

3D-QSAR studies of Dipeptidyl peptidase IV inhibitors using a docking based alignment

Raghuvir R. S. Pissurlenkar · Mushtaque S. Shaikh · Evans C. Coutinho

Received: 9 March 2007 / Accepted: 12 June 2007 / Published online: 4 August 2007
© Springer-Verlag 2007

Abstract Dipeptidyl peptidase IV (DPP-IV) deactivates the incretin hormones GLP-1 and GIP by cleaving the penultimate proline or alanine from the N-terminal (P1-position) of the peptide. Inhibition of this enzyme will prevent the degradation of the incretin hormones and maintain glucose homeostasis; this makes it an attractive target for the development of drugs for diabetes. This paper reports 3D-QSAR analysis of several DPP-IV inhibitors, which were aligned by the receptor-based technique. The conformation of the molecules in the active site was obtained through docking methods. The QSAR models were generated on two training sets composed of 74 and 25 molecules which included phenylalanine, thiazolidine, and fluorinated pyrrolidine analogs. The 3D-QSAR models are robust with statistically significant r^2 , q^2 , and r^2_{pred} values. The CoMFA and CoMSIA models were used to design some new inhibitors with several fold higher binding affinity.

Keywords CoMFA · CoMSIA · Dipeptidyl peptidase IV inhibitors · Docking/receptor based alignment · 3D-QSAR

Introduction

Dipeptidyl peptidase-IV (DPP-IV) is a multifunctional protein involved in many physiological processes: functioning as a binding protein, a receptor and a proteolytic enzyme. This

enzyme was discovered in the late 80's, as a serine peptidase belonging to the S9b protein family [1]. This ecto-enzyme exists in two forms, i.e., as a soluble homo-dimer and as a type II integral plasma membrane glycoprotein which is abundantly expressed on a variety of cell surfaces [2–6]. The enzyme is seen to have a strong correlation with many diseases such as diabetes, obesity and tumor progression, making it an attractive target in drug discovery research [7–12]. DPP-IV degrades endogenous peptides by cleaving the penultimate proline or alanine from the amino terminus. Several bioactive peptides like bradykinin, growth hormone-releasing hormone [13], and substance P [14, 15] are well known substrates for this enzyme.

The fact that DPP-IV deactivates the incretin hormones makes inhibitors of DPP-IV promising leads for designing new therapeutic agents in the treatment of type 2 diabetes. GLP is released from intestinal L-cells after a meal and potentiates both glucose stimulated insulin secretion [16] and inhibition of glucagon secretion [17]. GLP-1 has many actions like increased insulin gene expression and biosynthesis, islet neogenesis and inhibition of pancreatic β -cell apoptosis, leading to increased β -cell mass. GIP, another incretin hormone produced by the K-cells, is located in the duodenum and like GLP-1, is also extensively involved in glucose metabolism [18]. Since the intact N-terminal ends of both GLP-1 and GIP are essential for biological activity [19–21], cleavage of the N-terminal dipeptide segment by DPP-IV plays an important role in maintaining glucose homeostasis. Thus, shielding incretin hormones from the catastrophic effects of DPP-IV enzyme by the administration of inhibitors will maintain the concentrations of the incretin hormones and prolong their efficient antidiabetic action.

A number of reversible and irreversible inhibitors of DPP-IV have been synthesized and some recent reviews have appeared [22, 23]. A few molecules like NVP-

R. R. S. Pissurlenkar · M. S. Shaikh · E. C. Coutinho (✉)
Department of Pharmaceutical Chemistry,
Bombay College of Pharmacy,
Kalina, Santacruz (E),
Mumbai 400 098, India
e-mail: evans@bcpindia.org

DPP728, NVP-LAF237 (Vidagliptin) [24, 25], P93/01 and MK-0431 (Sitagliptin) [26] are under clinical studies [27].

A couple of 3D-QSAR studies of dipeptide analog inhibitors have been reported in the past decade [28, 29]. In one study the conformations of the inhibitors were derived from a theoretical conformational investigation of the substrate and inhibitors of DPP-IV [29]. The dipeptides had an activity range of 4–6 log units. The conformations of the dipeptides were superimposed manually, followed by

an energy minimization of their side chains while keeping the peptide backbone fixed. The CoMFA model has an r^2 of 0.99 with 5 PLS components. The contribution of the steric field is much higher compared to the electrostatic field. It was established from this study that a proline or an azetidine ring is important for activity. This group probably interacts with the aromatic ring of a Phe/Tyr residue in the active site of the enzyme. Recent crystallographic structures of DPP-IV with proline/azetidine inhibitor molecules and

Table 1 The structural classes of compounds used for the 3D QSAR study

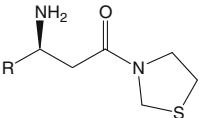
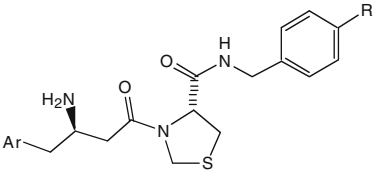
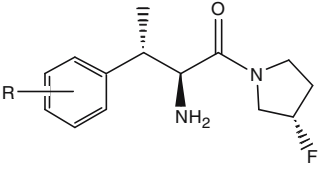
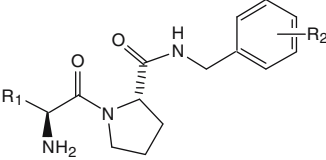
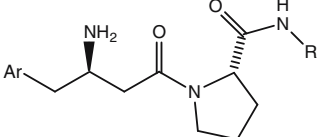
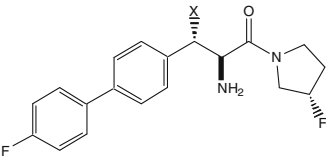
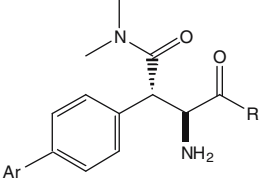
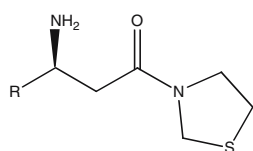
Sr. No.	Class of compounds	Representative structure	Activity range
1	Thiazolidine derivatives		4.35-6.57
2	1, 2-Disubstituted arylthiazolidine derivatives		6.79-9.52
3	Aryl fluoropyrrolidine derivatives		4.27-6.96
4	1, 2-Disubstituted pyrrolidine derivatives		5.14-5.44
5	1, 2-Disubstituted arylpyrrolidine derivatives		5.79-9.32
6	Biaryl β-substituted fluoropyrrolidine derivatives		5.46-8.18
7	Biaryl β-substituted aza-heterocyclic derivatives		6.26-8.66

Table 2 Structures of molecules along with their pIC₅₀ values used for 3D-QSAR studies**a) Thiazolidine derivatives**

Sr. No.	R	pIC ₅₀	Sr. No.	R	pIC ₅₀
D1T001		5.52	D1T015		5.05
D1T002		4.78	D1T016		5.44
D1T003		4.40	D1T017		5.48
D1T004		4.35	D1S018		5.52
D1T005		4.29	D1T019		4.97
D1T007		4.33	D1T020		5.11
D1T008		4.56	D1T021		4.83
D1T009		4.48	D1S022		5.81
D1T010		6.03	D1S023		5.77
D1T011		6.01	D1S024		6.23
D1T012		5.60	D1T025		6.57
D1T013		5.65	D1T026		5.69
D1S014		5.69	D1T027		6.92

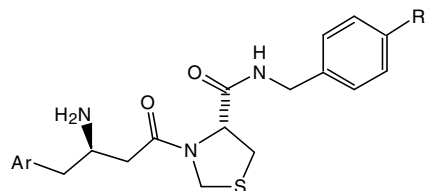
point mutation studies confirm that the proline or the azetidone ring resides in the hydrophobic pocket [30–33]. This proline is the basic core moiety on which several modifications have been made to derive better leads. Many new inhibitors of DPP-IV have been synthesized in the past couple of years, and the crystal structure of the enzyme in humans has also been published; there is thus a wide scope to improve upon these QSAR models.

Computational details

The molecular modeling packages *InsightII* v 2005L (Accelrys Inc., USA) and *Sybyl 7.1* (Tripos Inc., USA) were used in this study. The docking studies were carried out with *GOLD 3.0.1* (CCDC, UK) [34] running on a Pentium IV computer under the Windows OS. The 3D-QSAR techniques of CoMFA and CoMSIA were carried

Table 2 (continued)

b) 1, 2-Disubstituted arylthiazolidine derivatives



Sr. No.	Ar	R	pIC ₅₀
D2T057		-CH ₂ COOH	6.79
D2S058		-CH ₂ COOCH ₃	7.06
D2T059		-CH ₂ COOH	7.43
D2S060		-OCH(<i>i</i> -Pr)COOH (<i>S</i>)	9.52

out with *Sybyl 7.1* [35] running on a Pentium IV computer under the Linux RedHat Enterprise 2.3.1 OS.

Selection of training and test sets

The DPP-IV inhibitors were sub-classified on the basis of their structural diversity; the classes are depicted in Table 1. The dataset used for the development of CoMFA and CoMSIA models was divided into two sets based on the different binding modes seen in the crystal structures. The first dataset consisted of 102 molecules (molecules belonging to subclass a, c, f, g, and h in Table 2), while the second dataset had 34 molecules (molecules belonging to subclass b, d, and e in Table 2). The molecules span a 4 log unit activity range. The training and test set molecules were selected using the Library analysis/score option in the *Combi-chem-I* module in *Cerius2* (Accelrys Inc., USA) [36] running on a SGI workstation. The selection was done on the basis of structural, chemical and biological diversity using a similarity search technique [37]. The first dataset had 74 molecules in the training set (molecules labeled with prefix D1T in Table 2) and 28 in the test set (molecules labeled with prefix D1S in Table 2), while the second dataset was composed of a training set of 25 (denoted as D2T in Table 2) and a test set of nine molecules (marked as D2S in Table 2). The selection technique searches diverse molecules in 2D space by comparing the chemical nature of

the molecules. The dataset was split into training and test sets containing diverse molecules based on the D-Optimal design, Tanimoto similarity coefficient [38, 39] and the Euclidian distance matrix criteria. The remaining parameters *viz.* normalization criteria, number of Monte Carlo steps, Monte Carlo temperature, Monte Carlo seed, Gaussian alpha, bucket solution for K_d tree and the number of termination steps were set to their default values. Other parameters that were considered for the diverse set selection were the default molecular properties defined in *Cerius2* like molecular weight, AlogP, H-donor, H-acceptor, accessible surface area and the experimental pIC₅₀ values.

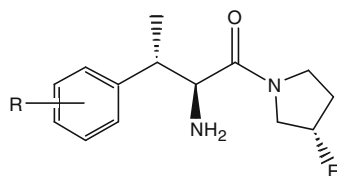
Ligand preparation

The DPP-IV inhibitors used in the 3D-QSAR study were taken from recent series of papers [40–43]. The first dataset was built from the inhibitors in the crystal structures (PDB codes: 1N1M, 2AJL and 1X70) [30–32] as the templates. The second dataset of 34 molecules was built from the closely related inhibitor in the crystal structure with PDB code 2BUB [33]. The template structures are shown in Fig. 1.

