DOI: 10.1002/cmdc.200800213

Discovery of DPP IV Inhibitors by Pharmacophore Modeling and QSAR Analysis followed by in silico Screening

Ihab M. Al-masri, Mohammad K. Mohammad, and Mutasem O. Taha*[a]

Dipeptidyl peptidase IV (DPP IV) deactivates the natural hypoglycemic incretin hormones. Inhibition of this enzyme should restore glucose homeostasis in diabetic patients making it an attractive target for the development of new antidiabetic drugs. With this in mind, the pharmacophoric space of DPP IV was explored using a set of 358 known inhibitors. Thereafter, genetic algorithm and multiple linear regression analysis were employed to select an optimal combination of pharmacophoric models and physicochemical descriptors that yield selfconsistent and predictive quantitative structure–activity relationships (QSAR) (r^2_{287} = 0.74, F-statistic = 44.5, r^2_{BS} = 0.74, r^2_{LOO} = 0.69, r^2_{PRESS} against 71 external testing inhibitors = 0.51). Two orthogonal pharmacophores (of cross-correlation r^2 = 0.23) emerged in the QSAR equation suggesting the existence of at least two distinct binding modes accessible to ligands within the DPP IV binding pocket. Docking experiments supported the binding modes suggested by QSAR/pharmacophore analyses. The validity of the QSAR equation and the associated pharmacophore models were established by the identification of new low-micromolar anti-DPP IV leads retrieved by in silico screening. One of our interesting potent anti-DPP IV hits is the fluoroquinolone gemifloxacin (IC_{50} = 1.12 μM). The fact that gemifloxacin was recently reported to potently inhibit the prodiabetic target glycogen synthase kinase 3β (GSK-3β) suggests that gemifloxacin is an excellent lead for the development of novel dual antidiabetic inhibitors against DPP IV and GSK-3β.

Introduction

Dipeptidyl peptidase IV (DPP IV) is a multifunctional protein engaged in many physiological processes. It functions as binding protein, receptor, and proteolytic enzyme. It is a serine peptidase belonging to the S9b protein family.[1] DPP IV exists in two forms: a soluble homodimer and ubiquitous type II integral plasma membrane glycoprotein.[2–6] It has 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 (N-terminal) proline or alanine. It cleaves several bioactive peptides, for example, growth hormone releasing hormone and substance P.[13,14] However, its most important substrates are incretins: glucagonlike peptide 1 (GLP-1) and glucose-dependent insulinotopic polypeptide (GIP). GLP-1 is released from intestinal L-cells in response to meals.[15] It stimulates insulin biosynthesis and secretion, reduces glucagon release, slows gastric emptying, reduces appetite, and stimulates regeneration and differentiation of islet β-cells.[16–18] On the other hand, GIP is produced by the duodenal K-cells and is extensively involved in glucose metabolism by enhancing insulin secretion.[19] Both peptides have very short half-lives because of their rapid degradation by DPP IV (approximately two minutes).[20,21] Therefore, inhibiting DPP IV should promote the hypoglycemic effects of GLP-1 and GIP. In fact, DPP IV inhibitors have been shown to reduce postprandial hyperglycemia.[22] Controlling postprandial glucose excursions is currently considered an essential objective for managing type II diabetes.[22]

Various crystallographic studies on DPP IV have uncovered features involved in its ligand binding and enzymatic catalysis, most notably, the catalytic triad Ser630-Asp708-His740, the oxyanion hole Tyr631-Tyr547, the hydrophobic S1 pocket Tyr631-Val656-Trp659-Tyr662-Tyr666-Val711, the P2 region Arg125-Asn710, and the N-terminal recognition region Glu205-Glu206-Tyr662.[23,24]

Many research groups were, and still are, involved in the discovery and optimization of new DPP IV inhibitors as leads for new hypoglycemic therapeutic agents for the treatment of type II diabetes.[25–29] These efforts culminated in the discovery and optimization of a number of reversible and irreversible DPP IV inhibitors, some of which are under clinical evaluation (for example, NVP-DPP728, Vildagliptin, P93/01).[30–34] In fact one of the potent inhibitors, MK-0431 (Sitagliptin), has been recently approved by the FDA for treating type II diabetes.[35] Clinically useful DPP IV inhibitors have several potential advantages over classical antidiabetic therapies, including: their adaptability for oral administration in a once-per-day dosing regimen,[25] being less prone to causing dangerous hypoglycemia, and being able to promote regeneration and differentiation of β-cells.[36–38]

[a] *I. M. Al-masri, M. K. Mohammad, M. O. Taha*
Department of Pharmaceutical Sciences, Faculty of Pharmacy
University of Jordan, Amman (Jordan)
Fax: (+962) 653-39-649
E-mail: mutasem@ju.edu.jo

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.200800213.

As there is a limited number of successful inhibitors available on the market, there is an urgent need for developing additional therapeutically useful DPP IV inhibitors. A large number of patents and leading research articles have appeared in the recent past that deal with this issue.[25–29,39,40] However, nearly all computer-aided efforts directed towards developing new DPP IV inhibitors were focused on structure-based (docking) techniques and quantitative structure–activity relationship (QSAR) methodologies, for example, comparative molecular field and molecular similarity indices analyses (CoMFA and CoMSIA).[41–43] Despite the excellent predictive potential of QSAR-based methodologies (for example, CoMFA and CoMSIA), they generally lack the ability to act as effective search queries to mine virtual three-dimensional (3D) databases for new hits.[44] Accordingly, we were prompted to develop robust, ligand-based 3D pharmacophore(s) integrated within a predictive QSAR model. Pharmacophore model(s) can be effectively used as 3D search queries to mine 3D compound libraries for new DPP IV inhibitors, whereas the associated QSAR model(s) can help to prioritize the selected hits for subsequent in vitro assays. Therefore, we employed the HYPOGEN module from the CATALYST package[45] to explore the pharmacophoric space of a large and diverse set of reported DPP IV inhibitors. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analyses were employed to search for an optimal QSAR model(s) that combine high-quality, binding pharmacophores with other molecular descriptors capable of explaining bioactivity variation across a collection of diverse DPP IV inhibitors. The optimal pharmacophores were later used as 3D search queries to screen virtual compound libraries for new DPP IV inhibitors. We recently reported similar strategies for the discovery novel lead inhibitors against GSK-3β,[46] pseudomonal quorum sensing,[47] h-PTP 1B,[48] and bacterial MurF enzyme.[49]

Results and Discussion

CATALYST-HYPOGEN models drug–receptor interaction using information derived from ligand structures.[45,50–57] HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups, and aromatic planes. The conformational flexibility of training ligands is modeled by creating multiple conformers, sensibly prepared to emphasize representative coverage over a specified energy range. CATALYST pharmacophores have been used as 3D queries for database searching and in 3D-QSAR studies.[52–54,58–62]

Data mining and conformational coverage

The literature was extensively surveyed to collect a large group of diverse DPP IV inhibitors (**1**–**358**, see Figure 1, and table A in the Supporting Information).[63–73] The 2D structures of the training inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations with the rule-based methods implemented within the package. The resulting single conformer 3D structures were then used as starting points for conformational analysis and in the determination of various molecular descriptors for QSAR modeling.

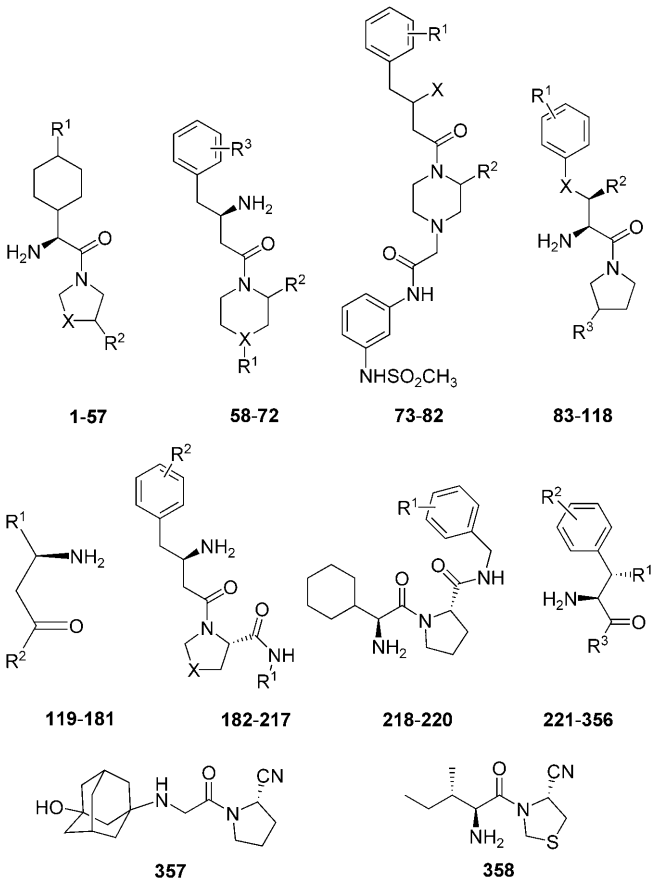

Figure 1. The chemical scaffolds of training compounds (see table A in the Supporting Information for further details).

The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within CATALYST. Poling promotes conformational variation by employing a molecular mechanical force field algorithm that penalizes similar conformers.[54] Conformational coverage was performed employing the "Best" module to ensure extensive sampling of conformational space to guarantee minimal conformation-related noise during the pharmacophore generation and validation stages. Pharmacophore modeling and pharmacophore-based in silico search procedures are sensitive to inadequate conformational sampling within the training compounds.[74]

Exploration of DPP IV pharmacophoric space

CATALYST-HYPOGEN enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over four orders of magnitude.[45,53,75] HYPOGEN implements an optimization algorithm that evalu-

ates a large number of potential models within the pharmacophoric "space" of a particular target through fine perturbations to pharmacophore hypotheses that survived the subtractive and constructive phases (see Experimental Section).[53] The extent of the evaluated space is reflected by the configuration (conFigure) cost calculated for each modeling run.

The size of the investigated pharmacophoric space is a function of training compounds, selected input chemical features, and other CATALYST control parameters.[54] Restricting the explored pharmacophoric space should improve the efficiency of optimization by allowing effective evaluation of a limited number of pharmacophoric models.[59] However, imposing more limitations on the "pharmacophoric space" might reduce the possibility of converging upon optimal pharmacophoric hypotheses, particularly if they occur outside the "boundaries" of the pharmacophoric space.[59]

Therefore, it was decided to explore the pharmacophoric space of DPP IV inhibitors under reasonably imposed "boundaries" through four carefully selected training subsets from the collected inhibitors: subsets A, B, C, and D (Table 1) and 32

Table 1. The training subsets employed in exploring the pharmacophoric space of DPP IV inhibitors.

Training subsets[a]	Most active[b]	Moderately active	Least active	Number of compounds
A	**32**, **102**, **265**, **269** **340**, **341**, **342**, **351**	**2**, **13**, **22**, **23**, **29**, **31**, **88**, **95**, **99**, **103**, **148**, **221**, **222**, **244**, **245**, **248**, **285**, **286**, **290**, **297**, **302**, **303**, **304**, **318**, **336**, **337**, **358**	**83**, **85**, **120**	38
B	**265**, **267**, **269**, **281**, **340**, **341**, **342**, **351**	**119**, **221**, **222**, **244**, **248**, **285**, **301**, **302**, **318**, **324**	**121**, **128**	20
C	**80**, **178**, **179**, **201**	**58**, **60**, **62**, **65**, **70**, **73**, **119**, **130**, **148**, **164**, **175**, **181**, **183**, **186**, **187**, **190**, **191**, **217**	**76**, **78**, **121**, **128**	26
D	**199**, **200**, **201**	**65**, **119**, **129**, **182**, **186**, **187**, **189**, **190**, **191**, **192**, **212**, **213**, **217**	**120**, **121**, **128**	19

[a] Numbers correspond to compounds in Figure 1, and table A in the Supporting Information. [b] Potency categories as defined by equations (2) and (3).

automatic HYPOGEN runs (see Table 2). The software was allowed to explore pharmacophoric models incorporating from zero to three features of any particular selected feature type (that is, HBA, HBD, hydrophobic, or aromatic ring), as shown in Table 2. Furthermore, only 4- and 5-featured pharmacophores were explored, that is, models with fewer features were ignored (Table 2). This restriction has the dual advantage of narrowing the investigated pharmacophoric space and best-representing the feature-rich nature of the DPP IV binding pocket (CATALYST-HYPOGEN can produce pharmacophore hypotheses with a maximum of five features[53,54]).

