



Institute of Materia Medica, Chinese Academy of Medical Sciences  
Chinese Pharmaceutical Association

Acta Pharmaceutica Sinica B

[www.elsevier.com/locate/apsb](http://www.elsevier.com/locate/apsb)  
[www.sciencedirect.com](http://www.sciencedirect.com)



ORIGINAL ARTICLE

# 3D-QSAR and docking studies of arylmethanamine-based DPP IV inhibitors

Chaoyi Jiang, Shuang Han, Tiegang Chen, Jianzhong Chen\*

ZJU-ENS Joint Laboratory of Medicinal Chemistry, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

Received 14 February 2012; revised 30 May 2012; accepted 12 June 2012

## KEY WORDS

Diabetes;  
DPP IV;  
Inhibitor;  
QSAR;  
Docking

**Abstract** The present work was focused on the study of the three-dimensional (3D) structural requirements for the highly potent bioactivity of dipeptidyl peptidase (DPP) IV's inhibitor. At first, molecular dynamic and mechanic (MD/MM) simulations were performed to research the conformations of the potent DPP IV's inhibitor 5-(aminomethyl)-6-(2,4-dichlorophenyl)-2-(3,5-dimethoxy-phenyl)pyrimidin-4-amine. Using the MD/MM-determined molecular conformers as templates, the 3D quantitative structure activity relationship (QSAR) studies were carried out based on a set of arylmethanamine DPP IV inhibitors with the comparative molecular field analysis (CoMFA) approach. The best 3D-QSAR model was constructed with good statistic values of  $r_{cv}^2$  and  $R^2$  using PLS analyses (CoMFA:  $r_{cv}^2=0.660$ ,  $R^2=0.953$ ). The generated 3D-QSAR model was proved to be reliable by internal and external validations. Docking studies were further performed to analyze the interaction mode between the highly potent or low potent arylmethanamine derivatives and DPP IV. Our flexible docking results also confirmed the possible bioactive conformation obtained from the 3D-QSAR model, of arylmethanamine-based DPP IV inhibitors. The 3D-QSAR model may provide information of pharmacophoric features for further design and optimization of new scaffold compounds with high inhibitory activity to DPP IV.

© 2012 Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association. Production and hosting by Elsevier B.V. All rights reserved.

\*Corresponding author. Tel.: +86 571 8820 8659.

E-mail address: [chjz@zju.edu.cn](mailto:chjz@zju.edu.cn) (Jianzhong Chen).

2211-3835 © 2012 Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association. Production and hosting by Elsevier B.V. All rights reserved.

Peer review under the responsibility of Institute of Materia Medica, Chinese Academy of Medical Sciences and Chinese Pharmaceutical Association.

<http://dx.doi.org/10.1016/j.apsb.2012.06.007>



Production and hosting by Elsevier

## 1. Introduction

Diabetes mellitus is a principle and growing public health problem in the world. It was estimated that the number of people suffering from diabetes would increase to 300 million by 2025<sup>1</sup>. Currently, the oral monotherapy or combination therapeutics with other drugs are main methods to aid in the control of diabetes by using all kinds of available antidiabetic agents in the clinical anti-diabetes therapy<sup>2</sup>. However, these agents are considered to be induction of severe adverse effects and chronic complications<sup>3,4</sup>. Such a situation promotes people to develop novel antidiabetic drug with good potency and low toxicity.

Dipeptidyl peptidase IV (DPP IV) is a serine protease that cleaves an endogenous oligopeptide at the second residue, which is the typical amino acid of alanine or proline, from its N-terminus<sup>5</sup>. The peptidic hormone glucagon-like peptide 1 (GLP-1) is one kind of such oligopeptides degraded by DPP IV protease. GLP-1 plays an important role in the regulation of insulin release to control the level of blood sugar in human body<sup>6,7</sup>. Several studies have demonstrated that the inhibition of DPP IV can increase the amount of circulating GLP-1 to improve the secretion of insulin in the body<sup>8,9</sup>. Therefore, it has been regarded as an attractive and promising target to develop novel drug for the treatment of type 2 diabetes. So far, a couple of identified DPP IV inhibitors, such as sitagliptin and saxagliptin, have been approved to be used clinically as anti-diabetic drugs by FDA<sup>10-13</sup>. However, there is still a need for more potent, selective and safer DPP IV inhibitor, which does not have the inspecificity and side effect possessed by the presently available inhibitors<sup>14</sup>, because of worldwide problem of type 2 diabetes. Therefore, it is reasonable for researcher to put a lot of efforts to study the new DPP IV ligands for the development of the novel anti-diabetes drug.

In the past few years, a large number of compounds were synthesized and evaluated as DPP IV inhibitors, including peptidomimetic series and non-peptidomimetic series<sup>2</sup>. So far, the SAR and QSAR studies were mainly focused on the peptidomimetic series. Actually, a couple of ligand-based models have been constructed to clarify the structure-activity relationship of peptidomimetic inhibitors of DPP IV. For example, Zeng et al.<sup>15</sup> developed 3D-QSAR models on a series of fluoropyrrolidine amides to investigate the interaction between DPP IV's inhibitors and their receptor by using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) methods. 3D-QSAR method was also applied for building a predictive model based on a series of triazolopiperazine amides as DPP IV inhibitors<sup>16</sup>. In these 3D-QSAR studies, the lowest energy conformation of compound was directly searched to be a template for structural alignment. Although this methodology offers an appropriate molecular superimposition, it is still a doubt whether the conformation with global minimum energy was an exactly bioactive one. Moreover, these generated 3D-QSAR models were not applicable for ligands with different binding modes. On the other hand, QSAR studies have not been reported to the non-peptidomimetic series of DPP IV's inhibitors, which can be an attractive point guiding us to design novel DPP IV inhibitors.

