.445	76.962	0.601	—	68.8	—	31.2	—
EA	6	0.288	0.723	0.763	0.418	90.991	0.647	—	69.6	—	—	30.4
HD	7	0.089	0.821	0.722	0.453	62.783	0.605	—	—	68.4	31.6	—
HA	9	0.240	0.754	0.849	0.336	104.674	0.625	—	—	66.4	—	33.6
DA	3	0.024	0.840	0.290	0.716	23.579	0.344	—	—	—	49.0	51.0
SEH	9	0.491	0.618	0.927	0.234	234.452	0.818	20.0	44.1	35.9	—	—
SED	10	0.345	0.702	0.904	0.269	156.302	0.758	28.2	49.3	—	22.5	—
SEA	10	0.404	0.670	0.923	0.242	197.650	0.789	28.6	48.2	—	—	23.2
SHD	7	0.160	0.788	0.743	0.436	69.856	0.670	27.2	—	47.3	25.5	—
SHA	9	0.306	0.721	0.873	0.308	127.614	0.726	27.3	—	46.6	—	26.2
SDA	6	0.135	0.798	0.645	0.511	51.439	0.586	35.8	—	—	29.4	34.7
EHD	10	0.424	0.659	0.928	0.233	214.427	0.779	—	43.9	38.7	17.4	—
EHA	9	0.492	0.617	0.930	0.229	246.151	0.806	—	42.4	39.1	—	1.85
EDA	10	0.272	0.740	0.871	0.311	112.477	0.621	—	54.1	—	24.1	21.8
HDA	10	0.203	0.775	0.861	0.323	102.990	0.660	—	—	52.3	21.4	26.3
SEHD	10	0.450	0.644	0.931	0.227	224.984	0.811	17.6	37.3	30.4	14.7	—
<b>SEHA</b>	<b>9</b>	<b>0.522</b>	<b>0.598</b>	<b>0.935</b>	<b>0.221</b>	<b>265.182</b>	<b>0.839</b>	<b>17.7</b>	<b>35.6</b>	<b>30.4</b>	—	<b>16.2</b>
SEDA	10	0.368	0.690	0.908	0.263	163.641	0.765	23.9	40.4	—	17.9	17.8
SHDA	10	0.253	0.750	0.890	0.288	134.163	0.736	22.7	—	38.7	16.8	21.8
EHDA	10	0.451	0.643	0.930	0.229	221.023	0.791	—	36.6	33.3	14.6	15.5
SEHDA	10	0.465	0.635	0.937	0.217	248.591	0.824	15.7	31.5	26.7	12.5	13.6

Q<sup>2</sup>, cross-validated correlation coefficient after the LOO procedure; R<sub>ncv</sub><sup>2</sup>, non-cross-validated correlation coefficient; SEE, standard error of estimate; F, ratio of R<sub>ncv</sub><sup>2</sup> explained to be unexplained = R<sub>ncv</sub><sup>2</sup>/(1 - R<sub>ncv</sub><sup>2</sup>); R<sub>pred</sub><sup>2</sup>, predicted correlation coefficient for the test set of compounds; SEP, standard error of prediction; N, optimal number of principal components.

Q<sup>2</sup> of 0.387 with an optimal number of components of 10, a non-cross-validated R<sub>ncv</sub><sup>2</sup> of 0.945, an estimated F value of 283.530 and a low SEE of 0.204. The steric component of these carboxamide analogs on the inhibitory potency described by this model is 50.4%, whereas the electrostatic portion is 49.6%. Generally, a Q<sup>2</sup> > 0.5 is considered to be the proof of acceptable internal predictive ability. Moreover, the high R<sub>ncv</sub><sup>2</sup> and F values along with the low SEE values should also be considered as the foundation of a reliable QSAR model. Therefore, we developed CoMSIA models by a combinational use of five descriptor fields: steric, electrostatic, hydrophobic, H-bond donor and acceptor with an attempt to build the optimal model.

Taking the same alignment method, training and test sets that were used to derive the best CoMFA model, various CoMSIA models were generated considering all possible combinations of CoMSIA field descriptors, *i.e.*, steric, electrostatic (E), hydrophobic, hydrogen bond donor and acceptor fields. The statistical significance of different CoMSIA models involving 31 possible combinations of descriptors is briefly shown in Table 1. From Table 1 we can see that the combination of steric, electrostatic,

hydrophobic and H-bond acceptor leads to the best CoMSIA model with larger Q<sup>2</sup>, R<sub>ncv</sub><sup>2</sup>, R<sub>pred</sub><sup>2</sup> values, which was then selected for further analysis. It gives an optimal number of components of 9 with a Q<sup>2</sup> of 0.522 which is better than the one of the previous CoMFA model (0.387). A non-cross-validated R<sub>ncv</sub><sup>2</sup> of 0.935, a SEE of 0.221, and an F value of 265.182 show a good statistical correlation between the predicted and the experimental pIC<sub>50</sub> values for the non-cross-validated form of the CoMSIA model. The CoMSIA analysis shows that the relative field contributions are 17.7%, 35.6%, 30.4% and 16.2% for steric, electrostatic, hydrophobic and H-bond acceptor fields, respectively. All these results indicate that the CoMSIA model is statistically satisfactory.

To test the robustness of the predictive capabilities of the models, the activities of the test set of 59 compounds excluded in the construction of the 3D-QSAR models were calculated to validate the reliabilities of the built models. A vital test for the predictability of QSAR analysis is to check the ability of a model to predict the biological activity of those compounds that are not included in the training set. As shown in Table 1, the CoMSIA

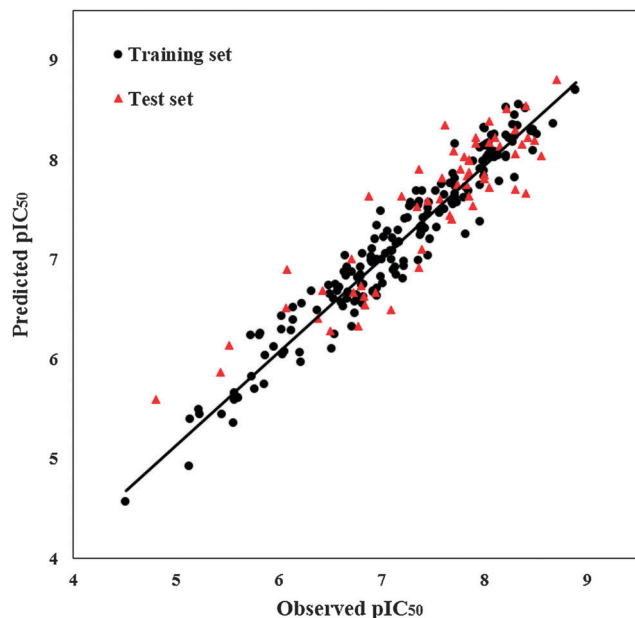


Fig. 2 The ligand-based correlation plots of the predicted *versus* the actual  $pIC_{50}$  values using the training (filled black dot) and the test (filled red triangles) set compounds based on the CoMSIA model from random splitting.

model is able to predict the affinity of the compounds with an  $R_{pred}^2$  of 0.839, suggesting that the CoMSIA model is able to accurately predict the affinity of compounds of  $\sim 84\%$ . Fig. 2 shows the linearity trend in the graphs of actual *vs.* predicted activity for training and test set molecules. According to these plots, the good agreement between experimental and predicted values for the test set molecules suggests that the CoMSIA model is reliable and can be used in the design of novel TRPV1 antagonists within this structural class.

