

Chenzhong Liao · Aihua Xie · Jiaju Zhou ·
Leming Shi · Zhibin Li · Xian-Ping Lu

3D QSAR studies on peroxisome proliferator-activated receptor γ agonists using CoMFA and CoMSIA

Received: 29 September 2003 / Accepted: 11 November 2003 / Published online: 12 March 2004
© Springer-Verlag 2004

Abstract The peroxisome proliferator-activated receptors (PPARs) have increasingly become attractive targets for developing novel anti-type 2 diabetic drugs. We employed comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) to study three-dimensional quantitative structure–activity relationship (3D QSAR) based on existing agonists of PPAR γ (including five thiazolidinediones and 74 tyrosine-based compounds). Predictive 3D QSAR models with conventional r^2 and cross-validated coefficient (q^2) values up to 0.974 and 0.642 for CoMFA and 0.979 and 0.686 for CoMSIA were established using the SYBYL package. These models were validated by a test set containing 18 compounds. The CoMFA and CoMSIA field distributions are in general agreement with the structural characteristics of the binding pockets of PPAR γ , which demonstrates that the 3D QSAR models built here are very useful in predicting activities of novel compounds for activating PPAR γ .

Keywords 3D QSAR · CoMFA · CoMSIA · PPAR · Thiazolidinedione · Insulin sensitizer · Drug design

Introduction

The peroxisome proliferator-activated receptors (PPARs) belong to the nuclear receptor superfamily of ligand-activated transcription factors. [1] There are three known subtypes, namely, PPAR α , PPAR γ , and PPAR δ (first

reported as PPAR β) distributed in a spatial fashion among different tissues. The PPARs form heterodimers with retinoid X receptor (RXR) and activate upon binding to a hormone response element located in the promoter region of target genes by their cognate ligands. The activation of PPAR/RXR heterodimers results in specific induction of subsets of genes in controlling lipids, carbohydrates, and energy homeostasis. [2] Thus, inappropriate activation or inactivation of PPARs can be directly linked to pathological processes such as cases of type 2 diabetes, cardiovascular diseases, obesity, and dyslipidemia.

Recently, a class of compounds termed thiazolidinediones (TZD) has been developed as treatment for type 2 diabetes to reduce hyperglycemia by promoting insulin action without additional insulin secretion. [3] Their effects are proposed to be a result of initiation and modulation of adipocyte differentiation by agonist activity of PPAR γ . Although TZD type treatments improve insulin resistance, they offer little protection from eminent cardiovascular risk associated with type 2 diabetes. Side effects such as inductions of obesity, edema, and anemia in treated patients further hamper their extended use in the management of the diseases. Therefore, development of new treatments with insulin sensitizing and cholesterol/triglyceride-lowering effects is of general interest.

Fibrates that are known to have triglyceride- and cholesterol-lowering activity activate another member of this family, the PPAR α , which is mainly expressed in tissues such as liver. PPAR α stimulates peroxisomal proliferation that enhances fatty acid oxidation, leading to reduced fatty acid levels in blood. [4] Most recently, PPAR δ was reported to modulate lipid metabolism in which PPAR δ serves as a widespread regulator of fat burning in tissues where PPAR α is much less expressed. Acute treatment of obese mice with a PPAR δ agonist leads to improved lipid profiles and reduced adiposity. [5] The biological data thus point to a therapeutic rationale for the design of agonists that could modulate three subtypes of PPAR receptors in a selective fashion.

C. Liao · A. Xie · L. Shi · Z. Li · X.-P. Lu (✉)
Chipscreen Biosciences Ltd.,
Research Institute of Tsinghua University, Suite C301,
P.O. Box 28, 518057 Shenzhen, Guangdong, China
e-mail: xplu@chipscreen.com
Tel.: +86-755-26711889
Fax: +86-755-26957291

C. Liao · A. Xie · J. Zhou
Institute of Process Engineering,
Chinese Academy of Sciences,
P.O. Box 353, 100080 Beijing, China

PPAR α , γ , and δ possess domain structures common to all other members of the nuclear hormone receptor superfamily, most notably the DNA-binding domain (DBD) and ligand-binding domain (LBD). The structures of the PPAR γ LBD with different agonists bound have been determined by X-ray crystallography. [6, 7, 8] A series of tyrosine-based PPAR γ agonists, [9, 10, 11] exemplified by GI 262570 (farglitazar), and other structurally diverse PPAR γ agonists [12, 13, 14, 15, 16] have been developed. Therefore, these data can be used as a training set to build proper QSAR models in an attempt to develop a methodology for better and quicker design of PPAR ligands.

CoMFA (comparative molecular field analysis) [17] and CoMSIA (comparative molecular similarity indices analysis) [18] are commonly used 3D QSAR methods. The widely used CoMFA method calculates steric and electrostatic properties according to Lennard-Jones and Coulomb potentials. The more recently reported CoMSIA approach calculates similarity indices in the space surrounding each of the aligned molecules in the data set. Similarity is expressed in terms of five different physicochemical properties: steric occupancy, electrostatic field, local hydrophobicity, and hydrogen-bond donor and acceptor properties. [19] Markus Böhm's study has shown that CoMSIA possesses better predictive power and greater robustness compared to CoMFA in some cases. [20] Another strength of CoMSIA is in visualization: it can bring forth five different contour maps, but CoMFA only can reveal two different contour maps.

Kurogi has developed a 3D QSAR model for antidiabetic TZDs using the Apex method. [21] This model identifies molecular features from only seven selected TZDs, and results in an inability to explain more complex compounds such as tyrosine-based PPAR γ agonists. Kulkarni et al. also have made a 3D QSAR model by using 53 TZDs as training set and 15 TZDs as test set, [22] but the method only contained CoMFA and also did not include more complex compounds.

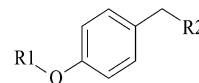
In this paper, we investigate the local physicochemical properties involved in the interaction between the agonists and the PPAR γ LBD by using CoMFA and CoMSIA. The results reveal that such models can be used to predict the affinity of new, untested PPAR γ agonists and, through analysis of contribution maps, to identify new opportunities in design and synthesis of high-affinity agonists.

Materials and methods

Molecular modeling and CoMFA, CoMSIA analyses were performed using SYBYL [23] version 6.81 running on a Silicon Graphics O2+ (R12000) workstation with the IRIX 6.5 operating system.