In the present study, a series of arylmethylamino derivatives, which were developed to be potent DPP IV inhibitors<sup>17</sup>, were used to generate a 3D-QSAR model for non-peptidomimetic

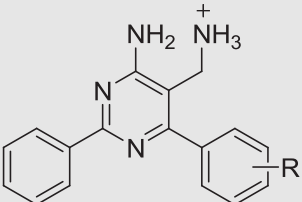
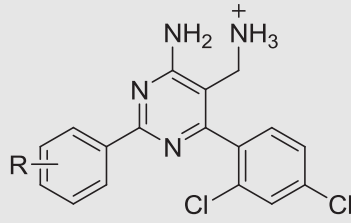
compounds using the CoMFA approach. For such a purpose, 33 compounds were selected from literature<sup>17</sup> and divided into a training database and a test database. Among them, a highly potent arylmethylamino compound 5-(aminomethyl)-6-(2,4-dichlorophenyl)-2-(3,5-dimethoxyphenyl)pyrimidin-4-amine (**27**) was first chosen to do MD/MM modeling for its possible local minimum conformations. Each of MD/MM simulated conformation was applied to be a starting conformational template for structural alignments of the compounds in both training and test databases. Partial Least-Squares (PLS) analyses were then performed to get cross-validated  $r_{cv}^2$  and no-cross-validated  $R^2$  describing statistical correlation between inhibitors' bioactivities and their CoMFA-calculated electrostatic and steric fields based on their conformations. The bioactive conformation of ligand was derived by improving the squared correlation coefficient  $r_{cv}^2$  on the basis of Partial Least-Squares analyses<sup>18</sup>. A good 3D-QSAR model was produced with  $r_{cv}^2$  bigger than 0.6. The generated 3D-QSAR model was also evaluated by its prediction of bioactivities of compounds in both training and test databases. Our result elucidated structural requirements for enhancing ligand's bioactivity to inhibit DPP IV. Furthermore, the flexibly docking simulations were performed to reveal the interaction mode between inhibitors and DPP IV. The docking results demonstrated that the binding conformation of arylmethylamino derivative was congruent with the one obtained from the CoMFA studies. Our research indicated that the established QSAR model could be reliable in identifying potential lead compounds with DPP IV inhibitory activity. The CoMFA contoured trends for the steric and electrostatic fields can be used as guides for the generation of a consistent pharmacophore model employed for *in silico* search new chemical scaffold of DPP IV's inhibitor. The corresponding 3D-QSAR model provides a means for predicting the bioactivity of untested compound.

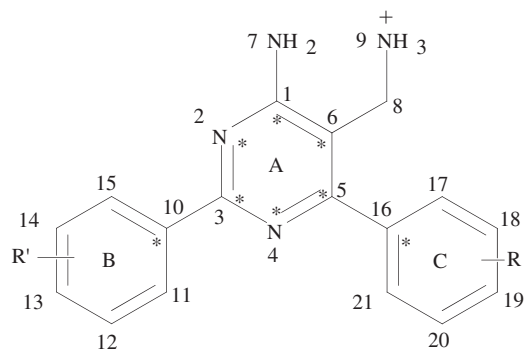
## 2. Methods

### 2.1. Data preparation

The quality of the biological data under investigation as well as the structural diversity of the data set is important foundations for successful 3D-QSAR studies. In the present work, total 33 DPP IV inhibitors were collected from literature published by Peters et al.<sup>17</sup> to do the 3D-QSAR studies using the method of comparative molecular field analysis. Table 1 lists their chemical structures and corresponding inhibitory IC<sub>50</sub> data. Fig. 1 illustrates the common structure of pyrimidine core of the selected arylmethylamino derivatives with different substituent groups, including methyl, methoxyl, halogenic, trifloromethyl groups, on the phenyl rings B or C. These substituent groups have different hydrophobic and electrostatic properties. Furthermore, all of the compounds have significant variation of their IC<sub>50</sub> values ranging seven orders of magnitude, which would be useful to generate a good 3D-QSAR model. Their inhibitory values were converted into the corresponding  $pIC_{50}$  (-logIC<sub>50</sub>) values to be used as dependent variables in the CoMFA study. The compounds of the training and test sets were carefully selected in order to ensure appropriate property coverage on the entire range of  $pIC_{50}$  values. As listed in Table 1, the data set was divided into a training set of 27 compounds, enclosing compounds **1-27**, for the 3D-QSAR model construction and a test

**Table 1** Molecular structures and bioactivity  $IC_{50}$  values of arylmethylamine-based DPP IV inhibitors in the training set (compounds **1–27**) and test set (compounds **28–33**) used to construct the 3D-QSAR model.

Compound	R	$IC_{50}$ ( $\mu$ M)	Compound	R	$IC_{50}$ ( $\mu$ M)
 <b>1-13,28-30</b>			 <b>14-27,31-33</b>		
<b>1</b>	H	42	<b>18</b>	<i>m</i> -Cl	0.24
<b>2</b>	<i>o</i> -Me	1.5	<b>19</b>	<i>m</i> -OMe	0.34
<b>3</b>	<i>o</i> -Cl	2.5	<b>20</b>	<i>m</i> -CF <sub>3</sub>	0.13
<b>4</b>	<i>o</i> -OMe	1.5	<b>21</b>	<i>p</i> -Cl	0.053
<b>5</b>	<i>o</i> -F	14	<b>22</b>	<i>p</i> -OMe	0.10
<b>6</b>	<i>m</i> -Me	20	<b>23</b>	<i>p</i> -F	0.0002
<b>7</b>	<i>m</i> -Cl	31	<b>24</b>	<i>p</i> -CF <sub>3</sub>	0.18
<b>8</b>	<i>m</i> -OMe	80		–	10
<b>9</b>	<i>m</i> -F	40	<b>25</b>		0.003
<b>10</b>	<i>m</i> -CF <sub>3</sub>	170	<b>26</b>		0.0001
<b>11</b>	<i>p</i> -OMe	47	<b>27</b>		14
<b>12</b>	<i>p</i> -F	18	<b>28</b>	<i>o</i> -CF <sub>3</sub>	1.0
<b>13</b>	<i>p</i> -CF <sub>3</sub>	1.1	<b>29</b>	<i>p</i> -Me	1.4
<b>14</b>	H	0.01	<b>30</b>	<i>m</i> -Me	0.0009
<b>15</b>	<i>o</i> -Me	1.75	<b>31</b>	<i>m</i> -F	0.0002
<b>16</b>	<i>o</i> -OMe	0.35	<b>32</b>	<i>p</i> -Me	0.09
<b>17</b>	<i>o</i> -F	0.047	<b>33</b>		

**Figure 1** Common structure of arylmethylamino derivatives. The atoms for alignment are marked with an asterisk.

set of 6 compounds, including compounds **28–33**, for external model validation based on the principle of structural diversity. Since the 8-amino group was reported to have direct interaction with two acid residues, E205 and E206, of DPP IV, the amino group of all compounds in both training and test sets was applied to be in protonation state for our generation of 3D-QSAR model. The highly potent compound **27** was chosen as the template molecule for the later structural alignment in our CoMFA analysis.

## 2.2. Computer molecular modeling

Molecular modeling was carried out using Tripos Sybyl molecular modeling package<sup>19,20</sup>. MD/MM simulations were first

performed to sample the compound **27**'s conformations at the local minima of the energy landscape with the Tripos force field and Gasteiger–Hückel atomic charges. Initially, the Sybyl/Sketch module was applied for building a starting conformation of compound **27**. The energy optimization process was then completed to get its local stable conformer with a distance-dependent dielectric function and a convergence criterion of 0.001 kcal/mol/Å. MD simulations were performed to mimic the movement of molecular fragments at the temperature of 2,000 K with the time length of 300 ps and the step of 1 fs. Total 300 snapshots were collected at a rate of 1 snapshot per ps. Each snapshot was finally optimized with the procedures and parameters as mentioned in the previous publication<sup>21</sup> to get the conformation at the local minima energy.

