

Three-dimensional Quantitative Structure Activity Relationships (3-D-QSAR) of Antihyperglycemic Agents

Santosh S. Kulkarni, Lalji K. Gediya and Vithal M. Kulkarni*

Pharmaceutical Division, Department of Chemical Technology, University of Mumbai, Mumbai 400 019, India

Received 20 January 1999; accepted 4 March 1999

Abstract—A three-dimensional quantitative structure activity relationship study (3-D-QSAR) was performed on a set of thiazolidinedione antihyperglycemic agents using the comparative molecular field analysis (CoMFA) method. The CoMFA models were derived from a training set of 53 compounds. Fifteen compounds, which were not used in model generation were used to validate the CoMFA models. All the compounds were superimposed to the template structure by atom-based and shape-based strategies. The SYBYL QSAR rigid body field fit was also used for aligning the ligands. A total of twelve different alignments were generated. The resulting models exhibited a good cross-validated r^2_{cv} values (0.624–0.764) and the conventional r^2 values (0.689–0.921). A more robust cross-validation test using cross-validation by 2 groups (leave half out method) was performed 100 times to ascertain the predictiveness of the CoMFA models. The mean of r^2_{cv} values from 100 runs ranged from 0.611–0.690. Few models exhibited good external predictivity. These models were then used to define a hypothetical receptor model for antihyperglycemic agents. © 1999 Published by Elsevier Science Ltd. All rights reserved.

Introduction

Non-insulin dependent diabetes mellitus (NIDDM) is characterized by insulin resistance in the liver and peripheral tissues together with a pancreatic β -cell defect.¹ The insulin resistant state at the peripheral level causes impaired glucose utilization leading to hyperglycemia.^{2–4} This insulin resistance together with the hyperinsulinemia that is often present, may also play a role in the etiology of a wider spectrum of metabolic disorders such as obesity, hypertension and atherosclerosis.⁵ A new series of 5-(4'-alkoxy benzyl)-2,4-thiazolidinediones are reported by Sohda as antihyperglycemic agents.⁶ The representative agent ciglitazone lowered elevated glucose, plasma insulin and triglyceride levels in insulin-resistant animals, but showed no hypoglycemic effect in non-diabetic or streptozotocin induced diabetic rats.⁷ Because of the several attractive features, such as improvement of insulin sensitivity and lack of hypoglycemia, development of agents of this type is becoming one of the major concerns in the field of antidiabetic drugs.

The exact mechanism of action of this class of compounds is not fully understood. A member of the nuclear receptor superfamily, peroxisome proliferator activated receptor γ (PPAR- γ), is proposed as a possible molecular

target for this type of compounds.^{8,9} The structure–activity relationships of antihyperglycemic 2,4-thiazolidinediones reported so far is shown in Figure 1. The compounds of this class have few essential pharmacophore elements. These are an acidic group linked to a central flat ring and a large lipophilic substructure. Various structural modifications containing these essential features have been reported. In the case of acidic fragment, thiazolidinedione was replaced with oxazolidinedione,¹⁰ tetrazole,¹¹ oxathiadiazole¹² and α -alkoxy carboxylic acid derivatives.¹³ The structural features override the acid strength of these fragments in determining the potency, as there was no correlation observed in the pK_a and potency of the compounds. The acidic fragment was attached to the central aromatic fragment by a carbon atom spacer. This carbon atom could be sp^2 or sp^3 hybridized. Chirality at the attachment position shows that both the isomers are active. The central phenyl linker fragment was replaced with thiophene,¹⁴ naphthalene,¹⁵ benzofuran,¹³ and benzoxazole.¹⁶ Central linker region contains the fragments, which restricts the conformations of terminal lipophilic fragment. The central aromatic region was linked to the lipophilic fragment with a 2 or 3 atom spacer. Various variations to this spacer have been reported, such as ethers,¹⁷ amides and ketones.¹⁴ This spacer has to present the lipophilic fragment in the proper spatial orientation to exhibit high activity. Numerous variations to the lipophilic fragment have been reported. This fragment had 2-pyridyl,¹⁸ phenyl,¹⁷ branched cycloalkyl,¹⁹ benzoxazolyl,²⁰ oxazolyl²¹ and

Key words: 3-D-QSAR; CoMFA; thiazolidinedione antihyperglycemics; receptor model.

* Corresponding author.

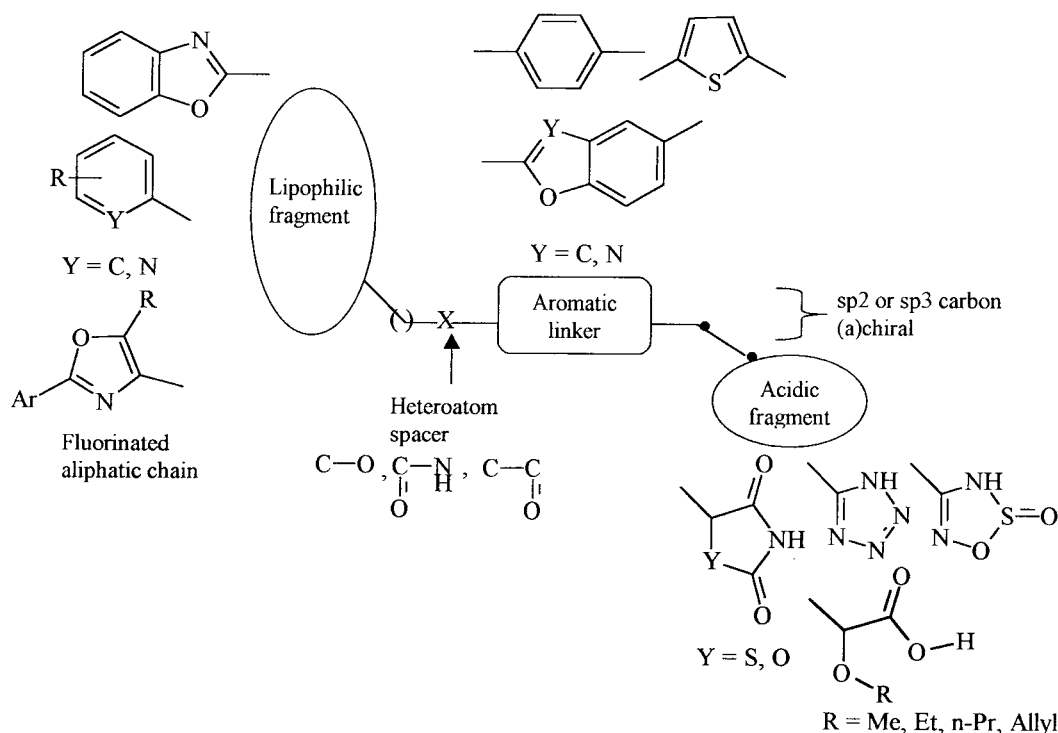


Figure 1. Summary of empirically derived structure–activity relationships of antihyperglycemic agents.

fluorinated aliphatic chain.¹¹ The structural fragments in this region also influence the nonreceptor events such as pharmacokinetics and toxicity profiles. In a limited study,¹⁰ a correlation was observed between logP of the compounds and potency. Increase in logP increases the potency but there was no optimum logP observed. Generally steric properties are more important than the lipophilicity in this region.

