

Molecular docking and 3D-QSAR studies on β -phenylalanine derivatives as dipeptidyl peptidase IV inhibitors

Yan-Ke Jiang

Received: 31 October 2009 / Accepted: 5 December 2009 / Published online: 13 January 2010
© Springer-Verlag 2010

Abstract Three-dimensional quantitative structure-activity relationship (3D-QSAR) and molecular docking studies were carried out to explore the binding of 73 inhibitors to dipeptidyl peptidase IV (DPP-IV), and to construct highly predictive 3D-QSAR models using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA). The negative logarithm of IC_{50} (pIC_{50}) was used as the biological activity in the 3D-QSAR study. The CoMFA model was developed by steric and electrostatic field methods, and leave-one-out cross-validated partial least squares analysis yielded a cross-validated value (r_{cv}^2) of 0.759. Three CoMSIA models developed by different combinations of steric, electrostatic, hydrophobic and hydrogen-bond fields yielded significant r_{cv}^2 values of 0.750, 0.708 and 0.694, respectively. The CoMFA and CoMSIA models were validated by a structurally diversified test set of 18 compounds. All of the test compounds were predicted accurately using these models. The mean and standard deviation of prediction errors were within 0.33 and 0.26 for all models. Analysis of CoMFA and CoMSIA contour maps helped identify the structural requirements of inhibitors, with implications for the design of the next generation of DPP-IV inhibitors for the treatment of type 2 diabetes.

Keywords β -phenylalanine · DPP-IV inhibitors · CoMFA · CoMSIA · Molecular docking · 3D-QSAR

Introduction

Dipeptidyl peptidase IV (DPP-IV, CD26, EC 3.4.14.5) is a type II membrane glycoprotein that is expressed in a variety of cell types, such as T cells, B cells, natural killer cells and monocytes [1]. DPP-IV modulates the biological activity of several peptide hormones, chemokines and neuropeptides by cleaving amino-terminal dipeptides from polypeptides with either L-proline or L-alanine at the penultimate position [2].

Glucagon-like peptide-1 (GLP-1, 7–36) stimulates insulin secretion in a glucose-dependent manner, increasing β -cell mass and function, and suppressing glucagon secretion and appetite [3, 4]. Chronic infusion of GLP-1 to patients with type 2 diabetes results in significant decreases in both blood glucose and hemoglobin A_{1c} levels. However, in plasma, GLP-1 is degraded rapidly by DPP-IV with a half-life of less than 2 min after intravenous administration as it has L-proline in the penultimate position [5]. Inhibition of DPP-IV would increase the half-life of GLP-1 and the levels of endogenous intact circulating GLP-1, prolonging the beneficial effects of this incretin hormone. Animal studies and initial clinical trials have shown that DPP-IV inhibitors are capable of significantly lowering fasting and postprandial glucose concentration, as well as reducing HbA_{1c} levels, with good tolerability and minimal risk of hypoglycemia [6–12]. Thus, there is great interest in developing DPP-IV inhibitors as a new class of drugs for the treatment of type 2 diabetes [13–17].

Several classes of DPP-IV inhibitors have been reported, including cyanopyrrolidines, noncyano-pyrrolidines and -thiazolines, xanthine-derived compounds and β -phenylalanine derivatives, all of which can be divided into two different kinds: reversible inhibitors and irreversible inhibitors [18–26]. Much attention has been paid to

Y.-K. Jiang (✉)
Research Center of Medical Chemistry and Chemical Biology,
Chongqing Technology and Business University,
Chongqing 400067, People's Republic of China
e-mail: yanke2046@yahoo.com.cn

reversible inhibitors in recent years. The class of β -phenylalanine derivatives was designed and synthesized as reversible inhibitors of DPP-IV to improve the chemical stability of the nitrile-containing compounds. The phenyl ring of β -phenylalanine was replaced by a 2-,4-,5-fluoro phenyl ring with a significant gain in potency [26, 27]. Fluorinated β -phenylalanine is the basic core moiety on which several modifications have been extended on the other parts of the molecule to derive better lead compounds with improved inhibitory activities (Table 1).

Two comparative molecular similarity analysis (CoMFA) models obtained in 1993 and 1995 on early dipeptide analogs as DPP-IV inhibitors successfully demonstrated that three-dimensional quantitative structure-activity relationship (3D-QSAR) was a useful tool for obtaining more effective inhibitor structures [28, 29]. The conformations of the inhibitors were constructed mainly from theoretical conformational investigation of the substrate. As selection and alignment of the active conformation are the key to successful 3D-QSAR modeling, improved CoMFA and comparative molecular similarity indices analysis (CoMSIA) have been performed for different series of five-membered heterocycle derivatives including arylthiazolidine derivatives, noncyano-pyrrolidine derivatives and, more recently, azaheterocyclic derivatives [30, 31]. The selected inhibitors were docked into the binding site of DPP-IV using molecule-docking software, and the docked conformations of the inhibitors were adopted to be the “active” conformation for the construction of 3D-QSAR models.

Here, we present 3D-QSAR studies using CoMFA and CoMSIA methods on β -phenylalanine derivatives as DPP-IV inhibitors by considering the steric, electrostatic, H-bond and hydrophobic influences. The selected ligands were docked into the binding site of the 3D model of DPP-IV using AutoDock software, and possible interaction models between DPP-IV and its inhibitors were obtained. Based on the binding conformations of these inhibitors and their alignment inside the binding pocket of DPP-IV, 3D-QSAR models were established by CoMFA and CoMSIA analyses. The predictive ability of these models was validated using 18 compounds in the test set. The results obtained in this study could provide a powerful tool for predicting the affinity of related compounds with DPP-IV, and useful information for guiding further design and synthesis of novel DPP-IV inhibitors.

