PharmTech

A Structure-Based 3D-QSAR and Docking Studies on a Series of Indolealkanoic Acid Derivatives as Diabetes Mellitus Inhibitors

Rohith Kumar. A*, Shravan Kumar. G, Mahmood SK

Bioinformatics Division, Osmania University, Hyderabad-500007, India

Corres.author : rohit.casper@gmail.com,
Mobile No. +91-9985114009

Abstract: Comparative molecular field analysis and comparative molecular similarity indices analysis (CoMSIA) based on three dimensional quantitative structure-activity relationship (3D-QSAR) studies were conducted on a series (28 compounds) of indolealkanoic acid derivatives as potent diabetes mellitus inhibitors. The best prediction was obtained with a CoMFA standard model ($q^2= 0.850$, $r^2= 0.983$) and with CoMSIA combined steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor fields ($q^2= 0.856$, $r^2= 0.977$). CoMFA and CoMSIA contour maps were then used to analyze the structural features of ligands to account for the activity in terms of positively contributing physiochemical properties such as steric, electrostatic, hydrophobic and hydrogen bond donor fields. The resulting contour maps produced by the best CoMFA and CoMSIA models were used to identify the structural features relevant to the biological activity in this series of analogs. The information obtained from CoMFA and CoMSIA 3-D contour maps can be used for the design of indolealkanoic acid derivatives as potent inhibitors of diabetes mellitus. The binding mode of the high active compound at the active site of Novel Benzothiazepine Inhibitor in Complex with human Aldose Reductase (PDB id: 3P2V) was explored using FlexX docking program and hydrogen-bonding interactions were observed between the inhibitor and the target.
Keywords: CoMFA; CoMSIA; QSAR; Diabetes mellitus; FlexX.

Introduction

Diabetes mellitus often referred to simply as diabetes; it is a syndrome characterized by disordered metabolism and abnormally high blood sugar (hyperglycemia) resulting from low levels of the hormone insulin with or without abnormal resistance to insulin's effects[1]. The characteristic symptoms are excessive urine production (polyurea), excessive thirst and increased fluid intake (polydipsia), blurred vision, unexplained weight loss and lethargy. Chronic hyperglycemia and other metabolic disturbances of DM lead to long-term tissue and organ damage as well as dysfunction involving the eyes, kidneys, and nervous and vascular systems[2]. These symptoms are likely to be absent if the blood sugar is only mildly elevated. Diabetes is a disease in which the body does not produce or properly use insulin. Insulin is a hormone that is needed to convert sugar, starches and other food into energy needed for daily life[3].

It is a chronic disease with long-term macrovascular and microvascular complications, including diabetic nephropathy, neuropathy, and retinopathy. It is a leading cause of death, disability, and blindness in the United States for person's 20−74 years of age. Approximately 80 percent of blindness in this age group is related to diabetic retinopathy (DR) [4,5]. At least 50,000 Americans are legally blind from this condition. Diabetes is also responsible for 5,800, or 10 percent, of the new cases of blindness reported annually[6] . In patients with diabetes mellitus and hereditary spherocytosis (or any condition that results

in a shortened erythrocyte life span), difficultiers can arise with interpretation of hemoglobin A_{1C} (HbA_{1C})[7]. Quantitative structure–activity relationship (QSAR) enables the investigators to establish a reliable quantitative structure–activity and structure–property relationships to derive an in silico QSAR models to predict the activity of novel molecules prior to their synthesis. The overall process of QSAR model development can be divided into three stages namely; the data preparation, data analysis, and model validation, representing a standard practice of any QSAR modeling[8]. 3D-QSAR methodologies have been successfully used to generate models for various chemotherapeutic agents. We have carried out 3D-QSAR studies employing comparative molecular field analysis (CoMFA)[9] and comparative molecular similarity indices analysis (CoMSIA)[10] techniques in order to study and deduce a correlation between structure and biological activity of indolealkanoic acid derivatives and related compounds.