The molecules were built with the *builder* module in *Sybyl 7.1*. The ligand geometries were optimized by energy minimization using the Powell gradient method with the Tripos force field and Gasteiger Hückel charges for all atoms, until a gradient 0.01 kcal/mol/Å was reached. The dielectric constant was set to 1.0.

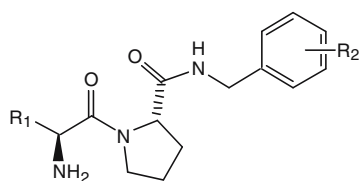
Table 2 (continued)

c) Aryl fluoropyrrolidine derivatives



Sr. No.	R	pIC ₅₀	Sr. No.	R	pIC ₅₀
D1S061	H	5.85	D1T075	4- CF ₃	6.22
D1S062	2-F	5.96	D1T076	4-SO ₂ CH ₃	5.44
D1S063	2-F, 4-Cl	6.32	D1T077	4-OCH ₃	6.11
D1S064	2- Cl, 4- F	5.38	D1S078	4-Ph	6.92
D1T065	2- CF ₃ , 4- F	4.27	D1S079	3-F, 4-Ph	5.70
D1T066	3- F	5.43	D1T080		5.10
D1T067	3- Cl	5.96	D1T081		5.77
D1T068	3,5- di-F	5.43	D1T082		5.82
D1T069	3,4- di- F	5.96	D1T083		6.62
D1S070	3-Cl, 4-F	5.55	D1S084		6.96
D1T071	4-F	6.55	D1S085		6.96
D1T072	4-Br	6.60	D1T086		6.24
D1T074	4-SCH ₃	6.44			

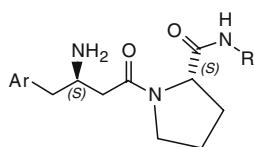
d) 1, 2-Disubstituted pyrrolidine derivatives



Sr. No.	R ₁	R ₂	pIC ₅₀
D2T028		2-CH ₂ COOH	5.44
D2T029		3-CH ₂ COOH	5.48
D2T030		4-CH ₂ COOH	5.38
D2T159			5.14


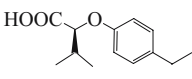

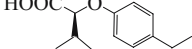
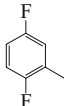
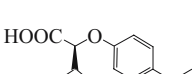
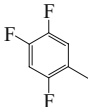
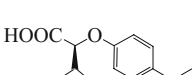
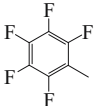
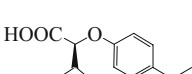
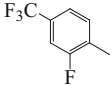
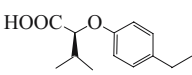
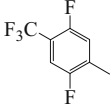
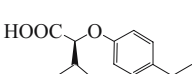
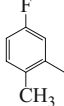
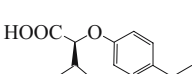

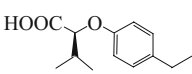
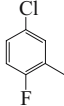
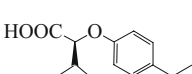
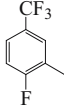
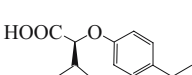
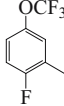
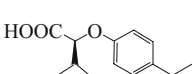
Table 2 (continued)

e) 1, 2-Disubstituted arylpyrrolidine derivatives



Sr. No.	Ar	R	pIC ₅₀
D2S031			5.92
D2T032			5.92
D2T033			6.29
D2T034			7.13
D2T035			7.27
D2S036			6.27
D2T037			6.14
D2S038			6.77
D2T039			7.41
D2T040			5.79
D2T041			6.93
D2T042			7.09
D2T043			7.92
D2T044			7.89

Table 2 (continued)

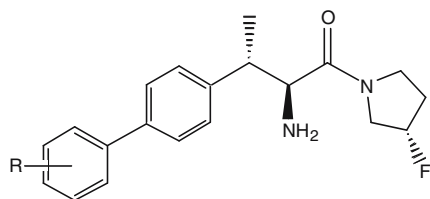
Sr. No.	Ar	R	pIC ₅₀
D2S045			8.75
D2T046			7.75
D2S047			9.32
D2T048			8.80
D2T049			8.09
D2S050			9.08
D2T051			8.02
D2T052			7.48
D2S053			8.77
D2T054			8.8
D2T055			7.92
D2T056			6.53

Receptor preparation for docking

The crystal structure of DPP-IV enzyme with PDB code 2AJL [32] was used for the docking studies. 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. The atom types were corrected for the enzyme and hydrogen atoms were added to the protein. The side chains

Table 2 (continued)

f) Biaryl β -methyl fluoropyrrolidine derivatives

Sr. No.	R	pIC ₅₀	Sr. No.	R	pIC ₅₀
D1T087	2-F	7.00	D1T103	4-NHSO ₂ CH ₃	6.15
D1T088	2-Cl	7.07	D1T104	4-COOC ₂ H ₅	6.36
D1T089	2-OCH ₃	6.96	D1T105	4-COOH	6.43
D1T090	2-CF ₃	6.89	D1T106	4-SO ₂ CH ₃	6.38
D1T091	2-CH ₃	6.82	D1S107	3,5-di-F	6.62
D1T092	3-COOC ₂ H ₅	7.12	D1S108	2,5-di-F	6.82
D1T093	3-NHSO ₂ CH ₃	6.82	D1T109	3,4-di-F	7.22
D1T094	3-COOH	7.25	D1S110	2,4-di-F	7.43
D1T095	3-SO ₂ CH ₃	6.96	D1T111		6.85
D1T096	3-SO ₂ NH ₂	6.85	D1T112		7.16
D1T097	3-NHSO ₂ CF ₃	7.57	D1T113		7.02
D1T098	3-CONH ₂	7.00	D1T114		6.96
D1S099	4-F	7.19	D1T115		7.01
D1T100	4-CF ₃	6.29	D1T116		6.86
D1S101	4-OCH ₃	6.89	D1T117		8.22
D1T102	4-CH ₃	6.41	D1T118		8.52

of aspartate, glutamate, lysine and arginine residues were kept in their ionized state corresponding to pH 8.5, while histidine was unprotonated (corresponding to the δ tautomer) at this pH.

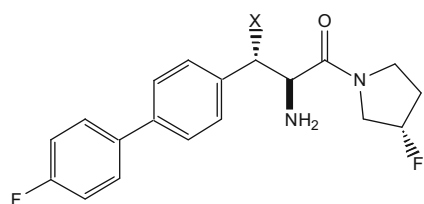
Docking protocol

The docking studies were carried out with the program *GOLD 3.0.1*. The parameters in *GOLD* modified/optimized for the docking studies were: (a) the dihedral angles of the ligand

rotatable bonds; (b) the ligand ring geometries (flipping ring corners); (c) the dihedral angles of protein OH and NH₃⁺ groups; and (d) mappings of the H-bond fitting points. At the start of a docking run, all these variables were randomized. The docking was carried for 20 GA runs, which was optimum to validate the crystal structure ligand. Most of the other GA parameters like the population size and the genetic operators were left at their default values.

All the molecules in the two datasets were docked into the receptor active site shaped by residues in a 10.0 Å

Table 2 (continued)

g) Biaryl β -substituted fluoropyrrolidine derivatives

Sr. No.	X	pIC ₅₀	Sr. No.	X	pIC ₅₀
D1T119	-H	6.01	D1T130	-CONHCH(CH ₃) ₂	7.00
D1S120	-CH ₂ OH	6.41	D1T131		7.15
D1T121	-CH ₂ N(CH ₃) ₂	5.80	D1T132	-CONH(CH ₂) ₂ OH	7.03
D1T122		5.46	D1T133		7.96
D1T123	-COOH	8.18	D1T134	-CON(CH ₃)C ₂ H ₅	7.70
D1T124	-COOCH ₃	6.66	D1T135	-CON(C ₂ H ₅) ₂	7.23
D1S125	-CONH ₂	6.96	D1T136		7.36
D1S126	-CONHCH ₃	7.43	D1T137		8.04
D1S127	-CON(CH ₃) ₂	7.92	D1S138	-CONCH ₃ (OCH ₃)	7.35
D1T128		6.72	D1S139	-CONHN(CH ₃) ₂	7.17
D1T129	-CONHC ₂ H ₅	6.96			

vicinity of the ligand. This active site comprised of Arg¹²⁵, Glu²⁰⁵, Glu²⁰⁶, Phe³⁵⁷, Arg³⁵⁸, Tyr⁵⁴⁷, Ser⁶³⁰, Trp⁶⁵⁹, Tyr⁶⁶⁶, Asn⁷¹⁰ and His⁷⁴⁰ residues.

Poses matching the conformation of the ligand in the crystal structure (PDB code 2BUB) were not reproduced by *GOLD* for molecules in the second dataset. These molecules were thus docked by superimposition on the ligand-FPB (Fig. 1c) in the crystal structure (code 2BUB) followed by minimization that also included enzyme residues within a 10 Å radius of the ligand. The minimizations for the inhibitors were carried out with CFF forcefield under the *Discover 3* module of *InsightII* (v 2005L, Accelrys Inc., USA) running on Pentium IV computer under the Linux RedHat Enterprise 2.3.1 OS [44]. The ligands were optimized to an energy gradient 0.001 kcal/mol/Å.

Docking based alignment

The poses of the ligands with good scores and conformations similar to the crystal structures were taken for the 3D-QSAR

studies. The two dataset were aligned separately by overlaying the active site residues; which is depicted in Fig. 2. This technique is often referred to as the ‘receptor based alignment’ or ‘docking based alignment’ [45].

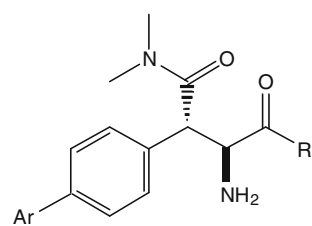
Atom-fit based alignment

The inhibitors in dataset I were also subjected to an atom-fit based alignment, by superimposition on the most active DPP-IV inhibitor, i.e., molecule D1T155 (Table 2h). The atoms chosen for superimposition were those common to all the inhibitors. The alignment based on atom-fit strategy is shown in Fig. 3.