The input pharmacophore features were selected in agreement with published SAR studies and crystallographic data. For example, the fact that crystallographic information suggested the involvement of Ser630 and Tyr547 in hydrogen bonding with some ligands of DPP IV prompted us to select hydrogen bond donor and acceptor (HBA and HBD) functionalities as possible pharmacophoric features. Similarly, the involvement of Glu205 and Glu206 in electrostatic interactions with most of the co-crystallized ligands characterized by X-ray crystallography forced us to input the positive ionizable (PI) function as a potential feature. Similarly, the following features were fed into HYPOGEN as possible pharmacophoric features: aromatic rings (Ring), negative ionizable (NI), and hydrophobic features (Hbic).

It remains to be mentioned that the software was instructed to explore pharmacophoric models of interfeature distances ranging from 1–3 Å in an attempt to access pharmacophoric models of optimal interfeature distances (Table 2).

Each automatic HYPOGEN run generated ten pharmacophoric hypotheses, yielding 320 models from 32 automatic runs.

Table 2. Training sets and CATALYST run parameters employed for exploring DPP IV pharmacophoric space.

Run No.	Training subsets[a]	Selected Input Features: Types and ranges[b,c]							Interfeature Spacing[d] (Å)
		HBA	HBD	Hbic	Hbicaro	Ring	PI	NI	
1	A	0-3[e]	0-3	0-3	-	0-3	-	0-3	3
2		0-3	0-3	0-3	-	0-3	-	0-3	1
3		0-3	0-3	-	0-3	-	0-3	0-3	3
4		0-3	0-3	-	0-3	-	0-3	0-3	1
5		0-3	-	0-3	-	0-3	0-3	0-3	3
6		0-3	-	0-3	-	0-3	0-3	0-3	1
7		2-3	0-1	0-2	0-1	-	0-1	-	3
8		2-3	0-1	0-2	0-1	-	0-1	-	1
9	B	0-3	0-3	0-3	-	0-3	-	0-3	3
10		0-3	0-3	0-3	-	0-3	-	0-3	1
11		0-3	-	0-3	-	0-3	0-3	0-3	3
12		0-3	-	0-3	-	0-3	0-3	0-3	1
13		0-3	0-3	0-3	0-3	-	0-3	-	3
14		0-3	0-3	0-3	0-3	-	0-3	-	1
15	C	1-3	0-1	1-3	-	0-1	0-1	-	3
16		1-3	0-1	1-3	-	0-1	0-1	-	1
17		0-3	0-3	0-3	-	0-3	-	0-3	3
18		0-3	0-3	0-3	-	0-3	-	0-3	1
19		0-3	-	0-3	-	0-3	0-3	0-3	3
20		0-3	-	0-3	-	0-3	0-3	0-3	1
21		0-3	0-3	0-3	0-3	-	-	0-3	3
22		0-3	0-3	0-3	0-3	-	-	0-3	1
23	D	1-2	0-1	0-2	-	0-1	-	0-1	3
24		1-2	0-1	0-2	-	0-1	-	0-1	1
25		0-3	0-3	0-3	0-3	-	-	0-3	3
26		0-3	0-3	0-3	0-3	-	-	0-3	1
27		0-3	0-3	0-3	-	0-3	-	0-3	3
28		0-3	0-3	0-3	-	0-3	-	0-3	1
29		0-3	-	0-3	-	0-3	0-3	0-3	3
30		0-3	-	0-3	-	0-3	0-3	0-3	1
31		0-2	-	0-3	-	0-1	0-1	0-1	3
32		0-2	-	0-3	-	0-1	0-1	0-1	1

[a] Corresponds to training sets in Table 1. [b] HBA: hydrogen bond acceptor; HBD: hydrogen bond donor; Ring: aromatic ring; Hbic: hydrophobic; Hbicaro: hydrophobic aromatic; PI: positive ionizable; NI: negative ionizable. [c] The number of output features was allowed to vary from four to five features. Other parameters were set to their default values. [d] Permitted interfeature spacing in each run. [e] Range of input features allowed.

The binding hypotheses from each run were automatically ranked according to their corresponding "total cost" values. Total cost is defined as the sum of error cost; weight cost, and configuration cost (see Experimental Section). Error cost provides the highest contribution to total cost and it is directly related to the intrinsic capacity of the particular hypothesis in correlating the molecular structures to the corresponding biological responses.[45,53–55] HYPOGEN also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table 3), the more likely that the hypothesis does not reflect a chance correlation, that is, statistical significance.[45,53–55] An additional validation technique based on Fisher's randomization test was recently introduced into CATALYST, that is, Cat.Scramble.[45,76] In this test the biological data and the corresponding structures are scrambled several times, and the software is challenged to generate pharmacophoric models from the randomized data (see Experimental Section).

than-optimal regression models,[77] we clustered the successful pharmacophores (of confidence $\geq$85%, 280 models) into 56 groups and only considered the best representatives for GFA-MLR-QSAR analysis (models shown in Table 3). The selected representatives were fitted against the training list of inhibitors (287 inhibitors; Figure 1, and table A in the Supporting Information) and were enrolled as independent variables (genes) in a cycle of GFA-MLR-QSAR analysis over 30,000 iterations (see Experimental Section).[77,78] However, as it is essential to assess the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected 71 molecules (marked with asterisks in table A in the Supporting Information) and employed them as external test molecules for validating the QSAR models (r^2_{PRESS}). Moreover, the models were cross-validated automatically using the leave-one-out cross--validation.[77,78]

Equation (1) shows the best possible QSAR model as judged from its predictive ability against the external testing list. Figure 2 shows the scatter plots of experimental versus estimated bioactivities for the training and testing inhibitor subsets, respectively, as calculated by equation (1).

$$\begin{aligned}\text{Log}\ (1/IC_{50}) = &-2.290+0.024(\text{Hypo32}/8)^2+0.126[\text{Hypo4}/10-3.569]-17.099[\text{JursRNCG}-0.121]+6\times10^{-4}[1.022\times10^4-\text{Apol}]\\ &+0.181(\kappa_{3\alpha})^2-0.159(\kappa_3)^2+0.009(^1\chi)^2-0.195[12.86-{}^1\chi]-0.187[\text{SssCH2}-1.037]+0.082[\text{SssssC}+11.00]\\ &-0.104(\text{AtypeO60})2-0.293(\text{AtypeF81})^2-0.544[2-\text{AtypeO58}]+0.708[6.00-\text{AtypeH47}]-1.508[\text{AtypeC2}-7.00]\\ &-0.167(\text{AtypeC1})\end{aligned} \quad (1)$$

$r^2_{287} = 0.74$, F-statistic $= 44.5$, $r^2_{BS} = 0.74$, $r^2_{LOO} = 0.69$, $r^2_{PRESS(71)} = 0.51$

Most pharmacophoric models illustrated excellent success criteria (Table 3) in particular they have significant Fisher confidence levels ($\geq$85%). The emergence of numerous high-quality pharmacophore models is probably related to the ability of DPP IV ligands to assume multiple pharmacophoric binding modes within the binding pocket. Therefore, it is quite challenging to select any particular pharmacophore hypothesis as a sole representative of the binding process.

QSAR modeling

Despite the significance of pharmacophoric models in understanding ligand macromolecule recognition and as 3D search queries, their predictive value as 3D-QSAR models is generally limited by steric shielding and bioactivity-modulating auxiliary groups.[56] The general limitations of pharmacophore models combined with the abundance of plausible DPP IV pharmacophores (280 models of Fisher confidence $\geq$85%) prompted us to employ classical QSAR analysis to search for the best combination of orthogonal pharmacophores and other structural descriptors capable of explaining bioactivity variation across the whole range of collected compounds (**1**–**358**; Figure 1, and table A in the Supporting Information). This task was performed with a genetic function algorithm-multiple linear regression QSAR analysis (GFA-MLR-QSAR).

However, to avoid overloading GFA-MLR with many independent variables, which may allow the emergence of less-

where, r^2_{287} is the squared correlation coefficient, r^2_{LOO} is the leave-one-out squared correlation coefficient, r^2_{BS} is the bootstrapping regression coefficient, and $r^2_{PRESS(71)}$ is the predictive r^2 determined for the 71 test compounds. Incidentally, removing two outliers from the testing set improved the r^2_{PRESS} value to 0.60. Hypo32/8 and Hypo4/10 are the fit values of the training compounds against the 8th and 10th pharmacophores from the 32nd and 4th automatic HYPOGEN runs, respectively (Table 3 and Table 4). Jurs-RNCG is obtained by dividing the charge of most negative atom by the total negative charge of the particular molecule. Jurs descriptors are an electrostatic set of descriptors that combines shape and electronic information to characterize molecules.[79] They are calculated by mapping atomic partial charges on solvent-accessible surface areas of individual atoms. Apol is the sum of atomic polarizabilities.[77] $\kappa_{3\alpha}$ and κ_3 are the third order and third order alpha-modified Kier's shape indices, respectively.[77] $^1\chi$ is the first order Kier and Hall connectivity index.[80] SssCH2 and SssssC are electrotopological state sum indices for methylene and quaternary carbon atoms, respectively.[77] AtypeC1, AtypeC2, AtypeO58, AtypeO60, AtypeF81, and AtypeH47 are atom-type-based descriptors encoding for the hydrophobic contributions of individual atoms (carbon, oxygen, fluorine, and hydrogen atoms, respectively).[81,82]

Several descriptors emerged from equation (1) in spline format. The spline terms employed herein are "truncated power splines" and are denoted by bolded brackets ([]). For

Table 3. The performance of the best representatives of clustered pharmacophore hypotheses generated for DPP IV.