Data set

The structures of 97 agonists including six TZDs and 91 tyrosine-based compounds used in this study and their binding affinities (pK_i) with PPAR γ originated from the same laboratory. [9, 10, 11]



Scheme 1

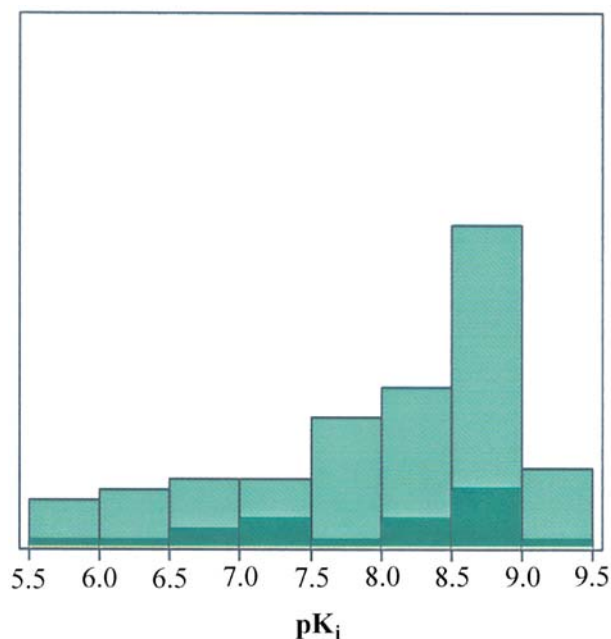


Fig. 1 Distribution of pK_i . The higher the column is, the more compounds it contains. The *dark green* represents test set

All compounds have been shown to be potent agonists of PPAR γ . Structural variations of the parent structure, present in all molecules, were allowed at positions R1 and R2 (see Scheme 1).

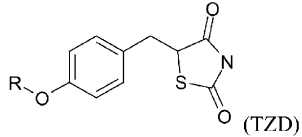
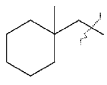
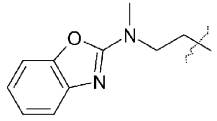
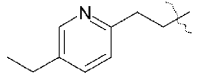
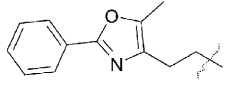
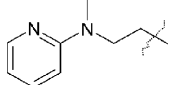
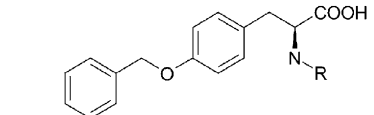
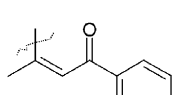
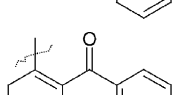
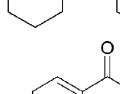
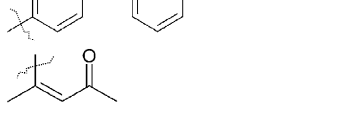
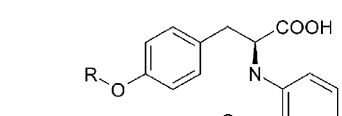
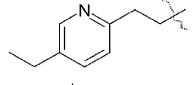
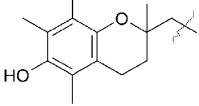
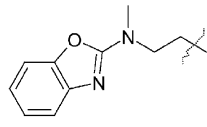
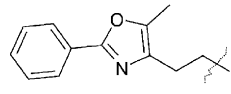
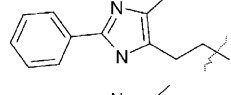
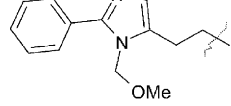
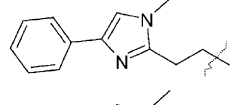
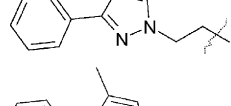
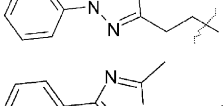
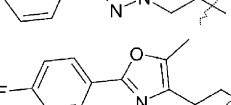
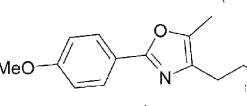
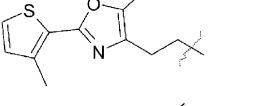
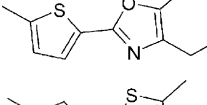
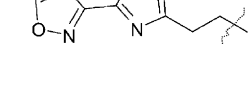

All compounds were treated as uncharged except for compounds **66** and **96** because they are dicarboxylic acids and a carboxy is connected to a phenyl ring directly, which results in stronger acidity than normal organic acids. These two compounds were treated as deprotonated and negatively charged in this study. The remaining tyrosine-based compounds all have a carboxy and an amino, so they were treated as neutral. A training set including 79 PPAR γ agonists was used for all CoMFA and CoMSIA analyses. These compounds are summarized in Table 1. Because only the (*S*)-enantiomers of the TZDs and tyrosine-based compounds bind to the receptor with high affinity, [24] the pK_i values of the racemates are augmented with log 2, i.e., 0.30 and the structures of the racemates used in the modeling were treated as the *S*-conformation. The alkene configurations of **78**, **79** and **83** were treated as *Z*, *E*, and *Z* respectively.

Fig. 1 shows the distribution of pK_i for both the training set and test set. The distribution of the data is not ideal for modeling since most of the data centralize in the region 8.5–9, meaning that most of the compounds have high affinity to PPAR γ .

Alignment

The high flexibility of the tyrosine-based compounds makes it much more difficult to obtain meaningful CoMFA and CoMSIA analyses than for rigid molecules. For example, through random conformational search in SYBYL by default parameters, we obtained the low energy conformers of farglitazar (**15**), and this conformation differs from the conformation in the crystal structure

Table 1 The structure of 79 agonists in the training set and their actual activities

no.	R	pK _i
		
1		5.81
2		7.87
3		6.21
4		8.67
5		7.63
6		7.93
7		6.10
8		5.90
9		5.71
10		6.79
11		7.94
12		8.19
13		8.28
14		8.83
15		8.94
16		8.59
17		8.70
18		9.16
19		8.75
20		8.90
21		8.32
22		8.80
23		8.96
24		8.72
25		9.07
26		9.05

Continuation of table 1

no.	R	pK _i	no.	R	pK _i
27		8.85	40		7.07
28		7.56	41		7.93
29 ^a		6.77	42		8.29
30		9.11	43		8.89
31		8.36	44		6.91
32		6.98	45		7.11
33		7.41	46		8.43
34		8.03	47		7.67
35		7.73			
		7.55			
36		7.55			
37		8.17			
38		7.49			
39		8.36			

no.	R	pK _i
48	2'-CF ₃	8.87
49	3'-CF ₃	8.88
50	4'-CF ₃	8.59
51	2'-CH ₃	8.95
52	4'-CH ₃	8.85
53	2'-OCH ₃	9.06
54	3'-OCH ₃	8.94
55	3'-OCH ₂ Ph	7.57
56	4'-Ph	7.61