CoMFA and CoMSIA studies

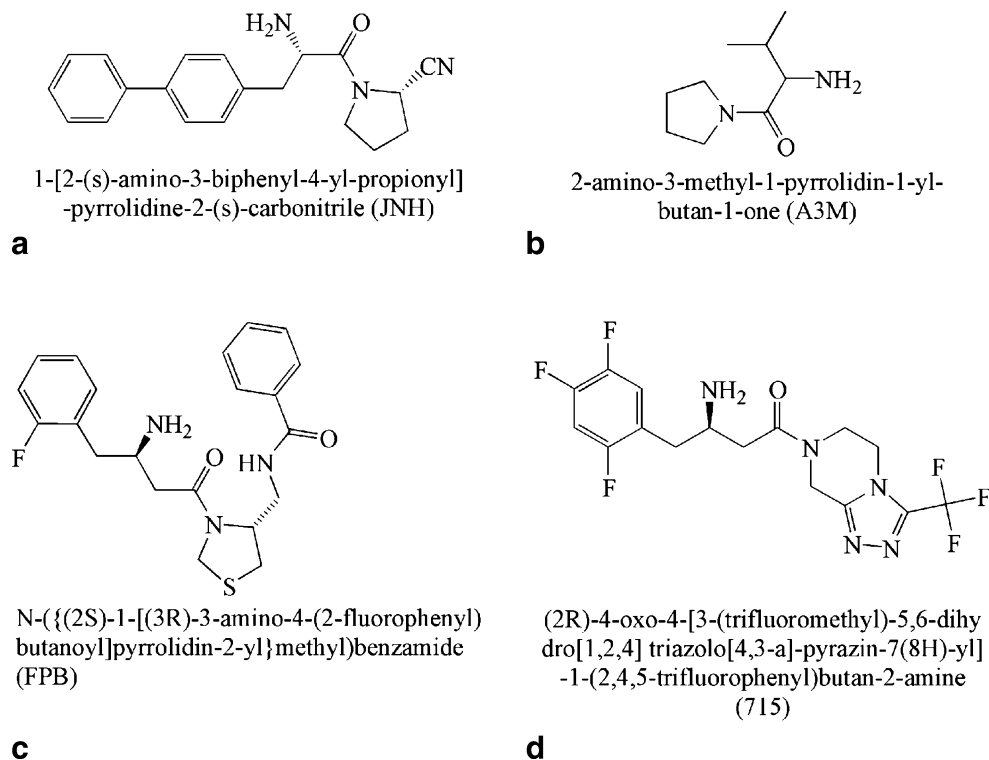
Comparative molecular field analysis (CoMFA) [46] and comparative molecular similarity indices analysis (CoMSIA) [47–49] studies were carried out with the default settings for the 3D cubic lattice, the grid spacing, the probe atom, and the cutoff for the interaction energy.

Table 2 (continued)

h) Biaryl β -substituted aza-heterocyclic derivatives

Sr. No.	Ar =	R =	pIC ₅₀
D1S140			7.00
D1S141			7.77
D1T142			7.54
D1T143			8.13
D1S144			7.52
D1T145			7.57
D1T146			8.04
D1T147			6.64
D1T149			6.26
D1T150			6.28
D1T151			7.89
D1T152			7.68
D1T153			7.25
D1T154			7.37
D1T155			8.66
D1T156			7.68
D1T157			7.72
D1T158			7.70

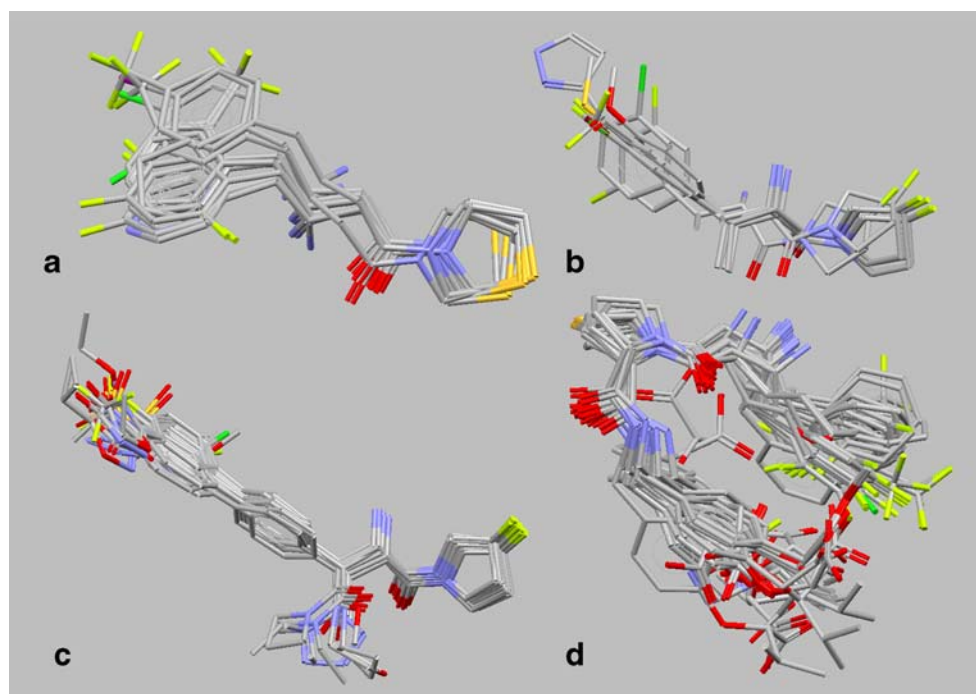
Fig. 1 a–d Structures of DPP-IV inhibitors used as templates to build the molecules in datasets I and II



For CoMSIA, five physicochemical properties (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were explored, and the attenuation factor was set to the default value of 0.3. The CoMFA and CoMSIA

descriptors were used as independent variables and the pIC_{50} values as the dependent variable in a partial least squares (PLS) regression analysis [50] to derive 3D-QSAR models. The models were internally evaluated by leave-

Fig. 2 Receptor based alignment of (a) thiazolidine subclass (b) fluoropyrrolidine subclass (c) biphenyl fluoropyrrolidine subclass and (d) bisubstituted pyrrolidine/thiazolidine subclass (these classes have been shown separately for clarity)



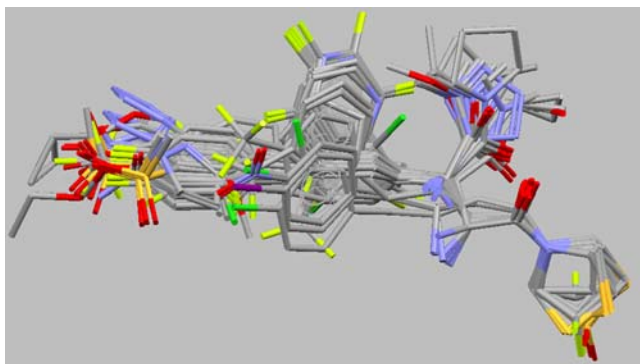


Fig. 3 Atom-fit alignment of DPP-IV inhibitors

one-out (LOO) cross-validation. The optimal number of components was determined by the SAMPLS [51] method which was subsequently used to derive the final QSAR models. In addition to the q^2 , the conventional correlation coefficient r^2 , the standard errors (SE) and the F-value were also computed. The robustness of the models was gauged by cross validation using leave-group-out (LGO) of 10 groups and bootstrapping [52] carried out with 100 runs.

Results and discussion

A CoMFA study was carried out on a set of dipeptides in the last decade by Brandt et al. as mentioned earlier. The study proposed that a proline or an azetidine ring was important for activity; this group probably interacts with the aromatic ring of a Phe/Tyr residue in the active site of the enzyme. It was also suggested that bulky substitutions are

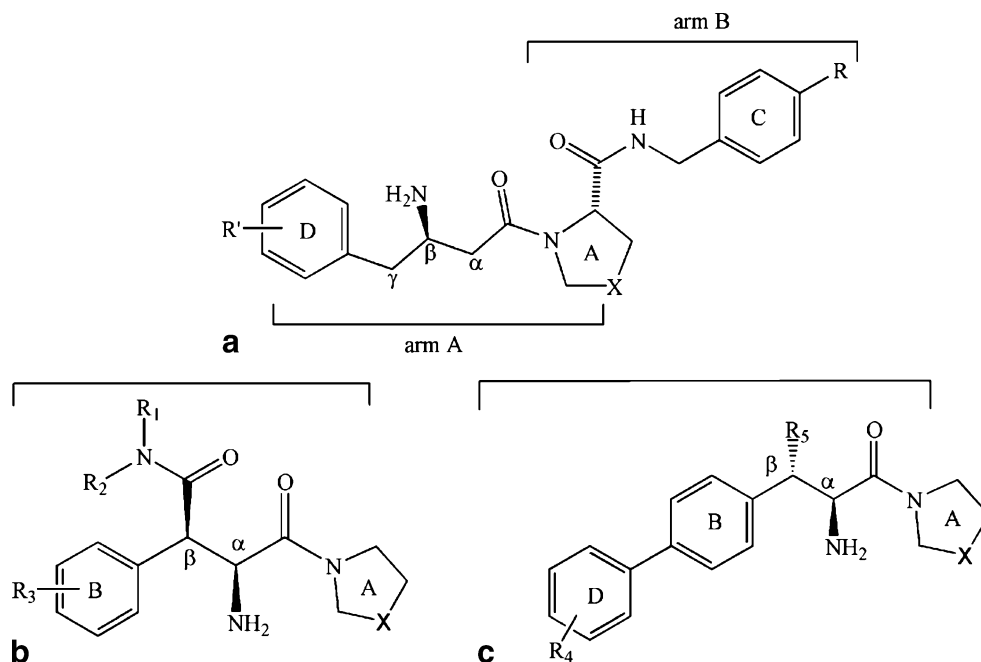
favoured at arm A (Fig. 4a), which has now been confirmed by recent research [53]. However, it was proposed that substitution on the proline residue (Arm B, Fig. 4a) would diminish the activity, which contradicts recent evidence that substitution on the proline actually enhances the activity. The advent of substituted prolines as DPP-IV inhibitors with 9 log unit activity, the release of several DPP-IV crystal structures, the short-comings of the previous CoMFA study mentioned above, and the altered binding mode for some of the potent DPP-IV inhibitors, prompted us to investigate the large variation in the activity and binding of these molecules.

Docking studies

The catalytic site of DPP-IV is located near the *C-terminal* extracellular end of the enzyme [54]. The active site of the enzyme lies amidst five-amino acids (Gly⁶²⁸, Trp⁶²⁹, Ser⁶³⁰, Tyr⁶³¹, and Gly⁶³²) positioned around Ser⁶³⁰, which is referred to as the S₁ pocket [55, 56]. Several other amino acids including Tyr⁵⁴⁷, Glu²⁰⁵ and Glu²⁰⁶ are also involved in ligand binding [7, 57]. The cavity near Arg¹²⁵, Glu²⁰⁵, and Glu²⁰⁶ residues is referred to as the S₂ pocket.

The DPP-IV inhibitors considered for the CoMFA/CoMSIA studies are quite diverse with respect to structure and activity. Any commendable correlation in a QSAR study is immensely sensitive to the conformation of the ligands. In general, the global minimum energy structure or the bioactive conformation is used in 3D-QSAR studies. Since in the present case, several crystal structures of enzyme-inhibitor complexes are available, the 3D structures

Fig. 4 a–c General structure of inhibitors with a description of the nomenclature used in the text



of inhibitors were built using ligands co-crystallized with the enzyme as the template. For dataset I, docking was carried out with the program *GOLD* and the best docked poses were analyzed, and those that matched the crystal structure were used for the alignment.

However docking with the program *GOLD* for molecules in dataset II could not reproduce the binding pose found in the crystal for the closely related ligand. In the crystal, the analogous molecule has an altered binding mode as compared to molecules in dataset I. These molecules were therefore docked by superimposition on the binding pose of the ligand in the crystal (PDB code 2BUB). The ligands were then minimized in the active site of the protein keeping the proline ring fixed while the ligand arms A and B were set free for minimization. The energy minimized ligands in their binding poses were then used for the alignment.