In order to obtain further insights into the structural requirements of antihyperglycemic compounds, we have performed a three-dimensional quantitative structure activity relationship study (3-D-QSAR) using comparative molecular field analysis (CoMFA) method.²² The CoMFA method of 3-D-QSAR was introduced by Cramer in 1988, in which an assumption is made that the interaction between an inhibitor and its molecular target is primarily noncovalent in nature and shape-dependent. Therefore, a QSAR may be derived by sampling the steric and electrostatic fields surrounding a set of ligands and correlating the differences in these fields to biological activity. This method can be used to develop a 3-D model, pharmacophore²³ describing the structure–activity relationship for a series of compounds. One advantage of CoMFA when the 3-D structure of the receptor is unknown is graphical representation of the results of the analysis as 3-D grids where the steric and electrostatic contributions of the activities are displayed.

In this paper we present a CoMFA study on thiazolidinedione antihyperglycemic agents. The *in vivo* antidiabetic activity was used as a dependent variable. The ligands were aligned using different pharmacophore elements as superimposition points. Overall 12 different

alignments were considered. The partial least square (PLS) method was used to derive the correlation. Extensive statistical analysis²⁴ such as leave half out cross-validation and test set predictions were performed to establish the predictiveness of the models. A hypothetical receptor model was proposed based on the analysis of the CoMFA contour maps.

Results

The CoMFA technique was used to derive a 3-D-QSAR model for the thiazolidinedione antihyperglycemic agents. The in vivo activity data of the thiazolidinedione antihyperglycemic agents was used as the dependent variable in the present study. All the molecules were aligned using six different recognition elements. These alignments were then realigned using SYBYL QSAR rigid body field fit. In all, this study involved 12 different alignments.

A preliminary study was performed on the atom-based alignment RMS I to decide upon the relevant CoMFA parameters that can have an influence on the final results. The results from the PLS analyses with varying column filtering value (σ min) are reported in Table 6. This analysis shows that a high r^2_{cv} value was obtained with σ min of 1.0 kcal/mol. A σ min value of 2.0 kcal/mol is widely used in most of the CoMFA studies reported but we have observed a decrease of 0.21 in r^2_{cv} with this σ min value. Setting the σ min values to still lower values did not improve the r^2_{cv} value but it does increase the required computational time for the PLS analysis. The comparison of F values shows that for the

analysis using σ min of 1.0 kcal/mol a good amount of variance can be explained using this model. Thus σ min of 1.0 kcal/mol was set in all further calculations as threshold column filtering value. To study the importance of each CoMFA fields, analyses using each field individually were performed. The cross-validated r^2_{cv} value (see Table 7) from the steric field only was higher than from electrostatic field only analysis. But both the values were lower than when both the fields were used for PLS analysis, suggesting that steric and electrostatic fields are equally important for CoMFA studies in this series of compounds. All further analyses were performed with steric and electrostatic fields calculated at each grid point simultaneously.

The CoMFA models were generated using the optimized parameters and then these models were validated using a test set. External predictions were used to select a putative pharmacophore model. Tables 8 and 9 depict the results obtained from atom-based and shape-based alignments, respectively. The atom-based alignment RMS I shows a r^2 cross-validated of 0.745 with first three components. A non-cross-validated r^2 of 0.917 and F value of 181.177 was observed with this model. The steric and electrostatic contributions were 59.7 and 40.3%, respectively. A graph depicting the actual versus fitted activities of the training set is shown in Figure 2. The leave one out method of cross-validation is rather obsolete and it generally gives a high r^2 value. To ascertain the true predictivity of the model a harder test using leave half out method of cross-validation was performed for all the analyses. This analysis was performed 100 times and the mean and standard deviation of the r^2 value for all the analyses are reported in Table 10. The mean r^2_{cv} value shows a good internal predictivity of the models. The RMS I model exhibited a good external prediction with a r^2_{pred} of 0.751. In fact this was the highest r^2_{pred} value observed in this study. Figure 3 shows a graph of actual versus predicted activities of the test set compounds. This alignment had only three points for superimposition.

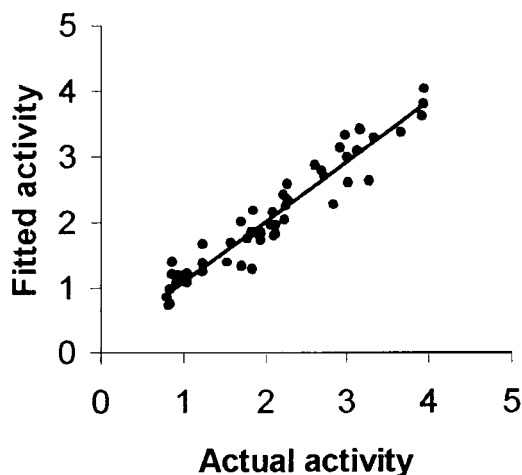


Figure 2. A graph of actual versus fitted activities of the training set molecules from RMS I analysis.

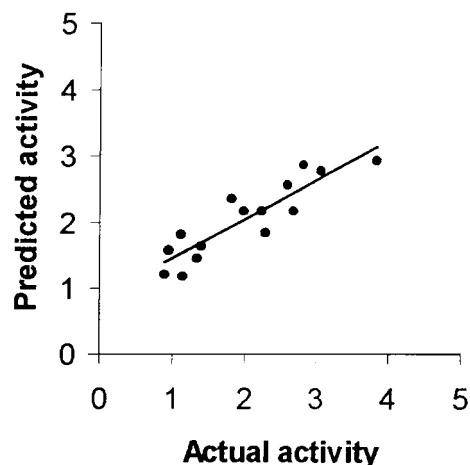


Figure 3. A graph of actual versus predicted activity of the test set molecules from RMS I analysis.