### 3.2 CoMSIA model based on SOM

In order to verify the random selection for splitting the dataset into training and test sets, the modeling set was then subdivided into a training set (80% of the modeling set) and a test set (20%) using SOM division for comparison. The purpose of this study is also to determine whether the random division method leads to more predictive models when compared with the SOM division approach. To do this, we reselected the training and test sets and rebuilt the QSAR models by using the SOM division. The CoMSIA model shows good predictivity with low residuals for these chemicals, indicating that the CoMSIA model was reliable. The experimental and predicted  $pIC_{50}$  values for the training set and the test set based on the SOM division are listed in the ESI.†

Similar to random selection, the ligand-based alignment of the CoMSIA analysis from the SOM division leads to the models with larger  $Q^2$ ,  $R_{ncv}^2$  and  $R_{pred}^2$  than those obtained from another alignment. The results from the CoMSIA analysis are given in Table 2. The cross-validated  $Q^2$  is 0.521, suggesting that the CoMSIA model is statistically significant. The non-cross validated regression gives an  $R_{ncv}^2$  value of 0.929 with 9 components, suggesting that the model would be able to predict the activity

Table 2 Summary of ligand-based CoMSIA models built based on both the random division and SOM splitting methods

PLS statistics	Random division	SOM
$N$	9	9
$Q^2$	0.522	0.521
SEP	0.598	0.612
$R_{ncv}^2$	0.935	0.929
SEE	0.221	0.236
$F$	265.182	247.856
$R_{pred}^2$	0.839	0.829
Contribution (%)		
Steric	17.7	17.5
Electrostatic	35.6	36.2
Hydrophobic	30.4	29.5
HB acceptor	16.2	16.7

of a compound 93% of the time. In addition, the CoMSIA mode is able to predict the affinity of the compounds with an  $R_{pred}^2$  of 0.829, suggesting that the CoMSIA model is able to accurately predict the affinity of compounds of  $\sim 83\%$ . The steric, electrostatic, hydrophobic and HB acceptor field contributions are 17.5%, 36.2%, 29.5% and 16.7%, respectively. When compared with the results obtained by the random selection method, the CoMSIA model of SOM splitting exhibits almost the same proper statistical results ( $Q^2 = 0.522$ ,  $R_{ncv}^2 = 0.935$ ,  $R_{pred}^2 = 0.839$ ), with the steric, electrostatic, hydrophobic and H-bond acceptor contributions of 17.7%, 35.6%, 30.4% and 16.2%, respectively. In addition, the graph of actual activity *versus* predicted  $pIC_{50}$  of the training and test sets is illustrated in Fig. 3. According to this plot, the predicted  $pIC_{50}$  values are in correlation with the experimental ones within the tolerable error range. All these results indicate that models based on either the random division or SOM methods generate significant statistical results for the test sets, proving the reasonability of both division methods for building the models. A close observation of the results in fitting to the statistical parameters in the two splitting methods provides a conclusion that the model developed has a respectable predictive ability based on the structural perception of known TRPV1 antagonists.<sup>57</sup>

### 3.3 Interpretation of structure-based CoMSIA contour maps

The 3D-coefficient contour maps generated based on the optimal ligand-based CoMSIA model give us  $StDev \times Coeff$  fields which enable us to spot the structural requirements of these antagonists and best explain the relationship between variations in the space and changes in the activity. Based on the best possible internal and external predictive powers of 3D-QSAR CoMSIA models, different colors give information of the regions where appropriate substituent groups on a molecule will boost the activity or the regions where disfavored substituents are detrimental to the activity. The default values of 80% contribution for favored and 20% for disfavored regions were set for the visualization of the contour maps. To aid visualization, the most active compound **197** and least active compound **61** in the whole dataset are shown as template molecules superimposed with the CoMSIA contour maps. This could help in discovering novel scaffolds with good biological activity. The most and least active

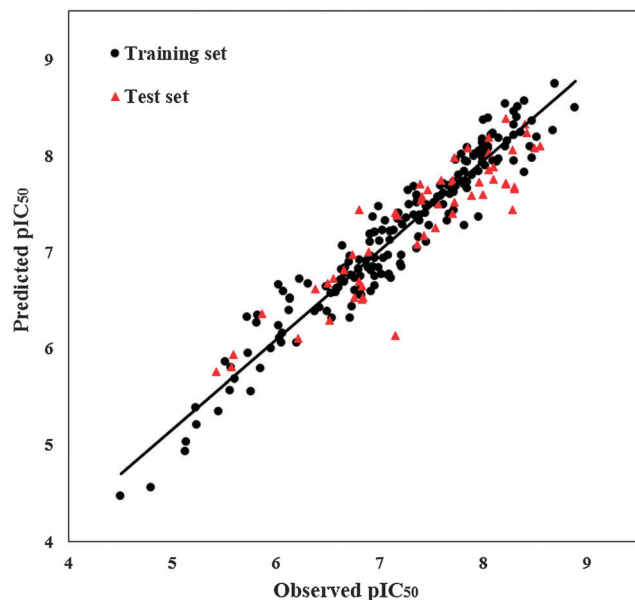


Fig. 3 The experimental  $pIC_{50}$  values versus the training and test sets predicted  $pIC_{50}$  values for the ligand-based CoMSIA model from the SOM division.

ligand contour maps were generated and are shown in Fig. 4 and 5. The two figures depict the contours of CoMSIA steric, electrostatic, hydrophobic, and H-bond acceptor fields, respectively.

As shown in Fig. 4A, green regions designate areas where steric bulk is predicted to enhance the TRPV1 antagonist activity; whereas yellow contours indicate regions where steric bulk is not favored. From this figure we can see that two large positive green isopleths around the tetrahydronaphthalene ring show favorable steric interactions in these positions, indicating that molecules carrying bulky substituents near these positions should be more active than those without or with smaller substituents at the same location. This result can be used to explain that compound **84** ( $pIC_{50} = 8.420$ ) fits into the green big contour around the oxazole ring and thus shows significantly improved activity compared to another structural analog **86** ( $pIC_{50} = 6.807$ ). In order to improve the antagonistic effect of the compound, new analogs with increasing steric substituents in these regions should be developed. In contrast, the imidazolidine moiety of the R1 group is encompassed by a large yellow contour meaning that the substitution of a too large group may decrease the activity. As a result, compounds **28**, **29**, **32**, **38** with a unfavorable bulkier substituent showed lower activity compared to the template. If the antagonist structure is too close to or directly in contact with the yellow contours, there will be an overall decrease of the antagonistic activity. By comparison, a bulk substituent of  $Me_2NCH_2$  and benzene was located within the steric disfavoring region, suggesting a lower biological activity due to its deviation from the contour map of the most inactive TRPV1 antagonist **61** (Fig. 5A).

Fig. 4B presents the electrostatic contour maps, where the blue and red contours represent positive and negative favored regions, respectively. In the positive electrostatic field, a visual

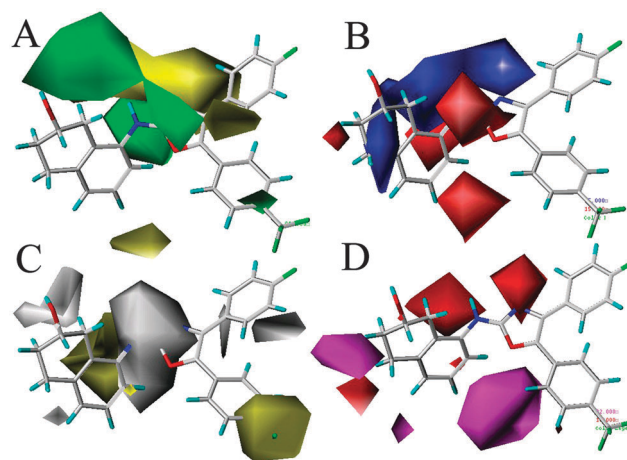


Fig. 4 CoMSIA StDev  $\times$  Coeff contour maps around the most potent TRPV1 antagonist **197**. (A) Steric contour map, green and yellow contours illustrate regions where steric bulk has favorable and unfavorable effects on the activity, respectively. (B) Electrostatic contour map in combination, in which blue contours indicate regions where electropositive groups increase the activity, while red contours indicate regions where electronegative groups increase the activity. (C) Hydrophobic contour map. Yellow contours indicate regions where hydrophobic substituents enhance the activity; white contours indicate regions where hydrophilic substituents enhance the activity; (D) HB acceptor contour map. Magenta contours indicate regions where HB acceptors on the receptor; red contours indicate regions where HB acceptors on the receptor demote the affinity.