Continuation of table 1

no.	R	pK _i
57	cyclopentyl	8.49
58	cyclohexyl	8.39
59 ^b	cyclohexyl	5.81
60	cycloheptyl	7.70
61	3-thienyl	8.93
62	1-naphthyl	8.79
63	4'-CH ₂ OH	8.68
64	4'-CH ₂ NMe ₂	8.11
65	3'-CH ₂ OH	8.77
66 ^c	3'-COOH	6.39
67	3'-CH ₂ NMe ₂	6.24
68	3'-NH ₂	8.79
no.	R	pK _i
69	cyclohexyl	8.79
70 ^b	cyclohexyl	6.72
71	3-pyridyl	9.03
72	4-pyridyl	8.74
73	NHMe	8.11
74	NMe ₂	6.90
75	OMe	8.43
76	OEt	8.52
77	OiPr	9.01
no.	Other Structure	pK _i
78		6.12
79		6.79

^a This compound was an oxazole in the original of reference 10 (compound **21**), but a careful reading suggests that it is actually a thiazole, like **28** and **30**

^b This compound is (*R*)-enantiomer

^c This compound was treated as deprotonated and negatively charged

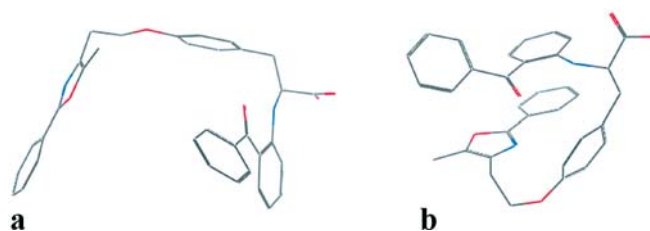


Fig. 2 **a** The conformation of farglitazar (**15**) in 1FM9 crystal. **b** The lowest energy conformation of farglitazar obtained by random conformational search in SYBYL

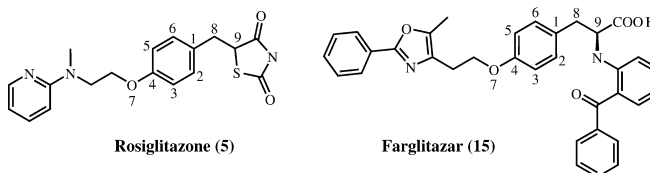


Fig. 3 Compounds used as templates for molecular alignment; atoms labeled in bold face and numbered 1–9 are used as the reference atoms in alignment protocol of the SYBYL program

(1FM9) significantly (see Fig. 2). This is due to some aromatic rings in farglitazar that induce π - π stacking.

To obtain a consistent alignment, two crystal structures of PPAR γ LBD from the Protein Data Bank (PDB) [25] were used as references (1FM6, 1FM9). [26] The first contains a TZD, rosiglitazone (**5**), and the second contains a tyrosine-based compound, farglitazar (**15**). We used these two compounds as templates to align all molecules. The reference atoms used as the alignment points were as follows: (i) six aromatic carbon atoms of the common phenyl alkyl ether moiety in the middle part of these molecules; (ii) an oxygen atom and a carbon atom connected this aromatic ring; (iii) the chiral carbon atoms situated at the ring of thiazolidinedione or the α -carbon atoms in tyrosine-based compounds (see atoms numbered 1–9 in Fig. 3).

The R1 and R2 substituents were constructed from the SYBYL fragment database and were relaxed to local minima using the TRIPOS force field. [27] Energy minimizations were performed for the molecules using the Tripos force field with a distance-dependent dielectric and the Powell conjugate-gradient algorithm with a convergence criterion of 0.01 kcal mol⁻¹. Partial atomic charges were calculated using the semiempirical program MOPAC 6.81 [28] and applying the AM1 Hamiltonian. The structural diversity of the aligned ligands obtained is shown in Fig. 4. The finally accepted superposition showed reasonable fit to the binding pockets.

CoMFA and CoMSIA 3D QSAR modeling

In both CoMFA and CoMSIA analyses, a 3D cubic lattice with grid spacing of 2 or 1 Å was created to encompass the aligned molecules. CoMFA descriptors were calculated using an *sp*³-carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard-Jones 6–12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. The SYBYL default energy cutoff of 30 kcal mol⁻¹ was used.

CoMSIA calculates similarity indices at the intersections of a surrounding lattice. The similarity index A_F for a molecule j with i atoms at the grid point q is determined as follows: [29]

$$A_{F,k}^q(j) = -\sum \omega_{probe,k} \omega_{ik} e^{-\alpha r_{iq}^2}$$

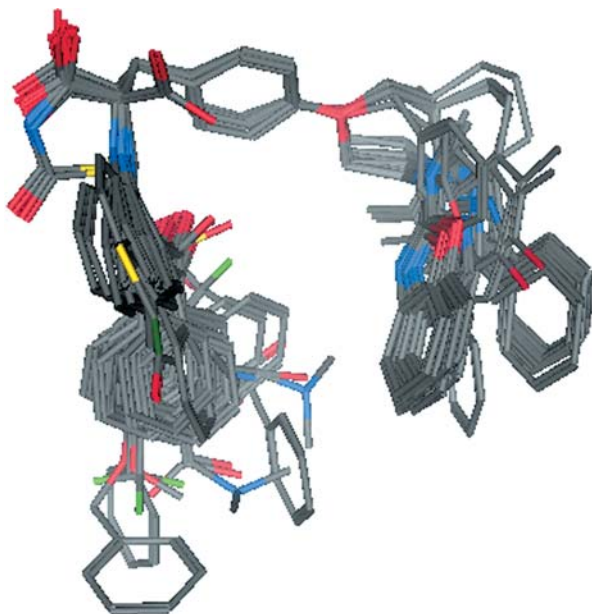


Fig. 4 Alignment of the 79 agonists in the training set

Five physicochemical properties k (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated, using a common probe atom with 1-Å radius and charge, hydrophobicity, and hydrogen-bond property of +1. A Gaussian type distance dependence was considered between the grid point q and each atom i of the molecule. The value of the so-called attenuation factor α was set to the default value, 0.3.