Inhibitors with a fluoropyrrolidine ring (e.g., molecules D1S061-64, D1T065-069, D1S070, D1T071-077, D1S078-079, D1T080-083, D1S084-85, and D1T086) were found to fit snugly in the S_1 pocket in the vicinity of Ser⁶³⁰, which explains their higher affinity when compared with the non-fluorinated ligands. In the atom-fit alignment approach with the fluoropyrrolidine ring superimposed on the thiazolidine ring, the activity of the fluoropyrrolidine analogs were found to be predicted with a lower activity; while the superimposition of the thiazolidine ring over the fluoropyrrolidine ring resulted in a higher prediction of activity of the thiazolidines analogs in the CoMFA models (*vide infra*). The slightly displaced overlay of these two classes of molecules - the pyrrolidines and thiazolidines in the

receptor-based alignment, reveals a crucial relationship between their activity and position in the active site, which cannot be accounted for by the traditional atom-fit alignment. The QSAR model, generated with the receptor based alignment was seen to exhibit the best correlation between the predicted and experimental pIC₅₀ values of the inhibitors. Such a good correlation could not be achieved with the traditional atom fit alignment, as these conformations/alignments may not be the true compliments of their interaction with the DPP-IV enzyme.

CoMFA and CoMSIA models

The experimental and X-ray crystallographic data dictated the splitting of the molecules into two datasets. Molecules in dataset II are closely related to FPB [N-({(2S)-1-[(3R)-3-amino-4-(2-fluorophenyl)butanoyl]pyrrolidin-2-yl}methyl)benzamide - PDB code 2BUB] which has a snug fit in the enzyme site compared to molecules of dataset I. The tight fitting of these molecules in the active site is reflected in their higher binding affinity as revealed in the experimental data. The CoMFA and CoMSIA studies were initially started with training sets of 74 and 25 molecules aligned by the receptor based technique. However, the activities of a few molecules were predicted by the CoMFA models with residuals greater than 0.80. Incorrect prediction of activities for some molecules could be attributed to their inaccurate conformations. These molecules were reintroduced into the dataset with alternate conformations and the QSAR models regenerated. The models were improved upon till the cross-

Table 3 PLS statistics of CoMFA and CoMSIA 3D-QSAR models derived by receptor based and atom fit alignments

PLS statistics	Receptor based alignment (dataset I)		Receptor based alignment (dataset II)		Atom fit alignment (dataset I)	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
n	74	74	25	25	74	74
r ²	0.95	0.96	0.91	0.87	0.94	0.94
q ² (LOO)	0.58	0.71	0.62	0.62	0.55	0.68
q ² (LGO)	0.63	0.70	0.54	0.64	0.56	0.68
r _{bootstrap} ²	0.96	0.97	0.94	0.90	0.96	0.96
r _{pred} ²	0.69	0.66	0.73	0.58	0.68	0.65
F	201.62	252.50	66.95	72.49	165.08	200.26
SE	0.23	0.21	0.35	0.41	0.25	0.25
PLS components	6	6	3	2	6	5
Field contribution						
Steric	0.40	0.076	0.58	0.103	0.43	0.087
Electrostatic	0.60	0.343	0.42	0.218	0.57	0.317
Hydrophobic	–	0.187	–	0.175	–	0.190
Donor	–	0.190	–	0.209	–	0.205
Acceptor	–	0.205	–	0.295	–	0.200

Table 4 The experimental and predicted pIC₅₀ values for the training set molecules of dataset I

Molecule ID	Experimental pIC ₅₀	Predicted pIC ₅₀ (CoMFA) Receptor based model	Predicted pIC ₅₀ (CoMSIA)	Predicted pIC ₅₀ (CoMFA) Atom fit model	Predicted pIC ₅₀ (CoMSIA)
D1T001	5.52	5.45	5.69	5.27	5.46
D1T003	4.40	4.58	4.59	4.83	4.57
D1T004	4.30	4.16	3.92	4.28	4.04
D1T005	4.29	4.66	4.19	4.59	4.30
D1T008	4.56	4.61	4.71	4.81	4.76
D1T010	6.03	5.89	5.73	5.70	5.50
D1T011	6.01	5.77	5.89	5.52	5.91
D1T013	5.64	5.63	5.65	5.41	5.56
D1T016	5.44	5.60	5.44	5.41	5.36
D1T017	5.48	5.35	5.68	5.32	5.60
D1T019	4.97	4.98	5.20	5.34	5.25
D1T020	5.11	5.31	5.15	5.47	5.70
D1T025	6.57	6.55	6.64	6.50	6.50
D1T026	5.69	5.67	5.54	5.45	5.48
D1T027	6.92	6.97	6.97	6.85	6.96
D1T066	5.43	5.51	5.68	5.68	5.69
D1T067	5.96	6.16	6.08	6.07	6.07
D1T068	5.43	5.50	5.74	5.67	5.73
D1T069	5.96	5.48	5.67	5.73	5.73
D1T072	6.60	6.48	6.49	6.42	6.59
D1T073	6.77	6.46	6.40	6.40	6.51
D1T074	6.44	6.29	6.29	6.38	6.45
D1T075	6.22	6.52	6.49	6.50	6.63
D1T077	6.11	6.03	6.25	6.11	6.16
D1T081	5.77	5.74	5.76	5.78	5.90
D1T083	6.62	6.36	6.42	6.44	6.43
D1T086	6.24	6.46	6.37	6.42	6.56
D1T087	7.00	7.32	6.86	7.21	6.84
D1T088	7.07	6.94	6.66	7.10	6.74
D1T089	6.96	7.04	6.80	7.10	6.80
D1T091	6.82	6.99	6.67	6.83	6.67
D1T092	7.12	7.09	7.00	7.26	6.90
D1T093	6.82	6.94	7.07	7.00	6.91
D1T094	7.25	7.01	7.36	7.03	7.33
D1T095	6.96	6.83	6.96	7.08	6.72
D1T096	6.86	6.80	6.89	6.91	6.86
D1T097	7.57	7.37	7.54	7.29	7.30
D1T098	7.00	7.00	6.98	6.88	7.03
D1T100	6.29	6.20	6.22	5.95	6.43
D1T102	6.41	6.49	6.40	6.60	6.57
D1T104	6.36	6.24	6.42	6.11	6.42
D1T105	6.43	6.36	6.37	6.54	6.40
D1T106	6.38	6.18	6.24	6.19	6.36
D1T109	7.22	6.76	6.75	6.97	6.66
D1T111	6.85	7.55	7.08	7.42	7.47
D1T112	7.16	7.14	7.16	7.06	7.11
D1T113	7.02	6.85	7.22	6.86	6.93
D1T114	6.96	6.83	6.99	7.08	6.82
D1T115	7.01	6.90	7.05	6.89	6.89
D1T116	6.85	6.97	6.96	6.85	6.89
D1T118	8.52	8.32	8.46	8.05	8.64
D1T119	6.01	6.46	6.48	6.19	6.41
D1T121	5.80	6.13	6.22	6.05	6.47
D1T124	6.66	6.94	6.56	6.69	6.50

Table 4 (continued)

Molecule ID	Experimental pIC ₅₀	Predicted pIC ₅₀ (CoMFA) Receptor based model	Predicted pIC ₅₀ (CoMSIA)	Predicted pIC ₅₀ (CoMFA) Atom fit model	Predicted pIC ₅₀ (CoMSIA)
D1T128	6.72	6.66	6.75	6.78	6.61
D1T129	6.96	7.29	7.23	7.06	6.82
D1T130	7.00	6.96	6.89	7.09	7.00
D1T131	7.15	7.14	7.04	7.15	6.94
D1T132	7.03	7.17	7.26	7.18	7.07
D1T134	7.70	7.31	7.59	7.05	7.66
D1T135	7.23	7.51	7.44	7.44	7.58
D1T136	7.36	7.35	7.37	7.10	7.18
D1T137	8.04	8.07	7.81	7.89	8.13
D1T143	8.12	7.81	7.81	7.81	7.77
D1T145	7.57	7.45	7.74	7.73	7.65
D1T146	8.04	8.01	8.12	8.42	7.96
D1T147	6.64	7.04	6.50	6.71	6.49
D1T151	7.89	7.61	7.80	7.58	7.72
D1T152	7.68	7.45	7.61	7.79	7.61
D1T153	7.53	7.72	7.83	7.71	8.11
D1T155	8.66	8.54	8.29	8.49	8.53
D1T156	7.68	7.91	7.72	7.97	7.83
D1T157	7.72	7.59	7.75	7.97	7.52
D1T158	7.70	7.92	7.78	7.87	7.66

Table 5 The experimental and predicted pIC₅₀ values for the test set molecules of dataset I

Molecule ID	Experimental pIC ₅₀	Predicted pIC ₅₀ (CoMFA) Receptor based model	Predicted pIC ₅₀ (CoMSIA)	Predicted pIC ₅₀ (CoMFA) Atom fit model	Predicted pIC ₅₀ (CoMSIA)
D1S014	5.69	4.62	5.26	4.93	5.41
D1S018	5.52	5.11	5.29	5.37	5.28
D1S022	5.81	5.28	5.70	5.30	5.35
D1S023	5.77	6.58	6.905	6.52	6.81
D1S024	6.23	5.64	6.22	5.93	6.29
D1S061	5.85	5.52	5.70	5.69	5.71
D1S062	5.96	5.52	5.69	5.70	5.70
D1S063	6.32	6.85	6.56	6.52	6.51
D1S064	5.38	5.99	6.23	5.84	6.22
D1S070	5.55	5.36	5.65	5.74	5.71
D1S078	6.92	7.11	6.75	6.82	6.64
D1S079	5.70	6.61	6.45	6.83	6.59
D1S084	6.96	6.69	6.11	6.78	6.43
D1S085	6.96	6.90	6.64	7.04	6.50
D1S099	7.19	6.62	6.52	6.98	6.50
D1S101	6.89	6.94	6.64	6.68	6.58
D1S107	6.62	6.85	6.66	7.08	6.82
D1S108	6.82	7.07	6.74	7.20	6.86
D1S110	7.43	7.03	6.70	7.04	6.68
D1S120	6.41	6.80	5.89	6.92	6.12
D1S125	6.96	6.80	7.17	6.74	7.45
D1S126	7.43	7.44	7.48	7.36	7.47
D1S127	7.92	7.25	7.52	7.10	7.64
D1S138	7.35	7.15	6.88	7.59	7.32
D1S139	7.17	7.48	7.41	7.43	6.99
D1S140	7.00	7.37	7.18	7.69	7.68
D1S141	7.77	7.48	7.59	7.75	7.66
D1S144	7.52	7.40	7.52	7.78	7.63

Table 6 The experimental and predicted pIC₅₀ values for the training set molecules of dataset II

Molecule ID	Experimental pIC ₅₀	Predicted pIC ₅₀ (CoMFA)	Predicted pIC ₅₀ (CoMSIA)
D2T028	5.44	5.59	5.13
D2T029	5.48	5.07	5.19
D2T030	5.38	5.07	5.44
D2T032	5.92	5.99	6.45
D2T033	6.29	6.89	7.03
D2T034	7.12	7.30	7.15
D2T035	7.27	7.44	7.22
D2T037	6.14	6.68	6.37
D2T039	7.41	7.25	7.51
D2T040	5.78	6.20	6.52
D2T041	7.33	7.09	7.01
D2T042	7.09	7.35	7.18
D2T043	7.92	7.58	7.53
D2T044	7.89	7.99	7.52
D2T046	7.74	7.74	7.29
D2T048	8.80	8.56	8.20
D2T049	8.09	8.37	8.26
D2T051	8.02	8.33	8.28
D2T052	7.48	8.22	8.04
D2T054	8.80	8.51	8.18
D2T055	7.92	8.51	8.19
D2T056	6.53	6.87	6.74
D2T057	6.79	6.73	6.67
D2T059	7.43	7.41	7.12
D2T159	5.14	4.91	5.02

validation and regression statistics were in the acceptable range. The statistical parameters for the 3D-QSAR models (Table 3) generated with the atom-fit alignment are not as robust as the models based on the receptor based alignment as evident from Table 3.