Computational details

Biological data

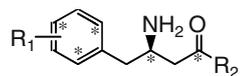
The structures of β -phenylalanine derivatives and the corresponding data on biological activity were obtained

from the literature [23, 26, 32–36]. Of the ~200 known inhibitors available in the literature, 91 were collected (Table 1). Other reported inhibitors were excluded because they lacked well-defined stereochemistry. The datasets were divided into a training set and a test set. The test set, which was selected randomly, comprised 18 compounds. The compounds in the training set were used to build 3D-QSAR models, and those in the test set were used to evaluate the 3D-QSAR models. The negative logarithm of IC_{50} (pIC_{50}) was used as the biological activity in the 3D-QSAR study (Table 1). Accordingly, the pIC_{50} values of the training set described in this manuscript span 3.6 log units distributed over this range.

Molecular modeling and molecular docking

Three-dimensional structure building and all modeling were performed using the Sybyl 8.1 program package running on a Pentium IV computer under the Linux RedHat Enterprise 2.3.1 OS [37]. The crystal structure of DPP-IV in complex with its inhibitor MK-0431 (PDB ID: 2QOE) was recovered from the Brookhaven Protein Database (<http://www.rcsb.org/pdb>). After correcting atom types and adding all the hydrogen atoms, Kollman all-atom charges were assigned for enzyme and molecular building was done for all inhibitors with molecular sketch program based on the structure of MK-0431 [38]. Geometry optimization was carried out using MAXIMIN molecular mechanics and Tripos force field, Gasteiger–Hückle charge supplied with the convergence criterion set at $0.05 \text{ kcal}\text{\AA}^{-1}\text{mol}^{-1}$ for ligands [39, 40].

The docking studies were carried out with Autodock 3.0.5 (CCDC, UK; <http://www.ccdc.cam.ac.uk/products/csd/>) running on a Pentium IV computer under the Linux system [41]. The enzyme exists as a dimer in the crystal. The docking study was done with the monomeric unit of the enzyme, as the active site of the enzyme resides deep within the receptor protein and not on the enzyme surface. The water molecules in the crystal were not considered during docking since no water molecule was found conserved in different crystal structures. Docking was continued for 20 Genetic Algorithm runs. The known crystal structure of complex (2QOE) was used to investigate and validate the docking protocol. At first, all variables for docking were set at their default values. Poses matching the conformations of the ligand in the crystal structure were not reproduced, that was probably caused by the important halogen-bonding interaction between fluorinated phenyl on the ligands and residues in the active site, which is not included in the AutoDock software. The halogen-bonding interaction could be compared with the hydrogen-bonding interaction [42, 43]. For the second docking protocol, one part of ligands

Table 1 Structures and biological activities of training set and test set compounds

Compd	R ₁	R ₂	pIC ₅₀	Compd	R ₁	R ₂	pIC ₅₀
01	2,5-F		7.57	47	2,4,5-F		6.94
02	2,5-F		6.99	48#	2,4,5-F		7.23
03	2,5-F		6.89	49	2,4,5-F		7.0
04	2,5-F		6.82	50	2,4,5-F		7.15
05	2,5-F		6.74	51	2,4,5-F		7.17
06	2,5-F		6.82	52	2,4,5-F		6.56
07	2,5-F		6.46	53#	2,4,5-F		7.60
08	2,5-F		6.80	54	2,4,5-F		8.40
09	2,5-F		6.66	55	2,4,5-F		8.40
10#	2,5-F		6.52	56	2,4,5-F		8.15
11	2,5-F		6.11	57	2,4,5-F		8.70
12#	2,5-F		6.82	58	2,4,5-F		6.21
13	2,5-F		6.89	59	2,4,5-F		7.22
14#	2,5-F		7.14	60	2,4,5-F		6.42
15#	2,5-F		6.89	61	2,4,5-F		7.30
16	2,5-F		7.07	62	2,4,5-F		6.47
17	2,4,5-F		6.23	63	2,4,5-F		8.82

Table 1 (continued)

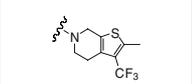
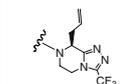
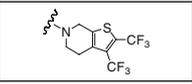
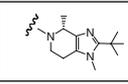
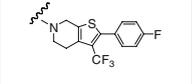
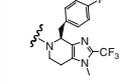
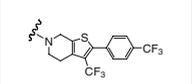
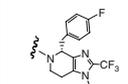
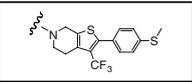
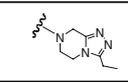
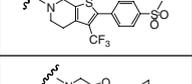
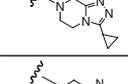
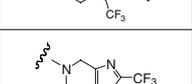
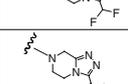
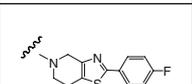
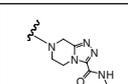
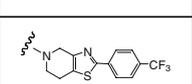
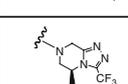
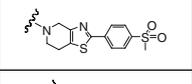
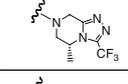
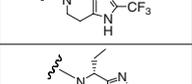
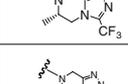
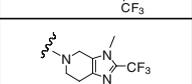
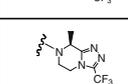
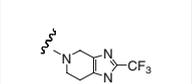
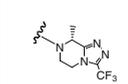
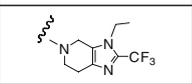
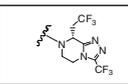
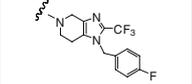
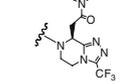
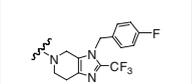
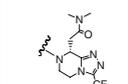
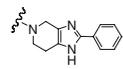
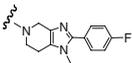
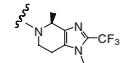
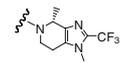
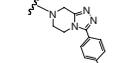
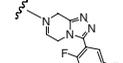
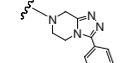
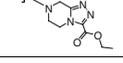
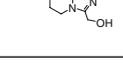
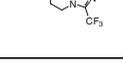
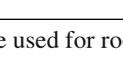
18	2,5-F		6.32	64	2,4,5-F		7.94
19	2,5-F		6.51	65#	2,4,5-F		7.43
20	2,5-F		6.44	66#	2,4,5-F		6.11
21	2,5-F		6.59	67#	2,4,5-F		8.24
22#	2,5-F		6.42	68#	2,4,5-F		7.43
23	2,5-F		6.89	69	2,4,5-F		7.52
24	2,5-F		6.54	70	2,4,5-F		7.54
25	2,5-F		7.24	71	2,4,5-F		6.66
26	2,5-F		7.59	72	2,4,5-F		6.63
27	2,5-F		7.89	73	2,4,5-F		7.64
28	2,5-F		8.85	74	2,4,5-F		7.05
29#	2,5-F		6.85	75	2,4,5-F		7.04
30	2,5-F		6.82	76	2,4,5-F		7.38
31#	2,5-F		6.89	77#	2,4,5-F		7.06
32	2,5-F		6.74	78#	2,4,5-F		8.37
33	2,5-F		6.64	79	2,4,5-F		8.24
34	2,5-F		6.60	80	2,4,5-F		6.42
35	2,5-F		6.60	81	2,4,5-F		8.55