Based on each of compound **27**'s conformations obtained from MD/MM simulations, conformations of other compounds in both training set and test set were constructed by modifying corresponding substitution group R listed in Table 1 using the molecular fragments library provided within Sybyl × 1.3. These obtained conformers of each compound were then energy-minimized using the same parameters as compound **27**'s minimization.

### 2.3. Structural alignment

Since there is a critical requirement of structure alignment in CoMFA analysis to generate a 3D-QSAR model, the alignment rule remains to be a crucial and controversial process in 3D-QSAR analyses. Because of the closely structural similarity of compounds in our CoMFA studies, all of the compounds in both training set and test set were assumed to interact with DPP IV through the same binding motifs. Each MD/MM simulated conformation of compound **27** was regarded as a structural template for the molecular superimposition because of its highly inhibitory activity to DPP IV. On the basis of the common structure and the reported SAR analyses of the arylmethylamino derivatives as DPP IV inhibitors<sup>17</sup>, it was chosen of the heavy atoms on the aromatic rings A, B, and C, respectively, illustrated in Fig. 1, for the structural alignment. It is reasonable to assume that inhibitors do not necessarily binding to DPP IV in their global minimum energy conformations because some degree of bond rotation may be required to adapt electrostatic and H-bonding distances that would have their good interaction. Therefore, MD/MM simulated conformation of all compounds were regarded as a starting points and single bond rotations were allowed to all of compounds in our studies for a good structural alignment. On the other hand, it is important to note that the permitted pharmacophoric conformations of different compounds must be restricted to those that can be obtained upon bind within reasonable energy limits. It was acceptable of a 10 kcal/mol cutoff difference between the local minimum and the aligned conformational energy of each compound upon single bond rotation for superimposition<sup>21</sup>.

### 2.4. CoMFA Partial Least-Squares analysis

After structural alignment of molecules in the training set, the 3D-QSAR models were generated using the CoMFA program of Sybylx1.3<sup>20</sup>. CoMFA analyses were performed for each combination of steric and electrostatic fields calculated according to the

molecular conformation. All of the molecules in the training set were placed in a rectangular grid box with the size of extension of 4 Å to all compounds in the X, Y, and Z directions of Cartesian coordinate system. The steric and electrostatic (AM1 charge) field energies were computed using a probe atom of  $sp^3$  hybridized carbon atom with +1 charge and a distance-dependent dielectric constant in all of regularly grid space (2.0 Å). Both steric and electrostatic cutoffs were set to the default 30 kcal/mol. PLS analysis was then carried out to generate quantitative relationship between bioactive values ( $pIC_{50}$ ) and steric and electrostatic fields using default parameters in Sybyl/CoMFA module. The minimum column filtering was set as 2.0 kcal/mol to improve the signal-to-noise ratio. The optimum number of components was determined through Leave-One-Out procedure. The final model (non-cross-validated conventional analysis) was developed to produce the no validated correlation coefficient  $R^2$  using the optimum number of component obtained from the model with the highest cross-validated  $r_{cv}^2$ .

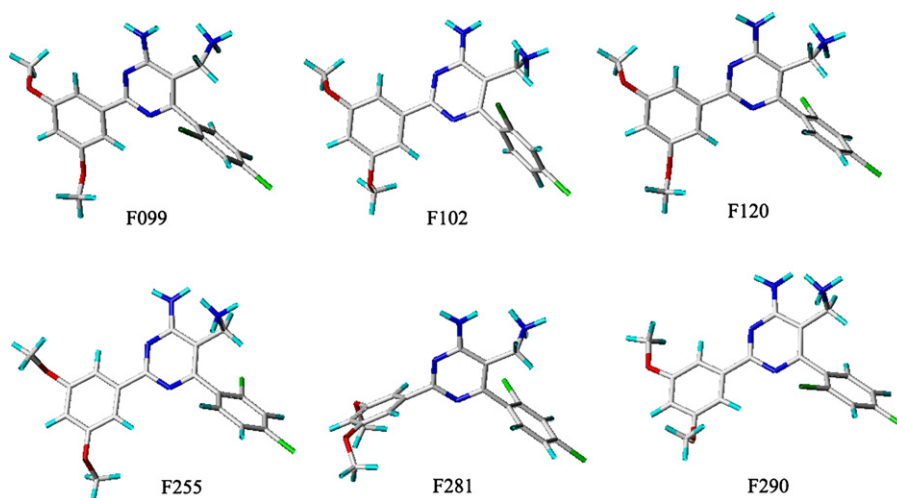
### 2.5. Docking studies of interaction between arylmethylamino derivative and DPP IV

Flexible docking calculation was further carried out to simulate the interaction mode between arylmethylamino derivative and DPP IV. For such a purpose, two compounds, the highly potent compound **23** and the low potent compound **10**, were chosen to do docking simulation using the FlexiDock module of Sybyl 6.9<sup>19</sup>. The X-ray co-crystal structure of the compound **27**-DPP IV complex was obtained from the Protein Data Bank (entry 1RWQ). The arylmethylamino derivative **23** or **10** was individually put into the binding cavity of DPP IV to replace the bound ligand **27** in the co-crystal structure of complex. A binding pocket was then defined to enclose all of the amino acid residues within 4 Å radius sphere centered by the docked compound in the initial compound-DPP IV complex. In the continued docking process, all of the single bonds of residues' side chains within the defined DPP IV binding pocket were regarded as rotatable or flexible bonds, and the docked compound **23** or **10** was allowed to rotate on all single bonds and move flexibly within the tentative binding pocket. The atomic charges were recalculated using the Kollman all-atom approach for DPP IV and the Gasteiger–Hückel approach for the docked compound **23** or **10**. The H-bonding sites were marked for suitable atoms, which are able to act as H-bond donors or acceptors, of both ligand and residues within the defined DPP-IV active site region. According to Sybyl/Flexidock method<sup>19</sup>, the binding interaction energy was calculated to include the terms of van der Waals, electrostatics, and torsion energy defined in the Tripos force field. The structure optimization was performed for 50,000-generations using a genetic algorithm, and the 20 best-scoring ligand–protein complexes were kept for further analyses. The structure with lowest energy was selected as a final model to analyze the composition of key amino acid residues of DPP IV involved in the interaction with its inhibitor.