Training set[a]	Run[b]	Hypotheses[c]	Pharmacophoric features in generated hypotheses	Total cost	Cost of null hypothesis	Residual cost[d]	R[e]	F-statistic[f]	Cat scramble
A	1	9[g]	3xHBA, Hbic	190.2	239.1	48.9	0.770	64.9	95
		10	2xHBA, HBD, Hbic	190.4	239.1	48.7	0.770	99.5	95
	2	1	3xHBA, Hbic	181.9	239.1	57.2	0.825	73.5	95
		8	3xHBA, Hbic	194.4	239.1	44.7	0.750	81.2	95
	4	3	3xHBA, Hbicaro	200.6	239.1	38.5	0.690	58.4	95
		10[h]	**3xHBA, PI**	**207.1**	**239.1**	**32.0**	**0.640**	**76.0**	**95**
	5	2	2xHBA, 2Hbic	186.6	239.1	52.5	0.790	64.5	95
		6	2xHBA, Hbic, PI	189.4	239.1	49.7	0.780	47.2	95
		8	2xHBA, Hbic, PI	191.4	239.1	47.7	0.766	53.5	95
	6	8	2xHBA, 2xHbic	192.3	239.1	46.8	0.765	54.4	95
		9	2xHBA, 2xHbic	192.5	239.1	46.6	0.764	99.5	95
		10	2xHBA, 2xHbic	192.6	239.1	46.5	0.764	47.7	95
	7	1	3xHBA,Hbic	180.2	239.1	58.9	0.827	74.8	95
		2	3xHBA, Hbic	180.6	239.1	58.5	0.830	78.0	95
		7	3xHBA, Hbic	185.1	239.1	54.0	0.800	76.0	95
	8	9	3xHBA, Hbic	188.8	239.1	50.3	0.780	70.3	95
B	10	1	HBA, HBD, Hbic, Ring	94.9	137.4	42.5	0.94	85.8	95
		9	HBA, HBD, 2xHbic	97.6	137.4	39.8	0.92	82.6	95
	12	2	HBA, 2xHbic, PI, Ring	95.3	137.4	42.1	0.93	57.1	90
		3	HBA, 3xHbic, Ring	95.8	137.4	41.6	0.93	71.6	90
		5	HBA, 3xHbic, PI	97.3	137.4	40.1	0.92	56.3	90
		7	HBA, 3xHbic, PI	99.7	137.4	37.7	0.90	52.1	90
		10	2xHBA, Hbic, Ring	101.0	137.4	36.4	0.930	87.6	90
	13	4	HBA, HBD, Hbicaro, 2xHbic	94.6	137.4	42.8	0.930	72.8	95
		6	2xHBA, HBD, Hbic	94.7	137.4	42.7	0.940	96.2	95
		10	2xHBA, 2xHbic, Hbicaro	96.7	137.4	40.7	0.920	120.9	95
	14	8	3xHBA, Hbicaro	97.9	137.4	39.5	0.91	55.0	95
		10	HBA, HBD, 3xHBic	98.3	137.4	39.1	0.92	67.0	95
	15	2	HBA, HBD, Hbic, Hbicaro	93.1	137.4	44.3	0.955	73.7	95
	16	1	HBA, 2xHbic, PI, Ring	92.1	137.4	45.3	0.960	70.2	95
		7	2xHBA, Hbic, Ring	99.6	137.4	37.8	0.909	79.3	95
		10	HBA, 3xHbic, PI	100.1	137.4	37.3	0.905	25.0	95
C	18	5	HBA - HBD - NI - Ring	138.9	169.9	31.0	0.800	35.3	85
	19	1	HBA, PI, NI, Ring	137.5	169.9	32.4	0.825	32.5	90
	21	1	2xHBA, HBD, Hbicaro	132.8	169.9	37.1	0.858	11.2	95
		2	HBA, HBD, 2xHbic, Hbicaro	140.9	169.9	29.0	0.766	19.2	85
	23	1	HBA, HBD, NI, Ring	134.4	169.9	35.5	0.825	34.0	90
		3	HBA, HBD, Hbic, Ring	137.0	169.9	32.9	0.803	15.0	90
		10	HBA, HBD, NI, Ring	141.3	169.9	28.6	0.790	33.9	85
	24	10	HBA, HBD, Hbic, Ring	141.2	169.9	28.7	0.792	22.8	85
D	25	3	HBD, 2xHbicaro, Hbic, NI	91.5	122.8	31.3	0.937	47.7	95
		6	HBA, Hbic, HBD, NI, Hbicaro	92.3	122.8	30.5	0.930	53.6	95
		9	HBA, 2xHBD, Hbic	93.3	122.8	29.5	0.926	25.5	95
	26	4	2xHBA, HBD, Hbic	92.2	122.8	30.6	0.940	65.5	95
		10	HBA, HBD, 2xHbicaro, NI	94.5	122.8	28.3	0.915	33.8	95
	27	3	HBD, 2xHbic, NI, Ring	91.2	122.8	31.6	0.939	41.8	95
		9	HBD, NI, 2xRing	93.7	122.8	29.1	0.920	34.2	95
	28	9	HBA, HBD, NI, Ring	91.6	122.8	31.2	0.930	32.6	95
	29	10	2xHBA, Hbic, Ring	93.5	122.8	29.3	0.913	68.1	95
	30	5	2xHbic, NI, PI, Ring	93.5	122.8	29.3	0.921	43.1	95
		7	HBA, 3xHbic, NI	94.6	122.8	28.2	0.912	35.7	95
		9	HBA, 2xHbic, PI	95.5	122.8	27.3	0.900	70.6	95
	31	1	2xHBA, PI, Ring	89.1	122.8	33.7	0.955	37.2	95
		6	3xHbic, PI, Ring	93.5	122.8	29.3	0.920	36.7	95
		9	2xHbic, NI, PI, Ring	94.2	122.8	28.6	0.916	56.6	95
	32	**8[h]**	**2xHBA, 2xHbic, NI**	**96.6**	**122.8**	**26.2**	**0.895**	**48.4**	**95**

[a] Corresponds to training sets in Table 1. [b] Corresponds to runs in Table 2. [c] Best models from their respective clusters, as judged based on F-statistic. [d] The difference between the total cost and the cost of the corresponding null hypotheses. [e] The correlation coefficients between bioactivity estimates and bioactivities of corresponding training set compounds. [f] Fisher statistic calculated based on the linear regression between the fit values of collected inhibitors (**1**–**358**; Figure 1, and table A in the Supporting Information) against the pharmacophore hypothesis [employing the "best fit" option and equation (5)] and their respective anti-DPP IV bioactivities. [g] Rank of each hypothesis in each particular run by CATALYST. [h] Bolded pharmacophores emerged from the best QSAR equations.

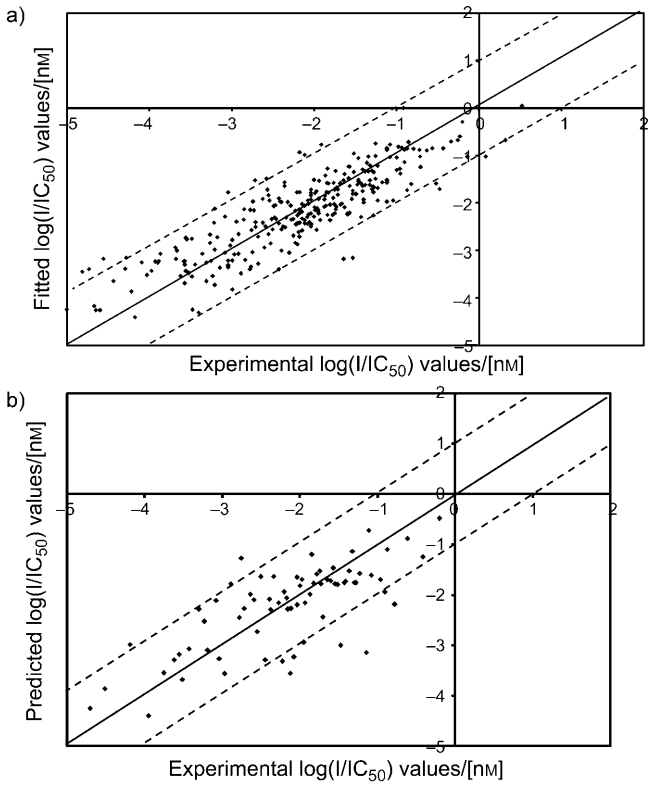

Figure 2. a) Experimental versus fitted log values ($1/IC_{50}$), 287compounds, $r^2_{LOO} = 0.69$; b) Experimental versus predicted, 71 compounds, $r^2_{PRESS} = 0.51$, bioactivities calculated from the best QSAR model [equation (1)]. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1.0 log point error margins.

example, [*f*(x) - a] equals zero if the value of (*f*(x) - a) is negative; otherwise, it equals (*f*(x) - a).[77]

Despite the mediocre statistical performance of equation (1) (that is, r^2_{287}, F-statistic, r^2_{BS}, r^2_{LOO}, r^2_{PRESS}), the sheer size of the training and testing sets significantly enhances its statistical weight allowing consequent informative inferences: Two pharmacophore hypotheses emerged in this model and recurred in other higher-ranking QSAR equations, namely, Hypo4/10 and Hypo32/8. Interestingly, both binding models are of low rank among other models generated in the corresponding HYPOGEN run, that is, they ranked 10th and 8th within runs four and 32, Table 3. However, it must be remembered that the pharmacophore ranking procedure implemented in CATALYST is dependent on the selected subset of training compounds employed in pharmacophore generation, and therefore, doesn't necessarily reflect the behavior of the pharmacophore models against the whole list of collected compounds, which explains the apparently odd selection of low-ranking models in the best QSAR equation. Table 4 shows the three-dimensional co-ordinates of the two pharmacophores.

The emergence of two orthogonal pharmacophoric models (Hypo32/8 and Hypo4/10, cross-correlation $r^2 = 0.23$) in the optimal QSAR equation, suggests they represent two complementary binding modes accessible to ligands within the binding pocket of DPP IV, that is, one of the pharmacophores can optimally explain the bioactivities of some training inhibitors, whereas the other inhibitors are more appropriately explained by the second pharmacophore.[59] Figure 3 shows how Hypo4/10 maps two potent DPP IV inhibitors **201** and **351** ($IC_{50} = 1.80$ and 4.30 nM, respectively) and Figure 4 shows how Hypo32/8 maps **201**, whereas Table 4 shows the X, Y, and Z co-ordinates of the two pharmacophores.

Emergence of negative contributions of Apol and JursRNCG in equation (1) suggests that strongly negative and/or polarizable moieties generally adversely influence ligand–DPP IV affinity. This trend is explainable by the proposition that ionized ligands, in particular negatively charged ones, favor hydration instead of docking into the binding site, particularly if they are misaligned with their corresponding counterparts in the binding pocket. A similar analysis was recently used to explain the general nonspecific enhancement in ligand–receptor affinity concomitant to increases in ligand lipophilicity.[83]

On the other hand, emergence of connectivity indices and electrotopological and shape descriptors (for example, $\kappa_{3\alpha}$ and κ_3, $^1\chi$, SssCH2, and SssssC) in equation (1) illustrate certain roles played by the ligands' topology in the binding process. However, despite the predictive significance of these descriptors, their information content is quite obscure. A similar conclusion can be extended to atom-type descriptors (for example, AtypeO60, AtypeF81, and AtypeO58) in equation (1).

Table 4. Pharmacophoric features and corresponding weights, tolerances, and 3D co-ordinates of Hypo4/10[a] and Hypo32/8[b].

Model	definitions		Features							
Hypo4/10			HBA		HBA		HBA		PI	
	Weights		1.856		1.856		1.856		1.856	
	Tolerances		1.6	2.2	1.6	2.2	1.6	2.2	1.6	
	Coordinates	X	-6.82	-6.39	1.67	2.34	1.25	-0.85	1.58	
		Y	5.71	8.64	-0.85	1.57	-0.17	1.13	-3.19	
		Z	-0.93	-0.46	1.55	3.18	-1.44	-3.15	-1.20	
Hypo32/8			HBA		HBA		Hbic		Hbic	NI
	Weights		1.608		1.608		1.608		1.608	1.608
	Tolerances		1.6	2.2	1.6	2.2	1.6		1.6	1.6
	Coordinates	X	-1.28	-4.12	1.30	3.02	-4.32		-0.46	-5.50
		Y	1.46	1.86	3.03	2.12	-3.58		4.08	-0.54
		Z	-0.23	0.66	-4.59	-6.92	6.14		-1.30	6.65

[a] Hypo4/10 is hypothesis number 10 generated in run number 4. [b] Hypo32/8 is hypothesis number 8 generated in run number 32.

Addition of exclusion volumes

Despite the significance of ligand-based pharmacophores as 3D QSAR models and search queries, they suffer from two

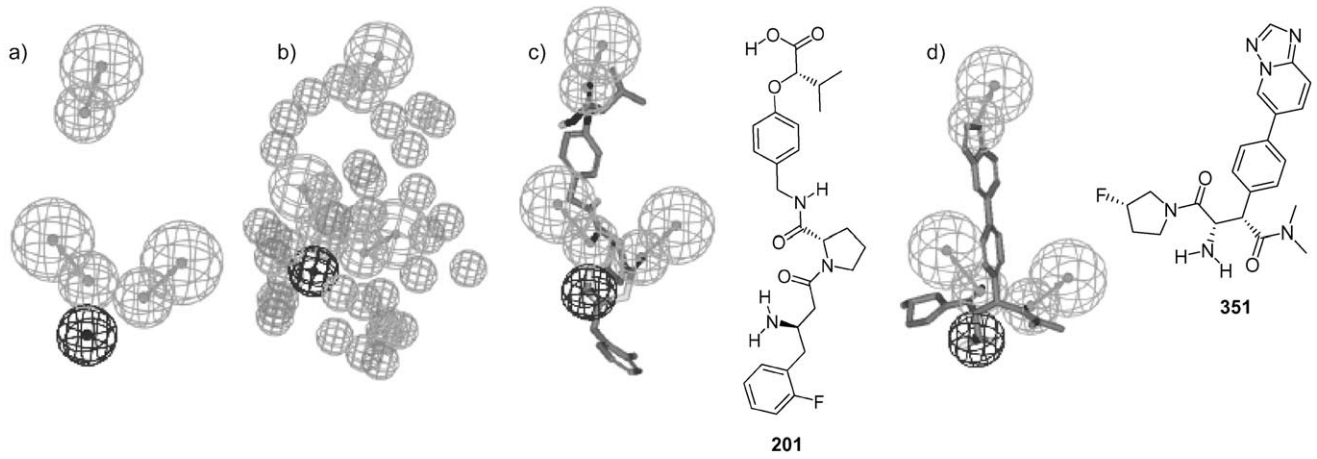

Figure 3. a) Hypo4/10: the pharmacophoric features of the binding model, hydrogen bond acceptors are represented as light grey vectored spheres, positive ionizable groups as black spheres; b) Hypo4/10 with added exclusion spheres (dark grey) as performed by HIPHOP-REFINE; c) Hypo4/10 fitted against potent DPP IV inhibitor **201** (IC_{50} = 1.80 nM); d) Hypo4/10 fitted against potent inhibitor **351** (IC_{50} = 4.30 nM),