Table 1 (continued)

36	2,5-F		6.89	82	2,4,5-F		6.85
37#	2,5-F		6.55	83	2,4,5-F		9.18
38#	2,5-F		6.72	84	2,4,5-F		6.49
39#	2,5-F		7.19	85	2,4,5-F		9.37
40	2,5-F		7.22	86	2,4,5-F		6.36
41	2,5-F		7.52	87	2,4,5-F		9.51
42	2,5-F		7.19	88	2,4,5-F		6.88
43	2,5-F		6.72	89	2,4,5-F		9.34
44	2,5-F		7.16	90	2,4,5-F		6.94
45#	2,4,5-F		7.74	91	2,4,5-F		9.74
46	2, 4,5-F		7.17				

* The alignment atoms were used for root mean square fitting corresponding to the atoms in the template.

Compounds those are used in the training set and not included in the construction of 3D-QSAR models.

(fluorinated phenyl) involved in halogen-bonding interaction was fixed as they share the same structural feature and bind in a similar way with DPP-IV, and the other parts of the ligands remain flexible (Table 1). The root-mean-square deviation (RMSD) between the conformation of the MK-0431 from the X-ray crystal structures and those from the AutoDock results were lower than 0.30 Å, indicating that the parameter set for the AutoDock simulation was reasonable to reproduce the X-ray structure. All of the compounds were docked into the active site of target protein using the second docking protocol.

3D-QSAR studies

CoMFA and CoMSIA analyses were carried out on a Pentium IV computer under the Linux RedHat Enterprise 2.3.1 OS in Sybyl 8.1. The molecular alignment of the studied compounds in Cartesian space was first required for 3D-QSAR techniques. The alignment generated during the docking study as a starting point was used for the QSAR analysis, and should be very useful as it ultimately allows a combination of both target- and ligand-based approaches in one general integrated model, confirming or disproving the

interaction model initially generated. Based on the alignments derived from molecular docking, CoMFA and CoMSIA studies were performed on these inhibitors to analyze the specific contributions of steric, electrostatic, hydrophobic, and hydrogen bond effects to the bioactivities of the inhibitors [44, 45]. Partial least-squares (PLS) analyses of CoMFA and CoMSIA were carried out with the default settings for the 3D cubic lattice, the grid spacing, the probe atom, and the cutoff for the interaction energy. The column filter was set to 2.0. The maximum number of PLS regression components was set to 15. The same grids constructed for the calculation of CoMFA fields were used for the CoMSIA fields' calculation. The CoMFA/CoMSIA descriptors served as independent variables and pIC_{50} values as a dependent variable in PLS analysis in deducing the 3D-QSAR models. The optimal number of components obtained from the cross-validated PLS analysis were used to derive the final QSAR model using the compounds in the training set without cross-validation. The non-cross-validated correlation coefficient r^2_{ncv} , the standard errors and the F-value were computed as a measure of the quality of the models. To test the predictive power of the model as a tool, an external set of compounds with known activities but not used in model generation (the test set) were predicted. The predictive r^2 , calculated using Eq. 1, was based on molecules in the test set and was used to evaluate the predictive power of the CoMFA and CoMSIA models

$$\text{predictive } r^2 = 1 - (\text{"press"}/\text{SD}) \quad (1)$$

where SD is the sum of the squared deviations between the actual activities of the compounds in the test set and the mean activity of the compounds in the training set and "press" is the sum of the squared deviations between predicted and actual activities for every compound in the test set.

Results and discussion

Docking studies

All selected compounds were docked into the active site of target protein using the same docking protocol. Twenty conformations were obtained for each ligand and the conformation with strongest predicted binding affinity to DPP-IV was selected as the possible binding conformation.

The specificity pocket S1 is composed of the side chains of Tyr631, Val656, Trp659, Tyr662, Tyr666, and Val711, which are involved in strong hydrophobic interactions with the derivatives [46]. The cavity near Glu205, Glu206 and Tyr662 residues is referred to as the S2 pocket. The

alignment of the docking conformation of all compounds with the binding pocket of DPP-IV was shown in Fig. 1a. All the derivatives share a similar phenylalanine moiety, which occupies a well-defined hydrophobic S1 pocket with a fluorinated benzene ring. The fluorine substitutions on the benzene ring are important for the improvement of inhibitor activity. The fluorine located in the 2-position of the benzene ring is available to form a halogen-bonding network with Asn710, Arg125 and Glu205, which provides a 3-fold boost in activity relative to the unsubstituted phenyl analog [27]. A fluorine located in the 4, 5-position of the benzene ring is within C-H...F bonding distance to Val656 and Tyr631, respectively. The (protonated) amine is likewise within hydrogen bonding distance to Glu205, Glu206 and Tyr662.