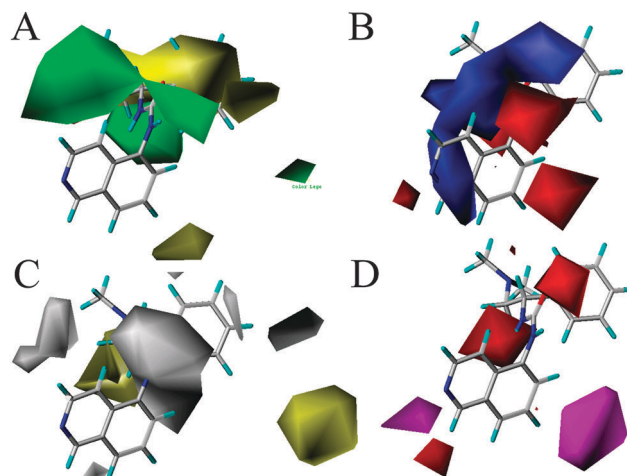


Fig. 5 CoMSIA StDev  $\times$  Coeff contour maps around the most inactive compound **61**. (A) Steric contour map. (B) Electrostatic contour map in combination. (C) Hydrophobic contour map. (D) HB acceptor contour map.

investigation reveals the blue contours located near the pyridine moiety of the core structure and a few blue contours scattered around the surrounding of the amide group, indicating that the presence of a positively charged group in this zone could help enhance the antagonistic activity. An electron-donating group ( $-NH_2$ ,  $-NHCH_3$ ,  $-NCH_3CH_3$  and  $-OCH_3$ ) at the position-3 of the oxazole ring and position-14 of the indole ring in compounds **2**, **3**, **51**, **55** and **57** results in significantly increased activity. At the same time, two red polyhedrons located in close proximity to

the pyridine moiety show that the electron-withdraw substituents at these positions are favored for activity. Most of the excellent derivatives (2–9, 17–22 and 24–43) possess an electron-withdrawing aromatic group (e.g. isoxazolyl, phenyl, thiazolyl, pyridinyl, pyrimidinyl, imidazolyl) at the terminal of R2 sites. In contrast, those without an electron-withdrawing substituent at this position (16, 23, 44–56, 58, 61) are all the inactive compounds. As shown in Fig. 5B, the substituent of Me<sub>2</sub>NCH<sub>2</sub> was located within the electron-withdrawing region in the most inactive compound 61, suggesting a lower biological activity.

A contour map for the hydrophobic property is displayed in Fig. 4C. Yellow and gray contours indicate regions where the hydrophobic feature favors and disfavors the activity, respectively. One large yellow polyhedron lying over an oxazole moiety of the R2 position and other big polyhedrons present in the vicinity of the naphthyl moiety at position-5 represent the favorable position for hydrophobic groups, suggesting that occupancy by hydrophobic groups would increase the activity. It is clear to see from the database that substituents (like –OMe, –OEt, –F, –Cl, –Br) can enhance the activity. Compounds 51, 52 with higher activities are found to follow this pattern for the existence of a hydrophobic group at the R2 position. In addition, there is a white unfavorable region stumbled on the R1 chain and amide group, indicating the favor for hydrophilic groups at this location. This is in agreement with the fact that compounds 9, 20, 21, 27, 28, 30, 32 and 43 bearing a relative hydrophilic group (e.g. *N*-thiazol-2-yl-acetamino, pyridinylmethyl, pyrimidinylmethyl, thiazolylmethyl, 5-methyl-iso-xazol-3-ylmethyl) are the higher potential derivatives. In contrast, no hydrophobic and hydrophilic groups are found in the favored region of the hydrophobic contour map (Fig. 5C) with the most inactive compound 61, further confirming that the compound 61 has the lowest biological activity.

In the CoMFA H-bond acceptor contour (Fig. 4D), the magenta and red regions mean that CoMSIA H-bond acceptor substituents are favorable and unfavorable for antagonist activity, respectively. A large magenta isopleth situates near the pyrazine core suggesting that the H-bond acceptor group may benefit the potency at this position, which on one hand is in accordance with the observations that most active compounds in the data set contain a H-bond acceptor group, on the other hand indicates that a H-bond acceptor group at this location may be a necessity for strong binding activities. This is illustrated by the O atom position-3 of the oxazole ring acting as a H-bond acceptor. Besides, the two red contours around the pyridine moiety and the top of the central phenyl moiety indicate that the presence of H-bond donor groups may increase the ligand activity. Most of the active derivatives (2–12, 17, 19–21, 24–43) all possess hydrogen bond acceptor groups such as carbonyl, oxygen or nitrogen atoms in the *N*-thiazol-2-yl-acetamino, pyridinylmethyl, pyrimidinylmethyl, thiazolylmethyl, 5-methyl-isoxazol-3-ylmethyl substituent at R2. Compounds 15, 16, 23 and 44–59 without a hydrogen bond acceptor substituent at this position exhibited significantly decreased potencies. In the least active compound as in Fig. 5D, two unfavorable red regions surrounded the tetrahydronaphthalene and benzene ring moiety, which indicated

that a decrease in the ring size could increase the biological activity of the compound.

### 3.4 Flexible docking analysis

The structural identification of the binding motif of the TRPV1 receptor is essential to determine in detail the mechanism of action of the TRPV1 ligand. TRPV1 is a tetrameric membrane protein with each monomer composed of six transmembrane helices (TM1–TM6) and cytosolic N- and C-terminal tails. The N-terminal region plays a key role in the channel function as it contains six ankyrin repeats which are 33-residue sequence motifs with functions including the signaling, cytoskeleton integrity, transcription, and cellular localization. The C-terminal has a coiled-coil segment just adjacent to the sixth transmembrane segment that has been implicated in the subunit tetramerization and channel gating.<sup>58</sup> The membrane region consists of two domains – a pore domain (TM5–TM6), containing a pore-forming loop between TM5 and TM6, and a voltage sensor domain (TM1–TM4) (Fig. 6).<sup>46</sup> The recent Moiseenkova-Bell's report of the single-particle electron cryomicroscopy structure revealed that TRPV1 has four monomers symmetrically arrayed to generate two distinct domains: a large open basket-like domain, likely corresponding to the cytoplasmic N- and C-terminal portions, and a more compact domain, corresponding to the transmembrane portion.<sup>59</sup>

Although an X-ray crystal structure has not been reported as yet, several research groups have proposed TRPV1 models and tried to predict the binding modes of some ligands in terms of their models.<sup>46</sup> Middleton and co-workers built a homology model for the TM3–TM4 (Ser505–Thr550) regions using the isolated voltage-sensor domain from KvAP; this model oriented the vanillyl moieties of capsaicin and RTX in opposite directions in TRPV1.<sup>60</sup> In this model, Met547 and Tyr555 interact with the vanillyl moiety and Tyr511 contacts the orthophenyl group of RTX.<sup>60</sup> Gavva *et al.* constructed a model limited to the TM3–TM4 regions and predicted the binding modes of capsaicin and RTX in which the same residues in TRPV1 interacted

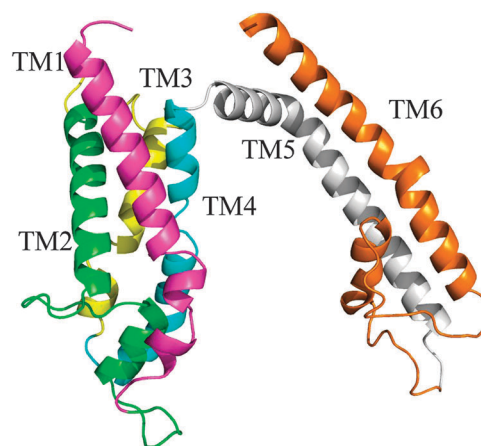


Fig. 6 The homology model of human TRPV1 developed with the obtained alignment using EasyModeller 4.0. It is represented in the secondary structure with six transmembrane helices (TM1–TM6).



with the vanillyl moieties of the two ligands.<sup>61</sup> This model proposed the interactions of the vanillyl moiety with Thr550 and the “tail end” hydrophobic group of capsaicin or RTX interacting with Tyr511.<sup>62</sup> In addition, based on chimeric and mutagenesis studies on TRPV1, vanilloid sensitivity and RTX binding are transferable with the TM3/4 region from rat TRPV1 to chicken and rabbit orthologues as well as to other TRP family members such as TRPV2 and TRPV4.<sup>61</sup> All these results indicate that the vanilloid binding site is defined by the TM3/4 region.