Methodology
Data set
In the present study, the vitro inhibitory activity data (IC_{50}) of the 28 indolealkanoic acid derivatives as potent inhibitors of diabetes mellitus was taken from the literature. The structures of the compounds and their biological data are given in Table 1 (see supplementary material). The IC_{50} values were converted to the corresponding pIC_{50} ($-logIC_{50}$) and used as dependent variables in CoMFA and CoMSIA analysis. The pIC_{50} values span a range of 3-log units providing a broad and homogenous data set for 3D-QSAR study. The 3D-QSAR models were generated using a training set of 20 molecules. Predictive power of the resulting models was evaluated using a test set of 8 molecules. The test set compounds were selected randomly such that a wide range of activity in the data set was included.

Molecular Modeling and Alignment
CoMFA and CoMSIA results may be extremely sensitive to a number of factors such as alignment rules, over all orientation of the aligned compounds, lattice shifting step size and probe atom type. The accuracy of the prediction of CoMFA and CoMSIA models and the reliability of the contour models depend strongly on the structural alignment of the molecules and thus we applied molecular alignment to align all the molecules used in present study in space. The molecular alignment was achieved by the SYBYL 6.7[11]. The initial structures were minimized at Tripos force field[12] with Gasteiger-Hückel charges[13] using conjugate gradient method, and convergence criterion was 0.005 kcal/mol. The most active compound (compound 28) was used as an alignment template and the rest of the molecules were aligned to it by using the common substructure as shown in Fig. 1.
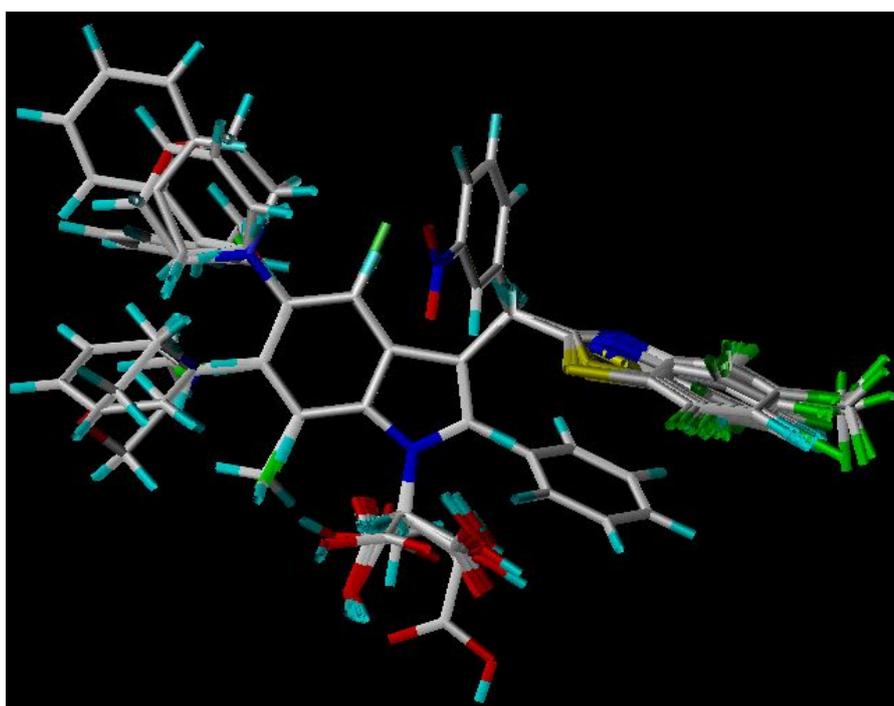

Figure 1.Superposition of 28 indolealkanoic acid molecules with common scaffold.

CoMFA and CoMSIA studies

For the CoMFA calculations, steric and electrostatic field energies were calculated using sp³ carbon as the steric probe atom and a +1 net charge as the electrostatic probe. The cutoff was set to 30kcal/mol. The minimum σ (column filtering) was set to 2.0kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold[14]. Regression analysis was performed using the cross validation of compounds leave-one-out method. CoMSIA calculates similarity indices at the intersections of a surrounding lattice. The charge, probe, and grid spacing used to construct the CoMFA best model were also used for the CoMSIA investigation. Five physicochemical properties of steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields were calculated. The lattice dimensions were selected with a sufficiently large margin (4Å) to enclose all aligned molecules[15]. The statistical evaluation for the CoMSIA analysis was carried out in the same way as described in CoMFA.