The CoMFA and CoMSIA models for datasets I and II were tested for their predictivity on an external dataset of 28 and 9 diverse molecules. The predictive r^2 (r_{pred}^2) for the two datasets are given in Table 3. The decent r_{pred}^2 values for the models generated using receptor based alignment is an indication that the receptor based alignment can adequately account for the interactions these diverse molecules make with the enzyme active site, which may not always be accounted by the simple atom-fit based alignment. Comparison of predicted pIC₅₀ values with experimental values for the training and test sets of datasets I and II are given in Tables 4, 5, 6, and 7 and correlation plots of the same are shown in Figs. 5, 6, and 7.

Graphical interpretation of the results

The CoMFA and CoMSIA contour maps were generated using scalar products of coefficients and standard deviation

(STDEV*COEFF) set at 80% and 20% for favored and disfavored levels, respectively.

Analysis of the contours for CoMFA and CoMSIA models

The CoMFA and CoMSIA contours around the most active molecule in the dataset I are shown in Figs. 8 and 9, and around the most active molecule in dataset II are shown in Figs. 10 and 11, respectively.

Dataset I

The electrostatic contribution to CoMFA model is greater than its steric counterpart. The electronegative fields are observed around the α -amino group and the β substituent in arm A (Fig. 4). Both electropositive and electronegative fields are seen in the vicinity of the di-substituted amide at the C_{β} position. The biaryl analogs (Fig. 4c) have electronegative contours near one of the *ortho* positions of ring B and the *meta* position of ring D. Electronegative and electropositive fields are observed at the *ortho* and *ortho'* positions of ring D respectively. A fluoro group at the C_3 position is a favored electronegative group that fits the requirement of the electrostatic contour; on the other hand, electropositive substituents are favored at the C_4 position. Steric bulk at the *ortho*- and *meta*-positions on ring D is predicted to improve the activity. For molecules with a C_{β} amide (Fig. 4b) the R_2 group should be bulky, while steric bulk at ring A, ring B, and R_1 are less preferred.

The CoMSIA contour maps shown in Fig. 9c to e, around the most active molecule (molecule D1T155, Table 2h), provide additional information on hydrophobic, H-donor and H-acceptor substitutions not available in the CoMFA maps. The steric contours of CoMSIA indicate that

Table 7 The experimental and predicted pIC₅₀ values for the test set molecules of dataset II

Molecule ID	Experimental pIC ₅₀	Predicted pIC ₅₀ (CoMFA)	Predicted pIC ₅₀ (CoMSIA)
D2S031	5.92	6.31	6.24
D2S036	6.27	6.03	6.44
D2S038	6.77	6.87	6.88
D2S045	8.74	8.37	8.08
D2S047	9.32	8.52	8.15
D2S050	9.38	8.17	8.24
D2S053	8.77	8.41	8.11
D2S058	7.06	6.90	6.89
D2S060	9.52	8.66	6.68

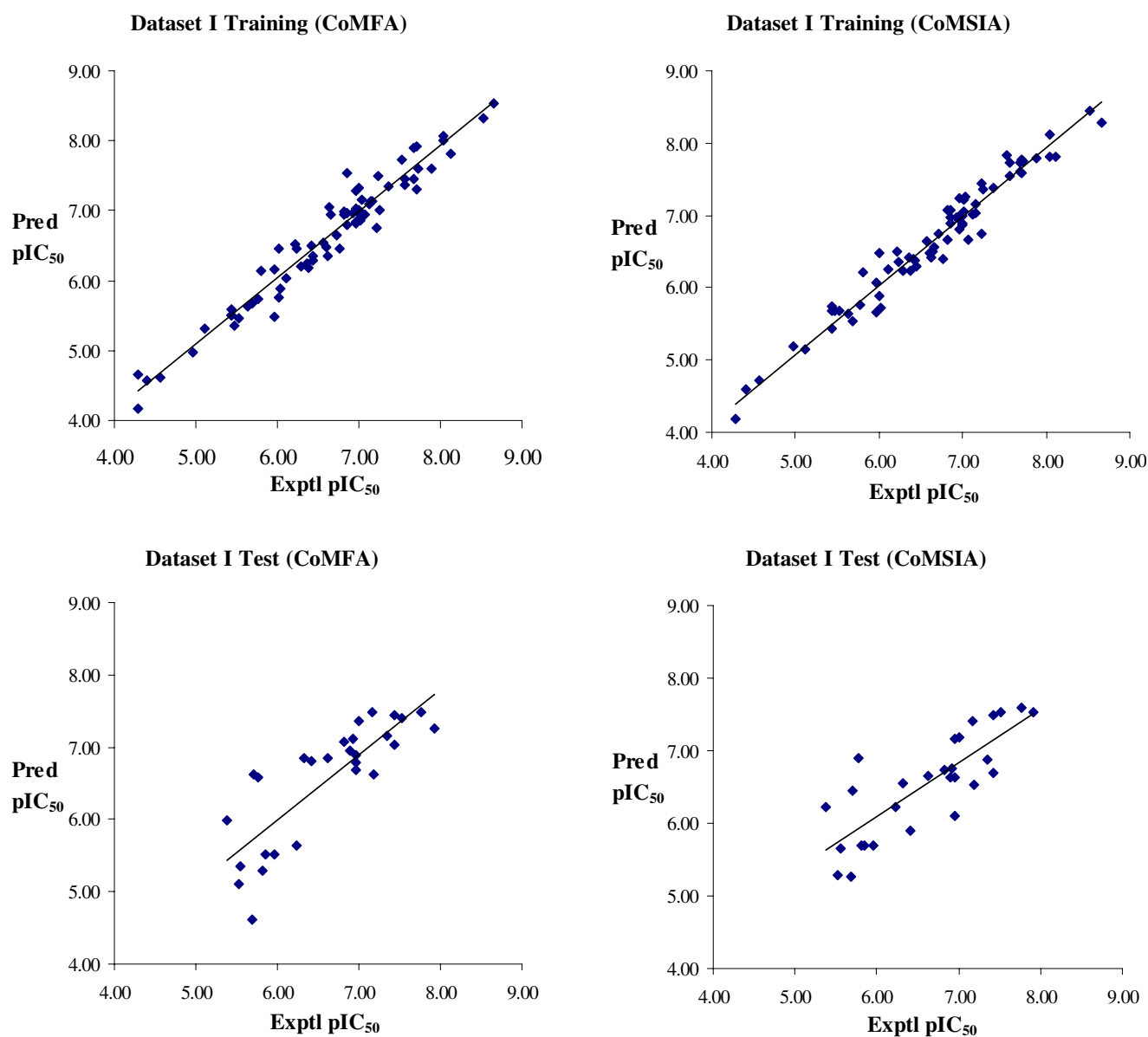


Fig. 5 The predicted vs experimental correlation plots for the training and test sets molecules of dataset I for CoMFA and CoMSIA models derived for receptor based alignment

any increase in bulk on the ligands will be detrimental to the activity, with exception that large substituents on ring D will improve the activity; which is also predicted by CoMFA (Fig. 8a). The CoMSIA contours (Fig. 9b) are in harmony with the CoMFA electrostatic fields (Fig. 8b).

The C_{β} substituent (R_2 group) at arm A should be hydrophilic (white color contour) (Fig. 9c) just as hydrophobic groups on some parts of ring A are disfavored. The hydrophobic contour surrounding most parts of ring B indicate that activity can be enhanced by increasing hydrophobicity, with the exception that hydrophobic substitution at the *meta* and *ortho* positions will diminish activity.

H-bond donor groups at the C_{β} position (R_2 group) and the C_5 position of ring A are disfavored, but favored at the C_2 position of ring A and on the *ortho* position of ring D (Fig. 9d). H-acceptor groups (magenta colored contour) are preferred at the *para* position and disfavored (red colored contour) around the *meta* position of ring D (Fig. 9e). Both the H-bond donor and acceptor characteristics are associated with the substituted amide at the C_{β} position as seen by the distribution of electrostatic contours at this position. H-acceptor groups are not favored around ring A, with the exception of the fluoro group at the C_3 position; any replacement of the fluoro atom by another halogen or an H-