To determine the potential binding conformations, the obtained compounds were docked into the binding site of TRPV1 protein using the GOLD docking algorithm. The side chains of the six residues (*i.e.* Tyr511, Ser512, Leu515, Met547, Thr550, and Asn551) in the binding site of the TM3/4 region were set to be flexible with ‘crystal mode’ (Fig. 7). The majority of hydrogen-bonds and hydrophobic modes are similar to those observed in the crystal structure of TRPV1, which validates the reliability of the docking model. All 236 compounds were docked into the binding site 10 times in order to find the possible binding conformation, and they have shown the similar binding pattern as that of the several docking studies in the homology model of TRPV1 antagonist. The best poses of the most active compound **197** were selected on the basis of Gold scores for detailed analyses and compound **197** fits well into the binding site. The binding mode of the ligand at the active site and the main residues involved in the interaction are shown in Fig. 8.

The topmost pose of compound **197** shows that its tetrahydronaphthalene group is inserted into the cavity formed by the TRPV1 protein interface. This bulky tetrahydronaphthalene group in the C-region extends toward Met547 in the upper hydrophobic area and the adjacent monomer's hydrophobic region, composed of Tyr511, Phe543 and Phe550, indicating that the molecule shows a perfect fit with the binding pocket. This finding is almost similar to the model of Jordt and Julius in which Tyr511 interacts with the vanillyl-moiety and the hydrophobic tail protrudes into the membrane.<sup>63</sup> The central oxazole moiety and two phenyl groups are involved in extensive hydrophobic contacts with residues Asp509, Ser512, Phe516,

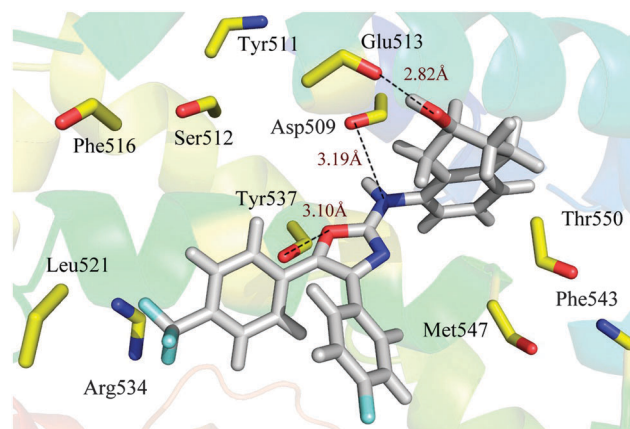


Fig. 8 Stereoview of docking interactions of compound **197** with active site amino acid residues of TRPV1.

Tyr537, Leu521, Arg534 and Met547, which are the important residues at the active site and are contributors to the ligand and receptor interaction. The central oxazole group in the C-region makes an additional hydrophobic interaction with the hydrophobic region composed of Met547.

Hydrogen bonds are analyzed during the simulation. The ligand core is anchored in the binding site *via* four hydrogen bonds with the TRPV1 protein. The tetrahydronaphthalen-2-ol ring of carbonyl oxygen participates in the hydrogen bonding with Glu513 ( $\text{O} \cdots \text{HO}$ , 2.82 Å), which perfectly fills the bottom of the TRPV1 cavity, and emerges as essential in determining the potency of the inhibitors. An oxygen atom of the oxazole group makes a hydrogen bond with Tyr537 ( $\text{O} \cdots \text{HO}$ , 3.10 Å) and also contributes to the appropriate positioning of the C-region for the hydrophobic interactions. In addition, a strong hydrogen bond is found between the Asp509 ( $\text{O} \cdots \text{HN}$ , 3.19 Å) and the position-5 of nitrogen atoms. The inhibitor appears to fit the binding pocket well. Stabilization of the ligand–protein complex is made mainly through a network of H-bonding interactions between the ligand and the amino acid residues defining the binding pocket of TRPV1.

### 3.5 Molecular dynamics simulation

MD simulation is a well-established method to computationally probe the structure and dynamics of biological macromolecules. When predicting the binding mode of the ligand to receptor, molecular docking provides a good starting to evaluate the stability of the predicted interactions involved in binding. While molecular dynamics was employed for further calculations considering the flexibility of the receptor and the effect of water salvation on the complex. In this study, firstly compound **197** was docked into the binding site of the receptors, whose complexes were further used to undertake MD simulations. The resulting docked complexes were embedded into a DOPC lipid bilayer, and solvated in a water box (TIP3 mode) considering the presence of counter ions. Then 10 ns simulations of the docked complex structure of TRPV1 with ligand **197** were carried out to obtain a dynamical picture of the conformational changes that take place in aqueous solution, with the main

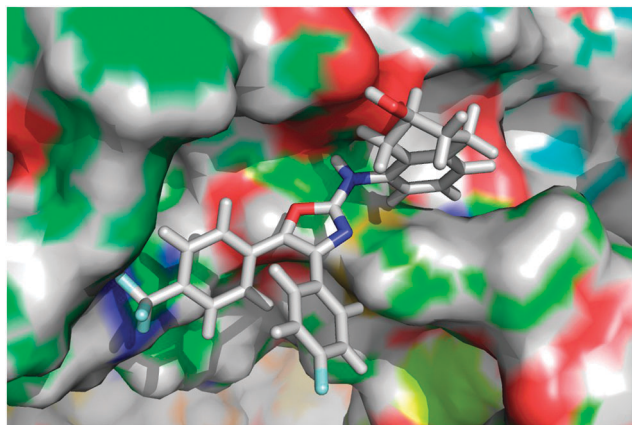
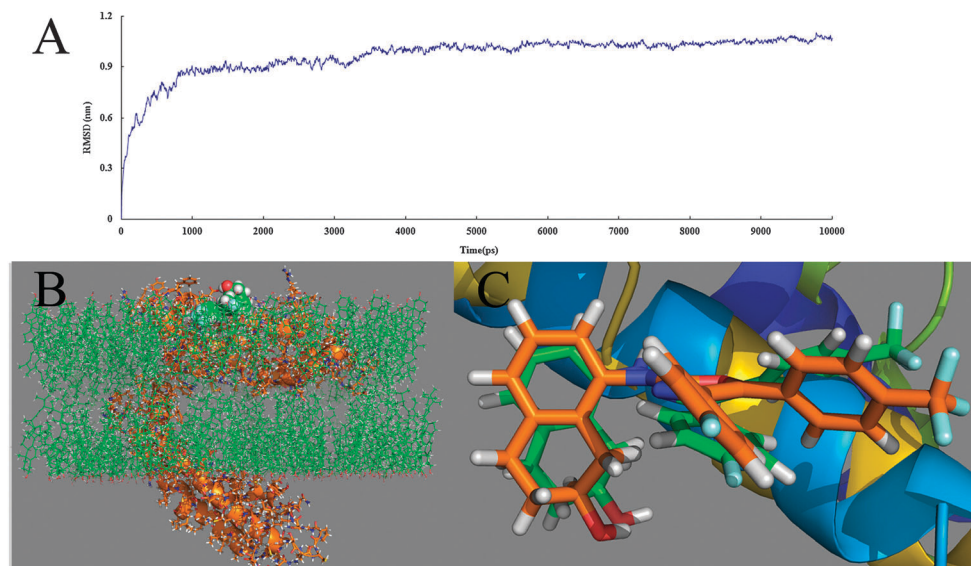


Fig. 7 Surface structure displayed with cavity depth potential of the binding site within compound **197**.