Partial Least Squares (PLS)

To quantify the relationship between the structural parameters (CoMFA and CoMSIA interaction energies) and the biological activities, the PLS algorithm[16] was used. The CoMFA descriptors were used as independent variables, and pID_{50} values as dependant variables in partial least square regression analysis. Cross-validation partial least square method of leave-one-out (LOO) was performed to obtain the optimal number of components used in the subsequent analysis. The minimum sigma (column filtering) was set to 2.0kcal/mol to improve the signal- to-noise ratio. The optimum number of principle components in the final non-cross-validated QSAR equations was determined to be that leading to the highest correlation coefficient (r^2) and the lowest standard error in the LOO cross validated predictions. The non-cross-validation was used in the analysis of CoMFA result and the prediction of the model. The same method was used for CoMSIA too, thereafter a full PLS was run using column filtering of 1.0 kcal/mol. Auto scaling was applied to all CoMSIA analysis[17].

Molecular Docking

Docking studies were carried out using the FlexX program[18], interfaced with SYBYL 6.7. In this automated docking program, the flexibility of the ligands is considered while the protein or biomolecules is considered as a rigid structure. The ligand is built in an incremental fashion, where each new fragment is added in all possible positions and conformations to a pre-placed base fragment inside the active site. All the molecules for docking were sketched in the SYBYL and minimized using Gasteiger-Hückel charges using conjugate gradient method and all the charges were removed[19]. For our studies, X–ray crystal structure of Novel Benzothiazepine Inhibitor in Complex with human Aldose Reductase was obtained from the protein data bank (PDB id: 3P2V) having resolution of 1.69 Å. Solvent molecules were deleted and bond order for crystal ligand and protein were adjusted. Formal charges were assigned to all the molecules and FlexX run was submitted.

Results and Discussions

3D QSAR Studies

The CoMFA method was used for deriving 3D-QSAR model for 28 indolealkanoic acid derivatives, which are reported as diabetes mellitus inhibitors. The leave-one-out partial least-squares (PLS) analysis of the obtained CoMFA model yielded high cross-validated q^2-value of 0.850 (five components) and non-cross-validated correlation coefficient r^2 of 0.983. The steric and electrostatic contributions are 64.1% and 35.9%. The CoMSIA study revealed $q^2 = 0.856$ (six components), and non-cross-validated correlation coefficient r^2 of 0. 977. The steric field descriptor explains 17.4 % of the variance and, the electrostatic descriptor explains 43.0 %, the hydrophobic field explains 28.4% while the hydrogen bond donor explains 11.4 % of the variance. Table 1 lists experimental activities, predicted activities and residual values of the training set and test set by CoMFA and CoMSIA models respectively. These correlation coefficients suggest that our model is reliable and accurate. Fig. 2(a) shows correlation between the experimental and predicted pIC_{50} values of training set and test set by the CoMFA and CoMSIA model.
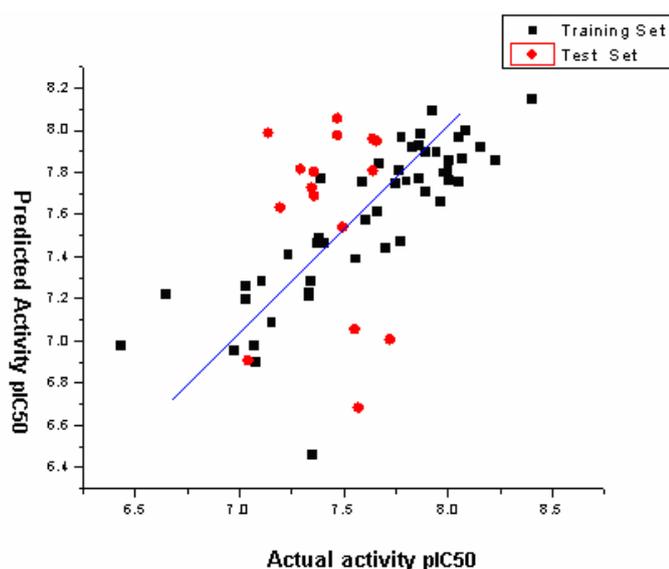
contribution, 80% for favored region and 20% for disfavored region was set during contour analysis.