The binding mode of the most active compound **91** is depicted generally in Fig. 1b. Compound **91** binds to the DPP-IV active site with a conformation similar to that of the

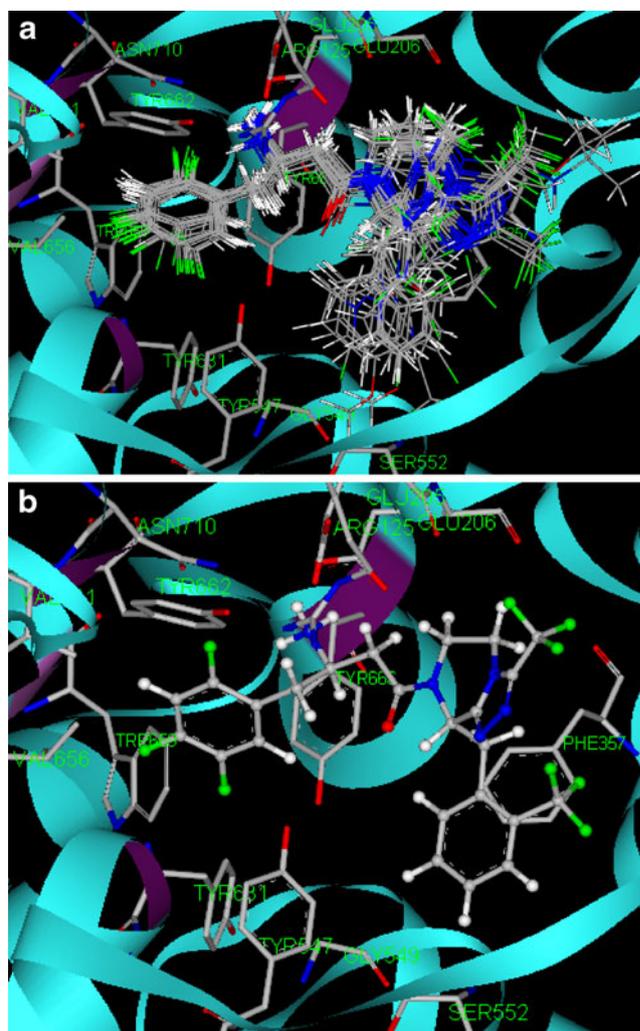


Fig. 1 **a** Binding conformations of docked compounds at the active site of dipeptidyl peptidase IV (DPP-IV). **b** The binding mode of the most active compound **91** with DPP-IV

known MK-0431 inhibitor, involving π - π stacking interaction with the Phe357 besides the interaction with pocket S1 and S2, as observed in most of these inhibitors. Four amino acids—Tyr547, Phe357, Gly549 and Ser552—form an adjacent peripheral cavity around the *para*-fluorophenyl ring at the R position of the chiral carbon, which creates significant hydrophobic interactions and enhances the IC₅₀ values by 2 orders of magnitude, whereas the *para*-fluorophenyl substitution at the corresponding S position (compound 48) leads to steric clashes with Phe357 that decrease inhibitory activity. These particular interactions play a very important role in DPP-IV inhibition and need to be present for good inhibition by the inhibitors.

CoMFA and CoMSIA statistical results

A training set of 73 compounds was selected from the existing database, representing the diversity of structures and activities (Table 1). CoMFA and CoMSIA analyses were performed for alignment models based on the conformational alignment obtained from the docking of 73 compounds. The statistical parameters of CoMFA and CoMSIA analysis with a 2.0 Å grid spacing are summarized in Table 2. The optimal components used to produce the cross-validation linear regression coefficient were used to produce the non-cross-validated model.

CoMFA models were developed by steric and electrostatic descriptors. Leave-one-out cross-validated PLS analysis gave rise to a cross-validated value (r_{cv}^2) of 0.759, suggesting that the model was a useful tool for predicting DPP-IV inhibitory activity [47]. The correlation coefficient

between the calculated and experimental activities was a non-cross-validated value (r_{ncv}^2) of 0.954 with a standard error of estimate 0.194 using six principal components. The relative contributions of steric and electrostatic fields were 0.778 and 0.222%, respectively, indicating that steric field was more predominant.

Three different CoMSIA models were developed by steric, electrostatic, hydrophobic and hydrogen-bond field. The first model with the combination of hydrophobic, hydrogen-bond fields yielded a cross-validated $r_{cv}^2 = 0.750$ with eight principal components, non cross-validated $r_{ncv}^2 = 0.959$, F value of 185.708. The contributions of hydrophobic, hydrogen bond donor and acceptor fields of this model were 0.548, 0.387 and 0.065%, respectively. Combinations of steric, electrostatic fields yielded a second model using five principal components with a cross-validated $r_{cv}^2 = 0.708$, non cross-validated $r_{ncv}^2 = 0.966$, F value of 175.340. The steric and electrostatic contributions were 0.508 and 0.492%, respectively. The third model with the combination of hydrophobic, steric and electrostatic fields using eight principal components yielded a cross-validated $r_{cv}^2 = 0.694$, non cross-validated $r_{ncv}^2 = 0.905$, F value of 185.708. The contributions of hydrophobic, steric and electrostatic fields of this model were 0.437, 0.306 and 0.258%, respectively.