**Fig. 9** (A) Plot of the root-mean-square deviation (RMSD) of docked complex/ligand versus the MD simulation time in the MD-simulated structures. (B) The receptor with a docked ligand within the lipid bilayer after 10 ns of MD simulation. Protein is shown as cartoons. Ligands are shown as spheres. Lipid molecules are shown as lines. (C) View of superimposed backbone atoms of the average structure for the MD simulations and the initial structure of the docking for the complex. Compound **197** is represented as a carbon-chain in green for the initial complex and a carbon-chain in maroon for the average structure, respectively.

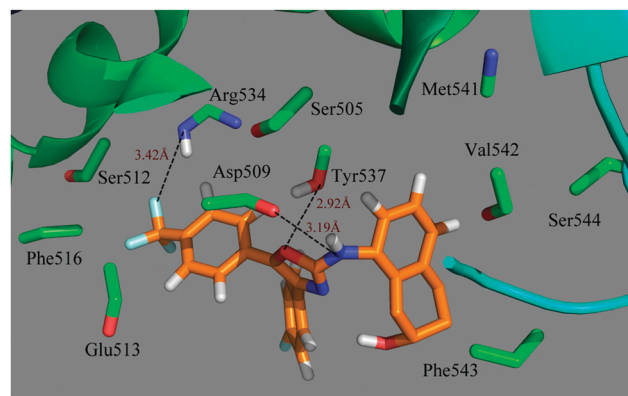
emphasis of investigating the conformational alterations that occur in the ligand and the protein. The binding mode of the ligand after 10 ns MD simulation in the lipid bilayer is displayed in Fig. 9B.

To explore the dynamic stability of the complexes and to ensure the rationality of the sampling method, root-mean-square deviations from the starting structure were analyzed. Fig. 9A depicts the RMSD of the trajectory with respect to the initial structure of the TRPV1 protein and the small molecule complex. The RMSD of the trajectory with respect to the initial structure of the protein–ligand complex ranges from 0 to 10.2 Å, and reaches a plateau about 9.5 Å after 4.2 ns, indicating that the structure of protein–ligand complex conformation thereafter is stable and reliable for the subsequent docking study. This clearly shows a metastable conformation after 4.2 ns of simulation for the docked complex structure.

Fig. 9C shows an overlay of the docked structure and the average structure of the last 1 ns, where the green ribbon represents the initial structure of the docked complex, the maroon ribbon shows the average structure of the MD simulations, with compound **197** represented as a stick and a line for the initial complex and the average complex, respectively. According to the analyzed interactions between compound **197** and the receptor, it is noted that the binding pocket does not experience a significantly larger-than-average conformational change between the average structure extracted from the MD simulations and the docked model of the complex. MD simulation results of compound **197** show a similar binding mode, lending credit to the reliability of the active conformations obtained by Gold.

To investigate the positional and conformational changes of ligand relative to the binding site, the principal native contact

and hydrogen bonding interaction were investigated. The native contact is a contact between the side chain of amino acid and the ligand, which is spatially close in the TRPV1. As shown in Fig. 10, the majority of key amino acids (within 4.5 Å distance from the ligand) from the average complex structure of MD simulation are almost similar to those obtained in the docking results, indicating that the docked complexes and the MD average structures appear at the same binding site. The binding site for the MD simulation, including Ser554, Phe543, Glu513, Phe516, Ser512, Arg534, Ser505, Tyr537, Met541, Val542, Asp509, has only a slight conformational drift from the docking results, which further verifies the reasonability of the docking model. The hydrogen bond interaction is one of the most important forces that maintain the binding between the ligand and the receptor (Fig. 10). The hydrogen bond was characterized by a distance



**Fig. 10** Plot of the MD-simulated structures of the binding site with compound **197**. H-bonds are shown as dotted red lines; active site amino acid residues are represented as sticks.

Table 3 Hydrogen bond analysis from docking and MD simulation

Method	Donor	Acceptor	Distance (Å)
Docking	Ligand O	Glu513 O	2.82
	Tyr537 O	Ligand O	3.10
	Ligand N	Asp509 O	3.19
MD	Ligand N	Asp509 O	3.19
	Tyr537 O	Ligand O	2.92
	Arg534 N	Ligand F	3.42

(<3.5 Å) and orientation (with the angle D-H...A > 120°). Table 3 shows the hydrogen bond analysis from docking and MD simulation.

It is shown that the oxazole group makes another hydrogen bond with Tyr537 (–O...HO, 2.92 Å) from the initial structure to the end of the simulation, which is consistent with the docking results. In addition, the compound is stabilized by a hydrogen bonding interaction (–O...HO, 3.19 Å) between the position-5 of nitrogen atoms and residue Asp509, which is also in accordance with the docking results. Compared with the MD and the original docking studies, we find that due to slight rotations of the ligand, one H-bond is formed between the tetrahydronaphthalen-2-ol ring of carbonyl oxygen and residue Glu513 breaks after the MD simulations. Furthermore, one new H-bond (–F...HN, 3.42 Å) is formed solidly between the *F* of the trifluoromethyl core of the ligand and residue Arg534, indicating that the trifluoromethyl plays a crucial role in the stabilization of the ligand at the active site. Based on these interactions analyzed from the MD simulations, it is summarized that the protein pocket is constituted by a series of residues including especially the Ser554, Phe543, Glu513, Phe516, Ser512, Arg534, Ser505, Tyr537, Met541, Val542 and Asp509. In this binding mode, three crucial amino acids, *i.e.* Asp509, Tyr537 and Arg534, form H-bond interactions in determining the active conformation of the molecule.

An analysis of the entire MD simulation trajectories for compound **197** in complex with the receptor 3D model further reveals that the overall conformation of the protein backbone undergoes only minimal global conformational changes upon the complex formation, while a rearrangement of the side chains of several residues lining the receptor binding site is required for the ligand binding.

## 4 Discussion

### 4.1 Validation of contours using docking results

To validate the reliability of generated contours, the results of docking analysis were used for comparison. The CoMSIA steric contour maps show two large positive green isopleths that favor bulky substitutions around the tetrahydronaphthalene ring, which is consistent with the docking results that bulky substitutions at this position will have an unfavorable steric interaction with Glu513, Asp509, Thr550 and Phe543. It seems that the pocket is deep enough and apparently bulkier groups at the positions of the tetrahydronaphthalene ring would be able to optimize the interactions with the residues of the pocket.

The active site of TRPV1 is mostly hydrophobic which is in well agreement with the CoMSIA model. Two yellow regions are

situated close to the tetrahydronaphthalene ring and the trifluoromethyl group. In addition, both positions point toward the hydrophobic pocket of the binding site, interacting with the Phe543, Met547, Tyr511 regions and the Leu512, Tyr537 regions, respectively. The groups at both positions should fit into the hydrophobic binding site to enhance the bioactivity. This is also consistent with the CoMSIA steric contour map, in which the sterically more bulky substituent is preferred for an increase in antagonistic activity. At the same time, hydrophilic amino acid residues Thr550 and Ser512 around the oxazole group indicate that compounds with hydrophilic groups in this position may reduce the activity. These conclusions are in well agreement with our previous CoMSIA hydrophilic contour maps.