## 3. Results and discussion

### 3.1. Molecular modeling

Fig. 2 illustrates six representative conformations of compound **27** obtained from MD/MM simulations. MD/MM

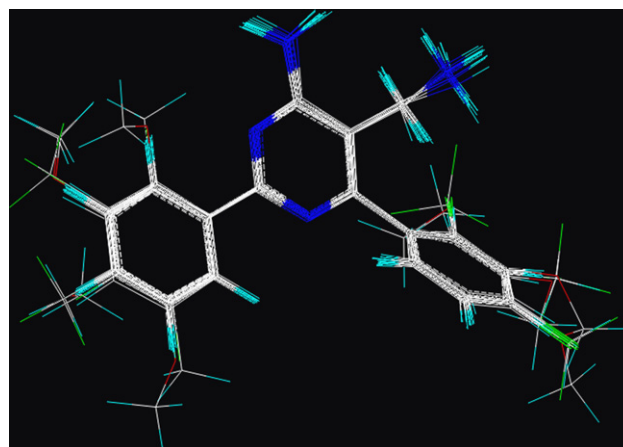


**Figure 2** Molecular graphic representation of six favored conformations of compound **27** on the basis of the energy minimization of structures occurring along the molecular dynamics trajectory.

simulations were carried out to search for the possible local low-energy conformations of compound **27**. Dynamic motions were simulated at temperature of 2,000 K to look for the probability of inducing conformational transitions past any possible high-energy barrier. The dynamic simulations were performed with time steps of 1 fs for 300 ps, and the data were recorded at 1 ps intervals to get a total of 300 frames. MM energy minimizations were then carried out for each of these 300 conformers. All 300 conformers were then superimposed with each other. The conformations with root mean square (RMS) less than 1.0 Å were classified into the same structural family. This operation resulted in a convergence of these 300 conformers into six families represented in Fig. 2 by six local minimum-energy conformations which were sampled during the MD simulation and then followed by energy minimization. The energy difference of these six conformers is less than 2 kcal/mol. All of these conformations are potentially present in the nature.

### 3.2. CoMFA analyses of arylmethylamine-based derivatives binding to DPP IV

On the basis of MD/MM simulated conformations of compound **27**, the CoMFA method was employed to build a good 3D-QSAR model for arylmethylamino derivatives based on their observed inhibitory activities ( $IC_{50}$  values) for the DPP IV receptor. Although the co-crystal structure of compound **27** with DPP IV has been resolved to determine its bioactive conformation, it would be needed of verifying that this bioactive conformation would be only one for the generation of a good 3D-QSAR model with the conformation-based CoMFA studies. For such a purpose, the MD/MM simulated conformers F099, F102, F120, F255, F281, and F290 of compound **27** were individually selected to be an initial template for the structural alignment of all compounds in the training set. In the alignment schemes, several variations were considered by superimposing the common or similar pharmacophore features, which were detailed in the section of Methods. Although it has been reported of the co-crystal structure of **27** with DPP IV, it would helpful to induce structural requirement for DPP IV inhibitor by comparing



**Figure 3** F120-based structural alignment of compounds in the training set and test set.

different 3D-QSAR model based on each of **27**'s potential conformation obtained from MD/MM simulation. Such protocol would also make us to characterize **27**'s bioactive conformation interacting with DPP IV by its conformation analyses and to lead more clues for the drug design of novel DPP IV's inhibitor. The best results involved not only a reasonably good overlap of the relevant pharmacophoric groups but also statistically significant 3D-QSAR models from CoMFA analyses. For example, Fig. 3 shows the 3D view of corresponding structural alignment of compounds in the training set of compounds using conformation F120 as a structural template.

In all instances, cross-validated PLS analyses were run to determine the optimal number of components in the model and to evaluate the robustness of the model based on how well it predicts data. Table 2 lists the cross-validation  $r_{cv}^2$  values and corresponding optimum component numbers in all CoMFA analyses on the basis of the six conformers, respectively, of compound **27**. There is a generally accepted criterion for CoMFA statistical validity of  $0.7 \geq r_{cv}^2 \geq 0.6$  and optimal component from 3 to 5 for the PLS method. Our results indicated that the PLS analysis only based on F120 exceeded

**Table 2** Cross-validated analyses of the CoMFA models based on six conformers of compound **27** as structural templates.

No.	Template conformation	$r_{cv}^2$	Optimal component
1	F099	0.757	6
2	F102	0.572	5
3	F120	0.660	4
4	F255	0.703	4
5	F281	0.665	1
6	F290	0.814	4

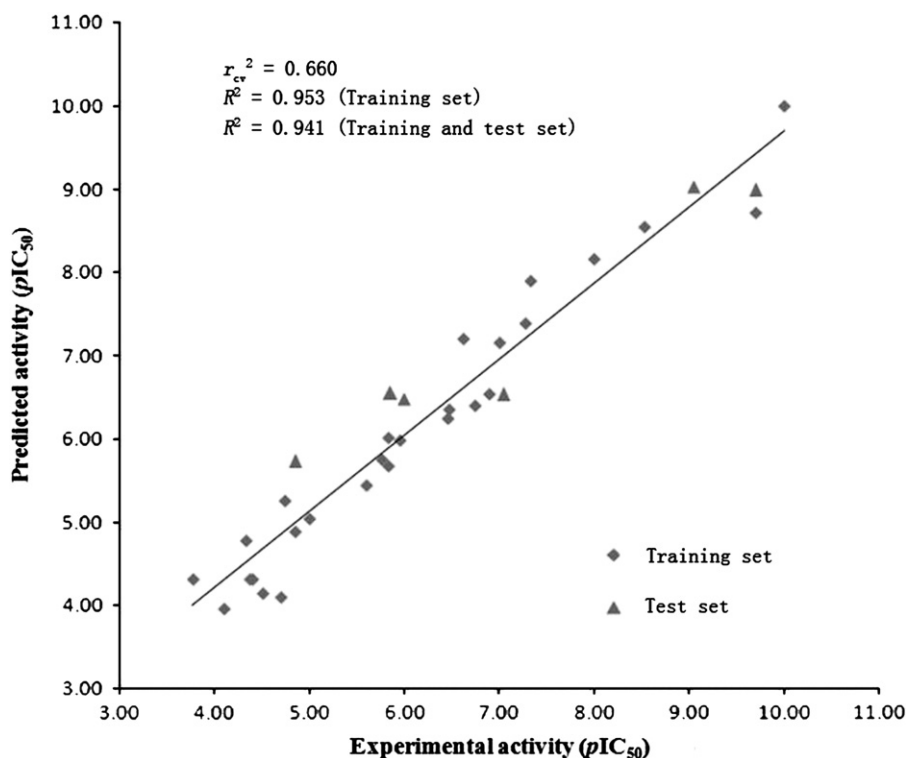
this criterion, thus the predictions obtained with this model was reliable. The non-cross-validated PLS analysis gave a good correlation coefficient  $R^2$  of 0.953 with the training set of compounds. In the generated CoMFA model, the contributions of steric and electrostatic fields were 0.819 and 0.181, respectively. Furthermore, other six compounds in test set were used to validate the CoMFA model. The non-cross-validated PLS analysis was also performed including all of compounds in both training set and test set. It conferred a good correlation coefficient  $R^2$  of 0.941 with a standard error of estimate (SEE) of 0.390.  $F$ -value stands for the degree of statistical confidence on the developed models and the model has good value of 111.18. At the same time, the generation of consistent statistical models depends on the proper selection of both training and test sets in terms of structural diversity and property values distribution. The values of  $pIC_{50}$  from training set and test set span approximately seven orders of magnitude and are acceptably distributed across the  $pIC_{50}$  range values. The derivative of arylmethylamine-based DPP IV inhibitors has substantial structural diversity. The ring B and ring C have variously different substituent groups. From the original data set of 33 inhibitors, 27 compounds (**1–27**, Table 1) were selected as members of the training set for model construction, and the other 6 compounds (**28–33**, Table 1) as members of the test set for external model validation, in the ratio of about 4:1 (approximately 20%). Table 3 lists the CoMFA-calculated  $pIC_{50}$  values of total 33 compounds, showing a good linear relationship. On the basis of the appropriate representation of chemical diversity and distribution of property values (Table 1), the training and test set meet the requirements for the purpose of internal and external model validation. Fig. 4 illustrates the good correlation between experimental and CoMFA model-predicted bioactivities of compounds in both training set and test set. The high  $R^2$  reflect robustness of the models, devoid of any chance factors. Therefore, it is a good 3D-QSAR model generated on the basis of conformer F120 of compound **27**, and it can be used to predict the bioactivity of unknown compound. F120 could be the preferred conformation of compound **27**, which is assumed to be a bioactive conformation of arylmethylamine compounds to interact with DPP IV. As discussed lately, F120 is actually consistent to the one of compound **27** in the crystal structure of complex with DPP IV. Moreover, when the conformer F120 with highest  $r_{cv}^2$  was generated and validated, it can be used as criteria to solve the problem of other compounds' spatial orientations to make sure that the results of 3D-QSAR are reliable and credible.