The activity data used in the present study may have contributions from other factors than just steric and electrostatic interactions. In order to consider such effects we have performed a limited study in which we have used an additional regressor in CoMFA table. The ClogP, the calculated logarithm of partition coefficient, was calculated and added to the CoMFA table. But inclusion of ClogP reduced the internal consistency of the model, r^2_{cv} of 0.704 (data not shown) and the external prediction was $r^2_{pred} = 0.719$. This additional descriptor did not increase the significance of the models. With good internal and external predictivity exhibited by steric and electrostatic fields alone, we performed all further studies using CoMFA fields only.

Realignment of the compounds using rigid body field fit (FF I) shows a r^2_{cv} of 0.634. The optimum number of components required was only one. The steric contributions override the electrostatic contributions with 69.2% contributions. But this alignment reduced the external predictivity, $r^2_{pred} = 0.504$. This clearly shows that the exact superimposition of the atoms is essential to exhibit good predictivity. The alignment of the compounds using atom-based selection RMS II shows a good internal predictivity with r^2_{cv} of 0.734. But this model exhibited a rather poor external predictivity with r^2_{pred} of -0.071 . This is because this alignment did not have 'O' atom of ethereal linkage as a superimposing element. Because of this the structural moieties of the test set ligands in the hydrophobic region were in the sterically unfavorable region. Also, the exact superimposition of the 'O' atom is essential to align central aromatic ring. Field fit alignment of this database reduced the internal prediction of the training set with r^2_{cv} of 0.628 with one optimum component. But there was an improvement in the external prediction from this alignment (r^2_{pred} was 0.384). The comparison of RMS II and FF II shows that the ligands were superimposed well on the training set when aligned using the rigid body field fit method. Considering the fact that the exact superimposition of 'O' atom is essential for good CoMFA predictions, alignments of the compounds using atom-based alignment RMS III with four points

of superimposition was performed. The training set ligands were predicted with a very good predictivity $r^2_{cv}=0.764$, but external predictivity of this model was very low with r^2_{cv} of 0.124. Realignment of the ligands with rigid body field fit reduced the internal predictivity ($r^2_{cv}=0.624$) but it increased the external predictivity $r^2_{pred}=0.438$.

Further analyses using shape-based alignments, i.e. use of ring centroids as superimposition element, were performed. The CoMFA results are depicted in Table 9. The RMS fitting of ring centroids and 'O' atom of ethereal linkage as RMS IV alignment shows high internal consistency in the dataset with r^2_{cv} of 0.759 with the optimum number of components of three. Non-cross-validated analysis showed a r^2 of 0.917. The steric and electrostatic fields contributed by 61.1 and 38.9%, respectively. This model exhibited a good external predictivity with r^2_{pred} of 0.625. Field fitting of this alignment (FF IV) reduced the internal predictivity ($r^2_{cv}=0.660$) and also the external predictivity ($r^2_{pred}=0.271$) of the test set. This decrease in predictivities from field fit alignment was also observed in case of atom-based alignment RMS I. Shape-based alignments using RMS V, RMS VI and subsequent field fit alignments FF V and FF VI did produce internally predictive models. All these alignments exhibit a moderate external predictivity. The statistical robustness of all the models was evaluated by bootstrapping analyses. All the CoMFA models exhibited a consistent r^2_{bs} value with low standard deviations during 100 runs of bootstrapping analysis. This shows the stability of the derived CoMFA models.

The atom-based RMS I alignment exhibited good internal as well as external predictivity hence all CoMFA contours were analyzed using this model. The CoMFA steric and electrostatic contour maps of RMS I analysis are shown in Figures 4 and 5. The field values were calculated at each grid point as the scalar product of the associated QSAR coefficient and the standard deviation of all values in the corresponding column of the data table (STDDEV*COEFF) and are plotted as percentage contributing to the QSAR equation.

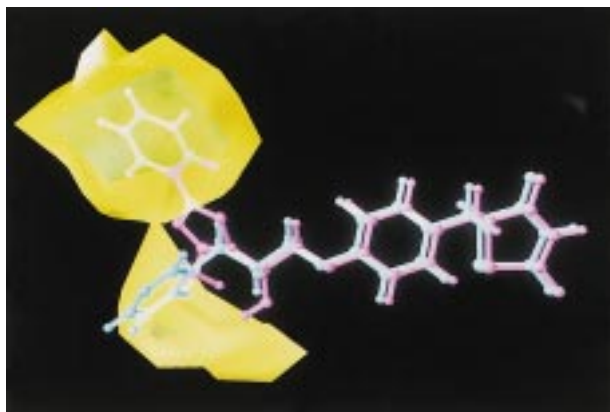


Figure 4. CoMFA steric STDEV*COEFF contour plots from the RMS I analysis. Sterically favored areas (contribution level of 70%) are represented by green polyhedra. Sterically disfavored (contribution level of 30%) are represented by yellow polyhedra. The active compound **45** (magenta) and low active compound **6** (cyan) are shown.

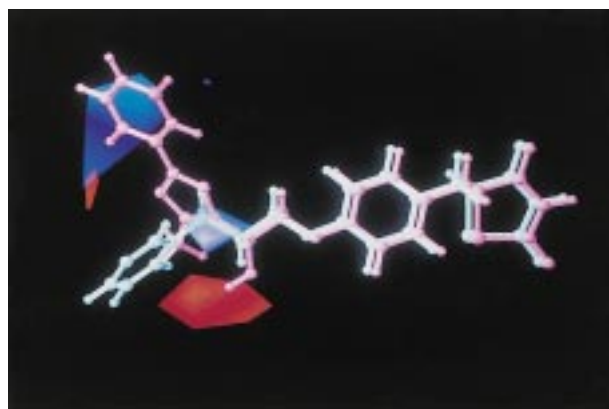


Figure 5. CoMFA electrostatic STDEV*COEFF contour plots from RMS I analysis. Positive charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level of 20%) are represented by red polyhedra. The active compound **45** (magenta) and low active compound **6** (cyan) are shown.

Figure 4 depicts the steric contour plot. The green contours represent regions of high steric tolerance (70% contribution) while yellow contours represent regions of unfavorable steric effect (30% contribution). The steric contour plot shows a large sterically favorable green area surrounded by a sterically unfavorable yellow area. The steric contours were observed only in one part of the molecule.

Figure 5 displays the electrostatic contour plot. The blue contours describe regions where positively charged groups enhance activity (80% contribution), and red contours describe regions where negatively charged groups enhance the activity (20% contribution). The electrostatic contour plot shows blue and red contours in and around the lipophilic region of the molecule.

These contour plots were used to derive a hypothetical receptor model of the receptor.