Validation of 3D-QSAR models

A structurally diversified test set of 18 compounds that were not included in the model generation were selected as test data set to validate the predictive abilities of the 3D-QSAR models

Table 2 Results of comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) for the compounds used in the training set. *SEE* Standard error of estimate, *SEP* standard error of prediction, *PLS* partial least squares

Parameters	CoMFA	CoMSIA		
	S, E ^f	H, D, A ^f	S, E ^f	H, S, E ^f
r_{ncv}^2 ^a	0.955	0.959	0.966	0.905
SEE	0.194	0.188	0.174	0.315
F _{test} ^b	233.271	185.708	175.340	161.313
r_{cv}^2 ^c	0.759	0.750	0.708	0.694
SEP	0.448	0.464	0.489	0.493
r_{pred}^2 ^d	0.737	0.636	0.791	0.901
PLS components ^e	6	8	5	4
Contribution:	0.778, 0.222	0.548, 0.387, 0.065	0.508, 0.492	0.437, 0.306, 0.258

^a Non-cross-validated correlation coefficient

^b Ratio of r_{ncv}^2 explained to unexplained = $r_{ncv}^2 / (1 - r_{ncv}^2)$

^c Cross-validated correlation coefficient after the leave-one-out procedure

^d Predicted correlation coefficient for the test set of compounds

^e Optimal number of principal components

^f CoMFA and CoMSIA with different field contributions such as steric (S), electrostatic (E), hydrophobic (H), donor (D), and acceptor (A) fields

(Table 1). The inhibitory activity (pIC50) and the calculated activities from CoMFA and CoMSIA models for the test set were listed in Table 3. The predictive r^2 value (r^2_{pred}) for the CoMFA model was 0.737. The corresponding r^2_{pred} values were 0.636, 0.791 and 0.901 for first, second and third CoMSIA model, respectively. All the test compounds were accurately predicted by those models. All the 3D-QSAR models predicted the 18 tested compounds well. The mean and standard deviation of prediction errors were within 0.33 and 0.26 for all models. The third CoMSIA model, with the combination of hydrophobic, steric and electrostatic fields, gave the best r^2_{pred} value (0.901), with less residual values of the test set compared to the other models; the mean and standard deviation of prediction errors were 0.19 and 0.13, respectively. The high predictive power of the CoMFA and CoMSIA training models for sets of diverse structural scaffolds suggested that these models possessed a high accommodating capacity and wide applicability in the development of novel DPP-IV inhibitors.

Analysis of contours for CoMFA models

The CoMFA steric and electrostatic fields from the non-cross-validated analysis were plotted as three-dimensional colored contour maps in Fig. 2a,b. To aid visualization, the most potent compound **91** was overlaid on the map. In the steric field, the green (sterically favorable) and yellow

(sterically unfavorable) contours represent 80% and 20% level contributions, respectively. A large green contour around *para*-fluorobenzyl substituent at the (8R)-position of the triazolopiperazine ring indicates that a sterically bulky group is favored in this region. In case of the highly active compounds **63**, **64**, **65**, **67**, **78**, **79**, **81**, **83**, **85**, **87**, **89**, **91**, the 8-position substituent touches the green contour, while many of the less active compounds including **66**, **80**, **82**, **84**, **86**, **88**, **90** tend to locate their 8-position substituent away from this contour and approach a large sterically unfavorable yellow contour. Several yellow contours seen in the vicinity of the 1-, 2- and 3-positions on triazolopiperazine ring indicate that occupancy of this sterically unfavorable contour will have a detrimental effect on inhibitory activity. Similarly, in the electrostatic field, the red (electronegative charge favorable) and blue (electropositive charge favorable) contours represent 80% and 20% level contributions, respectively. A red contour on the left side of the image indicates that an electronegative group at this position can improve inhibitory activity, as exemplified by comparing compounds **45** and **01**. There are several small red contours near the 1-, 2- and 3-positions of triazolopiperazine ring, suggesting that electronegative groups at this position are also favorable for inhibitory activity. Considering the analysis from steric field, it is concluded that the 1-, 2- and 3-positions of triazolopiperazine ring favor small and electronegative groups.

Table 3 Inhibitory activity (pIC50) and the calculated activities from CoMFA and CoMSIA models for the test set

Compound	Experiment (pIC50)	Prediction (pIC50)			
		CoMFA		CoMSIA	
		S, E ^a	H, D, A ^a	S, E ^a	H, S, E ^a
10	6.52	6.62	6.99	6.23	6.42
12	6.82	6.80	6.26	6.46	6.70
14	7.14	6.85	6.98	6.55	6.85
15	6.89	6.82	6.76	6.86	6.70
22	6.42	7.09	6.89	6.62	6.72
29	6.85	6.93	6.88	6.87	6.81
31	6.89	6.78	6.65	6.24	6.79
37	6.55	6.89	6.89	6.59	6.51
38	6.72	6.81	6.55	6.41	6.53
39	7.19	7.07	7.06	7.31	7.06
45	7.74	7.16	7.41	7.56	7.47
48	7.23	6.92	6.91	6.96	7.26
65	7.43	7.07	6.91	7.02	6.96
66	6.11	6.54	6.55	6.09	6.41
67	8.24	8.28	7.80	7.81	7.79
68	7.43	7.07	7.22	7.48	7.34
77	7.06	7.04	7.39	7.42	7.14
78	8.37	7.99	7.72	8.07	8.14

^a CoMFA and CoMSIA with different field contributions such as steric (S), electrostatic (E), hydrophobic (H), donor (D), and acceptor (A) fields