The CoMFA and CoMSIA descriptors were used as independent variables, and pK_i values were used as dependent variables in partial least squares (PLS) regression analyses to derive 3D QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated first by leave-

one-out (LOO) cross-validation. The optimal number of components was determined by selecting the smallest S_{PRESS} value. Usually this value corresponds to the highest q^2 value. The q^2 was calculated using

$$q^2 = 1 - \frac{\sum (Y_{\text{pred}} - Y_{\text{actual}})^2}{\sum (Y_{\text{actual}} - Y_{\text{mean}})^2}$$

where Y_{pred} , Y_{actual} , and Y_{mean} are predicted, actual, and mean values of the target property (pK_i), respectively. The same number of components was subsequently used to derive the final QSAR models. In addition to the q^2 , the corresponding S_{PRESS} , the number of components, the conventional correlation coefficient r^2 , and its standard error s were also computed.

Results and discussion

3D QSAR models

The statistical results are summarized in Table 2. The q^2 , S_{PRESS} , r^2 , F , and s values were computed as defined in SYBYL. As for the 79 compounds modeled by standard parameters, because compound **29** has a residual of -1.42 logarithm units in CoMFA and a residual of -1.53 logarithm unit in CoMSIA and compound **74** has a residual of -1.15 logarithm unit in CoMFA and a residual of -0.98 logarithm unit in the 2-Å grid spacing CoMSIA model, they appear to be outliers (the reason will be discussed later). After dropping them from the training set, we re-performed the 3D QSAR analyses on the data set of 77 compounds. The results are also shown in Table 2, which shows that the models based on the 77 compounds are much better than those based on 79 compounds. All the following results and discussion are based on the 77-compound models. The plots of predict-

Table 2 PLS statistics of CoMFA and CoMSIA 3D QSAR models

step size	PLS statistics	79 compounds model		77 compounds model	
		CoMFA	CoMSIA	CoMFA	CoMSIA
2 Å	q^2	0.514	0.580	0.642	0.686
	S_{PRESS}	0.795	0.750	0.716	0.708
	r^2	0.921	0.970	0.974	0.979
	F	108.0	248.3	288.2	358.2
	s	0.291	0.181	0.152	0.151
	PLS components	7	10	8	10
	Field contribution				
	Steric	0.593	0.173	0.563	0.168
	Electrostatic	0.407	0.256	0.437	0.241
	Hydrophobic		0.277		0.273
Donor		0.148		0.170	
Acceptor		0.145		0.148	
1 Å	q^2	0.509	0.573	0.597	0.677
	S_{PRESS}	0.780	0.755	0.756	0.714
	r^2	0.870	0.969	0.891	0.972
	F	89.1	246.2	107.4	361.3
	s	0.367	0.182	0.336	0.150
	PLS components	5	10	5	10
	Field contribution				
	Steric	0.504	0.173	0.509	0.272
	Electrostatic	0.496	0.255	0.491	0.167
	Hydrophobic		0.272		0.236
Donor		0.152		0.151	
Acceptor		0.148		0.175	

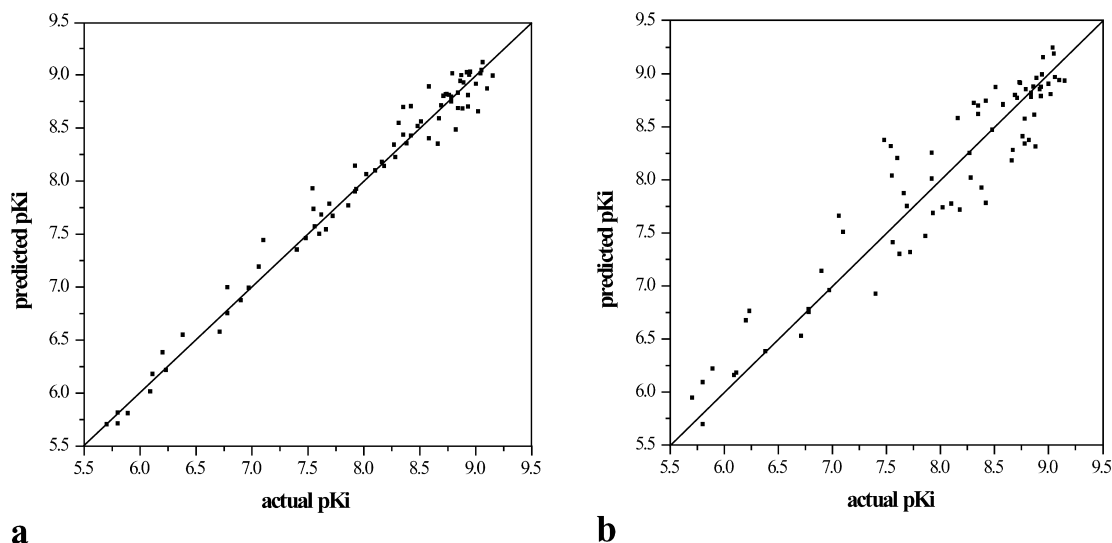


Fig. 5 Calculated predictions versus actual binding affinities for the 77 agonists of the training set (two outliers were picked out). The predicted values were obtained by PLS analyses using CoMFA with 2-Å (a) and 1-Å (b) grid spacing

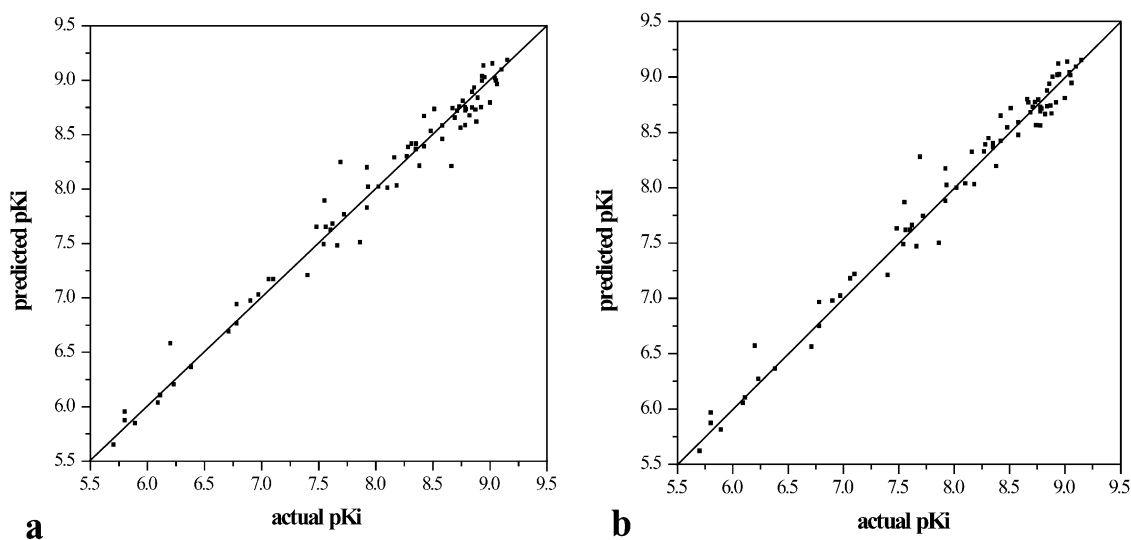


Fig. 6 Calculated predictions versus actual binding affinities for the 77 agonists of the training set (two outliers were picked out). The predicted values were obtained by PLS analyses using CoMSIA with 2-Å (a) and 1-Å (b) grid spacing

ed versus actual binding affinities for the fitted PLS analyses are shown in Figs. 5 and 6, which give the results of CoMFA and CoMSIA respectively. Also, augmenting the pK_i values of the racemic mixtures by $\log 2$ did not seem to result in over- or underprediction by these models.