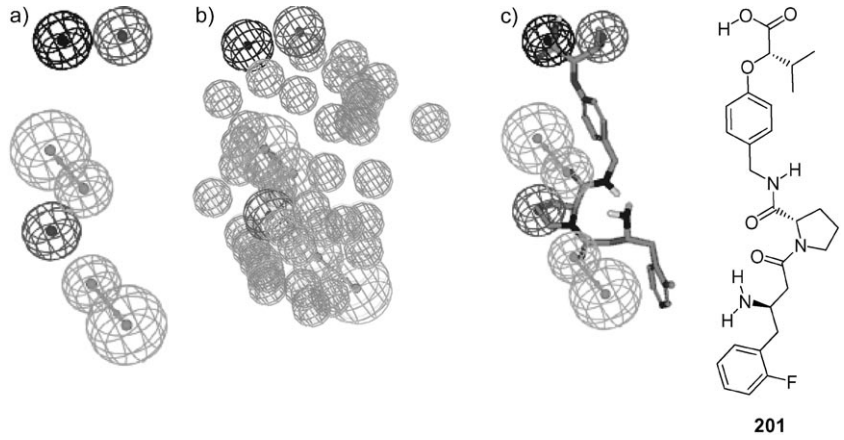

Figure 4. a) Hypo32/8: the pharmacophoric features of the binding model, hydrogen bond acceptors are represented as green vectored spheres, negative ionizable groups as black spheres, and hydrophobic features as dark grey spheres; b) Hypo32/8 with added exclusion spheres as performed by HIPHOP-REFINE; c) Hypo32/8 fitted against potent DPP IV inhibitor **201** (IC_{50} = 1.80 nM).

major drawbacks: Firstly, pharmacophore models lack specific information about the electronic contributions of different substituents. Nevertheless, QSAR analysis can identify statistically significant electronically-related descriptors, and therefore can partially solve this problem. Secondly, the generated models lack steric constraints necessary to define the size of the binding pocket. This fault renders pharmacophoric models rather promiscuous without steric limitations. Therefore, we decided to decorate the optimal models with exclusion spheres employing the HIPHOP-REFINE module implemented within CATALYST[45] to mark the steric requirements of the binding pocket. Excluded volumes resemble sterically inaccessible regions within the binding site.[45,75]

Accordingly, two structurally diverse training subsets were carefully selected from the collected compounds for HIPHOP-REFINE modeling: training subsets E and F to add exclusion spheres for Hypo32/8 and Hypo4/10, respectively (Table 5). The training compounds were selected in such a way that the bioactivities of weakly active compounds are explainable by steric clashes within the binding pocket. Figure 3 and Figure 4, shows the refined Hypo4/10 (40 added exclusion volumes) and refined Hypo32/8 (37 added exclusion volumes).

In silico screening of the NCI database and subsequent experimental evaluation

The fact that the emerging pharmacophores are orthogonal and complement each other in explaining bioactivity variations across the training compounds, that is, they represent different binding modes within the binding pocket, prompted us to combine them as 3D search queries. They were employed to screen the national cancer institute (NCI) list of compounds (238,819 compounds)[45] and our in-house built database of established drugs (1,490 compounds). Hypo32/8 and Hypo4/10 captured 383 and 3836 hits, respectively. These were subsequently filtered based on Lipinski's and Veber's rules[84,85] eventually leaving 1498 molecules. The selected hits were fitted against the two pharmacophores (see equation (5) in the Experimental Section) and their fit values were substituted in QSAR equation (1) to determine their predicted bioactivities. To minimize the effects of possible extrapolatory prediction errors[86] on decisions regarding hits that merit subsequent in vitro testing, we employed the predicted $Log(1/IC_{50})$ values merely to rank the corresponding hits. The highest ranking 56 NCI hits and seven drug molecules (gemifloxacin, famotidine, nizatidine, sildenafil, valacyclovir, iodipamide, and telmisartan) were requested for experimental validation. Only 28 NCI hits were available. All tested drug molecules

Table 5. Training subsets used for adding excluded spheres for Hypo32/8 and Hypo4/10 using HIPHOP-REFINE module of CATALYST.

Training subset E[a]				Training set F[a]			
Compd[b]	IC_{50} [nM]	Principal value	MaxOmitFeat[c]	Compd	IC_{50} [nM]	Principal value	MaxOmitFeat
194	7.10	2	0	**31**	26.00	2	0
195	10.00	2	0	**102**	8.00	2	0
196	3.70	2	0	**196**	3.70	2	0
197	5.10	2	0	**203**	0.48	2	0
199	12.00	2	0	**206**	0.83	2	0
200	13.00	2	0	**209**	1.70	2	0
201	1.80	2	0	**342**	8.00	2	0
203	0.48	2	0	**343**	13.00	2	0
204	1.60	2	0	**344**	4.10	2	0
206	0.83	2	0	**347**	8.00	2	0
207	9.50	2	0	**350**	8.60	2	0
209	1.70	2	0	**351**	4.30	2	0
211	12.00	2	0	**352**	2.70	2	0
214	0.30	2	0	**354**	6.30	2	0
76	65000	0	1	**355**	6.20	2	0
77	36000	0	1	**356**	7.50	2	0
85	12400	0	1	**76**	65000	0	0
89	15600	0	1	**77**	36000	0	0
120	100000	0	2	**78**	19000	0	1
121	16480	0	2	**85**	12400	0	1
128	33000	0	2	**289**	54000	0	2
185	1200	0	1	**304**	7900	0	2
191	1640	0	1				
218	4200	0	1				
219	3300	0	1				
289	54000	0	1				

[a] Training subset E for adding exclusion volumes to Hypo32/8 and subset F for adding exclusion volumes for Hypo4/10. [b] Compound numbers are as in Figure 1, and table A in the Supporting Information. [c] MaxOmitFeat: Maximum omitted features.

and 22 NCI hits were found to possess in vitro inhibitory actions against DPP IV at a concentration of 10 μM. Figure 5 shows the active hits, whereas Table 6 shows their names or NCI codes and their estimated and experimental DPP IV bioactivities.

Interestingly, gemifloxacin (**386**) presented the most potent DPP IV inhibitory activity ($IC_{50} = 1.12$ μM). Figure 6 shows how Hypo4/10 fits gemifloxacin. On the other hand, Figure 7 shows how Hypo32/8 fits **365** (the highest ranking inhibitor fitting Hypo32/8). To substantiate the anti-DPP IV activity of **386** (gemifloxacin), it was decided to test its ability to lower plasma glucose concentrations after oral glucose challenge (that is, OGTT) in Balb/c mice. Oral administration of gemifloxacin illustrated a dose-dependent inhibition of glucose excursion, as shown in Figure 8. Furthermore, it illustrated significant reduction in the area under the blood glucose/time curve (AUC) at three dosing levels (Figure 8B). The results suggest significant improvement in impaired glucose tolerance in Balb/c mice in response to gemifloxacin, probably by elevating intact GLP-1 and insulin levels in plasma.

Intriguingly, gemifloxacin has been recently reported to possess nanomolar inhibitory action against the prodiabetic target glycogen synthase kinase 3β (GSK-3β).[46] The combined inhibitory potential of gemifloxacin against two important prodiabetic targets suggests that this antibacterial agent could be an excellent lead for the development of a novel class of dual antidiabetic inhibitors against DPP IV and GSK-3β.

Comparison of Hypo4/10 and Hypo32/8 with the binding site of DPP IV

Despite the uncertainties of crystallographic data,[87] pharmacophore features obtained by pharmacophore/QSAR modeling can be compared with the structure of the DPP IV binding site to identify probable residues important for inhibition. The features in Hypo4/10 as well as the alignment of **386** (gemifloxacin) ($IC_{50} = 1.1$ μM), as proposed by this pharmacophore, were compared with the corresponding structure as it docks into the binding pocket of DPP IV (PDB code: 2G63, resolution 2.0 Å),[88] shown in Figure 6. The docking experiment was performed employing the FRED docking engine and Chemgauss2 scoring function.[89] A marked similarity was observed between the features proposed by the pharmacophore models and the ligand-binding features in the docked structures.

In the highest-ranking docked pose of **386** (Figure 6) the carboxylic acid moiety of Glu206 binds to the amino group of **386** via a hydrogen-bond-reinforced ionic interaction corresponding to mapping the amino group with a positive ionizable feature in Hypo4/10 (Figure 6). Moreover, the aromatic fluoro substituent of **386** seems to be hydrogen bonded to the guanidine side chain of Arg125, which agrees with mapping the same fluorine atom with the HBA feature in Hypo4/10, as in Figure 6. Similarly, mapping the methoxyimino substituent in **386** with a HBA in Hypo4/10 correlates nicely with the hydrogen-bonding interaction bridging the methoxyimino oxygen and Ser209. Finally, the docking experiment proposes a hydrogen-bond-reinforced ionic interaction between the carboxylic acid moiety of **386** and the terminal amino group of Gln553, agreeing with a HBA feature in Hypo4/10 mapping the carboxylic group (Figure 6).

On the other hand, the validity of Hypo32/8 can be further established by comparing the way it maps relevant hits with the docked poses of these hits inside the binding pocket. We selected compound **365** (inhibited DPP IV by 34% at 10 μM) for this purpose as it illustrated excellent fit against Hypo32/8, as shown in Figure 7. The HBA feature mapping the carboxylic acid group in **365** corresponds to hydrogen bonding with

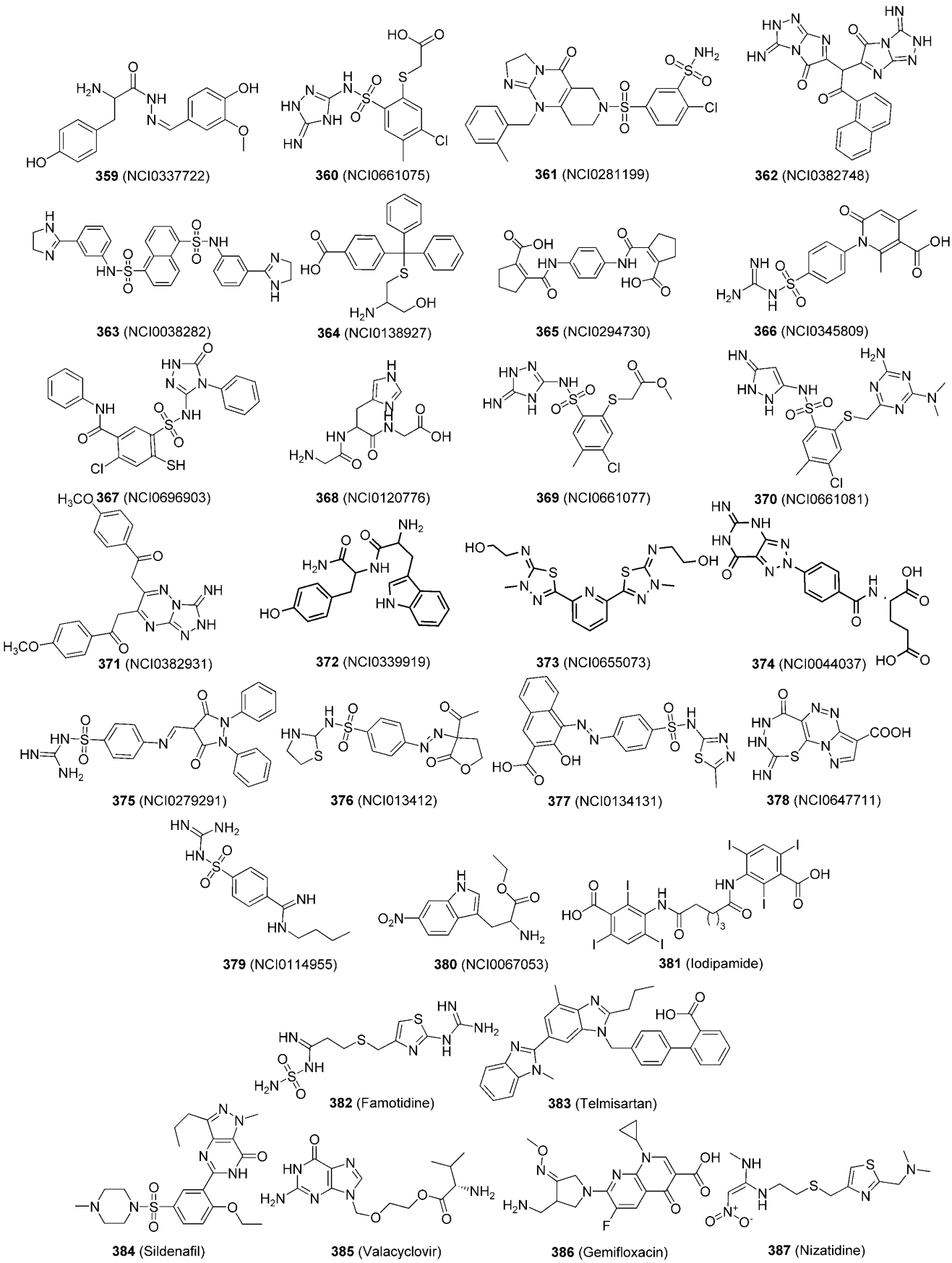

Figure 5. The chemical structures of the tested hits.