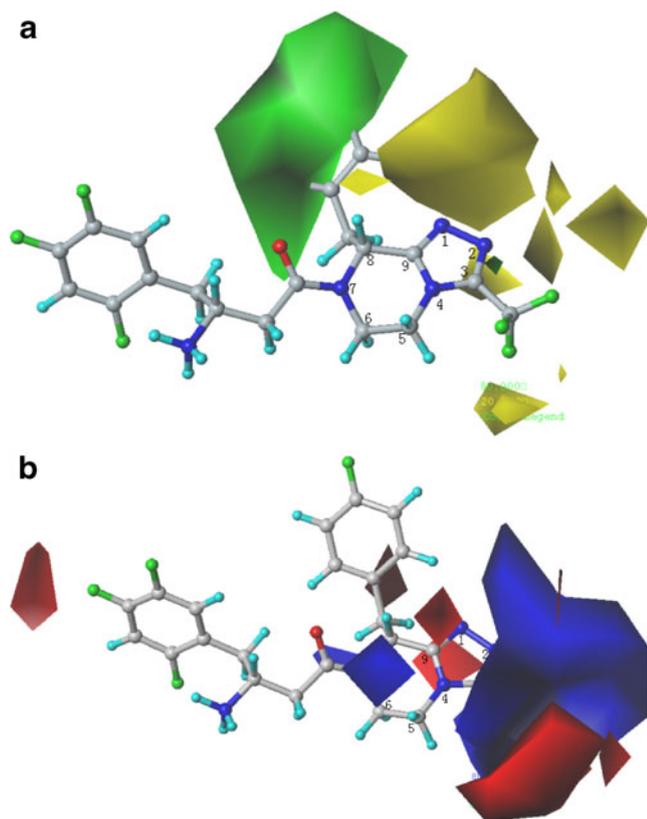


Fig. 2 a,b Comparative molecular field analysis (CoMFA) StDev*-Coeff contour plots. **a** Steric contour map: *green* regions where bulky groups increase activity, *yellow* regions where bulky groups decrease activity. **b** Electrostatic contour map: *red* regions where negative charges increase activity, *blue* regions where positive charges increase activity. Compound **91** is displayed as a reference

Analysis of the contours for CoMSIA models

In the CoMSIA hydrophobic field, the yellow (hydrophobic favorable) and cyan (hydrophobic unfavorable or hydrophilic favorable) contours represent 80% and 20% level contributions, respectively. Similarly, the purple (hydrogen bond donor favorable) and cyan (hydrogen bond donor unfavorable) contours represent 80% and 20% level contributions, respectively, in the hydrogen bond donor fields. In the CoMSIA hydrogen bond acceptor field, the magenta (hydrogen bond acceptor favorable) and red (hydrogen bond acceptor unfavorable) contours represent 80% and 20% level contributions, respectively.

The hydrophobic contour map of the CoMSIA model in the presence of the most active compound **91** is displayed in Fig. 3a. Three yellow contours around the 3-, 8R-position of the triazolopiperazine ring suggest that its occupancy by hydrophobic groups would increase inhibitory activity. Compounds **68** and **69** have hydrophobic groups such as ethyl and cyclopropyl at the 3-position, thus accounting for their relatively good activity. The (8R)-position corresponds to a pocket formed by sides of

residues Tyr547, Ser552, Pro550 of the receptor protein and its hydrophobic substitution inserts into the hydrophobic pocket, forming a favorable environment for the improvement of inhibitory activity. The substituent at this position could be a number of diverse hydrophobic residues, such as benzyl, cyclopropyl methyl, triazole methyl, etc. A polar group such as fluorine, trifluoromethyl, or methoxy on the aromatic ring of benzyl, could improve inhibitory activity (e.g., **54**, **55**, **87**, **89**, and **91**). A cyan contour on the 1- and 8-positions of the triazolopiperazine suggests that the hydrophilic group at the (8 S)-position would increase DPP-IV inhibitory effect. A cyan contour map under the 5- and 6-positions of the triazolopiperazine ring indicates that hydrophobic groups at (5R)- and (6 S)-positions are unfavorable. The less active compounds **74** and **75** orient the hydrophobic methyl at the (5R)- and (6 S)-positions toward the cyan contour, whereas the moderately active compounds **73** and **76** orient the hydrophobic methyl at the (5 S)- and (6R)-positions away from the cyan contour. Another cyan contour localizes in the vicinity under the areas formed by 2- and 3-positions and trifluoromethyl of the triazolopiperazine ring. It orients

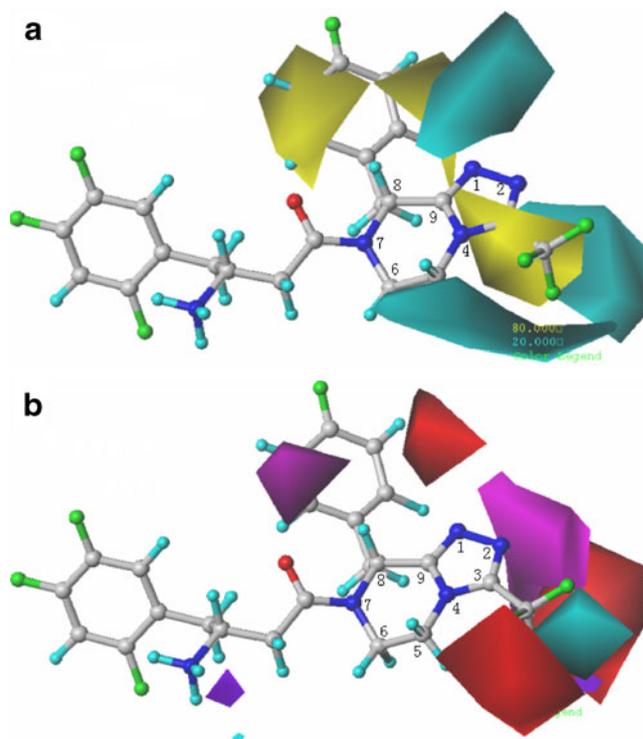


Fig. 3 a,b Comparative molecular similarity indices analysis (CoMSIA) StDev*-Coeff contour plots. **a** Hydrophobic contour maps: *yellow* regions where hydrophobic groups increase activity, *cyan* regions where hydrophobic groups decrease activity. **b** Hydrogen bond donor and hydrogen bond acceptor contour maps: *purple* and *cyan* contours favorable and unfavorable hydrogen bond donor groups, respectively; whereas *magenta* and *red* contours favorable and unfavorable hydrogen bond acceptor groups, respectively. Compound **91** is displayed as a reference