The CoMFA steric and electrostatic fields from the final non-cross-validated analysis were plotted as 3-D colored contour maps. The field energies at each lattice point were calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table (SD* coeff), always plotted as the percentages of the contributions of CoMFA equation. These maps show regions where differences in molecular fields are associated with differences in biological activity. The CoMFA contours for steric and electrostatic fields are shown in Fig. 3, while those of CoMSIA steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor are shown in Fig. 4, respectively. In the contour maps, each colored contour represents particular properties such as green contours for regions of high steric tolerance (80% contribution), yellow for low steric tolerance (20% contribution). Red contours for regions of decreased electrostatic tolerance for positive charge (20% contribution), blue for regions for decreased electrostatic tolerance for negative charge (80% contribution), yellow contours represent hydrophobically favored regions (80% contribution) and white contours for hydrophobically disfavored regions (20% contribution). The magenta and red contours denote favorable and unfavorable regions for hydrogen bond acceptor, respectively whereas cyan and purple contours represents favorable and unfavorable regions for hydrogen bond donor groups, respectively.
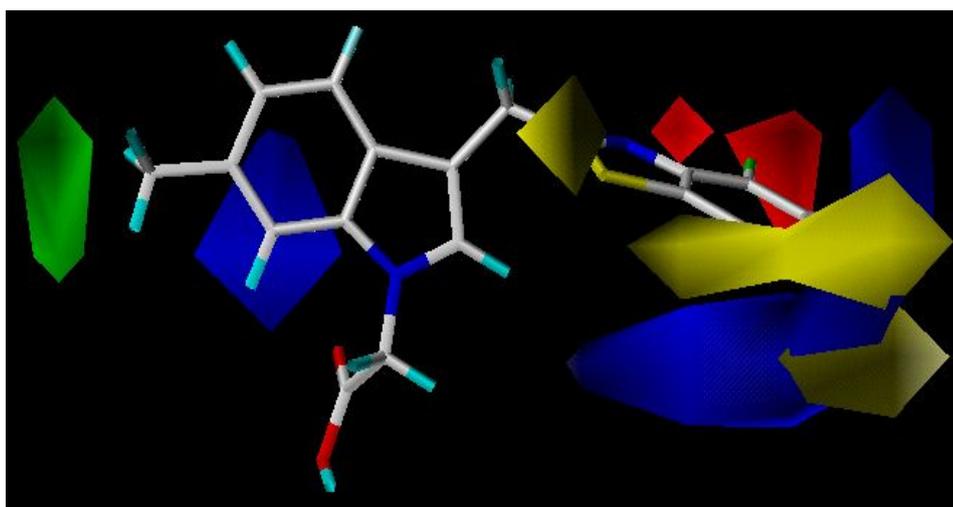

Figure 2.Plot of predicted versus actual pIC50 values of Training set and Test set molecules for CoMFA and CoMSIA model

Contour analysis

The visualization of the results of the CoMFA and CoMSIA models have been performed using the StDev*Coeff mapping option contoured by contribution. The default level of contour with