The CoMSIA H-bond acceptor contour map correlates with the involvement of H-bond interactions with the target. As shown in Fig. 8, the oxazole ring of carbonyl oxygen atoms participates in hydrogen bonding with Tyr537 (–O...HO, 3.10 Å), which correlates well with the magenta contour located in this position (the HB acceptor favored region) in previous CoMSIA H-bond acceptor contour maps (Fig. 4D). In addition, one H-bond (–O...HO, 2.82 Å) is formed between the tetrahydronaphthalen-2-ol ring of carbonyl oxygen and the residue Glu513, and another hydrogen bond is found between the Asp509 (–O...HN, 3.19 Å) and the position-5 of nitrogen atoms. Both the bonds correlate well with the red contour located around (representing the H-bond donor favored region) in previous CoMSIA hydrogen bond acceptor contour maps. This is also in agreement with the electrostatic contour maps which show a red contour near the same group that favors electronegative groups (Fig. 4B).

Consequently, the docking results reveal that the steric, electrostatic, hydrophobic and H-bond acceptor interactions in the binding pocket of TRPV1 correlate well with the contour maps, indicating that the QSAR model is reasonable and can offer constructive suggestions for the further design of potent TRPV1 antagonists.

### 4.2 Comparison with the binding modes of TRPV1 antagonists

To gain more insights into the binding mode of the TRPV1 antagonists and clarify whether they bind to TRPV1 through a similar mechanism to other TRPV1 antagonists, a comparison of all known binding pockets of the TRPV1 homology model with the docking model obtained from our results was conducted presently. Despite the fact that the X-ray crystal structure of human TRPV1 has not been identified, several research groups have used homology modeling of TRPV1 and docked some antagonists to predict the binding mode. To the best of our knowledge, in recent publications, molecular models for 15 kinds of TRPV1 antagonists including thiourea, 2-alkyloxy, evodiamine, heterocyclic, propanamide and capsaicin were proposed based on homology modeling of the transmembrane domain of TRPV1, and docking studies of these ligands were bound to the putative binding site. The 12 representative molecules are summarized in Fig. 11. Therefore, it is assumed that a comparison of the binding mechanism with other TRPV1 inhibitors should be of great help to comprehensively explore



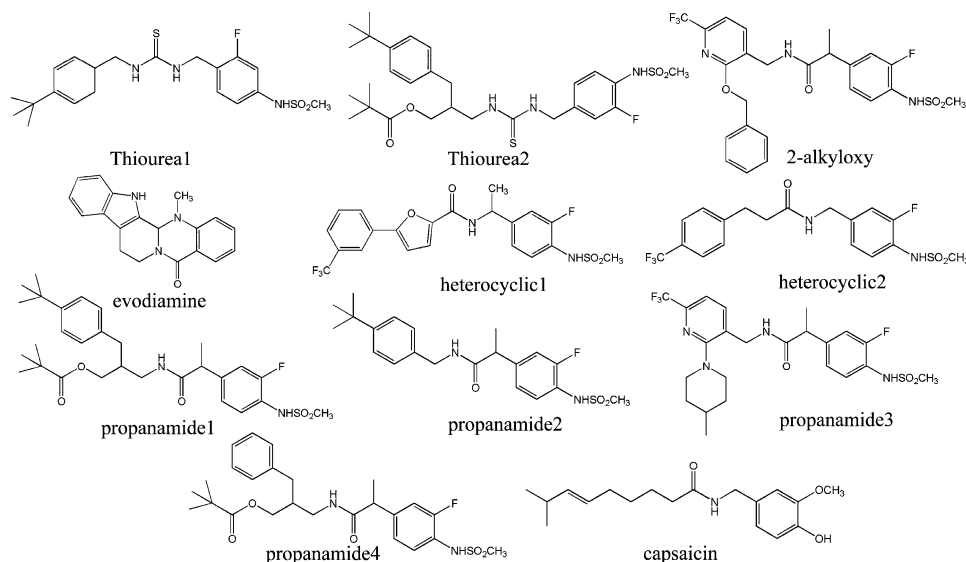


Fig. 11 Eleven TRPV1 antagonists binding in the TRPV1 homology model from the recent literature.

and understand their structural features for interactions with TRPV1, and thus was performed in the present work.

Reminiscent of the 15 antagonist binding in the TRPV1 cavity, according to our research, basically follows three different binding modes, as summarized in Fig. 12. In fact, the three binding modes share common characteristics: (1) that prominent interaction is H-bonds and all the compounds have been proven to occupy approximately the same region in the deep groove between the TM3–TM4 domains of the TRPV1 protein. Three models emphasize that residues in TM3–TM4 regions are critical for ligand recognition and assume that others in the N and C terminus play a modulatory role. These models are preliminary and require additional biochemical and structural information for validation and refinement of precise ligand orientation in the binding site.<sup>61</sup> Our model is dependent on further studies that evaluate the desensitization properties, and the paddle formation by the TM3–TM4 region of TRPV1, which will be important for validating our assumptions. (2) In this

configuration, the residue Tyr511 that seems to be involved in the three bindings are located at the beginning of TM3 exposed to the cytosolic environment, but close to the membrane interface. This demonstrates that residue Tyr511, a tyrosine, is critical for vanilloid sensitivity. Our observations agree with this result.

The difference between the three binding modes lies in the orientation of the active conformation of the small molecules, and the H-bond network including the residues referred to and the pattern of interaction (Table 4). As shown in Fig. 12A, as the long axes of the phenyl and  $\text{NHSO}_2\text{CH}_3$  system of this thiourea antagonist appear to rotate by almost  $120^\circ$  relative to the present docking results, the 3-fluoro atom engages in a hydrogen bond to the amide proton of Gly563 (1.77 Å) and the sulfur atom of the thiourea group makes a hydrogen bond to the side chain of Gln560 (2.26 Å). In addition, the hydrophobic 4-*tert*-butylbenzyl group contacts closely with the hydrophobic parts of Phe517, Trp549, Leu553, and Arg557.<sup>64</sup> Fig. 12B illustrates the two binding modes, in which due to the rotations of the

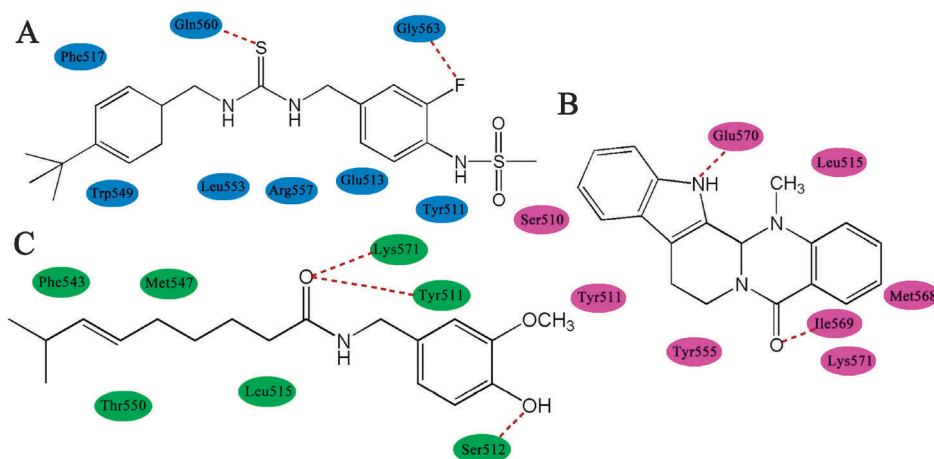


Fig. 12 Three binding modes in the TRPV1 pocket from the homology model. (A) Thiourea.<sup>64</sup> (B) Evodiamine.<sup>65</sup> (C) Capsaicin.<sup>46</sup>



**Table 4** Hydrogen bond analysis from the binding sites with thiourea, evodiamine, capsaicin and compound **197** of our docking results