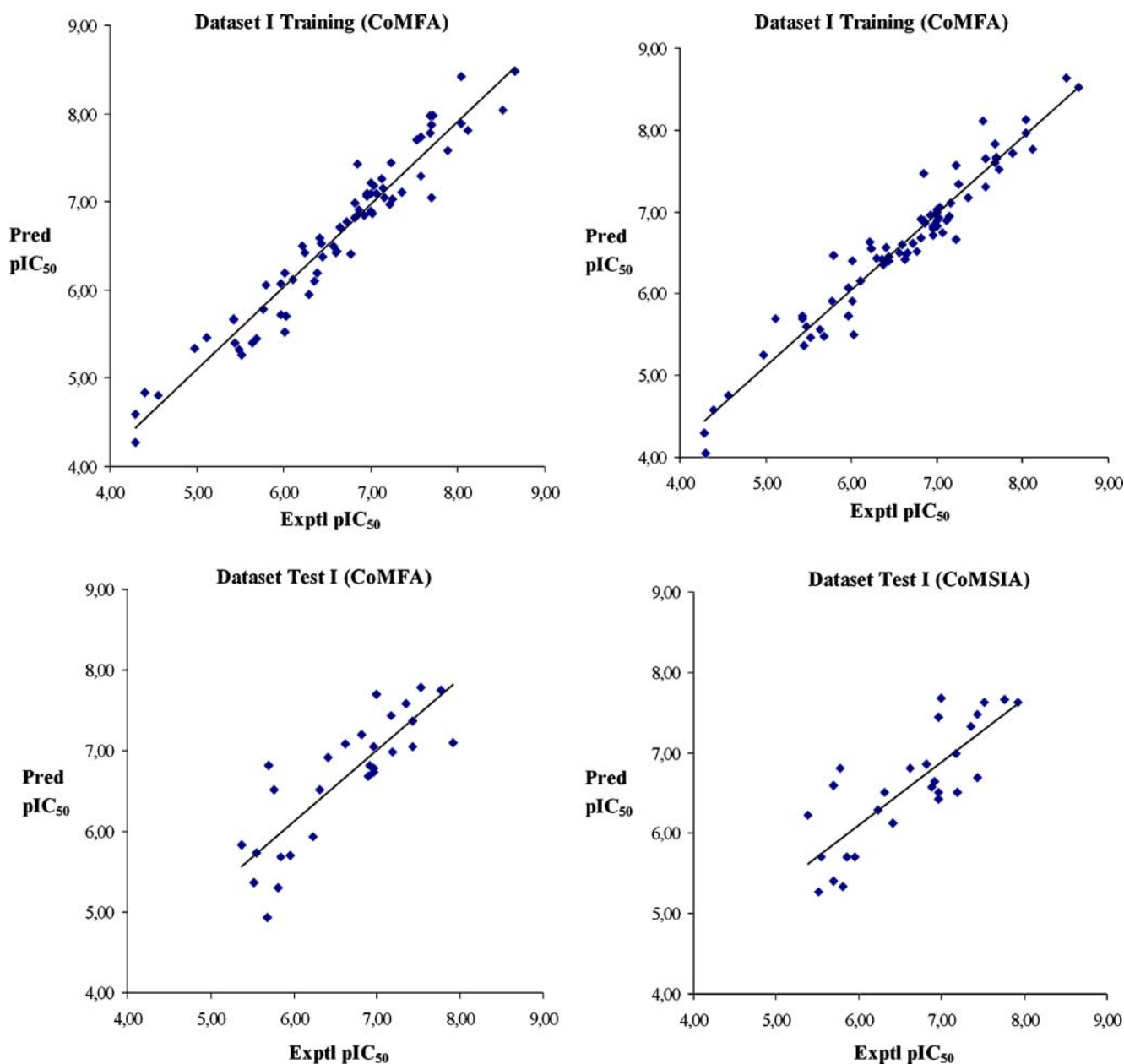


Fig. 6 The predicted vs experimental correlation plots for the training and test sets molecules of dataset I for CoMFA and CoMSIA models derived for atom fit alignment

acceptor group like $-\text{SO}_2-$ will result in a decrease in the activity.

Dataset II

The molecules of dataset II have a different binding mode compared to the molecules of dataset I which probably accounts for their 3 log unit higher activity. The CoMFA and CoMSIA contours for molecules of dataset II are

different due to the altered mode of binding for these molecules. The CoMFA and CoMSIA contours for dataset II were analyzed around the most active molecule (molecule D2T054; Table 2e) of the series.

Steric bulk (Fig. 10a) is observed to be disfavored around most positions of the molecules with exception at some sites on rings C and D. In most molecules of this type, the R group in arm B is 2-oxa-3-methylbutyric acid (Fig. 4a). A strong favor for steric bulk is observed at the C

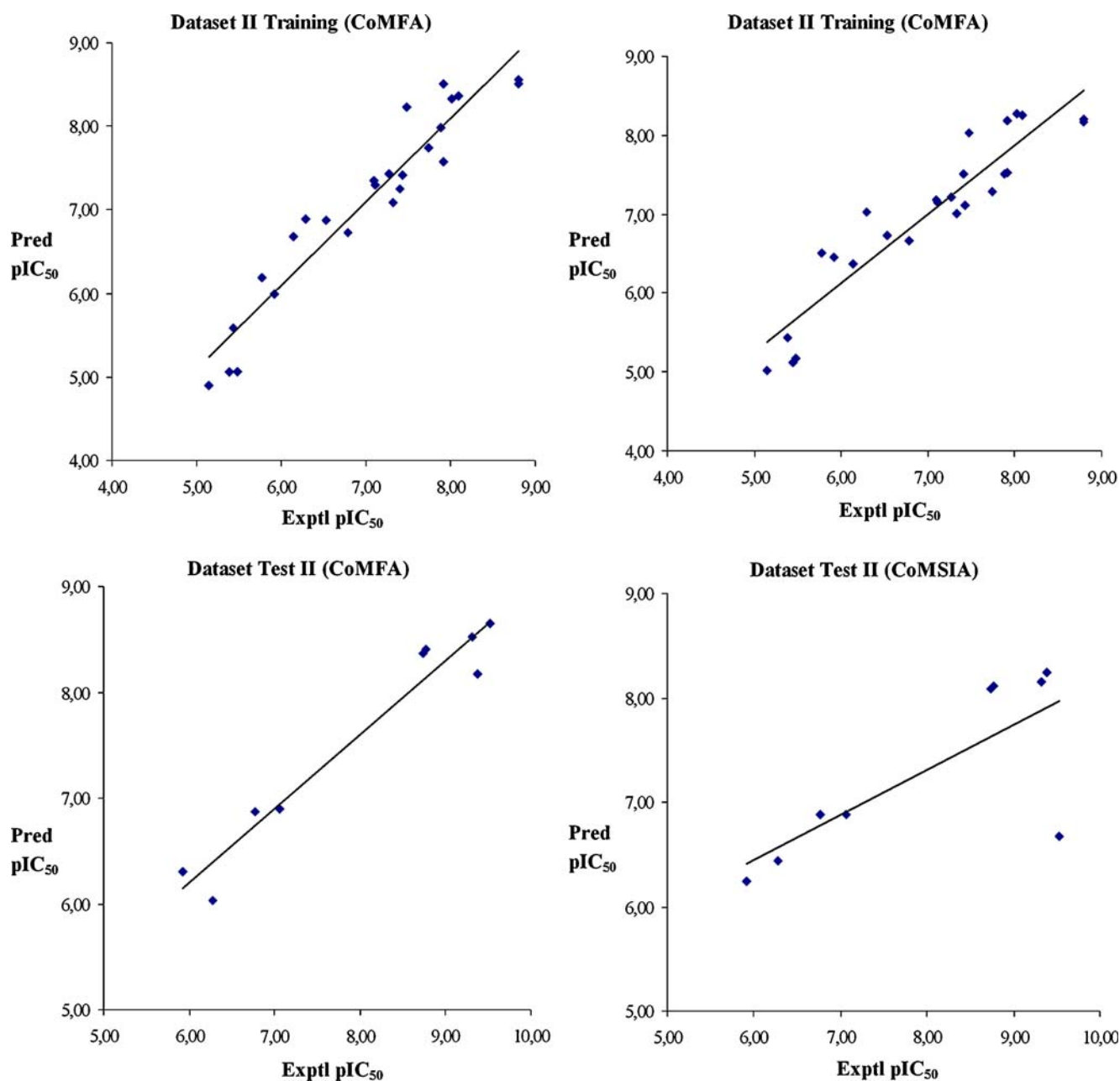


Fig. 7 The predicted vs experimental correlation plots for the training and test sets molecules of dataset II for CoMFA and CoMSIA models derived for receptor based alignment

ring; bulky esters replacing the carboxylate may have abbreviated potencies. The *meta* position in ring D is tolerant to some steric bulk.

Prominent electropositive contours are found around ring C and the benzylic position in arm B extending up to the carboxylate group; at the same time electronegative fields are seen to be favored over the isopropyl (R group) substituent (Fig. 10b).

The CoMSIA steric and electrostatic contours (Fig. 11a and b) are identical to their CoMFA counterparts with the exception that the electronegative group at the isopropyl (R) substituent is not seen, rather it points to the ether linkage in arm B.

The hydrophobic contours, as seen in Fig. 11c, exhibit an interesting pattern. Hydrophobic substituents are suggested at the C_α and *para* positions of ring B and

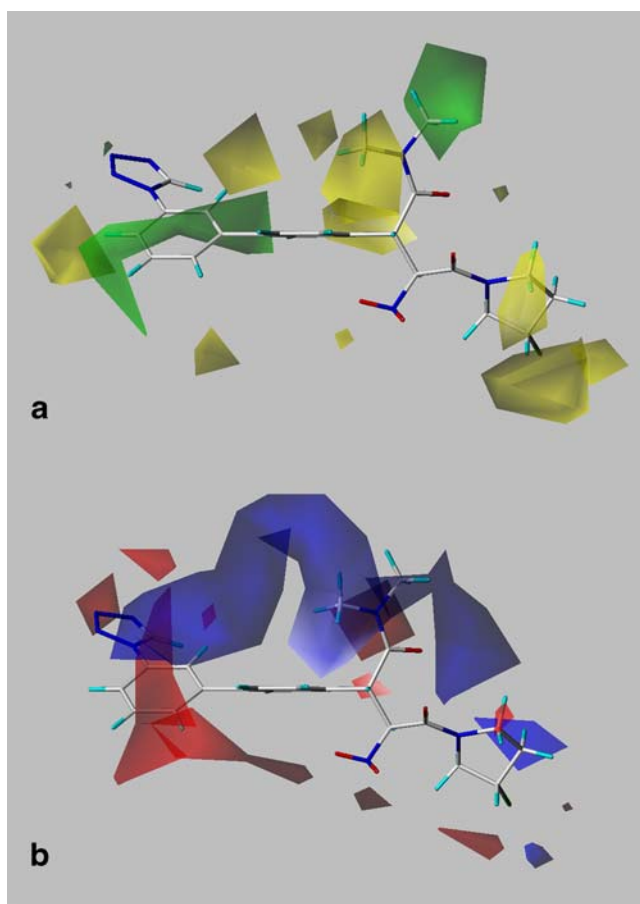
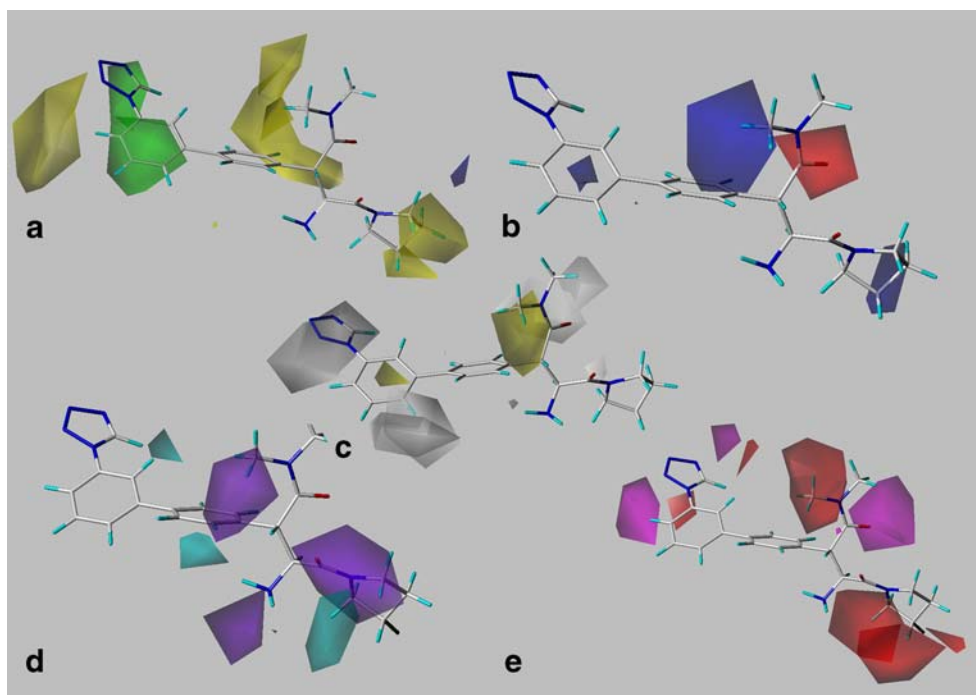


Fig. 8 The CoMFA contours around molecule D1T155 (a) steric contours - favored (green); disfavored (yellow) (b) electrostatic contours - electropositive (blue); electronegative (red)

Fig. 9 The CoMSIA contours around molecule D1T155 (a) steric contours - favored (green), disfavored (yellow) (b) electrostatic contours - electropositive (blue) electronegative (red) (c) hydrophobic contours - favored (yellow), disfavored (white) (d) H-Donor contours - favored (cyan) (e) H-Acceptor contours - favored (magenta), disfavored (red)



disfavored at the C_{β} - position in arm A and the benzylic methylene in arm B.