Table 6. The hit molecules captured by Hypo32/8 and Hypo10/4, their corresponding QSAR estimates from equation (1), and their in vitro bioactivities.

Tested Hits[a]	NCI codes/names	Fit values against[b]		QSAR-based estimates	In vitro Anti-DPP IV Activity	
		Hypo32/8	Hypo4/10	IC_{50} [μM][c]	Inhibition [%][d]	IC_{50} [μM]
359	337722	0	3.756	6.14	52.5	–
360	661075	0	4.534	6.2×10^{-4}	20.0	–
361	281199	0	3.776	1.54×10^{-3}	13.5	–
362	382748	0	4.129	2.1×10^{-3}	23.5	–
363	38282	0	4.925	8.7×10^{-3}	20.0	–
364	138927	0	6.958	9.4×10^{-3}	18.0	–
365	294730	6.044	0	5.4×10^{-4}	34.0	–
366	345809	0	7.272	4.1×10^{-2}	25.0	–
367	696903	0	3.091	6.5×10^{-2}	3.00	–
368	120776	0	4.097	3.6×10^{-4}	29.0	–
369	661077	0	4.551	7.1×10^{-2}	25.0	–
370	661081	0	4.941	7.4×10^{-2}	15.0	–
371	382931	0	5.793	8.2×10^{-2}	26.0	–
372	339919	0	5.748	8.2×10^{-2}	33.0	–
373	655073	0	6.651	2.9×10^{-1}	5.00	–
374	44037	0	6.394	4.0×10^{-3}	30.0	–
375	279291	0	7.021	4.6×10^{-2}	16.0	–
376	134121	0	7.285	4.0×10^{-2}	28.0	–
377	134131	5.186	0	2.8×10^{-2}	18.0	–
378	647711	0	2.851	9.5×10^{-6}	35.6	–
379	114955	0	2.753	23	19.8	–
380	67053	0	5.581	1.7	20.0	–
381	Iodipamide	7.468	0	1.5×10^{-11}	34.0	–
382	Famotidine	0	6.810	1.6×10^{-3}	26.2	–
383	Telmisartan	1.815	0	3.6×10^{-2}	5.00	–
384	Sildenafil	0	6.543	2.6×10^{-2}	5.00	–
385	Valacyclovir	0	6.924	4.5×10^{-3}	42.0	42.9
386	Gemifloxacin	0	6.1	8.9×10^{-3}	65.0	1.12
387	Nizatidine	0	4.874	25.2	49	6.06

[a] Structures as in Figure 5. [b] Best-fit values against each binding hypothesis calculated by equation (5). [c] As calculated by QSAR equation (1). [d] Percentage inhibition at 10 μM.

Trp 629 and/or Arg 125 in the docked pose (Figure 7). The close proximity of the carbonyl of **365** to the hydroxyl group of Tyr 547 in the docked structure suggests mutual hydrogen bonding. This proposition is supported by a HBA feature mapping the same carbonyl in Hypo32/8. Similarly, the second carboxyl of **365** is mapped by a negative ionizable (NI) feature in Hypo32/8 corresponding to ionic interaction with the guanidine moiety of Arg 125 in the docked pose (Figure 7). Finally, the cyclopentene of **365** is docked adjacent to Phe 357 suggesting the existence of significant mutual van der Waals' stacking interactions, which correlate with a hydrophobic feature mapping the cyclopentene ring in Hypo32/8 (Figure 7).

Conclusion

This work includes extensive exploration of the pharmacophore space of DPP IV inhibitors utilizing CATALYST-HYPOGEN. QSAR analysis was employed to select the best combination of molecular descriptors and pharmacophore models capable of explaining bioactivity variation across an informative list of training compounds. The best binding hypotheses were used to screen two of our structural databases for new DPP IV inhibitors. The resulting hits were prioritized for in vitro testing based on their predicted bioactivities. The pharmacophoric features of the optimal models agree with the binding features proposed by docking evaluation. One of our moderately potent anti-DPP IV hits, gemifloxacin, was recently reported to have potent inhibitory action against the prodiabetic target GSK-3β, which opens the door for the development of novel dual antidiabetic inhibitors against DPP IV and GSK-3β.

Experimental Section

Molecular modeling

Software and hardware

The following software packages were utilized in the present research: CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA; CERIUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA; OMEGA (Version 2.1.0), OpenEye Scientific Software (www. eyesopen.com), USA; FRED (Version 2.1.2), OpenEye Scientific Software, (www.eyesopen.com), USA; CS ChemDraw Ultra 7.01, Cambridge Soft Corp. (http://www.cambridgesoft.com), USA. Pharmacophore modeling and QSAR analysis were performed using CATALYST (HYPOGEN module) and CERIUS2 software suites installed on a Silicon Graphics Octane2 desktop workstation equipped with a dual 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system. Docking studies were performed using FRED and Omega software installed on a Pentium 4 PC.

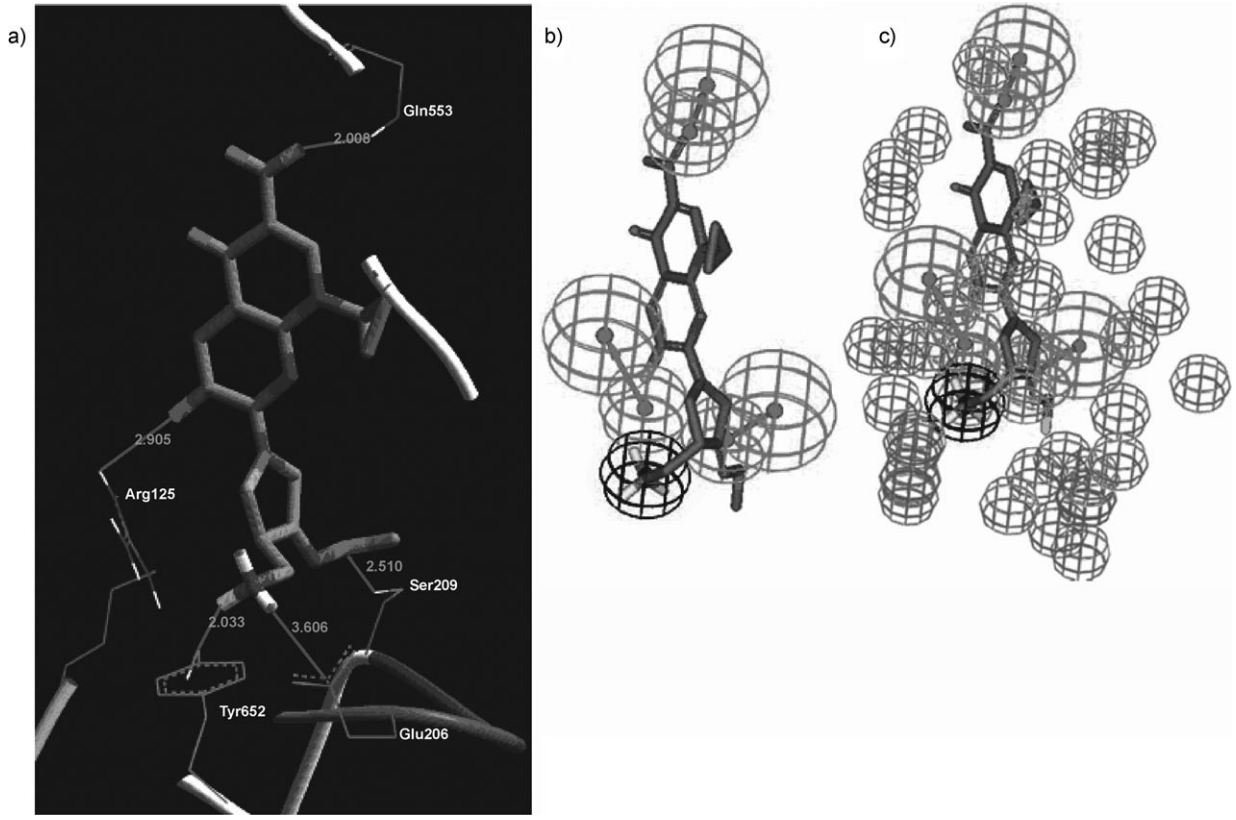

Figure 6. a) An optimal docked pose of inhibitor **386** (gemifloxacin, $IC_{50} = 1.12$ μM) into the binding pocket of DPP IV (PDB code: 2G63, resolution 2.00 Å), numbers indicate interatomic distances in Å; b) and c) Hypo4/10 fitted against inhibitor **386**, with and without exclusion spheres, respectively.

Data set

The structures of 358 DPP IV enzyme inhibitors (Figure 1, and table A in the Supporting Information) were collected from published literature.[63–73] The in vitro bioactivities of the collected inhibitors were expressed as the concentration of the test compound that inhibited the activity of DPP IV enzyme by 50% (IC_{50}). Figure 1 shows the structures and IC_{50} values of the considered inhibitors (see also table A in the Supporting Information). The logarithm of measured IC_{50} (nM) values were used in pharmacophore modeling and QSAR analysis, thus correlating the data linear to the free energy change. In one case in which the IC_{50} value was given as greater than 100 μM (compound **120**, table A in the Supporting Information), it was assumed that it equals that value. This assumption is necessary to allow statistical correlation and QSAR analysis. The logarithmic transformation of IC_{50} values should minimize any potential errors resulting from this assumption.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra. Subsequently, they were imported into CATALYST, converted into the corresponding standard 3D structures, and were utilized as starting conformers for conformational analysis.

Conformational analysis

Molecular flexibility was taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor (**1**–**358**; Figure 1, and table A in the Supporting Information) was explored adopting the "best conformer generation" option in CATALYST which is based on the generalized CHARMm force field implemented in the program. Default parameters were employed in the conformation generation procedure, that is, conformational ensembles were generated with an energy threshold of 20 kcal mol^{-1} from the local minimized structure with a maximum of 250 conformers per molecule. This search procedure will eventually identify the best three-dimensional arrangement of chemical functionalities explaining the activity variations across the training set.[45,56,57]

Exploration of the pharmacophoric space of DPP IV

All 358 molecules, with their associated conformational models, were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an "uncertainty" value of three, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor.[53–57] Subsequently, four structurally diverse training subsets (Table 1) were carefully selected from the collection for pharmacophore modeling. Typically, CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least four orders of magnitude.[53–57]

The selected training sets were utilized to conduct 32 modeling runs to explore the pharmacophoric space of DPP IV inhibitors as shown in Table 2. Different binding hypotheses were generated by altering the interfeature spacing and the type and ranges of allowed features in the resulting pharmacophores (Table 2). The max-