Reducing the lattice step size from 2 to 1 Å resulted in a change in the q^2 value from 0.642 to 0.597 in the CoMFA analysis, from 0.686 to 0.677 in the CoMSIA analysis in the 77 compounds model. The r^2 also declined from 0.974 to 0.891 in the CoMFA analysis, from 0.979 to 0.972 in the CoMSIA analysis. Thus, in this study, the models based on 2-Å lattice step size are somewhat better than the models based on 1-Å lattice step size. In all

cases, the CoMSIA analyses reveal better correlations expressed in terms of higher q^2 value (Table 2). CoMFA also appears to be more dependent on grid spacing than CoMSIA. [20] Table 2 also indicates that the numbers of PLS components are higher in CoMSIA than for CoMFA.

Although pEC_{50} of these compounds with PPAR γ were also reported, [9, 10, 11] we could not obtain good models of CoMFA and CoMSIA through pEC_{50} (the correlation coefficient between pK_i and pEC_{50} is 0.66). After all, the free energy of binding is related to pK_i but not to pEC_{50} .

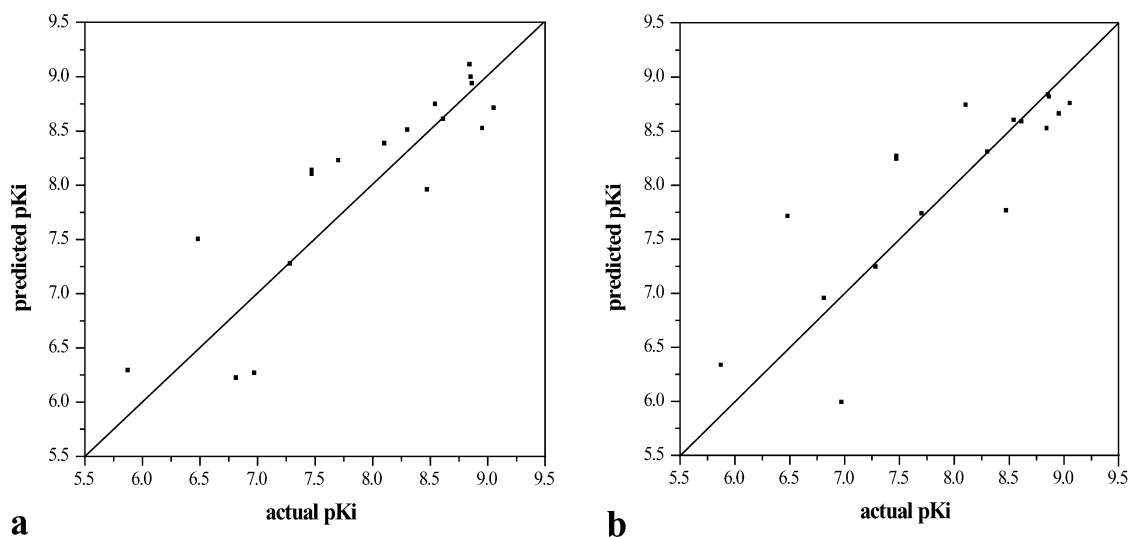
Table 3 The structure of 18 agonists in the test set and their actual activities

no.	Structure	pK _i	no.	Structure	pK _i
80		6.82	91		8.87
81		7.29	92		8.55
82		8.62	93		7.48
83		5.88	94		8.86
84		6.98	95		8.31
85		8.85	96 ^a		6.49
86		7.48	97		8.11
87		8.48			
88		8.96			
89		9.06			
90		7.71			

^a This compound was treated as deprotonated and negatively charged

Table 4 Residuals of the predictions of the test set by the CoMFA and CoMSIA models

Compound	pK_i	Residuals			
		CoMFA 2 Å	CoMFA 1 Å	CoMSIA 2 Å	CoMSIA 1 Å
80	6.82	0.60	-0.13	0.19	0.41
81	7.29	0.02	0.05	0.16	0.09
82	8.62	0.30	-0.43	-0.37	0.16
83	5.88	-0.41	-0.45	-0.23	-0.33
84	6.98	0.72	0.99	0.45	0.87
85	8.85	-0.26	0.33	0.09	0.35
86	7.48	-0.65	-0.78	-0.53	-0.43
87	8.48	0.53	0.72	0.65	0.36
88	8.96	-0.19	0.30	-0.20	0.14
89	9.06	0.35	0.31	0.16	0.20
90	7.71	-0.51	-0.02	-0.28	-0.41
91	8.87	-0.06	0.06	0.11	-0.03
92	8.55	-0.19	-0.05	-0.08	-0.21
93	7.48	-0.62	-0.76	-0.32	-0.53
94	8.86	-0.13	0.03	0.07	0.12
95	8.31	-0.20	0.01	-0.15	0.12
96	6.49	-1.01	-1.22	-1.25	-0.97
97	8.11	-0.27	-0.63	-0.26	-0.55
r^2_{pred}		0.779	0.730	0.886	0.821

**Fig. 7** Predicted versus actual binding affinities for the 18 agonists not included in the training set. The predicted values were obtained by PLS analyses using the CoMFA method with 2-Å (a) and 1-Å (b) grid spacing

Predictive power of the CoMFA and CoMSIA models

To test the predictive power of the CoMFA and CoMSIA models obtained, 18 additional agonists (including one TZD and 17 tyrosine-based compounds, see Table 3) were selected as a test set. The molecules were built and aligned by the same protocol as described for the training set. The criteria for choosing training set are (i) the pK_i values distribution of the test set should be consistent with that of the training set; (ii) the volume of oriented molecules in the test set should be within the volume of oriented molecules in the training set. Predictions were performed using the models based on both 2-Å and 1-Å lattice spacing. Table 4 summarizes the results obtained from the CoMFA and CoMSIA predictions. The plots of predicted versus actual binding affinity for the test set

molecules are shown in Figs. 7 and 8, which show models based on CoMFA and CoMSIA respectively. The predictive r^2 (r^2_{pred}) values are also shown in Table 4 (the outlier **96** was dropped when calculating these values). The r^2_{pred} will be based on molecules of the test set only and is defined as:

$$r^2_{\text{pred}} = \frac{SD - PRESS}{SD}$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activity of the training set molecules. $PRESS$ is the sum of the squared deviation between predicted and actual activity values from every molecule in the test set.