Discussion

The comparative molecular field analysis was performed on a series of thiazolidinedione antihyperglycemic agents. The biological activity of the ligands is a critical input for QSAR studies. For 3-D field based QSAR methods such as CoMFA, accurate biological data from the enzyme or receptor binding assays is required. We have used the in vivo activity data for the thiazolidinedione antihyperglycemic agents. Some of these compounds were tested on in vitro test system to detect the PPAR- γ agonist activity.⁹ A good correlation was observed between the in vitro activity and the in vivo activity from the animal models. Thus we felt confident of using in vivo activity data for the present series of compounds in the CoMFA studies.

The conformations of the molecules were generated from the systematic search of all the rotatable bonds with a uniform increment. The rotation of the linker chain produces many low energy conformations. The lowest energy conformation of most of the compounds

had an extended conformation in the central linker chain. Compound **2** had more low energy conformers than compound **4**. The lowest energy conformer of compound **2**, which had an extended conformation, was used for CoMFA studies. The conformations of some compounds was influenced by the substituents on the terminal hydrophobic fragment. In low energy conformation of compounds **14** and **15**, the 3-methyl and 'X' substituent were away from each other due to steric clashes. The highly active compound **45** had an extended conformation whereas compound **47**, which had less conformations, was less active. Compound **47** has one less atom in the linker chain because of which the terminal lipophilic fragment may not be in a proper spatial orientation to exhibit good activity. The compounds from the test set had less low energy conformers. Compounds **54** and **60** had less rotatable bonds, hence could not attain a proper conformation required for activity. Compound **64** exhibited conformation similar to the most active compound, **45**.

The alignment of the compounds is one of the critical inputs for the CoMFA studies. The alignments define the putative pharmacophore for the series of ligands. In the present study we have aligned these ligands onto a template structure using two different strategies. These strategies define the alternative ways in which these compounds can be superimposed. These superimpositions were validated using the CoMFA studies.

The PLS analyses on RMS I model was performed with varying parameters. Use of threshold column filtering (σ min) during partial least square analysis is known to improve the signal-to-noise ratio. In this study σ min value was selected based on the cross-validated r^2_{cv} obtained from a repeated leave one out PLS analysis with varying σ min values. As can be seen from the results reported in Table 6 the σ min value of 1.0 kcal/mol gives a significant model. Hence, all further analyses were performed using this value as the threshold column filtering value. The results of the PLS analyses on RMS I alignment using CoMFA fields calculated separately are reported in Table 7. Models with less correlative and predictive properties were obtained when CoMFA fields were used separately in the equation. This suggests the importance of both the fields for activity of the present series of compounds. These parameters were used to generate the CoMFA models. The CoMFA models were validated by predicting the activity of the external test set.

The results from PLS analyses of atom-based and shape-based alignments are reported in Tables 8 and 9, respectively. An internally predictive model was obtained from RMS I alignment. The PLS analyses were also performed using the leave half out cross-validation method. This analysis was performed 100 times, as random formation of the groups will have an effect on the final results. The mean r^2_{cv} value shows a good internal predictivity of the models. Some r^2_{cv} values observed during 100 cross-validation runs were less than those observed from leave one out analysis.

Also there were some r^2_{cv} values higher than those obtained from leave one out analysis. In no case were these values negative. This shows a good internal consistency of the underlying dataset. The RMS I alignment model shows a high external predictivity.

Additional regressor was added to the CoMFA table to study the influence of other factors on the CoMFA results. The inclusion of ClogP in the CoMFA table did not improve the statistical results of the models. This suggests that the activity data used in this study can be a direct measure of free energy of binding to the receptor. However it could also be possible that all other factors may have an equal influence for all the ligands used in this study.

The PLS analyses were performed on the models obtained from different atom-based alignments. The results obtained show that RMS I alignment produces a statistically significant model. These analyses show that all ligands have to be superimposed using the 3 point recognition elements (RMS I). The test set ligands are rigid analogues and contain an extended aromatic linker group (benzoxazole). The superimposition using the RMS I produced a good external prediction but field fit alignment decreased this predictivity whereas other two alignments showed an opposite trend in which field fit alignment produced an improved external predictions. The rigid test ligands place the hydrophobic appendage group in a different spatial disposition when superimposed using RMS II and RMS III alignment rules. The good internal and external predictions with RMS I alignments support our choice of these atoms for superimposition.

The results from PLS analyses with shape-based alignments are reported in Table 9. All the alignments exhibited correlative models with low to high external predictivity. The bootstrapping analyses shows that the models are stable and statistically robust.

The statistical results from shape-based alignments were slightly less predictive than those obtained from atom-based alignments. Also, one particular selection of atoms for superimpositions RMS I produced an externally predictive model. This shows that exact superimposition of the test set ligands is essential for good predictions. In fact slight variations in alignment rules lead to dramatic differences in the external predictions. This could be because the test set ligands were rigid and slight change in their orientation leads to placing of functional groups in unfavorable regions. Flexible fitting of these ligands using multi-fit algorithm did not improve the predictiveness of any alignments. Though present alignment RMS I cannot be considered as a binding modes of these ligands to the receptor. Many factors, which are ignored in this study, such as entropy, solvation and desolvation, etc., can have an effect on the binding modes. The alternative alignment rules used in this study did help us to define the minimum recognition elements required to exhibit the PPAR- γ agonist activity and subsequent antihyperglycemic activity.

The atom-based RMS I alignment was used to analyze the CoMFA contour maps as this model exhibited good internal as well as external predictivity. The CoMFA steric and electrostatic contour maps of RMS I analysis are shown in Figures 4 and 5. The contour plots are to be considered as a representation of the lattice points, where differences in field values are strongly associated with differences in receptor binding affinity. The absence of the lattice points does not indicate that a given substructure element has no influence in the biological activity. It is likely that all the compounds studied exert same steric and/or electrostatic influence in a certain area. Though CoMFA contour maps cannot be used as receptor maps, some useful interpretations can be derived from these contour maps.²⁵

Figure 4 depicts the steric contour plot. The lipophilic fragment of the molecules is surrounded by the steric contours. This sterically favorable green contour is embedded in the unfavorable yellow region. The most active compound **45** has a phenyl ring embedded in this green region and the less active compound **6** has a pyridine ring overlapping on the unfavorable steric yellow region. The sterically favored contours are embedded in the surrounding yellow contours, suggesting there is a definite requirement of a substructure with appropriate shape to exhibit high activity. The substructure fragments with less or very high steric bulk in this area reduces the activity. The steric contours in this region of the molecule underlay the importance of steric interactions of the ligands with the receptor.