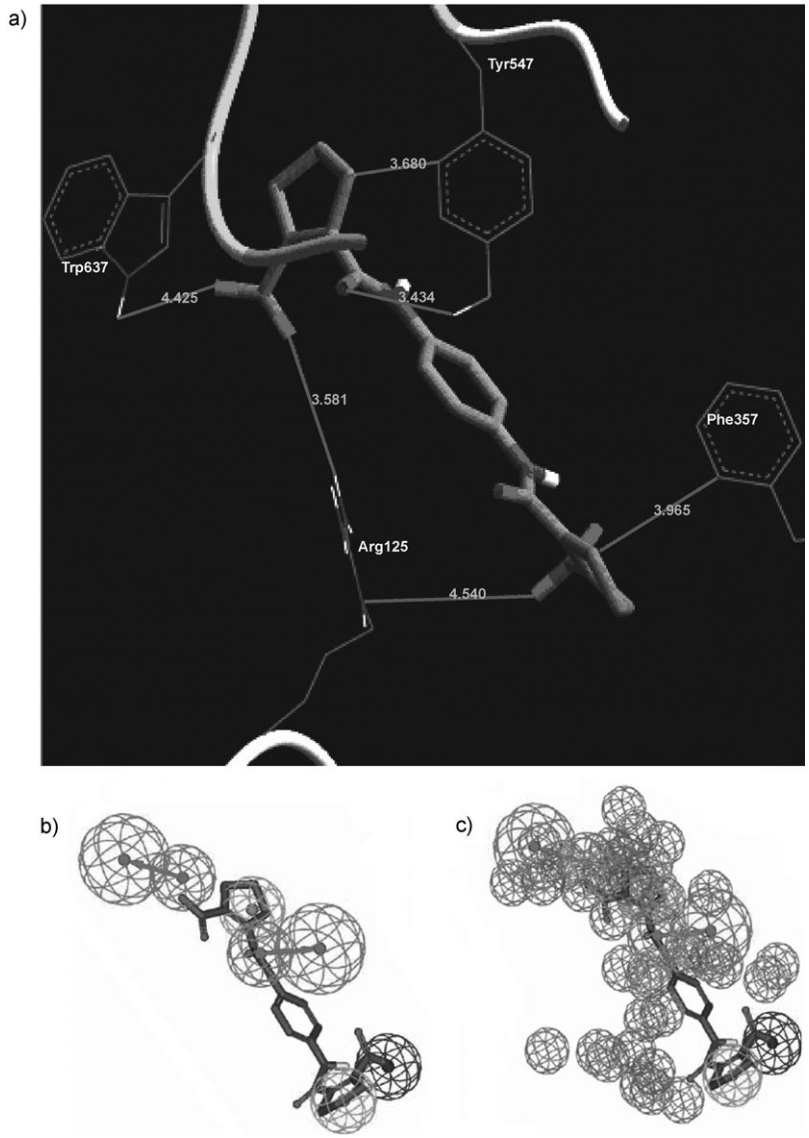

Figure 7. a) An optimal docked pose of inhibitor **365** (34% inhibitory at 10 μM) into the binding pocket of DPP IV (PDB code: 2G63, resolution 2.00 Å), numbers indicate represent interatomic distances in Å; b) and c) Hypo32/8 fitted against inhibitor **365**, with and without exclusion spheres, respectively.

imum number of features in each generated pharmacophore hypotheses was allowed to vary from four to five. CATALYST-HYPOGEN can produce pharmacophore hypotheses of a maximum of five features.[46,54]

Pharmacophore modeling employing CATALYST proceeded through three successive phases: the constructive phase, subtractive phase, and optimization phase.[53–57] During the constructive phase, CATALYST generates common conformational alignments among the most-active training compounds. Only molecular alignments based on a maximum of five chemical features are considered. The program identifies a particular compound as being within the most active category if it satisfies equation (2).[53–57]

$$(\mathrm{MAct} \times \mathrm{UncMAct}) - (\mathrm{Act}/\mathrm{UncAct}) > 0.0 \quad (2)$$

Where "MAct" is the activity of the most active compound in the training set, "Unc" is the uncertainty of the compounds, and "Act" is the activity of the training compounds under question. However, if there are more than eight most-active inhibitors, only the top eight are used. In the subsequent subtractive phase, CATALYST eliminates some hypotheses that fit inactive training compounds. A particular training compound is defined as being inactive if it satisfies equation (3):[53–57]

$$\mathrm{Log}\ (\mathrm{Act}) - \mathrm{log}\ (\mathrm{MAct}) > 3.5 \quad (3)$$

However, in the optimization phase, CATALYST applies fine perturbations in the form of vectored feature rotation, adding new features, and/or removing a feature to selected hypotheses that survived the subtractive phase to find new models of enhanced bioactivity-to-mapping correlations that is, improved 3D-QSAR properties. Eventually, CATALYST selects the highest-ranking models (ten by default) and presents them as the optimal pharmacophore hypotheses resulting from the particular automatic modeling run.[50]

Assessment of the generated hypotheses

When generating hypotheses, CATALYST-HYPOGEN attempts to minimize a cost function consisting of three terms: weight cost, error cost, and configuration cost.[50,53–57] Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of two. The deviation between the estimated activities of the training set and their experimentally determined values adds to the error cost.[53–57] The activity of any compound can be estimated from a particular hypothesis through equation (4).[50]

$$\mathrm{Log}\ (\mathrm{Estimated\ activity}) = \mathrm{I} + \mathrm{Fit} \quad (4)$$

Where, I = the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the fit values of the training compounds. The fit value for any compound is obtained automatically by employing equation (5).[53–55]

$$\mathrm{Fit} = \Sigma\mathrm{mapped\ hypothesis\ features} \times W\ [1\text{-}\Sigma\ (\mathrm{disp}/\mathrm{tol})^2] \quad (5)$$

Where, Σmapped hypothesis features represents the number of pharmacophore features that successfully superimpose (that is, map or overlap with) corresponding chemical moieties within the fitted compound, *W* is the weight of the corresponding hypothesis feature spheres. This value is fixed to 2.0 in HYPOGEN-generated

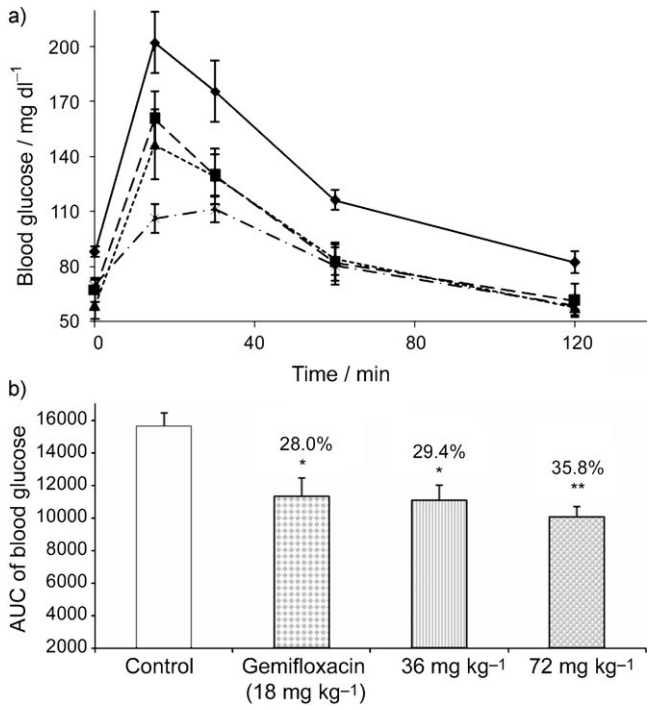

Figure 8. Oral glucose tolerance test of compound **386** (gemifloxacin) using male Balb/c mice. a) Various doses of compound **386** were administered intraperitoneally to Balb/c mice and glucose (2 g kg^{-1}, p. o.) was given 30 min latter (0 min). Plasma glucose concentration was measured at the indicated time. (◆) Control (normal saline); (■) gemifloxacin 18 mg kg^{-1}; (▲) gemifloxacin 36 mg kg^{-1}; (×) gemifloxacin 72 mg kg^{-1}. Data are represented as mean ± S.E.M. ($n = 5$–8); b) AUC of the glucose concentration curve (shown in panel a) was calculated and displayed. Percent inhibition values for each treatment (in parentheses) were generated from the AUC data. (*) P value < 0.05 versus control, (**) P value < 0.01 versus control.

models. disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound; tol is the radius of the pharmacophoric feature sphere (known as Tolerance, equals to 1.6 Å by default). $\Sigma(\text{disp/tol})^2$ is the summation of $(\text{disp/tol})^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound.[53–55]

The third term, the configuration cost, penalizes the complexity of the hypothesis. This is a fixed cost, which is equal to the entropy of the hypothesis space. The greater the number of features (a maximum of five) in a generated hypothesis, the higher the entropy, with a subsequent increase in this cost. The overall cost (total cost) of a hypothesis is calculated by summing the three cost factors. However, error cost is the main contributor to total cost.

CATALYST-HYPOGEN also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table 3), the more likely that the hypothesis does not reflect a chance correlation.[53–57] In a successful automatic modeling run, CATALYST® ranks the generated models according to their total costs.[53–55]

An additional approach to assess the quality of CATALYST-HYPOGEN pharmacophores is to crossvalidate them using the Cat-Scramble program implemented in CATALYST.[57,76] This validation procedure is based on Fisher's randomization test.[76] In this validation test, we selected a 95% confidence level, which instructs CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses.[90] Success in generating pharmacophores of comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training compounds and the unscrambled original pharmacophore models.[57,76]

Clustering of the generated pharmacophore hypotheses

Due to the large number of resulting pharmacophores (320 models) and to avoid overwhelming GFA/MLR with a large number of input descriptors, it was decided to reduce the number of input pharmacophore descriptors through the following: 1) only pharmacophore models of Fisher scrambling significance $\geq$ 85% were considered for subsequent QSAR modeling, and 2) the surviving models (280 hypotheses) were clustered into 56 groups and the best model was selected to represent its cluster in subsequent QSAR modeling. The clustering process was conducted as follows: every five models were clustered together utilizing the hierarchical average linkage method available in CATALYST. The models generated from each training set were clustered separately, for example, all 80 pharmacophore models generated from training subset A were clustered into 16 groups. To select the best representative pharmacophores from each cluster, each group member was fitted against the collected compounds (**1**–**358**; Figure 1, and table A in the Supporting Information) employing the "best" fit option implemented in CATALYST, and their fit values [calculated from equation (5)] were regressed against the corresponding experimental bioactivities. The highest-ranking models, based on their F-statistic, were selected to represent their corresponding clusters in subsequent QSAR modeling. Table 3 shows the pharmacophoric features of selected representative binding hypotheses, combined with their CATALYST success criteria.

QSAR modeling

A subset of 287 compounds from the total list of inhibitors (Figure 1, and table A in the Supporting Information) was utilized as a training set for QSAR modeling. However, as it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 71 molecules (approximately 20% of the dataset) were employed as an external test subset for validating the QSAR models. The test molecules were selected as follows: the inhibitors (**1**–**358**; Figure 1, and table A in the Supporting Information) were ranked according to their IC_{50} values, subsequently, every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set. The selected test inhibitors are marked with asterisks in table A in the Supporting Information.

The logarithm of measured $1/IC_{50}$ (nM) values was used in QSAR, thus correlating the data linearly to the free energy change. The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2 (100 terms). The calculated descriptors included various simple and

valence connectivity indices, electrotopological state indices, single point quantum-mechanical descriptors (with the AM1 model), and other molecular descriptors (for example, logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, etc…).[77] Furthermore, the fit values of training compounds against the representative pharmacophores (using the Best-fit option in CATALYST and equation 5)[45] were added as additional molecular descriptors.

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating variations in the biological activities of training compounds with variations in their molecular descriptors, that is, multiple linear regression modeling (MLR). GFA techniques rely on the evolutionary operations of "crossover and mutation" to select optimal combinations of descriptors (that is, chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations. Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF).[77]

Our preliminary diagnostic trials suggested the following optimal GFA parameters: Explore linear equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30000 and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), we decided to scan and evaluate all possible QSAR models resulting from 8 to 28 explanatory terms.

All QSAR models were validated employing leave one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}),[57,58] and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The predictive r^2_{PRESS} is defined as:

$$r^2_{PRESS} = \text{SD-PRESS}/\text{SD} \quad (6)$$

Where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set. The descriptor-scanning procedure identified equation 1 as the best equation. Figure 2 shows the plots of experimental versus fitted (training set) and predicted (testing set) bioactivities calculated from the best QSAR equation.

Addition of exclusion volumes

To account for the steric constraints of the binding pocket we decided to decorate Hypo4/10 and Hypo32/8 with exclusion volumes employing HIPHOP-REFINE module of CATALYST. HIPHOP-REFINE uses inactive training compounds to construct excluded volumes that resemble the steric constraints of the binding pocket. It identifies spaces occupied by the conformations of inactive compounds and spaces free from the active inhibitors. These regions are then filled with excluded volumes.[45]

As each pharmacophore resembles a separate binding mode, it was decided to select two separate training subsets for constructing appropriate exclusion spheres around Hypo32/8 and Hypo4/10, namely, subsets E and F, respectively (Table 5).