Figure 3. Compound 28 (most active) mapped on CoMFA showing steric contour and electrostatic contour maps.
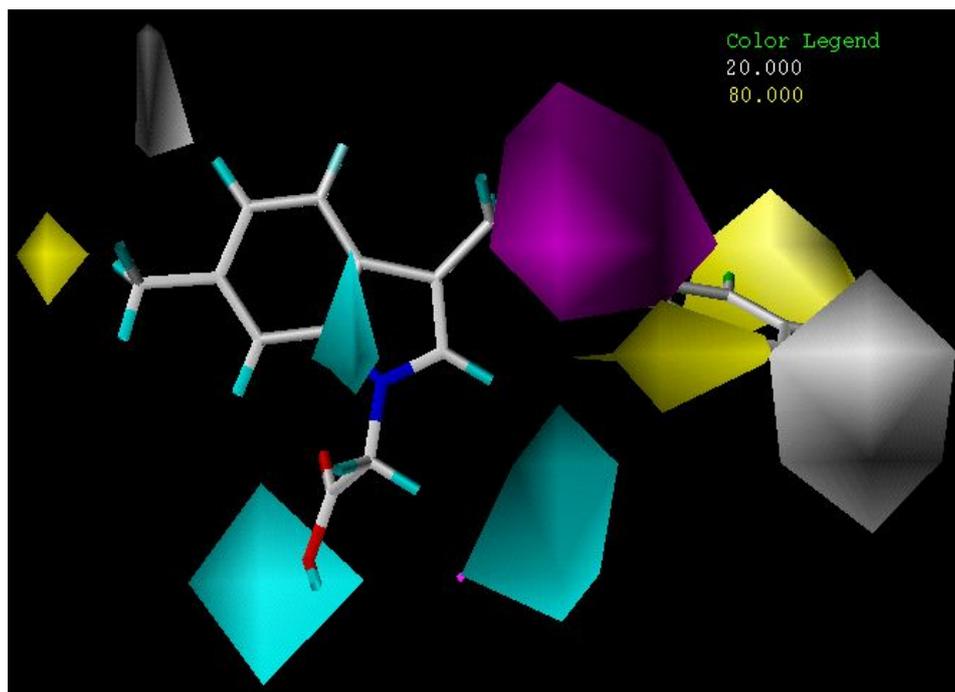
Figure 4.Compound 28 (most active) mapped on CoMSIA showing hydrophobic, hydrogen-bond donor and hydrogen-acceptor contour maps

The CoMFA contour maps are shown in Fig. 3, the green steric contours present in the CoMFA maps, indicates the area in which steric bulk might have a positive effect on activity while the yellow region is unfavorable for bulky groups. Green contours are mainly present near the indole ring, which indicates the substitution of bulky steric groups, increase the activity of the compound. . The yellow contours present at the –Cl and -S group of benzothiazole ring shows that bulky groups are unfavorable. This can be explained by the fact that the compounds with -Cl substitution in this area, show unfavorable bulky groups. The CoMFA electrostatic contour plots are displayed in Fig. 3. The blue contour indicates the region where positive groups are required for high activity while the red zone indicates a region favorable for negative groups. Two blue contours exist near the ring positioned at the –Cl group suggesting that there is a requirement of positive charge group at this position. Two red contours exist in the molecule, positioned above benzothiazole ring.

The CoMSIA contour plots are shown in Fig. 4; the CoMSIA steric and electrostatic contours are almost the same as the CoMFA-steric and electrostatic contours (Fig. 3). In Figure 4, two white contours were showing unfavorable hydrophobic interaction regions near indole and benzothiazole rings. This unfavorable region at the ring is due to the -Cl substituent which is similar to that of the yellow steric contour in Fig. 3. Two yellow contours are present near the

benzothiazole ring, which indicates that any bulk group present at this position will represent hydrophobically favored regions. The Hydrogen-bond donor contour maps signify the regions of hydrogen-bond donor favorable (cyan) and unfavorable (purple) regions. Cyan contours are seen near the indole ring attached to the nitrogen and oxygen groups, indicating that hydrogen bond donor functionalities in this region will enhance activity. There is no presence of purple contours on the molecule indicating the absence of unfavorable regions.

The CoMSIA hydrogen bond acceptor fields are denoted by magenta and red contours. Magenta contours represent regions where hydrogen bond acceptors on compounds are favorable, and red contours indicate regions where hydrogen bond acceptors on inhibitors are unfavorable for the activity. There are no red contours present on the molecule, but the large magenta contour present indicates that, in this position, any substituent containing an acceptor group increases the activity. The large magenta contour located on the thiazole ring, indicate that substituent's with the presence of sulphur groups are favored in these areas.