Figure 5 displays the electrostatic contour plot. A large blue contour was found to surround the terminal phenyl ring of compound **45**, suggesting that substructures with high electron density in this region reduce the activity. Another small blue contour was found to surround the substructure in the lipophilic region. A red contour at the bottom of the lipophilic region suggests that high electron density in this region increases the activity. The hydroxyl group of compound **45** is in the vicinity of this red contour. The blue contour in the lipophilic site overlaps the green contour in the steric plots, suggesting the substituents in this region should have bulky substituents.

A hypothetical receptor model can be proposed for the binding of these ligands based on the present CoMFA study (see Fig. 6). The thiazolidinedione antihyperglycemic agents bind with the PPAR- γ receptor using

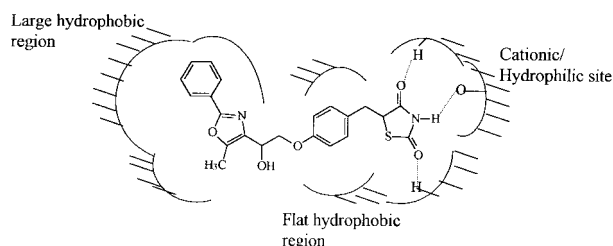


Figure 6. Proposed hypothetical receptor model of PPAR- γ binding site.

mainly a three-point recognition element, 'cationic site or hydrophilic site', 'flat aromatic region' and 'O' atom anchor'. There is also a large variable hydrophobic region, 'hydrophobic binding site' essential for activity. In the cationic/hydrophilic site, thiazolidinediones may form hydrogen bonds with the active site of the receptor. A large number of variations to this ring were reported but no correlation was observed with the acid strength and activity. Therefore, structural factors override the acid strength in this region. In the flat linker region the aromatic ring could be having a π - π stacking interactions with the receptor. Absolute planarity in this region is essential as slight variation in planarity as observed in some of our CoMFA alignments reduced the significance of the models. This suggests that a flat aromatic ring(s) is favorable in this region. Another important recognition element is 'O' atom of ether linkage. This could be forming a hydrogen bond with the receptor. Also, exact superimpositions of this atom help to place the terminal lipophilic fragment of the molecule in a proper spatial orientation. In hydrophobic region there are large steric interactions with the receptor. This region can accommodate various substructure fragments. Though this region is very large there is a limiting size and shape for the substituent that would be effective for tight binding. The substituents in this region should be such that they not only modulate the binding with the receptor but also influence the factors involved in non-receptor events such as pharmacokinetic and toxicity profiles. This is an important region of the molecule where wide choice of substituents can be made to design better antihyperglycemic agents.

Recently, when this manuscript was under preparation, Nolte et al. reported the X-ray crystallographic structures of the peroxisome proliferator-activated receptor γ in complex with an antidiabetic agent rosiglitazone (BRL49653).²⁶ In this crystal structure the thiazolidinedione ring in the cationic site forms hydrogen bonds with the receptor. The central benzene ring binds in a narrow pocket. The bridging oxygen atom between the central benzene ring and lipophilic substituent provides a vital geometry for the lipophilic substituent that occupies a large hydrophobic pocket. The ligand binding site from this study is consistent with the proposed receptor model generated from the present CoMFA study.

Conclusions

The CoMFA method was used to derive a 3-D-QSAR model of thiazolidinedione antihyperglycemic agents. The derived CoMFA models were evaluated by external predictions of a set of compounds, which were not represented in the training set. Various alignment rules were derived to superimpose these ligands. All the alignments show a high internal predictivity whereas only one alignment rule RMS I shows good external predictivity. This suggests that the PPAR- γ receptor has rigid requirements in the acidic and central aromatic binding region. Poor external predictions of some alignments supports this conclusion as these alignments

placed the acidic and aromatic linker fragments in the unfavorable region.

This CoMFA study also helped us to define a hypothetical receptor model of PPAR- γ . The receptor model contains an acidic binding pocket, where hydrogen bonding interactions may be involved. There is a central aromatic binding region. In this region hydrophobic and π - π stacking interactions may be involved in the ligand binding. Also this region has a limited space and hence large substructure fragments in the central aromatic region leads to reduced activity. A large hydrophobic region is present at the terminal end of the receptor. This region can accommodate different structural fragments. In the hydrophobic region steric interactions are more important than the hydrophobic interactions. The proposed receptor model is consistent with the recently published data on the X-ray crystal structure of PPAR- γ . The present study helped us to understand the structure–activity relationship of thiazolidinedione antihyperglycemic agents. This hypothetical receptor model and the CoMFA contour maps can be utilized for the design of more potent antihyperglycemic agents.

Experimental

Biological data


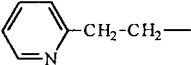
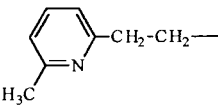
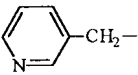
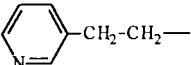
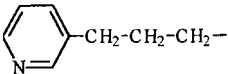
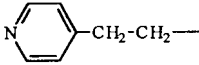
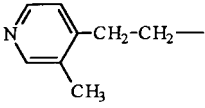
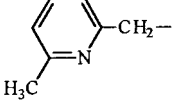
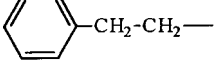
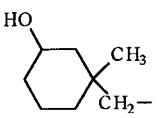
The compounds chosen for the present study were obtained from the literature.^{16,18,21} The structures and biological activity values of the compounds forming the training set are shown in Tables 1–3. The hypoglycemic activities of the compounds were tested using genetically obese and diabetic yellow KK mice. The response of the animals was measured as the reduction in the blood glucose level from the control animals. The biological activity was expressed as pED₂₅, which is $-\log$ of effective molar dose required to reduce blood glucose by 25%. Table 4 shows the structures and biological activity values of 15 thiazolidinediones containing benzoxazole moieties. These compounds were tested under similar conditions and their ED₂₅ values were reported recently. All the 15 compounds have a rigid benzoxazole linker that would restrict the conformational mobility of the lipophilic part of the molecules. Since these compounds have different substructure fragments, which were not present in the training set, we have used these compounds to assess the predictive ability of the CoMFA models.

Molecular modeling

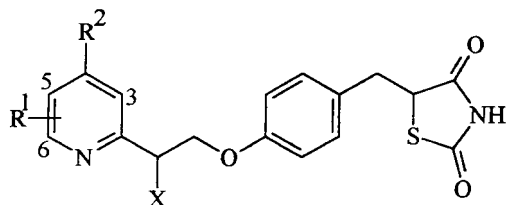
The geometries of the compounds were modeled with the standard bond lengths and angles using the molecular modeling package SYBYL 6.22²⁷ with standard TRIPOS forcefield.²⁸ The CoMFA/QSAR studies were performed using SYBYL running on a Silicon Graphics Indigo² computer.