In HIPHOP-REFINE the user defines how many molecules must map completely or partially to the hypothesis via the Principal and MaxOmitFeat parameters. Active compounds are normally assigned MaxOmitFeat parameter of zero and a principal value of two such that the software is instructed to map them against all the pharmacophoric features of a particular hypothesis. On the other hand, inactive compounds are allowed to miss one or two features by assigning them a MaxOmitFeat of one (or two) and a principal value of zero.

However, in the case of compounds that seem to be inactive because of steric clashes within the binding pocket, they were assigned MaxOmitFeat and principal values equal to zero. This combination instructs HIPHOP-REFINE to force inactive compound(s) to map all the pharmacophoric features of the binding model, and therefore permits the software to identify spaces occupied by excess structural fragments/features of such compounds and fill them with exclusion volumes.[45,91]

To identify an appropriate activity/inactivity cutoff value, we decided to consider compounds that have anti-DPP IV affinities $\geq$ 3.5 logarithmic cycles from the most potent active inhibitor (**214**, $IC_{50} = 0.3$ nM) as an appropriate activity/inactivity threshold. Accordingly, inhibitors of IC_{50} values $\leq$ 948 nM were regarded as ''actives'' and were assigned principal and MaxOmitFeat values of two and zero, respectively. On the other hand, inhibitors of $IC_{50} >$ 948 nM were considered inactive and were assigned principal values of zero.[45] However, each inactive compound was carefully evaluated to assess whether its low potency is attributable to missing one or more pharmacophoric features, that is, compared to active compounds, or related to possible steric clashes within the binding pocket, or due to both factors (that is, the MaxOmitFeat parameter was set to 0, 1, or 2). HIPHOP-REFINE was configured to allow a maximum of 100 exclusion spheres to be added to the generated pharmacophoric hypotheses. This represents the default value for the number of exclusion volumes in HIPHOP-REFINE. Table 5 shows the training compounds employed in this step and their corresponding principal and MaxOmitFeat parameters.

In silico screening of the NCI and drug databases for new DPP IV inhibitors

Hypo32/8 and Hypo4/10 combined with their respective excluded volumes were employed as 3D search queries against two multiconformer structural databases, namely, the national cancer institute list of compounds (NCI database, includes 238,819 compounds)[45] and our in-house multiconformer database of established drug molecules (includes 1490 compounds) via the "best flexible search" option in CATALYST. Only NCI hits were filtered based on Lipinski's and Veber's.[85]. Whereas we left the established drug hits without postscreening. The remaining hits (1498 molecules) were fitted against Hypo32/8 and Hypo4/10 using the "Best-Fit" option in CATALYST. Subsequently, their fit values together with other relevant molecular descriptors were substituted in equation 1 to determine their predicted anti-DPP IV activities and 35 high ranking compounds were acquired and tested in vitro.

Docking experiment

The chemical structure of **282** (NCI0294730) and **386** (gemifloxacin) were docked into the binding site of DPP IV (PDB code: 2G63, resolution = 2.00 Å)[88] employing FRED software (FRED, version 2.1.2 Users' Manual, 2006).[89] This docking engine takes a multiconformer database of the ligand(s) intended for docking, a target protein structure, a box defining the active site of the protein based on the co-crystallized ligand, and several optional parameters as input. The ligand conformers and protein structure are treated as rigid during the docking process. FRED's docking strategy is to exhaustively score all possible positions of each ligand in the active

site.[89] The exhaustive search is based on rigid rotations and translations of each conformer. Therefore, it avoids sampling issues associated with stochastic conformation-generation methods. The conformational space of the docked compounds was explored using OMEGA software. The software settings that best reproduced the co-crystallized poses of several ligand-DPP IV complexes (2G5T, 2G63, 1RWQ, 2AJL, 1X70, and 2HHA)[73,88,92–94] were employed in the docking experiment.

In vitro DPP IV enzyme inhibition assay

Each hit compound was dissolved in DMSO and diluted with Tris buffer (pH 7.5) for subsequent enzymatic assay. The assay was conducted using DPP IV drug discovery kit (Biomol, Germany), which is based on the cleavage of chromogenic substrate (H-Gly-Pro-*para*-Nitroaniline) by DPP IV to release *para*-nitroaniline (pNA) measured at 405 nm. Briefly, recombinant DPP IV was diluted in Tris buffer (pH 7.5, 50 mM) to a final enzymatic solution of 17.34 μU μL^{-1}. Subsequently, 15 μL aliquots of the enzymatic solution were pipetted into clear microplate wells and then appropriate volumes of the hits' stock solutions were added, completed to 50 μL with Tris buffer, and incubated at 37 °C for 20 min. Finally, 50 μL of the substrate solution (0.20 mM in Tris buffer) was added to each well. The plate was read at 405 nm in a microplate reader (BioTek, USA) and the rate of reduction of substrate absorbance was evaluated over 10 min and compared to a negative control (enzymatic solution without inhibition). A standard DPP IV inhibitor (P32/98 from Biomol, Germany) was employed as positive control.

Oral glucose tolerance test (OGTT) in mice

The animal experiments comply with the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health.[95] OGTT was performed on male BALB/c mice 25–30 g (~12 weeks old), obtained from Jordan University animal house, Jordan. The mice were housed in temperature controlled cages (20–22 °C) with a 12 h light–dark cycle, and given free access to water and formulated diet. Control and treated groups were matched for body weight in all experiments. Gemifloxacin (LG Life Sciences, Korea) was dissolved in sterile normal saline. Three doses of gemifloxacin were used; 18, 36, and 72 mg kg^{-1}. Mice were food deprived for 6 h at the end of the dark cycle before glucose tolerance tests were performed. The drug was intraperitoneally administered 30 min prior to oral glucose load evaluation. Controls were given normal saline i.p. The volume load was 10 μL/g body weight. Oral glucose was administered at 2 g kg^{-1} of body weight, thereafter, glucose levels were measured from tail bleeds with a glucometer (Arkray, Inc., Japan) at 0, 15, 30, 60, and 120 min after the oral glucose challenge.

Data are presented as means ± S.E.M. Statistical comparisons were performed using a one-way ANOVA or unpaired Student's t test. In all cases, $p < 0.05$ is considered statistically significant.

Acknowledgements

This project was partially sponsored by the Faculty of Graduate Studies (Ph.D. Thesis of Ihab M. AlMasri). The authors wish to thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds. The authors would like to thank also the OpenEye Scientific Software for providing us a free license of FRED software (FRED, version 2.1.2.).

Keywords: DPP IV · pharmacophore modeling · QSAR · in silico screening · in vivo and in vitro validation

[1] J. R. Bjelke, A. B. Kanstrup, H. B. Rasmussen, *Cell. Mol. Biol.* **2006**, *52*, 3.
[2] J. R. Bjelke, J. Christensen, S. Branner, N. Wagtmann, C. Olsen, A. B. Kanstrup, H. B. Rasmussen, *J. Biol. Chem.* **2004**, *279*, 34691.
[3] C. A. Abbott, M. D. Gorrell, in *Ectopeptidases: Cd13aminopeptidase N And Cd26dipeptidylpeptidase Iv In Medicine And Biology* (Eds: J. Langner, S. Ansorge), Kluwer/Plenum, **2002**, New York, pp. 171.
[4] A. M. Lambeir, P. J. F. Díaz, P. Chacón, G. Vermeulen, K. Heremans, B. Devreese, J. Van Beeumen, I. De Meester, S. Scharpé, *Biochim. Biophys. Acta.* **1997**, *1340*, 215.
[5] J. S. Duke-Cohan, C. Morimoto, J. A. Rocker, S. F. Schlossman, *J. Immunol.* **1996**, *156*, 1714.
[6] S. Hartel, R. Gossrau, C. Hanski, W. Reutter, *Histochemistry* **1988**, *89*,151.
[7] M. Engel, T. Hoffmann, L. Wagner, M. Wermann, U. Heiser, R. Keifersauer, R. Huber, W. Bode, H-U. Demuth, H. Brandstetter, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5063.
[8] R. A. Pederson, H. A. White, D. Schlenzig, R. P. Pauly, C. H. McIntosh, H-U. Demuth, *Diabetes* **1998**, *47*, 1253.
[9] J. A. Pospisilik, S. G. Stafford, H-U. Demuth, C. H. McIntosh, R. A. Pederson, *Diabetes* **2002**, *51*, 2677.
[10] J. D. Cheng, R. L. Dunbrack Jr. , M. Valianou, A. Rogatko, R. K. Alpaugh, L. M. Weiner, *Cancer Res.* **2002**, *62*, 4767.
[11] H. Kajiyama, F. Kikkawa, T. Suzuki, K. Shibata, K. Ino, S. Mizutani, *Cancer Res.* **2002**, *62*, 2753.
[12] L. Ho, U. Aytac, L. C. Stephens, K. Ohnuma, G. B. Mills, K. S. McKee, C. Neumann, R. LaPushin, F. Cabanillas, J. L. Abbruzzese, C. Morimoto, N. H. Dang, *Clin. Cancer Res.* **2001**, *7*, 2031.
[13] L. A. Frohman, T. R. Downs, E. P. Heimer, A. M. Felix, *J. Clin. Invest.* **1989**, *83*, 1533.
[14] S. Ahmad, L. Wang, P. E. Ward, *J. Pharmacol. Exp. Ther.* **1992**, *260*,1257.
[15] D. J. Drucker, M. A. Nauck, *Lancet* **2006**, *368*,1696.
[16] M. A. Nauck, U. Niedereichholz, R. Ettler, J. J. Holst, C. Orskov, R. Ritzel, W. H. Schmiegel, *Am. J. Physiol.* **1997**, *273*, E981.
[17] P. L. Brubaker, D. J. Drucker, *Endocrinology* **2004**, *145*, 2653.
[18] Q. Wang, L. Li, E. Xu, V. Wong, C. Rhodes, P. Brubaker, *Diabetologia* **2004**, *47*, 478.
[19] J. J. Meier, M. A. Nauck, W. Schmidt, B. Gallwitz, *Regul. Pept.* **2002**, *107*, 1.
[20] K. Adelhorst, B. B. Hedegaard, L. B. Knudsen, O. Kirk, *J. Biol. Chem.* **1994** *269*, 6275.
[21] B. Rolin, C. F. Deacon, R. Carr, B. Ahrén, *Eur. J. Pharmacol.* **2004**, *494*, 283.
[22] A. Ceriello, *Diabetes Metab. Res.* **2000**, *16*, 125.
[23] M. Engel, T. Hoffmann, S. Manhart, U. Heiser, S. Chambre, R. Huber, H-U. Demuth, W. Bode, *J. Mol. Biol.* **2006**, *355(4)*, 768.
[24] B. Kuhn, M. Hennig, P. Mattei, *Curr. Top. Med. Chem.* **2007**, *7*, 609.
[25] C. F. Deacon, B. Ahren, J. J. Holst, *Expert Opin. Invest. Drugs* **2004**, *13*, 1091.
[26] B. Hulin, S. Cabral, M. G. Lopaze, M. A. Van Volkenburg, K. M. Andrews, J. C. Parker, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4770.
[27] A. Mastracchio, E. R. Parmee, B. Leiting, F. Marsilio, R. Patel, N. A. Thornberry, A. E. Weber, S. D. Edmondson, *Heterocycles* **2004**, *62*, 203.
[28] D. Kim, J. E. Kowalchick, L. Brockunier, E. R. Parmee, G. J. Eiermann, M. H. Fisher, H. He, B. Leiting, K. Lyons, G. Scapin, S. B. Patel, A. Petrov, K. D. Pryor, R. S. Roy, J. K. Wu, X. Zhang, M. J. Wyvratt, B. Zhang, L. Zhu, N. A. Thornberry, A. E. Weber, *J. Med. Chem.* **2008**, *51*, 589.
[29] T. Kondo, T. Nekado, I. Sugimoto, K. Ochi, S. Takai, A. Kinoshita, A. Hatayama, S. Yamamoto, K. Kawabata, H. Nakai, M. Toda, *Bioorg. Med. Chem.* **2008**, *16*, 190.
[30] B. D. Green, P. R. Flatt, C. J. Bailey, *Expert Opin. Emerg. Drugs* **2006**, *11*, 525.
[31] D. M. Kendall, D. Kim D. Maggs, *Diabetes Technol. Ther.* **2006**, *8*, 385.
[32] B. Ahren, M. Landin-Olsson, P. A. Jansson, M. Svensson, D. Holmes, A. Schweizer, *J. Clin. Endocrinol. Metab.* **2004**, *89*, 2078.