By predicting the binding constants of 18 additional ligands not included in the training set, we appear to

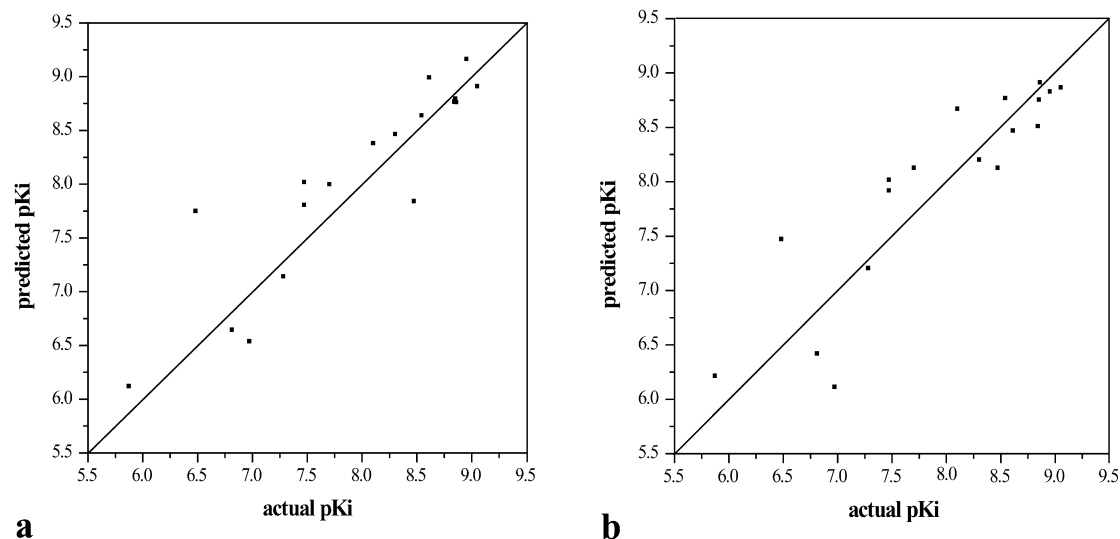


Fig. 8 Predicted versus actual binding affinities for the 18 agonists not included in the training set. The predicted values were obtained by PLS analyses using the CoMSIA method with 2-Å (a) and 1-Å (b) grid spacing

obtain 3D QSAR models with statistical significance. Both CoMFA and CoMSIA performed well in the prediction of the activities of the test compounds. In almost all the cases, the predicted values fall close to the observed pK_i values, deviating by not more than 1 logarithm unit in binding affinity to PPAR γ (except compound **96** which appears to be an outlier).

Outliers

The training set was checked initially for outliers. Empirically, agonists with a residual between experimental and predicted pK_i values above 1 logarithm unit were considered as outliers. According to these rules, compounds **29** and **74** were regarded as outliers. In the test set, compound **96** has residuals of more than 1 logarithm unit in almost all 3D QSAR models.

Omission of two compounds, **29** and **74**, resulted in an increase in the q^2 values from 0.514 to 0.642 for the CoMFA model, and 0.580 to 0.686 for the CoMSIA model when using 2-Å grid spacing. There are several reasons that may account for the outliers, for example, an incorrectly measured experimental value, a different binding conformation, a significant difference in the physiochemical properties, or structural uniqueness. Compound **29** appears to have lower activity than expected when compared to compounds **30** and **15**; and compound **74** appears to have lower activity than expected when compared to compounds **73** and **77**. In the training set, only compound **66** was deprotonated and negatively charged, as was compound **96** in the test set. Thus, there is only one counterpart, compound **66**, in the training set for compound **96**, which resulted in a larger residual when predicted by the QSAR models. The other reason of overprediction of compound **96** is the functional

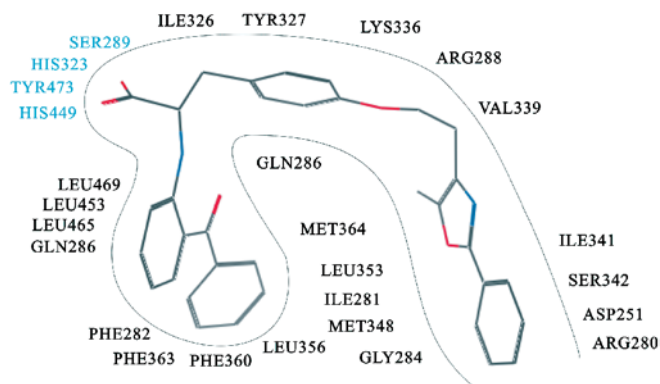


Fig. 9 The relevant amino acids surrounding the binding sites of the PPAR γ . The catalytic quadrant is formed by SER289, HIS323, TYR473 and HIS449 (in blue). Farglitazar (**15**) is exemplified as the agonist

carboxy group, which carries a full negative charge. Examination of the training set reveals two other molecules in the training set that are highly analogous (**63** and **64**) which possess slightly negatively charged functional groups in the 4-position of the lower part of the *N*-(2-benzoylphenyl) ring. These two compounds both have very high pK_i .

PPAR γ LBD structure

The PPAR γ LBD is a bundle of 13 α -helices and a small four-stranded β -sheet. [30] The crystal structure of the apo-PPAR γ (absence of ligand) LBD revealed a large ($\sim 1,300 \text{ \AA}^3$) Y-shaped ligand-binding site located within the bottom half of the LBD. [3, 4] Fig. 9 gives the catalytic quadrant of PPAR γ composed of SER289, HIS323, TYR473 and HIS449 [31] (the agonist is

exemplified by farglitazar). This figure can explain the difference in biological activity between TZD and tyrosine-based compounds. From Tables 1 and 3, we can see that the binding affinity of TZDs, **1–5**, **80** is lower than their tyrosine-based counterparts **81**, **14**, **12**, **15**, **85**, **13**. This is because there is a pocket in the bottom left of the active site where the *N*-(2-benzoylphenyl) of the tyrosine-based compounds can plug in (see Fig. 9), and this results in a more potent binding affinity (mainly through hydrophobic action) between the compound and PPAR γ . However, TZDs do not have such a group to plug into this pocket. Because the number of TZDs is much lower than that of tyrosine-based compounds, this activity difference cannot be revealed by the CoMFA or CoMSIA contour maps in this study.