Ligand	Thiourea	Evodiamine	Capsaicin	<b>197</b>
Residues	Gln560 Gly563	Glu570 Ile569	Lys571 Tyr511 Ser512	Glu513 Tyr537 Asp509

ligand, the evodiamine forms two H-bonds between the formyl carbonyl oxygen and the side chains of Lys571, and between the indole nitrogen and the side chains of Ile569, respectively.<sup>65</sup> In addition, the evodiamine ring pointing toward Tyr555 forms an aromatic  $\pi$ - $\pi$  interaction. Although the evodiamine derivative also has a rotation angle compared with our results, our results show that evodiamine docked into the binding pocket in a similar orientation. Fig. 12C illustrates the predicted binding mode of capsaicin with TRPV1. By comparing the binding mode of capsaicin with our docking results, the capsaicin derivative is found also possessing a rotation angle. As shown in this figure, the vanillyl moiety forms hydrophobic and  $\pi$ - $\pi$  stacking interactions with Tyr511 and H-bonding with Ser512; the carbonyl group forms two H-bonds with Tyr511 and Lys571.<sup>46</sup> Although the hydrophobic nonenyl tail orients toward the upper hydrophobic region of the binding site, it does not fully occupy the hydrophobic region of the two shallow hydrophobic areas composed of Phe543 and Met547 because it is linear and too short to reach both areas.<sup>46</sup>

The above comparisons between the results indicate the similarities and differences which may provide a new strategy for a ligand based technique exploiting the fragment features in combination with the structure-guided identification of novel scaffolds.

## 5 Conclusion

In the present study, we constructed a statistically reliable 3D-QSAR model using antagonistic activity of 236 carboxamides derivatives as potent TRPV1 antagonists. The models of random ( $Q^2 = 0.522$ ,  $R_{ncv}^2 = 0.935$ ,  $R_{pred}^2 = 0.839$ ) and SOM ( $Q^2 = 0.521$ ,  $R_{ncv}^2 = 0.929$ ,  $R_{pred}^2 = 0.829$ ) dataset splitting methods exhibit almost the same proper statistical results, indicating that models based on these two division methods both generate significant statistical results and possess strong predictive power for the test sets, proving their reasonability for building the models. The analysis of the contour maps for the optimal CoMSIA model provides crucial clues that can be used as a useful tool for the rational drug design process. The good consistency between the 3D-QSAR, the docking and MD modeling results indicates the robustness of the 3D-QSAR model. Overall, our findings are summarized as follows:

(1) Bulky groups around the tetrahydronaphthalene ring benefit the activity of the dataset. Electrostatic groups located at the pyridine moiety of the core structure increase the antagonistic activity. Hydrophobic groups near the oxazole moiety of the R2 position and naphthyl moiety of the position-5 are helpful to activity, which is validated by contour maps and docking results.

H-bond acceptor groups around the R2 position are favorable for binding activity.

(2) Based on the interactions analyzed from the MD simulations, it is summarized that the key amino acids impacting the receptor-ligand interactions are Asp509, Tyr537 and Arg534.

All the correlation of the results obtained from the *in silico* studies leads to a better understanding of the structural requirements for improving the activity of the antagonists and is helpful for the design of new and potent TRPV1 antagonists.

## Conflict of interest

The author(s) confirm that this article content has no conflicts of interest.

## Acknowledgements

Thanks for the financial support given by the National Natural Science Foundation of China (Grant No. 11201049). This research is also supported by the high-performance computing platform of Northwest A&F University, with financial support given by the National Natural Science Foundation of China (Grant No. 10801025 and 30973590), the National High Technology Research and Development Program ("863") of China (No. 2009AA02Z205).

## References

- 1 R. L. McLeod, C. C. Correll, Y. Jia and J. C. Anthes, *Lung*, 2008, **186**, 59–65.
- 2 M. J. Caterina, M. A. Schumacher, M. Tominaga, T. A. Rosen, J. D. Levine and D. Julius, *Nature*, 1997, **389**, 816–824.
- 3 N. R. Gavva, *Trends Pharmacol. Sci.*, 2008, **29**, 550–557.
- 4 A. Szallasi and P. M. Blumberg, *Life Sci.*, 1990, **47**, 1399–1408.
- 5 M. Leonelli, D. O. Martins and L. R. Britto, *Cell. Mol. Neurobiol.*, 2013, **33**, 379–392.
- 6 A. Forsby, K. Norman, J. E. Andaloussi-Lilja, J. Lundqvist, V. Walczak, R. Curren, K. Martin and N. Tierney, *Toxicol. Sci.*, 2012, **129**, 325–331.
- 7 D. van Veghel, J. Cleynhens, L. V. Pearce, I. A. DeAndrea-Lazarus, P. M. Blumberg, K. Van Laere, A. Verbruggen and G. Bormans, *ACS Chem. Neurosci.*, 2013, **4**, 624–634.
- 8 J.-W. Min, W.-H. Liu, X.-H. He and B.-W. Peng, *Toxicol.*, 2013, **71**, 66–75.
- 9 J. Klafke, M. Da Silva, G. Trevisan, M. Rossato, C. Da Silva, G. Guerra, J. Villarinho, F. Rigo, G. Dalmolin and M. Gomez, *Neuroscience*, 2012, **222**, 136–146.
- 10 B. Nilius, J. Prenen and G. Owsianik, *J. Physiol.*, 2011, **589**, 1543–1549.
- 11 M. S. Kim, H. Ryu, D. W. Kang, S.-H. Cho, S. Seo, Y. S. Park, M.-Y. Kim, E. J. Kwak, Y. S. Kim and R. S. Bhondwe, *J. Med. Chem.*, 2012, **55**, 8392–8408.
- 12 K. Valenzano and Q. Sun, *Curr. Med. Chem.*, 2004, **11**, 3185–3202.
- 13 M. J. Gunthorpe and B. A. Chizh, *Drug Discovery Today*, 2009, **14**, 56–67.