The conformational analysis was performed using the systematic search routine. All rotatable bonds were

Table 1. Structures and biological activities of training set molecules¹⁸

Compd no.	R	Biological activity ^a
1		0.92
2		1.83
3		1.93
4		0.80
5		1.52
6		0.84
7		0.82
8		0.84
9		1.04
10		1.04
11		1.70

^aExpressed as $\log[1/ED_{25}$ (mmol/kg/d)].

Table 2. Structures and biological activities of training set molecules¹⁸

Compd no.	R ¹	R ²	X	Biological activity ^a
12	H	H	OH	2.06
13	3-Me	H	H	1.23
14	3-Me	H	OH	2.12
15	3-Me	H	CH ₂ OH	1.57
16	H	CH ₃	H	1.23
17	5-Me	H	H	1.23
18	5-Et	H	OH	1.69
19	5-Et	H	H	1.77
20	5-Et	H	OH	1.84
21	5-Et	H	CH ₂ OH	1.89
22	6-Me	H	OH	2.21
23	6-Me	Me	H	1.93
24	6-Me	Me	OH	2.27
25	6-CH ₂ OH	H	H	0.86
26	6-Me	OH	H	0.86

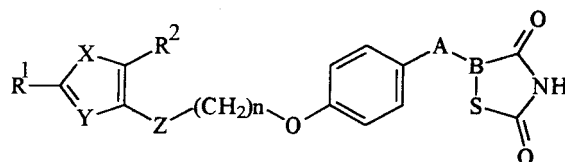
^a Expressed as log[1/ED₂₅ (mmol/kg/d)].

searched in 15° increments from 0° to 360°. Some conformers were rejected during the search due to non-bonded steric interactions. The van der Waals radii scaling factors were 0.95, 0.87 and 0.65 for general, 1–4 and H-bond interactions, respectively. Conformational energies were computed with electrostatic term. Low energy conformers were selected and used for superimpositions.

Alignment rules

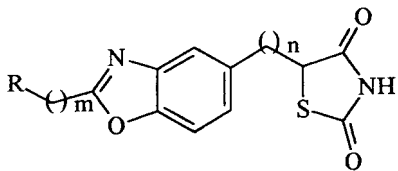
The alignment, i.e. molecular conformation and orientation, is one of the sensitive inputs for CoMFA. The conformation of the compounds identified by systematic search was used for superimposition. We have devised two different approaches to superimpose the molecules. One of the active molecules, compound **45**, served as a template structure to derive all alignments.

1. Atom-based alignments: In these alignments atoms of the molecules were used for RMS-fitting onto the corresponding atoms of the template structure.
2. Shape-based alignments: In these alignments centroids rather than the exact superimposition of the atoms of the rings were used for RMS-fitting to the template structure.

Table 3. Structures and biological activities of training set molecules²¹

Compd no.	R ¹	R ²	X	Y	Z	n	A-B	Biological activity ^a
27	Me	H	S	N	CH ₂	1	—	1.94
28	Cyclohexyl	H	S	N	CH ₂	1	—	2.72
29	Ph	H	S	N	CH ₂	1	—	2.91
30	Me	H	O	N	CH ₂	1	—	1.84
31	n-Pr	H	O	N	CH ₂	1	—	2.59
32	Me	Me	O	N	CH ₂	1	—	2.25
33	Me	Et	O	N	CH ₂	1	—	2.15
34	Cyclohexyl	H	O	N	CH ₂	1	—	2.87
35	Ph	H	O	N	CH ₂	1	—	3.10
36	Ph	Me	O	N	CH ₂	1	—	3.62
37	Ph	Et	O	N	CH ₂	1	—	3.42
38	Cyclohexyl	Me	O	N	CH ₂	1	—	3.29
39	2-furyl	Me	O	N	CH ₂	1	—	2.99
40	2-thienyl	Me	O	N	CH ₂	1	—	3.37
41	4-(MeO)-C ₆ H ₄ -	Me	O	N	CH ₂	1	—	3.41
42	3,4-(MeO) ₂ -C ₆ H ₃ -	Me	O	N	CH ₂	1	—	3.33
43	3-(Me)-C ₆ H ₄ -	Me	O	N	CH ₂	1	—	3.80
44	Me	Me	O	N	CH(OH)	1	—	2.64
45	Ph	Me	O	N	CH(OH)	1	—	4.02
46	Ph	H	S	N	CH ₂	0	—	1.83
47	Ph	H	O	N	CH ₂	0	—	1.79
48	Ph	Me	O	N	CH ₂	0	—	1.82
49	Ph	H	S	N	CH ₂	1	=	2.27
50	Ph	Me	O	N	CH ₂	1	=	2.61
51	Ph	Me	O	N	CH(OH)	1	=	2.79
52	Me	Me	O	N	CH(OH)	1	=	2.04
53	H	Me	N	S	CH ₂	1	—	1.18

^a Expressed as log[1/ED₂₅ (mmol/kg/d)].

Table 4. Structures and biological activities of test set molecules¹⁶


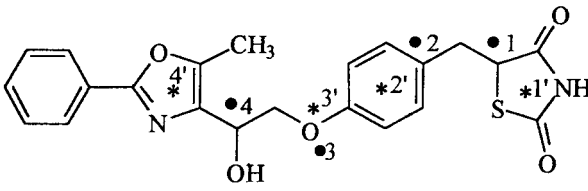
Compd no.	R	m	n	Biological activity ^a
54	Ph	0	1	0.88
55	Ph	1	1	2.28
56	Ph	2	1	1.10
57	4-(PhO)-C ₆ H ₄ -	1	1	2.21
58	4-(1-pyrrolidyl)-C ₆ H ₄ -	1	1	1.80
59	4-(2-morpholyl)-C ₆ H ₄ -	1	1	1.96
60	4-Cl-C ₆ H ₄ -	1	0	1.13
61	2-naphthyl	1	1	3.06
62	6-Cl-2-naphthyl	1	1	2.58
63	2-naphthyl	1	0	0.94
64	5-Me-2-Ph-4-oxazolyl	1	1	3.83
65	2-cyclohexyl-4-thiazolyl	1	1	2.80
66	2-(Me-S)-4-thiazolyl	1	1	2.66
67	2-thienyl	1	1	1.40
68	N-(cyclohexyl)-M-Me-amino	1	1	1.32

^a Expressed as log[1/ED₂₅ (mmol/kg/d)].