Fig. 5 illustrates CoMFA-generated contour maps of both steric (A) and electrostatic (B) fields around compound **27**. The contour maps reveal that essential regions in the steric and electrostatic fields around ligand might affect the binding of

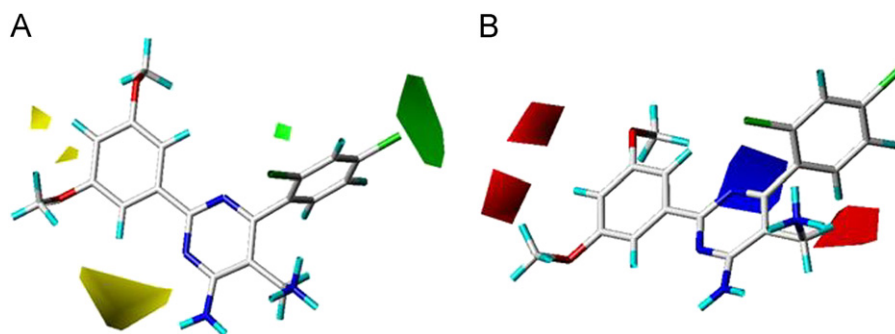
**Table 3** Experimental (obsd) and CoMFA-predicted (pred)  $pIC_{50}$  values of molecules in both training set and test set.

DPP IV CoMFA model			
$R^2$	0.953		
Standard error of estimate	0.390		
$F$	111.180		
Compound	$pIC_{50}$ (obsd)	$pIC_{50}$ (pred)	Residual
<b>1</b>	4.38	4.32	-0.06
<b>2</b>	5.82	6.03	0.20
<b>3</b>	5.60	5.45	-0.15
<b>4</b>	5.82	5.69	-0.14
<b>5</b>	4.85	4.89	0.04
<b>6</b>	4.70	4.11	-0.59
<b>7</b>	4.51	4.16	-0.35
<b>8</b>	4.10	3.96	-0.14
<b>9</b>	4.40	4.32	-0.08
<b>10</b>	3.77	4.32	0.55
<b>11</b>	4.33	4.78	0.45
<b>12</b>	4.74	5.27	0.52
<b>13</b>	5.96	5.99	0.03
<b>14</b>	8.00	8.16	0.16
<b>15</b>	5.76	5.76	0.00
<b>16</b>	6.46	6.25	-0.21
<b>17</b>	7.33	7.91	0.58
<b>18</b>	6.62	7.21	0.59
<b>19</b>	6.47	6.37	-0.10
<b>20</b>	6.89	6.54	-0.34
<b>21</b>	7.28	7.39	0.12
<b>22</b>	7.00	7.16	0.16
<b>23</b>	9.70	8.73	-0.97
<b>24</b>	6.74	6.40	-0.34
<b>25</b>	5.00	5.05	0.05
<b>26</b>	8.52	8.55	0.03
<b>27</b>	10.00	10.01	0.01
<b>28</b>	4.85	5.73	0.88
<b>29</b>	6.00	6.48	0.48
<b>30</b>	5.85	6.55	0.69
<b>31</b>	9.05	9.03	-0.02
<b>32</b>	9.70	8.99	-0.71
<b>33</b>	7.05	6.54	-0.51

arylmethylamine derivatives in the active pocket of DPP IV. As shown in Fig. 5A, the green polyhedrons characterize the regions where a steric bulky group would increase bioactivity, whereas yellow contours depict regions where steric substituent would not be tolerated. Favored and disfavored levels of steric field are



**Figure 4** Plots of predicted versus experimental  $pIC_{50}$  values of training and test set for the CoMFA model.



**Figure 5** CoMFA contour maps for the (A) steric field: Green/ yellow contours indicate regions where steric bulky groups increase/ decrease activity. Favored and disfavored levels of these displayed fields are fixed at 75% and 15%, respectively. (B) electrostatic field: Red/blue contours indicate regions where negative charge increase/ decrease activity. Favored and disfavored levels of these displayed fields are fixed at 93% and 7%, respectively.

fixed at 75% and 25%, respectively. On the other hand, the blue or red contours in Fig. 5B indicate the regions of favored positively or negatively charged, respectively, function group would increase inhibitory activity of ligand for DPP IV. Favored and disfavored levels of electrostatic field are fixed at 93% and 7%, correspondingly. The abundance of the red–blue region to the yellow–green region points out that electrostatic properties have more significant impact on DPP IV binding potency than the steric properties. This result is consistent with the statistical data of the contributions of steric and electrostatic fields mentioned above.