- 14 H. K. Rami, M. Thompson, G. Stemp, S. Fell, J. C. Jerman, A. J. Stevens, D. Smart, B. Sargent, D. Sanderson and A. D. Randall, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 3287–3291.
- 15 J.-D. Brederson, P. R. Kym and A. Szallasi, *Eur. J. Pharmacol.*, 2013, **716**, 61–76.
- 16 R. El Kouhen, C. S. Surowy, B. R. Bianchi, T. R. Neelands, H. A. McDonald, W. Niforatos, A. Gomtsyan, C.-H. Lee, P. Honore and J. P. Sullivan, *J. Pharmacol. Exp. Ther.*, 2005, **314**, 400–409.
- 17 P. Honore, C. T. Wismer, J. Mikusa, C. Z. Zhu, C. Zhong, D. M. Gauvin, A. Gomtsyan, R. El Kouhen, C.-H. Lee and K. Marsh, *J. Pharmacol. Exp. Ther.*, 2005, **314**, 410–421.
- 18 S. Mandadi and B. D. Roufogalis, *Curr. Neuropharmacol.*, 2008, **6**, 21–38.
- 19 J. E. Krause, B. Chenard and D. N. Cortright, *Curr. Opin. Invest. Drugs*, 2005, **6**, 48–57.
- 20 A. Szallasi, D. N. Cortright, C. A. Blum and S. R. Eid, *Nat. Rev. Drug Discovery*, 2007, **6**, 357–372.
- 21 E. Takahashi, N. Hirano, T. Nagahara, S. Yoshikawa, S. Momen, H. Yokokawa and R. Hayashi, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 3154–3156.
- 22 P. Li, J. Chen, J. Wang, W. Zhou, X. Wang, B. Li, W. Tao, W. Wang, Y. Wang and L. Yang, *J. Ethnopharmacol.*, 2014, **151**, 93–107.
- 23 W. Tao, X. Xu, X. Wang, B. Li, Y. Wang, Y. Li and L. Yang, *J. Ethnopharmacol.*, 2013, **145**, 1–10.
- 24 W. Zhou and Y. Wang, *J. Ethnopharmacol.*, 2014, **151**, 66–77.
- 25 X. Xu, W. Zhang, C. Huang, Y. Li, H. Yu, Y. Wang, J. Duan and Y. Ling, *Int. J. Mol. Sci.*, 2012, **13**, 6964–6982.
- 26 H. Yu, J. Chen, X. Xu, Y. Li, H. Zhao, Y. Fang, X. Li, W. Zhou, W. Wang and Y. Wang, *PLoS One*, 2012, **7**.
- 27 I. Drizin, A. Gomtsyan, E. K. Bayburt, R. G. Schmidt, G. Z. Zheng, R. J. Perner, S. DiDomenico, J. R. Koenig, S. C. Turner and T. K. Jinkerson, *Bioorg. Med. Chem.*, 2006, **14**, 4740–4749.
- 28 B. S. Brown, R. Keddy, G. Z. Zheng, R. G. Schmidt, J. R. Koenig, H. A. McDonald, B. R. Bianchi, P. Honore, M. F. Jarvis and C. S. Surowy, *Bioorg. Med. Chem.*, 2008, **16**, 8516–8525.
- 29 R. J. Perner, J. R. Koenig, S. DiDomenico, A. Gomtsyan, R. G. Schmidt, C.-H. Lee, M. C. Hsu, H. A. McDonald, D. M. Gauvin and S. Joshi, *Bioorg. Med. Chem.*, 2010, **18**, 4821–4829.
- 30 A. Gomtsyan, E. K. Bayburt, R. Keddy, S. C. Turner, T. K. Jinkerson, S. DiDomenico, R. J. Perner, J. R. Koenig, I. Drizin and H. A. McDonald, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 3894–3899.
- 31 B. S. Brown, R. Keddy, R. J. Perner, S. DiDomenico, J. R. Koenig, T. K. Jinkerson, S. M. Hannick, H. A. McDonald, B. R. Bianchi and P. Honore, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 3291–3294.
- 32 R. J. Perner, S. DiDomenico, J. R. Koenig, A. Gomtsyan, E. K. Bayburt, R. G. Schmidt, I. Drizin, G. Z. Zheng, S. C. Turner and T. Jinkerson, *J. Med. Chem.*, 2007, **50**, 3651–3660.
- 33 A. Gomtsyan, E. K. Bayburt, R. G. Schmidt, G. Z. Zheng, R. J. Perner, S. DiDomenico, J. R. Koenig, S. Turner, T. Jinkerson and I. Drizin, *J. Med. Chem.*, 2005, **48**, 744–752.
- 34 A. Gomtsyan, E. K. Bayburt, R. G. Schmidt, C. S. Surowy, P. Honore, K. C. Marsh, S. M. Hannick, H. A. McDonald, J. M. Wetter and J. P. Sullivan, *J. Med. Chem.*, 2008, **51**, 392–395.
- 35 R. G. Schmidt, E. K. Bayburt, S. P. Latshaw, J. R. Koenig, J. F. Daanen, H. A. McDonald, B. R. Bianchi, C. Zhong, S. Joshi and P. Honore, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1338–1341.
- 36 A. D. Rodgers, H. Zhu, D. Fourches, I. Rusyn and A. Tropsha, *Chem. Res. Toxicol.*, 2010, **23**, 724–732.
- 37 T. M. Martin, P. Harten, D. M. Young, E. N. Muratov, A. Golbraikh, H. Zhu and A. Tropsha, *J. Chem. Inf. Model.*, 2012, **52**, 2570–2578.
- 38 A. Golbraikh, M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee and A. Tropsha, *J. Comput.-Aided Mol. Des.*, 2003, **17**, 241–253.
- 39 J. Wang, Y. Li, Y. Yang, S. Zhang and L. Yang, *Curr. Med. Chem.*, 2013, **20**, 2032–2042.
- 40 Y. Yang, J. Wang, Y. Li, W. Xiao, Z. Wang, J. Zhang, W. Gao, S. Zhang and L. Yang, *Soft Matter*, 2013, **9**, 11054–11077.
- 41 T. Kohonen, *Self-organized formation of topologically correct feature maps, Neurocomputing foundations of research*, MIT Press, 1988, pp. 509–521.
- 42 A. Yan, X. Nie, K. Wang and M. Wang, *Eur. J. Med. Chem.*, 2013, **61**, 73–83.
- 43 T. Kohonen, *Biol. Cybern.*, 1982, **43**, 59–69.
- 44 J. Gasteiger and M. Marsili, *Tetrahedron*, 1980, **36**, 3219–3228.
- 45 M. Clark, R. D. Cramer and N. Van Opdenbosch, *J. Comput. Chem.*, 1989, **10**, 982–1012.
- 46 J. H. Lee, Y. Lee, H. Ryu, D. W. Kang, J. Lee, J. Lazar, L. V. Pearce, V. A. Pavlyukovets, P. M. Blumberg and S. Choi, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 317–327.
- 47 G. Kothandan, C. G. Gadhe, T. Madhavan and S. J. Cho, *Chem. Biol. Drug Des.*, 2011, **78**, 161–174.
- 48 L. Zhao, Y. Liu, S. Hu and H. Zhang, *J. Mol. Model.*, 2012, **18**, 3669–3694.
- 49 C. Muñoz, F. Adasme, J. H. Alzate-Morales, A. Vergara-Jaque, T. Kniess and J. Caballero, *J. Mol. Graphics Modell.*, 2012, **32**, 39–48.
- 50 P. Malla, R. Kumar and M. Kumar, *Chem. Biol. Drug Des.*, 2013, **82**, 71–80.
- 51 CCDC Software Ltd (2005) Cambridge CB2 1EZ, UK. <http://www.ccdc.cam.ac.uk>.
- 52 *GOLD User Guide & Tutorials*, CCDC Software Ltd, 2005, [http://www.ccdc.cam.ac.uk/support/documentation/gold/4\\_0/gold/toc.html](http://www.ccdc.cam.ac.uk/support/documentation/gold/4_0/gold/toc.html).
- 53 H. J. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
- 54 A. A. Skelton, Y. R. Maharaj and M. E. Soliman, *Cell. Mol. Bioeng.*, 2014, **7**, 45–57.
- 55 S. Jo, T. Kim and W. Im, *PLoS One*, 2007, **2**, e880.
- 56 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 57 J.-H. Lin, A. L. Perryman, J. R. Schames and J. A. McCammon, *J. Am. Chem. Soc.*, 2002, **124**, 5632–5633.

- 58 G. Fernandez-Ballester and A. Ferrer-Montiel, *J. Membr. Biol.*, 2008, **223**, 161–172.
- 59 V. Y. Moiseenkova-Bell, L. A. Stanciu, I. I. Serysheva, B. J. Tobe and T. G. Wensel, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7451–7455.
- 60 M. Z. Chou, T. Mtui, Y.-D. Gao, M. Kohler and R. E. Middleton, *Biochemistry*, 2004, **43**, 2501–2511.
- 61 N. R. Gavva, L. Klionsky, Y. Qu, L. Shi, R. Tamir, S. Edenson, T. Zhang, V. N. Viswanadhan, A. Toth and L. V. Pearce, *J. Biol. Chem.*, 2004, **279**, 20283–20295.
- 62 H. Ryu, M. K. Jin, S. Y. Kim, H. K. Choi, S. U. Kang, D. W. Kang, J. Lee, L. V. Pearce, V. A. Pavlyukovets, M. A. Morgan, R. Tran, A. Toth, D. J. Lundberg and P. M. Blumberg, *J. Med. Chem.*, 2008, **51**, 57–67.
- 63 S.-E. Jordt and D. Julius, *Cell*, 2002, **108**, 421–430.
- 64 J. Lee, M.-K. Jin, S.-U. Kang, S. Y. Kim, J. Lee, M. Shin, J. Hwang, S. Cho, Y.-S. Choi and H.-K. Choi, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 4143–4150.
- 65 Z. Wang, L. Sun, H. Yu, Y. Zhang, W. Gong, H. Jin, L. Zhang and H. Liang, *Int. J. Mol. Sci.*, 2012, **13**, 8958–8969.