A hydrogen donor group is not preferred at the C_5 position of ring A, the C_{α} and C_{β} positions in arm A and at the *ortho* positions in ring B. The β -amino arm A and the amide nitrogen in arm B satisfy the need of hydrogen donor groups (Fig. 11d).

The need of a hydrogen acceptor group is satisfied by the carboxylate moiety in the R substituent of arm B, while hydrogen acceptor groups are disfavored at the C_2 position of ring A, the C_{α} position in arm A and the *meta* position in ring C (Fig. 11e).

The least active inhibitor (molecule D2T159; 2 d) has a bulky *cyclohexyl* moiety in place of ring B and hence on minimization adopts a conformation different from the rest of the series. Consequently it does not satisfy the steric and electrostatic requirements essential for good activity.

The CoMFA and CoMSIA studies for each dataset cannot be used to explain SAR of its counterpart, because of their varied binding modes and alignments. Also the CoMFA and CoMSIA studies are unable to throw light on the large differences of the binding modes of the molecules in the two datasets and the consequent effect on their respective activities. Any tool that can treat the two datasets as a single entity would have the power to rationalize the differences in their binding and activity. Receptor based methods like comparative residue interaction analysis (CoRIA) have the functionalities that can provide insights

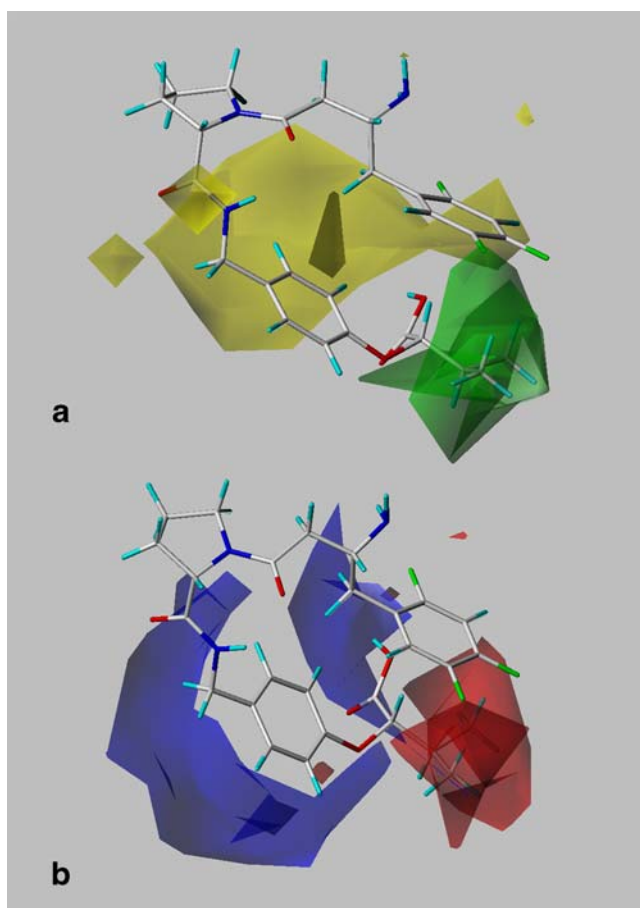
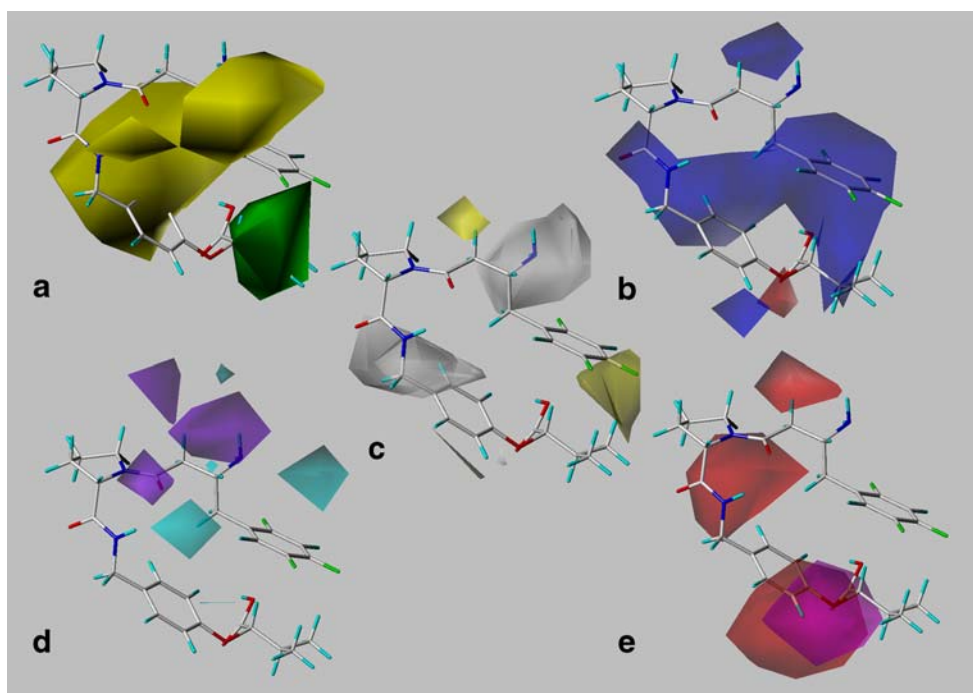


Fig. 10 The CoMFA contours around molecule D2T054 (a) steric contours - favored (green); disfavored (yellow) (b) electrostatic contours - electropositive (blue); electronegative (red)

Fig. 11 The CoMSIA contours around molecule D2T054 (a) steric contours - favored (green), disfavored (yellow) (b) electrostatic contours - electropositive (blue) electronegative (red) (c) hydrophobic contours - favored (yellow), disfavored (white) (d) H-Donor contours - favored (cyan) (e) H-Acceptor contours - favored (magenta), disfavored (red)



into the different attributes of these two classes of molecules [58].

Receptor topology and contour map correlation

CoMFA The electropositive (blue) contours and the electronegative (red) contours of the CoMFA model associated with the most active molecule, i.e., D1T155 are seen occupying the region enclosing Arg¹²⁵, His⁷⁴⁰, Ser⁶³⁰, Glu²⁰⁵ and Glu²⁰⁶. These residues being ionized at pH 8.5 nicely meet the requirements of electrostatic interactions as demanded by CoMFA.

In the context of steric interactions, the bulk favoring contours (green colored) are seen around residues Phe³⁵⁷ and Arg³⁵⁸ and also in the cavity filled by residues Arg¹²⁵, Arg⁷¹⁰, His⁷⁴⁰, and Ser⁶³⁰. On the other hand, bulk disfavoring interactions as indicated by the yellow colored contours, line the pocket composed of Glu²⁰⁵, Glu²⁰⁶, Arg¹²⁵, Asn⁷¹⁰, and Trp⁶⁵⁹ and a second pocket made by residues Ser⁶³⁰ and Asn⁷¹⁰.

CoMSIA Additional information in CoMSIA over CoMFA in terms of hydrogen bonding and hydrophobic interactions with the receptor. CoMSIA contours indicating hydrogen donor and acceptor aptitudes are seen in the proximity of Glu²⁰⁵ and Glu²⁰⁶ and near Tyr⁵⁴⁷, Ser⁶³⁰, and Arg³⁵⁸.

Contours favoring hydrophobic interactions (yellow color) are seen near Phe³⁵⁷ while disfavored hydrophobic interactions (white color) lie in the pocket made by Trp⁶⁵⁹,

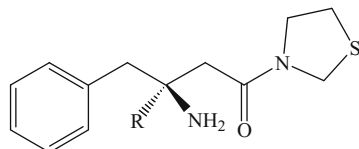
Ser⁶³⁰, and Asn⁷¹⁰. The hydrophobic (Phe) and polar (Trp, Ser, Asn) nature of these residues are in line with the character of the contours.

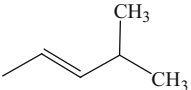
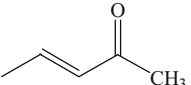
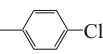



Design of new molecules based on CoMFA and CoMSIA studies

The CoMFA and CoMSIA models described in the previous section were used to design new molecules. The

design strategy involved modifications in the substitution pattern of some molecules selected from both datasets. Molecule D1T001 ($pIC_{50}=5.52$) one of the low active, and molecule D1T155 ($pIC_{50}=8.66$) the highest active molecule, were selected from dataset I and molecule D2S060 ($pIC_{50}=9.52$) was chosen from dataset II. Modifications were made on the β -amino group in arm A of molecule D1T001 (R group in Table 8) and molecule D1T155 (R group in Table 9) and in the *para* position of ring C of

Table 8 The CoMFA and CoMSIA predictions for newly designed DPP-IV inhibitors based on molecule D1T001 as the template



Sr. No.	R	Predicted pIC_{50}	
		CoMFA	CoMSIA
N1001	$-C_2H_4 N(CH_3)_2$	6.66	6.12
N1002		7.0	5.77
N1003		6.56	5.57
N1004	<i>n</i> Pentyl	6.51	5.86
N1005		6.79	5.69
N1006		6.99	5.61
N1007		6.54	5.53
N1008	$-C_2H_4-iPr$	6.61	5.99
N1009		6.75	6.41

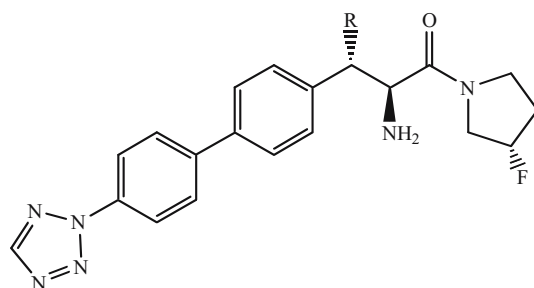
molecule D2S060 (R group in Table 10). The predicted activities of these new analogs are given in Tables 8, 9, and 10. Any other substituent replacing ‘-OCH(*i*Pr)COOH (*S*)’ group at the *para* position of ring C of molecule D2S060 did not result in any improvement in activity. It is noteworthy that the activities of the designed molecules are still better than the parent molecule (predicted pIC_{50} = 6.25, Table 10, molecule N3001). To sum up, the activity can be enhanced by at least one hundred fold through suitable substitutions. Thus, the present QSAR models are

powerful enough to suggest several fold improvement in the low affinity molecules.