As shown in Fig. 5A, the yellow polyhedra areas are primarily present around *para*-site in 3-phenyl ring (ring B) attaching to the pyrimidine ring A, indicating that these areas would prefer a small branch rather than a bulky substituent. It is cleared that compound **23** is more bioactive than compound

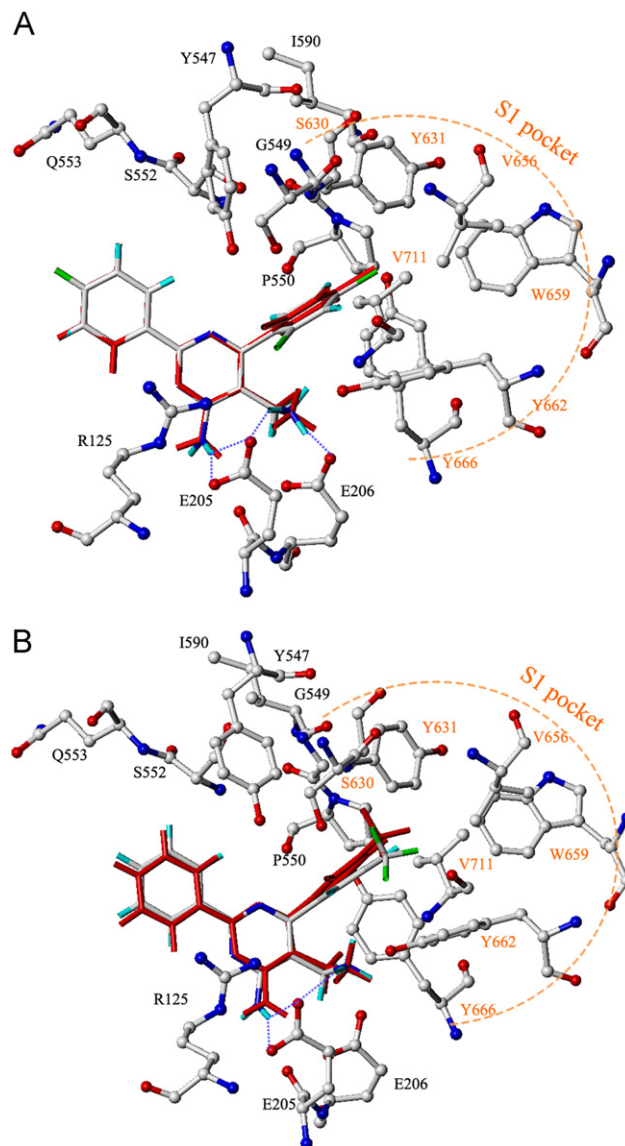
**21** because of the different size of the fluoro- and chloro-groups. We can also make the similar conclusion by comparing the bioactivities of compound **23** and compounds **22** and **24**. Fig. 5A also illustrates a large yellow contour around the *ortho* position of ring B. Comparing compounds **14** and **17**, introducing a bulky substituent such as methyl and methoxyl (compounds **15** and **16**) would decrease bioactivity. At mean time, this contour also indicates inducing a bulky group might lose a ligand's bioactivity, of 1-amino and 6-methylamino groups. On the contrary, a small green region around the *para*-position of the phenyl ring C indicates that a steric bulky group at the corresponding position would enhance the inhibitory activity. For example, compound **14** has the  $pIC_{50}$  value of 8.00 with the steric-favored 2,4-dichlorine group on the phenyl ring C but compound **1** has the  $pIC_{50}$  value of 4.38 without any substituent on the ring C. It also demonstrates

that this region would prefer a bulky group with the fact that compound **29** ( $pIC_{50}=6.00$ ) with a *p*-methyl group on the 5-phenyl ring shows better inhibitory activity for DPP IV than compound **12** ( $pIC_{50}=4.74$ ) with a *p*-F moiety. It is indeed consistent with the results obtained from the previous studies that the bulky group occupied a hydrophobic S1 pocket of DPP IV with a large volume<sup>22</sup>.

In the electrostatic contour maps shown in Fig. 5B, the red contour is next to the substituent around *meta* and *ortho* position of the ring B. It implies that a negative charge within these areas of molecules would increase the binding affinity. This could explain why compound **23** with a fluorine group on the 3-phenyl ring B is more potent than compound **33** with a *p*-CH<sub>3</sub> group on the ring B. On the other hand, a red contour near ring C indicates that introduction of the relatively negative groups would be favorable to the molecular bioactivity. This hypothesis is clearly demonstrated by a comparison of bioactivities of compounds **15–24**. Among these compounds, the ones having a halogen atom or -CF<sub>3</sub> group in ring C usually show higher bioactivity than ones with methyl or methoxyl group. In addition, it is obvious that a large blue contour is around on the pyrimidine ring A. We can deduce the similar results according to the later docking simulation of inhibitor-DPP IV interaction with the electron-deficient moieties, such as amino or methylamino group, can make good interaction with the important residues, such as Glu205 and Glu206. These results can make sure that the conformations of other compounds which were generated from compound **27** are reasonable and close to the bioactive conformations. Meanwhile, the good linear relationship and the CoMFA-generated contour maps also show that the 3D-QSAR methods can reflect the good relationship of bioactive data between known and unknown effectively.

### 3.3. Docking analyses of arylmethylamine-based derivatives binding to DPP IV

Fig. 6A illustrates the docking simulated 3D structural model of compound **23**-DPP IV complex by using Sybyl/FlexiDock module<sup>19</sup>. In this figure, the amino acid residues in DPP IV were displayed with a “ball-sticks” mode and the docked compound was exhibited with a “sticks-only” style. In the mean time, the important residues were labeled to have direct interaction with inhibitor. The docking results indicate that the best score is  $-4,493$  kcal/mol for the DPP IV-compound **23** interaction, incorporating the sum of the van der Waals, electrostatics, and torsional energy terms within the Tripos force field. Fig. 6A represents the most stable binding conformation of compound **23** well docked into the active pocket of DPP IV. The simulated model of compound **23**-DPP IV complex indicated the hydrophobic and hydrophilic interactions present between ligand and enzyme. The -NH<sub>2</sub> group of compound **23** can form two H-bonds to the oxygen atoms of carboxyl group of Glu205 with the bond lengths and angles of 2.252 Å, 160.87° and 2.924 Å, 146.43°, respectively. Another two H-bond interactions are also formed between protons of the methylamino group and oxygen atoms of the carboxyl functional group in both amino acids of Glu205 and Glu206 with the bond lengths and angles of 2.045 Å, 151.37° and 2.026 Å, 122.99°, respectively. These H-bond interactions reveal that the residues Glu205 and Glu206 play critical roles