in the direction of the substitution on the 2- and 3-positions. The poor inhibitory activity of compounds 02 and 05 is probably due to occupancy of the cyan contour by the hydrophobic trifluoroethyl group on the 3-position.

The hydrogen bond donor and acceptor contour maps of the CoMSIA models in the presence of the most potent compound **91** are shown in Fig. 3b. A purple contour distantly located from the $-\text{NH}_3$ suggests a possible positive effect on activity by having a hydrogen bond donor in this region. A cyan contour lies off the trifluoromethyl on the 3-position of the triazolopiperazine ring, indicating that the hydrogen bond donor group is unfavorable for activity. A magenta contour around the 2-position of the triazolopiperazine ring represents the higher activity of compounds having a hydrogen bond acceptor group at this position. The most potent compounds have trifluoromethyl (compound **91**) and bifluoromethyl (compounds **70**) as substituents at this position. All these groups possess hydrogen bond donor atoms that seem to play a key role in their predominantly positive effect on DPP-IV inhibitory activity. A red contour map between the 5- and 3-positions indicates that hydrogen bond acceptor groups are unfavorable (or hydrogen bond donor group favorable). A large red contour located not too close to any of the atoms of the compounds suggests that occupancy of this spatial region by a hydrogen bond acceptor group would cause a decrease in activity. The substituent on the 3-position in the poorly active compound **43** is a carboxylic ester whose two hydrogen bond acceptor oxygen atoms are closer to red contours. A small hydrogen bond acceptor unfavorable red contour map on the right side of the image resulted from the *meta*-position of the benzyl.

Implications for new inhibitor design

Acceptor-based methods of analysis such as CoMFA and CoMSIA are widely used not only because they are not very computationally demanding but also because they can lead to the rapid generation of QSARs from which the biological activity of newly designed compounds can be predicted. The findings derived from 3D QSAR studies presented here identified several key structural features important to the inhibitory activity of DPP-IV that could have some implications for the design of new inhibitors. First, having the 1-, 2- and 3-positions of the triazolopiperazine ring occupied by small and electronegative groups will significantly improve pIC₅₀ values due to better occupation of the electronegative favorable contour and avoidance of the sterically unfavorable contour. Second, the substituted group must be carefully introduced into the 3-position as it located between two hydrogen bond acceptor-unfavorable red contours and a hydrophobic unfavorable contour. Third, at the chiral carbon of the 5-, 6- and 8-positions, the favorable hydrophilic contour

on one side and a favorable hydrophobic contour on the other side can be exploited to yield potent analogues.

Conclusions

In this work, 3D-QSAR and molecular docking studies were carried out to explore the binding of 73 inhibitors to DPP-IV, and to construct highly predictive 3D-QSAR models for designing new DPP-IV inhibitors for the treatment of type 2 diabetes. A high r_{cv}^2 and r_{ncv}^2 value for CoMFA and CoMSIA models with a small standard deviation indicated the existence of a similar relationship among all of the compounds used to build the model. In addition to steric and electrostatic fields, hydrophobic and hydrogen-bond donor/acceptor fields were also found to be important for DPP-IV inhibitory activity. These models were further validated using a test set of 18 compounds that revealed high predicted values with a small standard deviation in the prediction of inhibitory activities. Interpretation of CoMFA and CoMSIA contour maps provided a better understanding of inhibitor–DPP-IV interactions. The consistency between the CoMFA/CoMSIA field distributions and the 3D topology of the protein structure showed the robustness of the 3D-QSAR models and will allow reliable quantitative models to be developed for rational drug design. Taken together, the present 3D QSAR models were found to accurately predict the DPP-IV inhibitory activity of a structurally diverse test set of compounds and to yield reliable clues for further optimization of β -phenylalanine derivatives in the data set.

References

1. Mentlein R (1999) Regul Pept 85:9–24
2. Hopsu-Havu VKGG (1966) Histochemie 7:197–201
3. Holst JJ, Deacon CF (2004) Curr Opin Pharmacol 4:589–596
4. Drucker DJ (2001) Curr Pharm Des 7:1399–1412
5. Zander M, Madsbad S, Madsen JL, Holst JJ (2002) Lancet 359:824–830
6. Weber AE (2004) J Med Chem 47:4135–4141
7. Deacon CF, Ahren B, Holst J (2004) J Exp Opin Investig Drugs 13:1091–1102
8. Drucker D (2003) J Exp Opin Investig Drugs 12:87–100
9. Augustyns K, Van der Veken P, Senten K, Haemers A (2003) Exp Opin Ther Patents 13:499–510
10. Wiedeman PE, Trevillyan JM (2003) Curr Opin Investig Drugs 4:412–420
11. Barlocco D (2004) Curr Opin Investig Drugs 5:1094–1100
12. McIntyre JA, Castaner J (2004) Drugs Fut 29:887–891
13. Holst JJ, Deacon CF (1998) Diabetes 47:1663–1670
14. Augustyns K, Bal G, Thonus G, Belyaev A, Zhang XM, Bollaert W, Lambeir AM, Durinx C, Goossens F, Haemers A (1999) Curr Med Chem 6:311–327
15. Villhauer EB, Coppola GM, Hughes TE (2001) Annu Rep Med Chem 36:191–200