Graphical interpretation of the results

Compared with CoMFA, CoMSIA is claimed to be less affected by changes in molecule alignment and to provide smoother and more interpretable contour maps as a result of employing Gaussian type distance dependence with the molecular similarity indices. [18] Furthermore, in addition to the steric and electrostatic fields, CoMSIA defines hydrophobic and hydrogen-bond donor and acceptor descriptor fields, which are not available with standard CoMFA. The use of CoMSIA along with CoMFA shows that in most instances the former performs similarly to CoMFA in terms of predictive ability, sometimes slightly better, and other times slightly worse than CoMFA. In this study, the steric and electrostatic PLS stdev*coeff contour maps for the CoMFA models (not shown here) were similarly placed as those of CoMSIA models, albeit the polyhedra volumes of the two models are somewhat different from each other.

The contribution maps obtained by CoMSIA show how 3D QSAR methods can identify features important for the interaction between small molecules and the protein. They allow identification of those positions that require a particular physicochemical property to improve binding affinity. The steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor contours of CoMSIA are shown in Fig. 10 (a–e, respectively). The steric fields (green, more steric bulk favored; yellow, more steric bulk disfavored) and electrostatic fields (blue, negative charge favored; red, positive charge favored) and hydrophobic fields (white, hydrophobic group favored; yellow, hydrophobic disfavored) and the H-bond donor (cyan, favored; purple, disfavored) and H-bond acceptor (magenta, favored; red disfavored) fields indicate areas around the molecules where changes increase or decrease activity.

The CoMSIA steric and electrostatic fields are consistent with the known structure–activity relationship for the compounds contained in the training set from experimental data. For example, the green contour in the region around the R1 group for this CoMSIA model indicates that the presence of bulky groups is expected to enhance

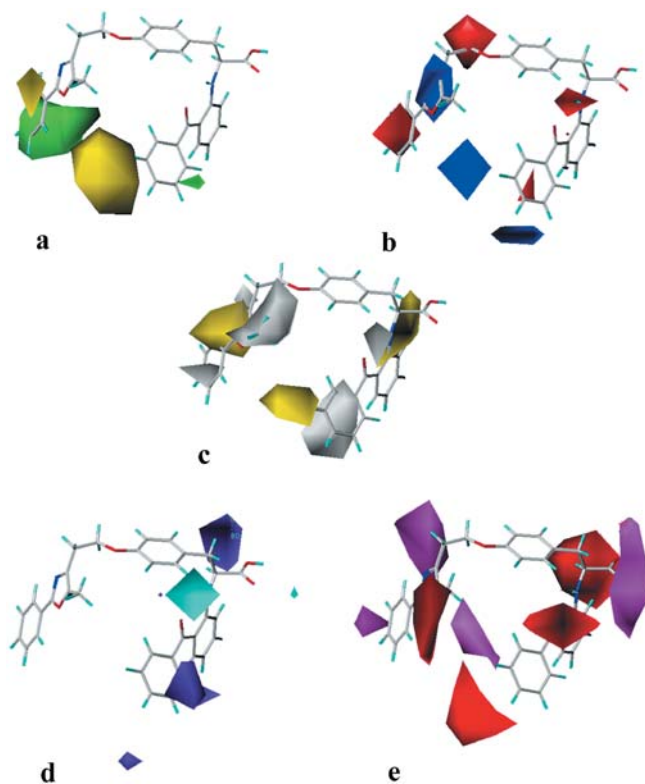


Fig. 10 CoMSIA stdev*coeff contour maps. Farglitazar (**15**) is shown inside the field. **a** Steric map. *Green* and *yellow* polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the affinity. **b** Electrostatic map. *Blue* and *red* polyhedra indicate regions where negative charge or positive charge will enhance the affinity. **c** Hydrophobic map. *Yellow* and *white* polyhedra indicate regions where hydrophobic or hydrophilic groups, respectively, will enhance the affinity. **d** H-bond donor ability map. *Cyan* and *purple* polyhedra indicate regions where hydrogen bond donor groups on the receptor will enhance or disfavor the binding. **e** H-bond acceptor ability map. *Magenta* and *red* polyhedra indicate regions where hydrogen bond acceptor groups on the receptor will enhance or disfavor the binding

binding affinity, examples as shown in the compounds **1–5**. Consistent with the presence of the rather large yellow contour close to the R2 side, diminished binding affinity is shown for compounds **55**, **56** in which bulky groups are substituted at *N*-(2-benzoylphenyl).

Fig. 10b shows that there is a blue contour in the 4-position of the lower part of the *N*-(2-benzoylphenyl) ring where negative charge would be favorable for high pK_i . For instance, compounds **63** and **64** have slightly negatively charged functional groups ($-\text{CH}_2\text{OH}$ and $-\text{CH}_2\text{NMe}_2$ respectively) at that site, so they have high pK_i value, as discussed before. Compound **96** places a fully charged carboxylate group in this position in the model and is thus heavily overpredicted due to the magnitude of the charge.

The hydrophobic field made the largest contribution to the CoMSIA QSAR models (Table 2), which suggests that among the descriptors considered, the hydrophobicity

is the most important factor influencing the PPAR γ activating ability of the compounds in the training set.

The contour map of H-bond properties should highlight the areas beyond the ligands where putative hydrogen-bond partners in the protein can form H-bonds that influence binding affinity significantly. In Fig. 10d, there is a cyan polyhedron near the oxygen atom of *N*-(2-benzoylphenyl). This means that hydrogen-bond donor groups near this oxygen atom on the receptor will enhance the binding ability. In fact, this oxygen forms two H-bonds with a water molecule and the nitrogen atom in *N*-(2-benzoylphenyl). [31] However, in this molecular modeling study, we did not consider the effect of water molecule. Therefore, in this situation, the water molecule should be a part of the receptor. Of course, if there is a H-bond acceptor near this oxygen atom, the biological activity will be reduced. This is consistent with Fig. 10e. In Fig. 10e, there is a magenta polyhedron and a red polyhedron in the upper right corner corresponding to the (*S*)-enantiomer and (*R*)-enantiomer of the tyrosine-based compounds, respectively.