[33] B. Ahren, E. Simonsson, H. Larsson, M. Landin Olsson, H. Torgeirsson, P. A. Jansson, M. Sandqvist, P. Bavenholm, S. Efendic, J. W. Eriksson, S. Dickinson, D. Holmes, *Diabetes Care* **2002**, *25*, 869.
[34] C. Triplitt, A. Wright, E. Chiquette, *Pharmacotherapy* **2006**, *26*, 360.
[35] A. Bergman, D. Ebel, F. Liu, J. Stone, A. Wang, W. Zeng, L. Chen, S. Dilzer, K. Lasseter, G. Herman, J. Wagner, R. Krishna, *Biopharm. Drug Dispos.* **2007**, *28*, 315.
[36] B. Gallwitz, *Drugs Today* **2007**, *43*, 801.
[37] K. Hermansen, L. S. Mortensen, *Drug Saf.* **2007**, *30*, 1127.
[38] B. Ahren, *Curr. Cancer Drug Targets Curr. Diab. Rep.* **2007**, *7*, 340.
[39] F. Himmelsbach, WO 2008/017 670.
[40] V. M. Patell, A. Mathur, A. Suman, V. Devadoss, Fredrick Robin. U. S. Pat. Appl. Publ., 2008/020971.
[41] R. A. Ward, T. D. J. Perkins, J. Stafford, *J. Med. Chem.* **2005**, *48*, 6991.
[42] W. Brandt, T. Lehmann, I. Thondorf, I. Born, M. Schutkowski, J-U. Rahfield, K. Neubert, A. Barth, *Int. J. Pept. Protein Res.* **1995**, *46*, 494.
[43] R. R. S. Pissurlenkar, M. S. Shaikh, E. C. Coutinho, *J. Mol. Model.* **2007**, *13*, 1047.
[44] M. Akamatsu, *Curr. Top. Med. Chem.* **2002**, *2*, 1381.
[45] Catalyst User Guide, Accelrys Software Inc., San Diego, **2005**.
[46] M. O. Taha, Y. Bustanji, M. Al-Ghussein, M. Mohammad, H. Zalloum, I. M. Al-Masri, N. Atallah, *J. Med. Chem.* **2008**, *51*, 2062.
[47] M. O. Taha, A. Al-Bakri, W. Zalloum, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5902.
[48] M. O. Taha, Y. Bustanji, A. Al-Bakri, M. Yousef, W. Zalloum, I. M. Al-Masri, N. Atallah, *J. Mol. Graphics Modell.* **2007**, *25*, 870.
[49] M. O. Taha, N. Atallah, A. Al-Bakri, C. Paradis-Bleau, H. Zalloum, K. Younis, R. Levesque, *Bioorg. Med. Chem.* **2008**, *16*, 1218.
[50] P. W. Sprague, R. Hoffmann, in *Computer Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry,* (Eds.: H. V. Waterbeemd, B. Testa, G. Folkers), VHCA, Basel, **1997**, pp. 230.
[51] D. Barnum, J. Greene, A. Smellie, P. Sprague, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563.
[52] A. Smellie, S. Teig, P. Towbin, *J. Comput. Chem.* **1995**, *16,* 171.
[53] H. Li, J. Sutter, R. Hoffmann, in *Pharmacophore Perception, Development, and Use in Drug Design*, (Ed.: O. F. Güner), International University Line, California, **2000**, pp. 173.
[54] J. Sutter, O. Güner, R. Hoffmann, H. Li, M. Waldman, in *Pharmacophore Perception, Development, and Use in Drug Design*, (Ed.: O. F. Güner), International University Line, California, **2000**, pp. 501.
[55] Y. Kurogi, O. F. Güner, *Curr. Med. Chem.* **2001**, *8*, 1035.
[56] I. B. Bersuker, S. Bahçeci, J. E. Boggs, in *Pharmacophore Perception, Development, and Use in Drug Design*, (Ed.: O. F. Güner), International University Line, California, **2000**, pp. 457.
[57] K. Poptodorov, T. Luu, R. Hoffmann, in *Methods and principles in Medicinal Chemistry, Pharmacophores and Pharmacophores Searches, Vol. 2* (Eds.: T. Langer, R. D. Hoffmann), WILEY-VCH: Weinheim, **2006**, pp.17.
[58] J. Singh, C. E. Chuaqui, P. A. Boriack-Sjodin, W-C. Lee, T. Pontz, M. J. Corbley, H-K. Cheung, R. M. Arduini, J. N. Mead, M. N. Newman, J. L. Papadatos, S. Bowes, S. Josiah, L. E. Ling, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4355.
[59] M. O. Taha, A. M. Qandil, D. D. Zaki, M. A. AlDamen, *Eur. J. Med. Chem.* **2005**, *40*, 701.
[60] P. A. Keller, M. Bowman, K. H. Dang, J. Garner, S. P. Leach, R. Smith, A. McCluskey, *J. Med. Chem.* **1999**, *42*, 2351.
[61] R. G. Karki, V. M. Kulkarni, *Eur. J. Med. Chem.* **2001**, *36*, 147.
[62] N. Bharatham, K. Bharatham, K. W. Lee, *Bull. Korean Chem. Soc.*, 2007, 28, 200.
[63] E. Parmee, J. He, A. Mastracchio, S. Edmondson, L. Colwell, G. Eiermann, W. Feeney, B. Habulihaz, H. He, R. Kilburn, B. Leiting, K. Lyons, F. Marsilio, R. Patel, A. Petrov, J. Di Salvo, J. Wu, N. Thornberry, A. Weber, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 43.
[64] C. Caldwell, P. Chen, J. He, E. Parmee, B. Leiting, F. Marsilio, R. Patel, J. Wu, G. Eiermann, A. Petrov, H. He, K. Lyons, N. Thornberry, A. Weber, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1265.
[65] L. Brockunier, J. He, L. Colwell, B. Habulihaz, H. He, B. Leiting, K. Lyons, F. Marsilio, R. Patel, Y. Teffera, J. Wu, N. Thornberry, A. Weber, E. Parmee, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4763.
[66] S. Edmondson, A. Mastracchio, M. Beconi, L. Colwell, B. Habulihaz, H. He, S. Kumar, B. Leiting, K. Lyons, A. Mao, F. Marsilio, R. Patel, J. Wu, L. Zhu, N. Thornberry, A. Weber, E. Parmee, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5151.
[67] S. Edmondson, A. Mastracchio, J. Duffy, G. Eiermann, H. He, I. Ita, B. Leiting, J. Leone, K. Lyons, A. Makarewicz, R. Patel, A. Petrov, J. Wu, N. Thornberry, A. Weber, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3048.
[68] S. Edmondson, A. Mastracchio, R. Mathvink, J. He, B. Harper, Y-J. Park, M. Beconi, J. Di Salvo, G. J. Eiermann, H. He, B. Leiting, J. F. Leone, D. Levorse, K. Lyons, R. A. Patel, S. B. Patel, A. Petrov, G. Scapin, J. Shang, R. Roy, A. Smith, J. K. Wu, S. Xu, B. Zhu, N. Thornberry, A. Weber, *J. Med. Chem.* **2006**, *49*, 3614.
[69] W. Ashton, R. Sisco, H. Dong, K. Lyons, H. He, G. Doss, B. Leiting, R. Patel, J. Wu, F. Marsilio, N. Thornberry, A. Weber, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2253.
[70] J. Xu, H. Ok, E. Gonzalez, L. Colwell, B. Habulihaz, H. He, B. Leiting, K. Lyons, F. Marsilio, R. Patel, J. Wu, N. Thornberry, A. Weber, E. Parmee, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4759.
[71] J. Xu, L. Wei, R. Mathvink, J. He, Y-J. Park, H. He, B. Leiting, K. Lyons, F. Marsilio, R. Patel, J. Wu, N. Thornberry, A. Weber, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2533.
[72] J. Xu, L. Wei, R. Mathvink, S. Edmondson, A. Mastracchio, G. Eiermann, H. He, J. Leone, B. Leiting, K. Lyons, F. Marsilio, R. Patel, A. Petrov, J. Wu, N. Thornberry, A. Weber, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1346.
[73] J. Xu, L. Wei, R. J. Mathvink, S. D. Edmondson, G. J. Eiermann, H. He, J. F. Leone, B. Leiting, K. A. Lyons, F. Marsilio, R. A. Patel, S. B. Patel, A. Petrov, G. Scapin, J. K. Wu, N. A. Thornberry, A. E. Weber, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5373.
[74] R. P. Sheridan, S. K. Kearsley, *Drug Discovery Today* **2002**, *7*, 903.
[75] O. O. Clement, A. T. Mehl, in *Pharmacophore Perception, Development, and Use in Drug Design*, (Ed.: F. O. Guner), International University Line, California, **2000**, pp.71.
[76] R. Fisher, *The Principle of Experimentation Illustrated by a Psycho*-Physical, 8th ed., ExpeHafner Publishing Co., New York, **1966**, Chapter II.
[77] CERIUS2 QSAR users' manual, Accelrys Inc. San Diego, CA, **2005**, pp. 43, 221, 237.
[78] L. F. Ramsey, W. D. Schafer, *The Statistical Sleuth*, 1 st ed. Wadesworth Publishing Company, USA, **1997**.
[79] D. T. Stanton, P. C. Jurs, *Anal. Chem.* **1990**, *62*, 2323.
[80] L. H. Hall, L. B. Kier, "The molecular connectivity chi indices and kappa shape indices in structure-property modeling" in *Reviews of Computational Chemistry*, (Eds.: K. B. Lipkowitz, D. B. Boyd), New York, VCH, **1991**, 367.
[81] A. K. Ghose, G. M. Crippen, *J. Comput. Chem.* **1986**, *7*, 565.
[82] A. K. Ghose, V. N. Viswanadhan, G. R. Revankar, R. K. Robins, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163.
[83] C. J. Manly, J. Chandrasekhar, J. W. Ochterski, J. D. Hammer, B. Warfield, *Drug Discovery Today* **2008**, *13*, 99.
[84] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **2001**, *46*, 3.
[85] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615.
[86] M. T. D. Cronin, T. W. Schultz, *J. Mol. Struct. (Theochem)* **2003**, *622*, 39.
[87] M. A. DePristo, P. I. de Bakker, T. L. Blundell, *Structure* **2004**, *12*, 831.
[88] Z. Pei, X. Li, K. Longenecker, T. W. Von Geldern, P. E. Wiedeman, T. H. Lubben, B. A. Zinker, K. Stewart, S. J. Ballaron, M. A. Stashko, A. K. Mika, D. W. Beno, M. Long, H. Wells, A. J. Kempf-Grote, D. J. Madar, T. S. McDermott, L. Bhagavatula, M. G. Fickes, D. Pireh, L. R. Solomon, M. R. Lake, R. Edalji, E. H. Fry, H. Sham, J. M. Trevillyan, *J. Med. Chem.* **2006**, *49*, 3520.
[89] FRED, **2006**. version 2.1.2 Users' Manual, OpenEye Scientific Software Inc., Santa Fe, New Mexico.
[90] E. M. Krovat, T. Langer, *J. Med. Chem.* **2003**, *46*, 716.
[91] The 9th European CATALYST User Group Meeting, Advanced Seminars in CATALYST, Frankfurt, Germany, March 24, Accelrys Inc.: San Diego, CA, **2006**.
[92] J. U. Peters, S. Weber, S. Kritter, P. Weiss, A. Wallier, M. Boehringer, M. Hennig, B. Kuhn, B. M. Loeffler, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1491.
[93] L. Qiao, C. A. Baumann, C. S. Crysler, N. S. Ninan, M. C. Abad, J. C. Spurlino, R. L. Desjarlais, J. Kervinen, M. P. Neeper, S. S. Bayoumy, R. Williams, I. C. Deckman, M. Dasgupta, R. L. Reed, N. D. Huebert, B. E. Tomczuk, K. J. Moriarty, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 123.

[94] D. Kim, L. Wang, M. Beconi, G. Eiermann, M. Fisher, H. He, G. Hickey, J. Kowalchick, B. Leiting, K. Lyons, F. Marsilio, M. McCann, R. Patel, A. Petrov, G. Scapin, S. B. Patel, R. S. Roy, J. Wu, M. Wyvratt, B. Zhang, L. Zhu, N. Thornberry, A. Weber, *J. Med. Chem.* **2005**, *48*, 141.

[95] Guide for the Care and Use of Laboratory Animals, U. S. National Institutes of Health (NIH Publication No. 85 – 23, revised 1985).

Received: July 8, 2008
Revised: August 20, 2008