16. Drucker DJ (2003) *Exp Opin Investig Drugs* 12:87–100
17. Wiedeman PE, Trevillyan JM (2003) *Curr Opin Investig Drugs* 4:412–420
18. Villhauer EB, Brinkman JA, Naderi GB, Dunning BE, Mangold BL, Mone MD, Russell ME, Weldon SC, Hughes TE (2002) *J Med Chem* 45:2362–2365
19. Hughes TE, Mone MD, Russell ME, Weldon SC, Villhauer EB (1999) *Biochemistry* 38:11597–11603
20. Kanstrup AB, Christiansen LB, Lundbeck JM, Sams CK, Kristiansen M (2002) *PCT Int Appl: WO2002002560*
21. Sorbera LA, Revel L, Castaner J (2001) *Drugs Fut* 26:859–864
22. Feng J, Gwaltney SL, Stafford JA, Zang Z (2005) *Jpn Kokai Tokkyo Koho: JP 2005263780*
23. Kim D, Wang L, Beconi M, Eiermann GJ, Fisher MH, He H, Hickey GJ, Kowalchick JE, Leiting B, Lyons K, Marsilio F, McCann ME, Patel RA, Petrov A, Scapin G, Patel SB, Roy RS, Wu JK, Wyvratt MJ, Zhang BB, Zhu L, Thornberry NA, Weber AE (2005) *J Med Chem* 48:141–151
24. Mcintyre JA, Castaner J (2004) *Drugs Fut* 29:887–891
25. Barlocco D (2004) *Curr Opin Invest Drugs* 5:1094–1100
26. Xu JY, Ok HO, Gonzalez EJ, Colwell LF, Habulihaz JB, He HB, Leiting B, Lyons KA, Marsilio F, Patel RA, Wu JK, Thornberry NA, Weber AE, Parmee ER (2004) *Bioorg Med Chem Lett* 14:4759–4762
27. Brockunier LL, He JF, Colwell LF Jr, Habulihaz JB, He HB, Leiting B, Lyons KA, Marsilio F, Patel RA, Teffera Y, Wu JK, Thornberry NA, Weber AE, Parmee ER (2004) *Bioorg Med Chem Lett* 14:4763–4766
28. Brandt W, Lehmann T, Barth A, Fittkau S (1993) *J Mol Graphics* 11:277–278
29. Brandt W LT, Thondor I, Born I, Schutkowski M, Rahfeld JU, Neubert K, Barth A (1995) *Int J Peptide Prot Res* 46:494–507
30. Zeng J, Liu GX, Tang Y, Jiang HL (2007) *J Mol Model* 13:993–1000
31. Pissurlenkar RRS, Shaikh MS, Coutinho EC (2007) *J Mol Model* 13:1047–1071
32. Ashton WT, Sisco RM, Dong H, Lyons KA, He HB, Doss GA, Leiting B, Patel RA, Wu JK, Marsilio F, Thornberry NA, Weber AE (2005) *Bioorg Med Chem Lett* 15:2253–2258
33. Kowalchick JE, Leiting B, Pryor KD, Marsilio F, Wu JK, He HB, Lyons KA, Eiermann GJ, Petrov A, Scapin G, Patel RA, Thornberry NA, Weber AE, Kim D (2007) *Bioorg Med Chem Lett* 17:5934–5939
34. Chen P, Caldwell CG, Mathvink RJ, Leiting B, Marsilio F, Patel RA, Wu JK, He HB, Lyons KA, Thornberry NA, Weber AE (2007) *Bioorg Med Chem Lett* 17:5853–5857
35. Kim D, Kowalchick JE, Edmondson SD, Mastracchio A, Xu J, Eiermann GJ, Leiting B, Wu JK, Pryor KD, Patel RA, He HB, Lyons KA, Thornberry NA, Weber AE (2007) *Bioorg Med Chem Lett* 17:3373–3377
36. Kim D, Kowalchick JE, Brockunier LL, Parmee ER, Eiermann GJ, Fisher MH, He HB, Leiting B, Lyons KA, Scapin G, Patel SB, Petrov A, Pryor KD, Roy RS, Wu JK, Zhang XP, Wyvratt MJ, Zhang BB, Zhu L, Thornberry NA, Weber AE (2008) *J Med Chem* 51:589–602
37. Sybyl Version 8.1 (2009) Tripos, St. Louis, MO
38. Clark M, Cramer RDI, Opdenbosch NV (1989) *J Comput Chem* 10:982–1012
39. Gasteiger J, Marsili M (1980) *Tetrahedron* 36:3219–3228
40. Weiner SJ, Kollman PA, Case DA, Singh C, Ghio G, Alagona S, Profeta P, Weiner P (1984) *J Am Chem Soc* 106:765–784
41. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) *J Comput Chem* 19:1639–1662
42. Auffinger P, Hays FA, Westhof E, Ho PS (2004) *Proc Natl Acad Sci USA* 101:16789–16794
43. Howard EI, Sanishvili R, Cachau RE, Mitschler A, Chevrier B, Barth P, Lamour V, Zandt MV, Sibley E, Bon C, Moras D, Schneider TR, Joachimiak A, Podjarny A (2004) *Proteins* 55:792–804
44. Cramer RD, Patterson III DE, Bunce JD (1988) *J Am Chem Soc* 110:5959–5967
45. Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130–4146
46. Kuhn B, Hennig M, Mattei P (2007) *Curr Top Med Chem* 7:609–619
47. Xu M, Zhang AQ, Han SK, Wang LS (2002) *Chemosphere* 48:707–715