Conclusions

Although there is a high degree of flexibility of compounds in the training set, we obtained 3D QSAR models with statistical significance and good predictive abilities by using CoMFA and CoMSIA. CoMSIA 3D QSAR models performed better than the CoMFA models in this study. Molecular surface property (steric, electrostatic, lipophilicity, and hydrogen-bonding potential) mapping has been integrated with CoMSIA 3D QSAR to refine what is known about the binding mode and highlight the cause of TZD and tyrosine-based compounds' enantioselectivity. TZDs such as rosiglitazone and pioglitazone enhance the sensitivity of target tissues to insulin and also reduce lipid and insulin levels in animal models of type 2 diabetes and clinically in type 2 diabetes. [32] However, TZDs have some side effects, for example, weight gain, edema, and anemia. [33] Some tyrosine-based compounds are also in clinical trials now. This study should provide further insights to support structure-based design of anti-type 2 diabetic drugs to develop novel PPAR γ agonists with improved activity profiles.

Acknowledgements The authors gratefully acknowledge financial supports from the national "863" project of People's Republic of China and project of "Most Important Biopharmaceutical Projects" by Guangdong province government in China. The authors also want to give special thanks to Drs Zhiqiang Ning and Weiming Hu for their support of this project.

References

1. Kersten S, Desvergne B, Wahli W (2000) *Nature* 405:421–424
2. Willson TM, Brown PJ, Sternbach DD, Henke BR (2000) *J Med Chem* 43:527–550
3. Lehmann JM, Moore LB, Smith-Oliver TA, Wilkison WO, Willson TM, Kliewer SA (1995) *J Biol Chem* 270:12953–12956
4. Issemann I, Green S (1990) *Nature* 347:645–650
5. Wang YX, Lee CH, Tiep SR, Yu T, Ham J, Kang H, Evans RM (2003) *Cell* 113:159–170
6. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, Rosenfeld MG, Willson TM, Glass CK, Milburn MV (1998) *Nature* 395:137–143
7. Uppenberg J, Svensson C, Jaki M, Bertilsson G, Jendeborg L, Berkenstam A (1998) *J Biol Chem* 273:31108–31112
8. Xu HE, Lambert MH, Montana VG, Plunket GM, Moore LB, Collins JL, Oplinger JA, Kliewer SA, Gampe RT, McKee DD, Moore JT, Willson TM (2001) *PNAS* 98:13919–13924
9. Henke BR, Blanchard SG, Brackeen MF, Brown KK, Cobb JE, Collins JL, Harrington WW, Hashim MA, Hull-Ryde EA, Kaldor I, Kliewer SA, Lake DH, Leesnitzer LM, Lehmann JM, Lenhard JM, Orband-Miller LA, Miller JF, Mook RA, Noble SA, Oliver W, Parks DJ, Plunket KD, Szweczyk JR, Willson TM (1998) *J Med Chem* 41:5020–5036
10. Collins JL, Blanchard SG, Boswell GE, Charifson PS, Cobb JE, Henke BR, Hull-Ryde EA, Kazmierski WM, Lake DH, Leesnitzer LM, Lehmann J, Lenhard JM, Orband-Miller LA, Gray-Nunez Y, Parks DJ, Plunkett KD, Tong WQ (1998) *J Med Chem* 41:5037–5054
11. Cobb JE, Blanchard SG, Boswell EG, Brown KK, Charifson PS, Cooper JP, Collins JL, Dezube M, Henke BR, Hull-Ryde EA, Lake DH, Lenhard JM, Oliver W, Oplinger J, Pentti M, Parks DJ, Plunket KD, Tong WL (1998) *J Med Chem* 41:5055–5069
12. Henke BR, Adkison KK, Blanchard SG, Leesnitzer LM, Robert A, Mook J, Plunket DK, Ray JA, Roberson C, Unwalla R, Willson TM (1999) *Bioorg Med Chem Lett* 9:3329–3334
13. Buckle DR, Cantello BCC, Cawthorne MA, Coyle PJ, Dean DK, Faller A, Haigh D, Hindley RM, Jecott LJ, Lister CA, Pinto IL, Rami HK, Smith DG, Smith SA (1996) *Bioorg Med Chem Lett* 6:2121–2126
14. Buckle DR, Cantello BCC, Cawthorne MA, Coyle PJ, Dean DK, Faller A, Haigh D, Hindley RM, Jecott LJ, Lister CA, Pinto IL, Rami HK, Smith DG, Smith SA (1996) *Bioorg Med Chem Lett* 6:2127–2130
15. Shinkai H, Onogi S, Tanaka M, Shibata T, Wao M, Wakitani K, Uchida I (1998) *J Med Chem* 41:1927–1933
16. Sauerberg P, Pettersson I, Jeppesen L, Bury PS, Mogensen JP, Wassermann K, Brand CL, Sturis J, Woldike HF, Fleckner J, Andersen AST, Mortensen SB, Svensson LA, Rasmussen HB, Lehmann SV, Polivka Z, Sindelar K, Panajotova V, Ynddal L, Wulff EM (2002) *J Med Chem* 45:789–804
17. Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959–5967
18. Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130–4146
19. Klebe G (1998) *Perspect Drug Discovery Des* 12:87–104
20. Böhm M, Stürzebecher J, Klebe G (1999) *J Med Chem* 42:458–477
21. Kurogi Y (1999) *Drug Des Discov* 16:109–118
22. Kulkarni SS, Gediya LK, Kulkarni VM (1999) *Bioorgan Med Chem* 7:1475–1485
23. SYBYL 681 (2001) Tripos Inc, 1699 Hanley Road, St Louis, MO 63144, USA
24. Parks DJ, Tomkinson NCO, Villeneuve MS, Blanchard SG, Willson TM (1998) *Bioorg Med Chem Lett* 8:3657–3658
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) *Nucleic Acids Res* 28:235–242
26. Gampe Jr RT, Montana VG, Lambert MH, Miller AB, Bledsoe RK, Milburn MV, Kliewer SA, Willson TM, Xu HE (2000) *Mol Cell* 5:545–555
27. Clark M, Cramer RD, van Opdenbosh N (1989) *J Comput Chem* 10:982–1012
28. Stewart JJ (1990) *J Comput-Aided Mol Des* 4:1–105

29. Kearsley SK, Smith GM (1990) *Tetrahedron Comput Methodol* 3:615–633
30. Weatherman RV, Fletterick RJ, Scanlan TS (1999) *Annu Rev Biochem* 68:559–581
31. Xu HE, Lambert MH, Montana VG, Plunket GM, Moore LB, Collins JL, Oplinger JA, Kliewer SA, Gampe RT, McKee DD, Moore JT, Willson TM (2001) *PNAS* 98:13919–13924
32. Grossman SL, Lessem L (1997) *Exp Opin Invest Drugs* 6:1025–1040
33. Henry RR (1997) *Endocrinol Metab Clin North Am* 26:553–573