**Figure 6** The docking simulated interaction modes of high potent compound **23** (A) and low potent compound **10** (B), respectively. In the mean time, the CoMFA-generated conformations (red colored) of these compounds were superimposed onto their docked conformations individually. The superimposition indicates the molecular conformations in 3D-QSAR model are consistent with the docking ones.

in the binding of arylmethylamine derivatives to the DPP IV enzyme. Furthermore, our docking results also indicate that the cation- $\pi$  interaction is present between Arg125 and the compound's pyrimidine ring A with a distance of 3.823 Å. The 2, 4-dichlorophenyl group insert into the hydrophobic S1 pocket, which is composed by the residues Ser630, Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711. Although the co-crystal structure of compound **27**-DPP IV has been published<sup>17</sup>, the interaction mode of compound **27**-DPP IV was also simulated by using Sybyl/FlexiDock to evaluate the docking results of arylmethylamine-based derivatives binding to DPP IV. The interaction score is  $-4,699$  kcal/mol for the interaction of compound **27**-DPP IV. The simulated interaction mode of compound **27**-DPP IV



is quite similar to their X-ray co-crystal structure. The inhibitory potency of compound **23** is close to **27**, since the docking results show that compound **23** fit very well to the binding pocket of DPP IV like compound **27**.

Fig. 6B illustrates the docking simulated 3D structural model of low potent compound **10** binding to DPP IV. The FlexiDock results indicate that the best score is -4,319 kcal/mol for the DPP IV-compound **10** interaction, incorporating the sum of the van der Waals, electrostatics, and torsional energy terms in the Tripos force field. Although the interaction mode between compound **10** and DPP IV is like to the docking results of compound **27**-DPP IV, the former interaction energy is obviously weaker than latter one from the score result of FlexiDock. Furthermore, there is also a noticeable difference between compound **27** or **23** and **10** in the S1 pocket position. The torsion of ring C of compound **27** is  $104.8^\circ$  at C6-C5-C16-C21 while the torsion of ring C of compound **10** is  $88.5^\circ$  (compound **23** is  $103.9^\circ$ ). Moreover, the ring C of compound **23** with a *p*-Cl substituent group on ring C can stretch into the S1 pocket, which encloses couples of aromatic ring-contained residues, like Tyr631, Trp659, Tyr662, and Tyr666, deeper than the ring C of compound **10**. Because *p*-Cl is a hydrophobic group, it can improve the hydrophobic interaction between the ring C and the S1 pocket. By contrast, there is no hydrophobic group in the *p*-position of ring C in compound **10** which may decrease the hydrophobic interaction between the ring C and S1 pocket. This result can explain why the potency of compound **10** is lower than compound **27**, even compound **23** reasonably. For the inhibition of DPP IV, it is vital of whether the relevant substituting group on the ring C can fit the S1 pocket very well. In the mean time, since a hydrophilic 12-CF<sub>3</sub> group is attached to the ring C of compound **10**, its phenyl ring C is pushed out from pocket S1 to have less hydrophobic interaction with lipophilic residues in the pocket S1 of DPP IV than compound **27** or **23** in their docking simulated complexes with DPP IV. It may influence the inhibitory activity of compound **10** to DPP IV. Comparing the docking simulated 3D structures of compounds **10** and **23**, the conformation of substituting group at S1 pocket should be paid more attention when to design new DPP IV inhibitors. From these results, it is concluded that the consistency between the QSAR and binding conformations is high. It is reasonable that the conformations of the modified compounds are advisable.

#### 3.4. Comparison of the generated 3D-QSAR model and the docking model

In order to examine the reliability of 3D-QSAR model obtained from the previous CoMFA studies of the DPP IV inhibitors, we superimposed the conformation of compound **23** from 3D-QSAR model (red one) onto its docked one (colored one) as shown in the Fig. 6A. The results indicate that the 3D-QSAR model is congruent with the results of docking simulation very well. As it is observed from the superimposition, there is a general correlation between the conformation of compound used in the generated QSAR model and its best docking-simulated binding conformation. For example, RMSD values between these two conformations were calculated to be  $0.246 \text{ \AA}$  for compound **23** when superimposing its conformation in the generated 3D-QSAR model and

docking-simulated one by all heavy atoms. However, it should be noted that some small differences have been observed to orientations of the phenyl ring C in compound **23** between these conformations. The torsion angle C6-C5-C16-C21 is  $84.7^\circ$  in QSAR conformation, while it is  $103.9^\circ$  in the binding conformation. It is reasonable to hypothesize that a ligand may not have its global minimum energy conformation to bind on its receptor protein. It is because the lowest energy ligand-protein complex would be formed under some degree of bond rotations required to form suitable electrostatic and H-bonding interaction between protein and ligand.

In Fig. 5A, the CoMFA yellow contours appearing around methylamino and amino groups indicate that a steric bulky substituent might have negative effect on the bioactivity of the ligands. As shown in Fig. 6A, our docking results illustrate a narrow space around the acidic residues of Glu205 and Glu206, which can form tight H-bond interactions with methylamino and amino groups as discussed in the previous section. Therefore, there is unnecessary modification for both methylamino and amino groups in arylmethylamine-based DPP IV inhibitors. Similarly, the green contour appearing around the phenyl ring C approaches the lipophilic region near S1 pocket, which plays an important role in suppressing the ability of DPP IV when it is occupied by hydrophobic group. It suggests that a bulkier group will increase the biological activity of an inhibitor. By comparing Figs. 5B and 6A, it can also be concluded that molecular electrostatic contribution in the generated 3D-QSAR model is consistent to the electrostatic interaction mode from the docking simulation of the complex between arylmethylamine derivatives and DPP IV. For example, blue colored contour maps are oriented towards the electronegative groups of DPP IV's acidic residues like Glu205 and Glu206. The blue contour around the methylamino and amino groups covers the space enclosing the residues of Glu205 and Glu206 which are acidic amino acids. The potency of inhibitor can be increased while electropositive groups are introduced, since the H-bond or salt bridge interaction can be formed between the -NH<sub>2</sub> groups and the carboxyl group of Glu205 and Glu206. Moreover, the blue contour around the ring A points to the residue of Tyr547 mainly because of the -OH group of Tyr547 which is a polar group. Equally no modification was made in the pyrimidine ring which is an alkalinescent group, so it is reasonable to present a large blue contour below the ring A. Furthermore, there are red contours present around the ring B. Considering the surrounding polar amino acids, like Ser552 and Gln553, it is reasonable that the same effect would be obtained whether electropositive or electronegative groups are added.