There are three different ways in which atoms or centroids can be aligned. The atoms/centroids used for alignments is defined in Table 5. These six alignments were subsequently realigned using SYBYL QSAR rigid body field fit. The field fit alignment uses a simplex algorithm in SYBYL to minimize the differences in steric and electrostatic fields averaged over all the lattice grid points to find the best fit. The test set ligands were superimposed similarly. Overall present study involved 12 different alignments. These alignments were subsequently used for CoMFA probe interaction energy calculations.

CoMFA interaction energy

The steric and electrostatic interaction energies were calculated at each lattice intersections of a regularly

Table 5. Atoms and centroids used for superimpositions for aligning the molecules


Alignment rule	Atoms/centroids
(a) Atom-based	
I	●1, ●2, ●3
II	●1, ●2, ●4
III	●1, ●2, ●3, ●4
(b) Shape-based	
IV	*1', *2', *3'
V	*1', *2', *4'
VI	*1', *2', *3', *4'

Table 6. Effect of column filtering 'σ' (minimum sigma) on CoMFA results

	σ kcal/mol				
	0.1	0.2	0.5	1.0	2.0
r ² _{cv} ^a	0.737	0.740	0.743	0.745	0.724
SEP	0.472	0.469	0.467	0.465	0.484
r ² _{ncv}	0.904	0.907	0.909	0.917	0.904
SEE	0.286	0.281	0.278	0.265	0.285
F value	153.363	159.049	163.481	181.177	153.608

^a Results from leave one out (LOO) cross-validation analysis using three components.**Table 7.** Summary of CoMFA results when steric and electrostatic fields were used separately^a

CoMFA fields	r ² _{cv}	Opt. comp.	SEP	r ² _{ncv}	SEE	F value
Steric	0.702	2	0.497	0.811	0.397	107.028
Electrostatic	0.584	3	0.594	0.823	0.387	76.187
Steric and electrostatic	0.745	3	0.465	0.917	0.265	181.177

^a All results from leave one out (LOO) cross-validation analysis.

spaced grid box. The lattice spacing was set to a value of 2.0 Å. The CoMFA region was defined automatically, which extends the lattice walls beyond the dimensions of each structure by 4.0 Å in all directions. The steric and electrostatic probe ligand interactions represented by Lennard–Jones and Coulomb potential function of the Tripos forcefield were calculated with a C sp³ probe atom carrying a +1 charge. A distance dependent dielectric expression $\epsilon = \epsilon_0 R_{ij}$ with $\epsilon_0 = 1.0$ was used during the calculations. The steric and electrostatic energies were truncated to ± 30.0 kcal/mol. The electrostatic contributions were ignored at lattice intersection with maximal steric interactions.

Partial least squares (PLS) analysis

The partial least squares (PLS) algorithm²⁹ was used in conjugation with leave one out (LOO) cross-validation. In this analysis one compound was dropped in turn and a model was generated from the remaining compound. This model was then used to predict the activity of the dropped compound. This procedure was repeated until all the compounds were predicted. This PLS analysis gave the optimum number of components that was used to generate the final models without cross-validation. The result from a cross-validation analysis was expressed as r²_{cv} value, which is defined as

$$r_{cv}^2 = 1 - \frac{\text{PRESS}}{\Sigma(Y - Y_{\text{mean}})^2}$$

where PRESS = $\Sigma(Y - Y_{\text{pred}})^2$

The r²_{cv} can take up values in the range from 1, suggesting a perfect model, to less than 0 where errors of prediction are greater than the error from assigning each

Table 8. CoMFA results for atom-based alignments

	I		II		III	
	RMS I	FF I	RMS II	FF II	RMS III	FF III
r^2_{cv}	0.745	0.634	0.734	0.628	0.764	0.624
Components	3	1	3	1	3	1
SEP	0.465	0.546	0.475	0.550	0.447	0.554
r^2_{ncv}	0.917	0.697	0.905	0.693	0.919	0.689
SEE	0.265	0.497	0.283	0.500	0.262	0.504
F value	181.177	117.374	156.073	115.176	185.735	112.835
P value	0.00	0.00	0.00	0.00	0.00	0.00
Contrib. steric	59.7	69.2	65.0	70.0	60.3	70.3
Elect.	40.3	30.8	35.0	30.0	39.7	29.7
r^2_{pred}	0.751	0.504	−0.071	0.384	0.124	0.438
$r^2_{bs}^a$	0.929	0.717	0.934	0.728	0.934	0.713
Standard deviation ^a	0.019	0.052	0.016	0.044	0.017	0.046

^a From 100 bootstrapping runs.**Table 9.** CoMFA results for shape-based alignments

	IV		V		VI	
	RMS IV	FF IV	RMS V	FF V	RMS VI	FF VI
r^2_{cv}	0.759	0.660	0.707	0.723	0.692	0.679
Components	3	3	3	3	3	1
SEP	0.452	0.537	0.499	0.485	0.511	0.511
r^2_{ncv}	0.917	0.908	0.880	0.921	0.877	0.730
SEE	0.266	0.279	0.319	0.259	0.323	0.469
F value	179.712	161.647	119.451	190.087	116.195	138.077
P value	0.00	0.00	0.00	0.00	0.00	0.00
Contrib. steric	61.1	59.1	65.1	58.2	63.2	63.9
Elect.	38.9	40.9	34.9	41.8	36.8	36.1
r^2_{pred}	0.625	0.271	0.253	−0.198	0.362	0.427
$r^2_{bs}^a$	0.936	0.926	0.906	0.919	0.910	0.753
Standard deviation ^a	0.016	0.021	0.023	0.020	0.025	0.044

^a From 100 bootstrapping runs.**Table 10.** Cross-validated r^2 value from leave half out (LHO) analysis

	Atom-based			Shape-based		
	I	II	III	IV	V	VI
	RMS I/FF I	RMS II/FF II	RMS III/FF III	RMS IV/FF IV	RMS V/FF V	RMS VI/FF VI
Mean ^a	0.674	0.671	0.690	0.683	0.670	0.649
	0.625	0.623	0.611	0.618	0.675	0.679
Standard deviation	0.058	0.054	0.057	0.057	0.042	0.035
	0.049	0.040	0.043	0.041	0.049	0.054
High	0.785	0.789	0.815	0.802	0.755	0.630
	0.726	0.721	0.716	0.727	0.761	0.772
Low	0.516	0.516	0.538	0.514	0.533	0.460
	0.429	0.473	0.473	0.501	0.536	0.449

^a From 100 leave half out runs.

compound mean activity of the model. The optimum number of components was taken as the number required to increase r^2_{cv} by ~5% from the model with one fewer component rather than the default SYBYL estimate which is that which gave the highest r^2_{cv} value. Equal weights were assigned to the steric and electrostatic descriptors using CoMFA_STD scaling option. All cross-validated PLS analyses were performed with a minimum σ (column filter) value of 1.0 kcal/mol which

minimized the influence of column noise and reduced the computation time. The cross-validation analysis was also performed by setting the number of cross-validation groups to 2 (leave half out (LHO)). In this case cross-validation groups were randomly selected and a model is derived. This was then used to predict the activity of the compounds from the other group. This analysis was repeated 100 times and the mean and standard deviation of r^2_{cv} values are reported. To obtain the statistical

confidence limits on the analysis, PLS analysis using 100 bootstrap groups³⁰ with optimum number of components was performed.