Docking results

The most active compound 28 along with the remaining 27 molecules was docked into receptor site by using FlexX. The crystal structure (PDB ID: 3P2V) was used. The ligand with all water molecules was

deleted and Gasteiger-Hückel charges were assigned. The structure was then minimized using the conjugate gradient algorithm for 5,000 steps with no initial optimization, using Tripos force field. The non-bonding cut-off was set to 15Å and a distance dependent dielectric constant was applied. All atoms of the protein were treated as aggregates, with the exception of those within the 15Å radius of the bound ligand. The ligand was pre-processed before docking

calculations by giving charges according to the Gasteiger-Hückel method followed by energy minimization with 5,000 iterations of conjugate gradient algorithm using Tripos force field. For both structures the active sites include all residues within 15Å radius of the bound ligand and metals. Other functions were set to default values. The most active compound 28 docked with the human Aldose Reductase protein is shown in Figure 5.
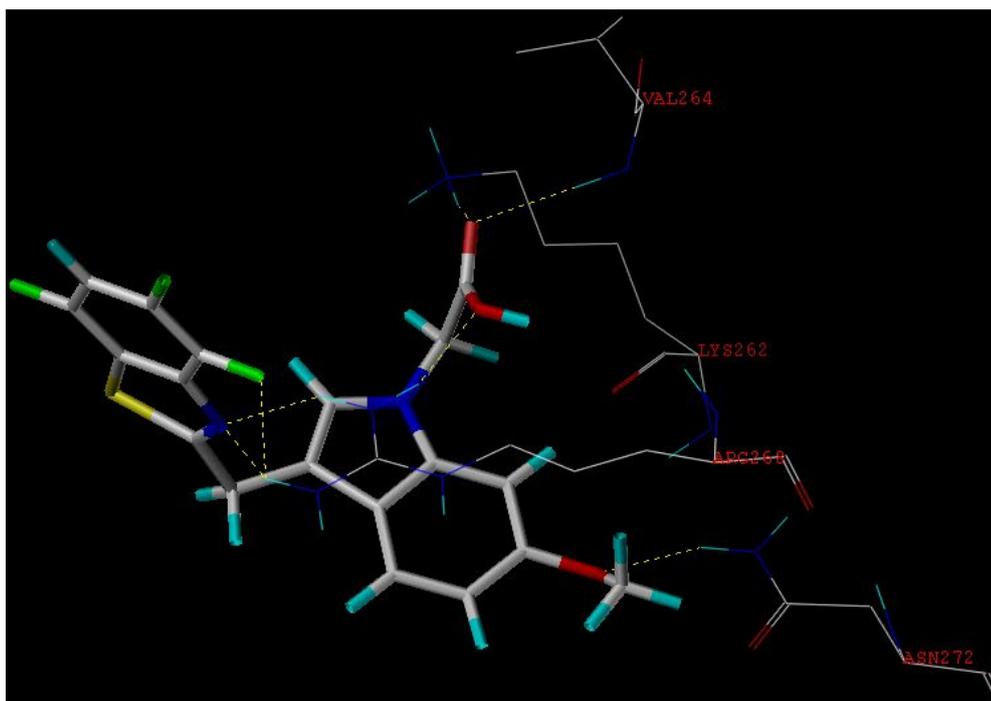
Fig. 5 The active site of Crystal Structure of human Aldose Reductase (PDB id: 3P2V) with important amino acid residues (shown as lines) and the docked ligand of most active compound 28 (shown in capped stick model)

Table 1. Structures and biological activities of molecules used in QSAR study representing both training set and test set (marked as *)