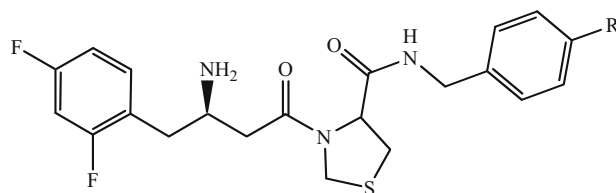
Conclusions

The QSAR analysis provides useful insights into the structural features that may be appended to the basic pharmacophore of the DPP-IV inhibitors so as to improve their potency. The docking based alignment in the present

Table 9 The CoMFA and CoMSIA predictions for newly designed DPP-IV inhibitors based on molecule D1T155 as the template



Sr. No.	R	Predicted pIC_{50}	
		CoMFA	CoMSIA
N2001	-OCH ₂ Ph	8.82	7.99
N2002	-O- <i>n</i> Bu	8.72	8.00
N2003	-C ₂ H ₄ CONH ₂	8.65	7.75
N2004	-SCH ₂ Ph	8.58	7.71
N2005	-CO- <i>i</i> Pr	8.51	8.22
N2006	-O- <i>s</i> Bu	8.51	8.02
N2007	-OCOMe	8.30	8.08
N2008	-SO ₂ Me	8.07	8.36
N2009	-O- <i>cyc</i> Pr	8.55	7.99
N2010	-OEt	8.46	7.94

Table 10 The CoMFA and CoMSIA predictions for newly designed DPP-IV inhibitors based on molecule D2S060 as the template

Sr. No.	R	Predicted pIC ₅₀	
		CoMFA	CoMSIA
N3001	-H	6.25	6.66
N3002	-OSO ₂ Ph	7.35	6.99
N3003	-NH- <i>t</i> Bu	7.24	7.00
N3004	-C ₂ H ₄ - <i>t</i> Pr	7.26	7.00
N3005	-C ₂ H ₄ - <i>t</i> Bu	7.31	7.04
N3006	-CH ₂ CONHPh	7.34	6.94
N3007	- <i>n</i> Octyl	7.19	6.98
N3008	-OC ₂ H ₄ OMe	7.16	7.04
N3009	-CH ₂ C ₆ H ₄ Cl	7.18	6.98
N3010	-O- <i>cyc</i> Hex	7.15	7.06

case is superior to the traditional atom-fit alignment, giving a better picture of the bioactive conformations that these diverse inhibitors adopt within the DPP-IV active site. The CoMFA and CoMSIA maps used together have guided the design of inhibitors whose predicted activities are several fold higher than the parent molecules.

Acknowledgements The computational facilities were jointly provided by the All India Council of Technical Education through grant (F.No. 8022/RID/NPROJ/RPS-5/2003-04) and the Department of Science and Technology through their FIST program (SR/FST/LSI-163/2003). R.R.S. Pissurlenkar thanks the Amrut Mody Research Foundation (AMRF) and M.S. Shaikh, the University Grants Commission (UGC) for financial support [Grant /F.No.7-16/2003(SR)].

References

- Ogata S, Misumi Y, Ikehara YS (1959) *J Biol Chem* 264:596–601
- Bjelke RJ, Christensen J, Branner S, Wagtmann N, Olsen C, Kanstrup AB, Rasmussen HB (2004) *J Biol Chem* 279:34691–34697
- Abbott CA, Gorrell MD (2002) Ectopeptidases: CD13/Aminopeptidase N and CD26/Dipeptidylpeptidase IV. In: Langner J, Ansoerge S (eds) *Medicine and biology*. Kulwer/Plenum, New York, pp 171–184
- Lambeir AM, Diaz Pereira JF, Chacon P, Vermeulen G, Heremans K, Devreese B, Van Beumen J, De Meester I, Scharpe S (1997) *Biochim Biophys Acta* 1340:215
- Duke-Cohan JS, Morimoto C, Rocker JA, Schlossman SF (1996) *J Immunol* 156:1714–1721
- Hartel S, Gossrau R, Hanski C, Reutter W (1988) *Histochemistry* 89:151–161
- Engel M, Hoffmann T, Wagner L, Wermann M, Heiser U, Keifersauer R, Huber R, Bode W, Demuth H-U, Brandstetter H (2003) *Proc Natl Acc Sci* 100:5063–5068
- Pederson RA, White HA, Schlentzig D, Pauly RP, McIntosh CH, Demuth H-U (1998) *Diabetes* 47:1253–1258
- Pospisilik JA, Stafford SG, Demuth HU, McIntosh CH, Pederson RA (2002) *Diabetes* 51:2677–2683
- Cheng JD, Dunbrack RL, Valianou M, Rogatko A, Alpaugh RK, Weiner LM (2002) *Cancer Res* 62:4767–4772
- Kajiyama H, Kikkawa F, Suzuki T, Shibata K, Ino K, Mizutani S (2002) *Cancer Res* 62:2753–2757
- Ho L, Aytac U, Stephens LC, Ohnuma K, Mills GB, McKee KS, Neumann C, LaPushin R, Cabanillas F, Abbruzzese JL, Morimoto C, Dang NH (2001) *Clinic Cancer Res* 7:2031–2040
- Frohman LA, Downs TR, Heimer EP, Felix AM (1989) *J Clin Invest* 83:1533–1540
- Nausch I, Heymann E (1985) *J Neurochem* 44:1354–1357
- Ahmad S, Wang L, Ward PE (1992) *J Pharmacol Exp Ther* 260:1257–1261
- Nauck MA, Niedereichholz U, Ettler R, Holst JJ, Ørskov C, Ritzel R, Schmiegel WH (1997) *Am J Physiol* 273:E981–E988
- Ahrén B (1998) *BioEssays* 20:642–651
- Meier JJ, Nauck MA, Schmidt WE, Gallwitz B (2002) *Regul Pept* 107:1–13
- Jörnvall H, Carlquist M, Kwauk S, Otte SC, McIntosh CH, Brown JC, Mutt V (1981) *FEBS Lett* 123:205–210
- Adelhorst K, Hedegaard BB, Knudsen LB, Kirk O (1994) *J Biol Chem* 269:6275–6278
- Rolin B, Deacon CF, Carr RD, Ahrén B (2004) *Eur J Pharmacol* 494:283–288
- Green BD, Flatt PR, Bailey CJ (2006) *Expert Opin Emerg Drugs* 11:525–539
- Kendall DM, Kim D, Maggs D (2006) *Diabetes Technol Ther* 8:385–396
- Ahren B, Landin-Olsson M, Jansson PA, Svensson M, Holmes D, Schweizer A (2004) *J Clin Endocrinol Metab* 89:2078–2084
- Pratley R, Galbreath E (2004) *Diabetes* 53:A83
- Herman GA, Zhao P-L, Dietrich G (2004) *Diabetologia* 47:794–805
- Triplitt C, Wright A, Chiquette E (2006) *Pharmacotherapy* 26:360–374
- Brandt W, Lehmann T, Barth A, Fittkau S (1993) *J Mol Graphics* 11:277–278
- Brandt W, Lehmann T, Thondor I, Born I, Schutkowski M, Rahfeld JU, Neubert K, Barth A (1995) *Int J Peptide Prot Res* 46:494–507
- Rasmussen HB, Branner S, Wiberg FC, Wagtmann NR (2003) *Nat Struct Biol* 10:19–25
- Qiao L, Baumann CA, Crysler CS, Ninan NS, Abad MC, Spurlino JC, Desjarlais RL, Kervinen J, Neeper MP, Bayoumy SS, Williams R, Deckman IC, Dasgupta M, Reed RL, Huebert ND, Tomczuk BE, Moriarty KJ (2006) *Bioorg Med Chem Lett* 16:123–128
- 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, Wyratt MJ, Zhang BB, Zhu L, Thornberry NA, Weber AE (2005) *J Med Chem* 48:141–151
- Nordhoff S, Cerezo-Galvez S, Feurer A, Hill O, Matassa VG, Metz G, Rummey C, Thiemann M, Edwards PJ (2006) *Bioorg Med Chem Lett* 16:1744–1748
- GOLD 3.0.1* CCDC Ltd., UK
- Sybyl* version 7.1 Tripos Inc., USA
- Cerius2* version 4.8, Accelrys Inc., USA
- Nayyar A, Malde A, Jain R, Coutinho EC (2006) *Bioorg Med Chem* 14:847–856
- Tanimoto TT, *Trans NY* (1961) *Acad Sci Ser* 2(23):576–578
- Nikolova N, Jaworska J (2003) *QSAR Comb Sci* 22:1006–1026
- Xu J, Ok HO, Gonzalez EJ, Colwell LF Jr, Habulihaz B, He H, Leiting B, Lyons KA, Marsilio F, Patel RA, Wu JK, Thornberry NA, Weber AE, Parmee ER (2004) *Bioorg Med Chem Lett* 14:4759–4762
- Edmondson SD, Mastracchio A, Duffy JL, Eiermann GJ, He H, Ita I, Leiting B, Leone JF, Lyons KA, Makarewicz AM, Patel RA, Petrov A, Wu JK, Thornberry NA, Weber AE (2005) *Bioorg Med Chem Lett* 15:3048–3052
- Xu J, Wei L, Mathvink R, He J, Park Y-J, He H, Leiting B, Lyons KA, Marsilio F, Patel RA, Wu JK, Thornberry NA, Weber AE (2005) *Bioorg Med Chem Lett* 15:2533–2536
- Hulin B, Cabral S, Lopaze MG, Van Volkenburg MA, Andrews KM, Parker JC (2005) *Bioorg Med Chem Lett* 15:4770–4773
- InsightII* version 2005L, Accelrys Inc, USA
- Datar PA, Coutinho EC (2004) *J Mol Graph Model* 23:239–251
- Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959–5967
- Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130–4146
- Klebe G (1998) *Drug Discov Des* 12:87–104
- Böhm M, Stürzebecher J, Klebe G (1999) *J Med Chem* 42:458–477
- Wold S, Johansson E, Cocchi M (1993) PLS-partial least squares projections to latent structures in 3D-QSAR. In: Kubinyi H (ed) *Drug design; theory methods and applications*. ESCOM Lieden, The Netherlands, pp 523–550
- Bush BL, Nachbar RB (1993) *J Comput Aided Mol Des* 7:587–619
- Legar C, Politis DN, Romano JP (1992) *Technometrics* 34:378–399
- Edmondson SD, Mastracchio A, Beconi M, Colwell LF, Habulihaz B, He H, Kumar S, Leiting B, Lyons KA, Mao A, Marsilio F, Patel RA, Wu JK, Zhu L, Thornberry NA, Weber AE, Parmee ER (2004) *Bioorg Med Chem Lett* 14:5151–5155
- Ludwig K, Yan S, Fan H, Reutter W, Bottcher C (2003) *Biochem Biophys Res Commun* 304:73–77
- Mentlein R (1999) *Regul Pept* 85:9–24
- Ogata S, Misumi Y, Tsuji E, Takami N, Oda K, Ikehara Y (1992) *Biochemistry* 31:2582–2587
- Abbott CA, McCaughan GW, Gorrell MD (1999) *FEBS Letters* 458:278–284
- Datar PA, Khedkar SA, Malde AK, Coutinho EC (2006) *J Comput Aided Mol Des* 20:343–360