Predictive r^2 values

The predictive ability of each analysis was determined from a set of 15 compounds that were not used in the training set. The activities of these compounds were predicted from each PLS analysis. The predictive r^2 (r^2_{pred}) will be based on molecules of the test set only and is defined as:

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

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 sum of the squared deviation between predicted and actual activity values from every molecule in the test set.

Acknowledgements

The authors thank the University Grants Commission (UGC), New Delhi for the financial assistance through its COSIST programme. L.K.G. thanks the UGC for the award of a senior research fellowship. S.S.K. thanks Dr. Tanaji and Mr. Vijay for useful discussions.

References

- DeFronzo, R. A.; Bonadonna, R. C.; Ferrannini, E. *Diabetes Care*, **1992**, *15*, 318.
- Reaven, G. M. *Metabolism*, **1980**, *29*, 445.
- DeFronzo, R. A.; Ferrannini, E.; Koivisto, V. P. *Am. J. Med.*, **1983**, *74*(Suppl. 1A), 52.
- Olefsky, J. M.; Kolterman, O. G. *Am. J. Med.*, **1981**, *70*, 151.
- Reaven, G. *Diabetes*, **1988**, *37*, 1595.
- Sohda, T.; Mizuno, K.; Imamiya, E.; Sugiyama, Y.; Fujita, T.; Kawamatsu, Y. *Chem. Pharm. Bull.*, **1982**, *30*, 3580.
- Fujita, T.; Sugiyama, Y.; Taketomi, S.; Sohda, T.; Kawamatsu, Y.; Iwatsuka, H.; Suzuoki, Z. *Diabetes*, **1983**, *32*, 804.
- Lehmann, J. M.; Moore, L. B.; Smith-Oliver, T. A.; Wilkinson, W. O.; Willson, T. M.; Kliewer, S. A. *J. Biol. Chem.*, **1995**, *270*, 12953.
- Willson, T. M.; Cobb, J. E.; Cowan, D. J.; Wiethe, R. W.; Correa, I. D.; Prakash, S. R.; Beck, K. D.; Moore, L. B.; Kliewer, S. A.; Lehmann, J. M. *J. Med. Chem.*, **1996**, *39*, 665.
- Dow, R. L.; Bechle, B. M.; Chou, T. T.; Clark, D. A.; Hulin, B.; Stevenson, R. W. *J. Med. Chem.*, **1991**, *34*, 1538.
- Kees, K. L.; Cheeseman, R. S.; Prozialeck, D. H.; Steiner, K. E. *J. Med. Chem.*, **1989**, *32*, 11.
- Ellingboe, J. W.; Alessi, T. R.; Dolak, J. M.; Nguyen, T. T.; Tomer, J. D.; Guzzo, F.; Bagli, J. F.; McCaleb, M. L. *J. Med. Chem.*, **1992**, *35*, 1176.
- Hulin, B.; Newton, L. S.; Lewis, D. M.; Genereux, P. E.; Gibbs, E. M.; Clark, D. A. *J. Med. Chem.*, **1996**, *39*, 3897.
- Hulin, B.; Clark, D. A.; Goldstein, S. W.; McDermott, R. E.; Dambek, P. J.; Kappeler, W. H.; Lamphere, C. H.; Lewis, D. M.; Rizzi, J. D. *J. Med. Chem.*, **1992**, *35*, 1853.
- Zask, A.; Jirkovsky, I.; Nowicki, J. W.; McCaleb, M. L. *J. Med. Chem.*, **1990**, *33*, 1418.
- Arakawa, K.; Inamasu, M.; Matsumoto, M.; Okumura, K.; Yasuda, K.; Akatsuka, H.; Kawanami, S.; Watanabe, A.; Homma, K.; Saiga, Y.; Ozeki, M.; Iijima, I. *Chem. Pharm. Bull.*, **1997**, *45*, 1984.
- Sohda, T.; Mizuno, K.; Tawada, H.; Sugiyama, Y.; Fujita, T.; Kawamatsu, Y.; *Chem. Pharm. Bull.*, **1982**, *30*, 3563.
- Sohda, T.; Momose, Y.; Meguro, K.; Kawamatsu, Y.; Sugiyama, Y.; Ikeda, H. *Arzeim-Forsch/Drug Des.*, **1990**, *40*, 37.
- Sohda, T.; Mizuna, K.; Imamiya, E.; Sugiyama, Y.; Fujita, T.; Kawamatsu, T. *Chem. Pharm. Bull.*, **1982**, *30*, 3580.
- Cantello, B. C. C.; Cawthorne, M. A.; Cottam, G. P.; Duff, P. T.; Haigh, D.; Hindley, R. M.; Lister, C. A.; Smith, S. A.; Thurlby, P. L. *J. Med. Chem.*, **1994**, *37*, 3977.
- Sohda, T.; Mizuno, K.; Momose, Y.; Ikeda, H.; Fujita, T.; Meguro, K. *J. Med. Chem.*, **1992**, *35*, 2617.
- Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.*, **1988**, *110*, 5959.
- Hariprasad, V.; Kulkarni, V. M. *J. Comput. -Aided Mol. Design*, **1996**, *10*, 284.
- Kulkarni, S. S.; Kulkarni, V. M. *J. Med. Chem.*, **1999**, *42*, 373.
- Gokhale, V. M.; Kulkarni, V. M., *J. Med. Chem.* Communicated.
- Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V. *Nature*, **1998**, *395*, 137.
- Tripos Associates, 1699, S. Hanley Road, Suite 303, St. Louis, Missouri, MO 63144-2913, USA.
- Clark, M.; Cramer, R. D. III; Van Opdenbosh, N. *J. Comput. Chem.*, **1989**, *10*, 982.
- Wold, S.; Ruhe, A.; Wold, H.; Dunn W. J. III; *SIAM J. Sci. Stat. Comput.*, **1984**, *5*, 735.
- Cramer, R. D. III; Bunce, J. D.; Patterson, D. E. *Quant. Struct. Act. Relat.*, **1988**, *7*, 18.