Therefore, the performed docking studies not only highlight the consistency between the QSAR and binding conformations but also provide the knowledge about crucial interactions with the active site of DPP IV for enhanced activity. In this study, we preferred to generate energy minimized conformation(s), get bioactive conformations by 3D-QSAR method and validate the method through the docking model. Because the conformations generated by QSAR methods are consistent with bound conformations. It is convinced that through QSAR and docking methods, we can find some relationship between energy minimized conformation and bound conformation. At the same time, these results also validated that the bioactive conformation can be generated by QSAR studies.

When QSAR and docking studies are integrated, it is useful for us to explain the difference between the conformation and inhibition, which can lead us to further design new compounds reported in our patent application 23, with high potentially inhibitory activity for DPP IV.

#### 4. Conclusions

The global purpose of this work is to produce a 3D-QSAR model for the prediction of arylmethylamine analogs. We report here the first establishment of 3D-QSAR model and docking model on a series of arylmethylamine-based non-peptidomimetic DPP IV inhibitor. The constructed models revealed statistical significance and good predictive abilities by using CoMFA and docking studies. Moreover, on the basis of the CoMFA model contour maps, significant regions for steric, electrostatic, hydrophobic, H-bond interactions were identified to enhance bioactivity. At the same time, CoMFA studies were combined with the docking results based on the known X-ray crystal structure of DPP IV to evaluate the reliability of the generated 3D-QSAR. The correlation of the results obtained from docking and QSAR studies lead to better understanding of the structural requirements for enhanced activity. The obtained results can be used as a guideline to design and predict new potent DPP IV inhibitors, which could be an effective way to find novel leads for the development of antidiabetic drug.

#### Acknowledgments

The project is supported by grants from National Key Tech Project for Major Creation of New Drugs (2009ZX09501-003 to Jianzhong Chen) and Science Foundation of Chinese University (2009QNA7024 to Jianzhong Chen)

#### References

- Zimmet P, Alberti KG, Shaw J. Global and societal implications of the diabetes epidemic. *Nature* 2001;**414**:782–7.
- Havale SH, Pal M. Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. *Bioorg Med Chem* 2009;**17**:1783–802.
- Barnett AH. Thiazolidinediones and cardiovascular outcomes. *Brit J Diabetes Vasc Dis* 2008;**8**:45–9.
- Mizuno CS, Chittiboyina AG, Kurtz TW, Pershadsingh HA, Avery MA. Type 2 diabetes and oral antihyperglycemic drugs. *Curr Med Chem* 2008;**15**:61–74.
- Sebokova E, Christ AD, Boehringer M, Mizrahi J. Dipeptidyl peptidase IV inhibitors: the next generation of new promising therapies for the management of type 2 diabetes. *Curr Top Med Chem* 2007;**7**:547–55.
- Pospisilik JA, Stafford SG, Demuth HU, Brownsey R, Parkhouse W, Finegood DT, et al. Long-term treatment with the dipeptidyl peptidase IV inhibitor P32/98 causes sustained improvements in glucose tolerance, insulin sensitivity, hyperinsulinemia, and  $\beta$ -cell glucose responsiveness in VDF (*fa/fa*) Zucker rats. *Diabetes* 2002;**51**:943–50.
- Ahrén B, Holst JJ, Mårtensson H, Balkan B. Improved glucose tolerance and insulin secretion by inhibition of dipeptidyl peptidase IV in mice. *Eur J Pharmacol* 2000;**404**:239–45.
- Zander M, Madsbad S, Madsen JL, Holst JJ. Effect of 6-week course of glucagon-like peptide 1 on glycaemic control, insulin sensitivity, and  $\beta$ -cell function in type 2 diabetes: a parallel-group study. *Lancet* 2002;**359**:824–30.
- Pospisilik JA, Hinke SA, Pederson RA, Hoffmann T, Rosche F, Schlenzig D, et al. Metabolism of glucagon by dipeptidyl peptidase IV (CD26). *Regul Pept* 2001;**96**:133–41.
- Patel M, Rybczynski PJ. Treatment of non-insulin-dependent diabetes mellitus. *Expert Opin Investig Drugs* 2003;**12**:623–33.
- Knudsen LB. Glucagon-like peptide-1: the basis of a new class of treatment for type 2 diabetes. *J Med Chem* 2004;**47**:4128–34.
- Wb AE. Dipeptidyl peptidase IV inhibitors for the treatment of diabetes. *J Med Chem* 2004;**47**:4135–41.
- Kim D, Kowalchick JE, Edmondson SD, Mastracchio A, Xu J, Eiermann GJ, et al. Triazolopiperazine-amides as dipeptidyl peptidase IV inhibitors: close analogs of JANUVIA (sitagliptin phosphate). *Bioorg Med Chem Lett* 2007;**17**:3373–7.
- Defronzo RA, Okerson T, Viswanathan P, Guan X, Holcombe JH, MacConnell L. Effects of exenatide versus sitagliptin on postprandial glucose, insulin and glucagon secretion, gastric emptying, and caloric intake: a randomized, cross-over study. *Curr Med Res Opin* 2008;**24**:2943–52.
- Zeng J, Liu G, Tang Y, Jiang H. 3D-QSAR studies on fluoropyrrolidine amides as dipeptidyl peptidase IV inhibitors by CoMFA and CoMSIA. *J Mol Model* 2007;**13**:993–1000.
- Saqib U, Siddiqi MI. 3D-QSAR studies on triazolopiperazine amide inhibitors of dipeptidyl peptidase-IV as anti-diabetic agents. *SAR QSAR Environ Res* 2009;**20**:519–35.
- Peters JU, Weber S, Kritter S, Weiss P, Wallier A, Boehringer M, et al. Aminomethylpyrimidines as novel DPP-IV inhibitors: a 105-fold activity increase by optimization of aromatic substituents. *Bioorg Med Chem Lett* 2004;**14**:1491–3.
- Dunn WJ, Wold S, Edlund U, Hellberg S, Gasteiger J. Multivariate structure–activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: the PLS method. *Quant Struct Act Rel* 1984;**3**:131–7.
- Sybyl molecular modeling software packages. Tripos Associates Inc., St. Louis, MO63144, 2002.
- Sybyl molecular modeling software packages. Tripos Associates Inc., St. Louis, MO63144, 2011.
- Chen JZ, Han XW, Liu Q, Makriyannis A, Wang JM, Xie XQ. 3D-QSAR studies of arylpyrazole antagonists of cannabinoid receptor subtypes CB1 and CB2. A combined NMR and CoMFA approach. *J Med Chem* 2006;**49**:625–36.
- Kuhn B, Hennig M, Mattei P. Molecular recognition of ligands in dipeptidyl peptidase IV. *Curr Top Med Chem* 2007;**7**:609–19.
- Chen J, Jiang C, Wu H, He Q, Yang B. Preparation and usage of novel DPP IV's inhibitors. China Patent Appl. 20121008855.X, 2012.