Comp. No	Comp. Name	Structure	pIC$_{50}$	CoMFA		CoMSIA	
				Predicted	Residual	Predicted	Residual
1*	3-(4,5,7-Trifluorobenzothiazol-2-yl) methyl-indole-N-acetic Acid		8.301	7.75	-0.551	8.282	0.018
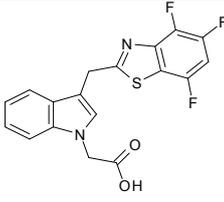
2	5-methyl-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.096	8.188	-0.088	8.036	0.064
3*	7-methyl-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.221	7.718	-0.503	8.201	0.020
4	6-chloro-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.096	7.991	0.109	7.991	0.109
5*	5-benzyloxy-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indol e-N-acetic acid		7.920	8.014	0.093	7.874	0.046
6	6-fluoro-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.154	7.74	0.41	8.123	0.024
7	6-methyl-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.301	7.917	0.383	8.415	0.155
8	3-(5-trifluoromethylbenzothiazol-2-yl) methyl-indole-N-acetic Acid		7.004	7.052	-0.052	7.057	-0.057
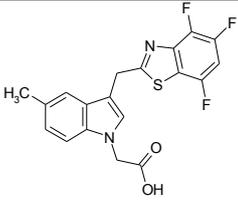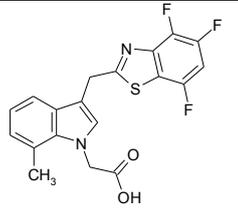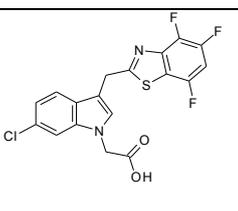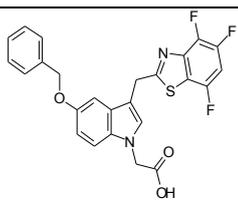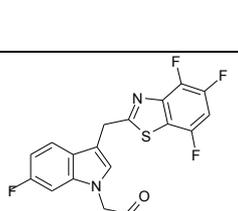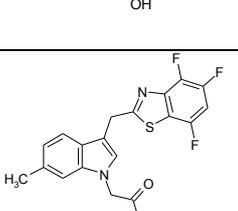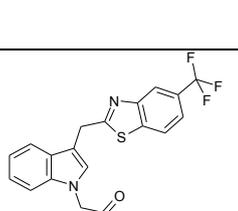
9	5-Methyl-3-(5 Trifluoromethylbenzothiazol-2-yl) methyl-indole -N- acetic acid		6.991	6.989	0.001	6.939	0.051
10	3-(3-nitrophenyl) methyl-indole-n acetic acid		7.136	7.127	0.009	7.14	-0.004
11*	2-phenyl-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		7.00	7.951	0.951	6.997	0.002
12	5-phenyl-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		7.782	7.768	-0.488	7.25	0.03
13	6-phenyl-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		7.602	7.687	-0.087	7.675	-0.075
14	5-morpholino-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indo le-N-acetic acid		8.096	7.944	0.156	8.064	0.036
15*	6-morpholino-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indo le-N-acetic acid		7.82	8.135	0.315	7.889	-0.069
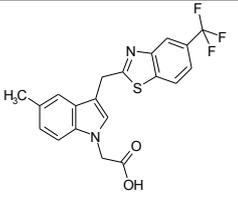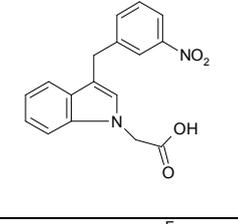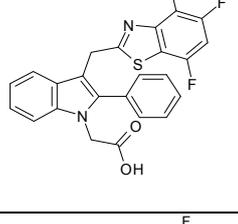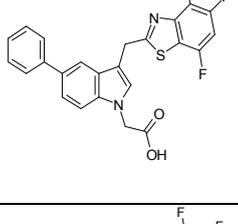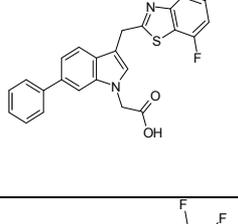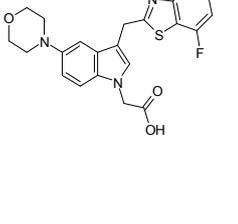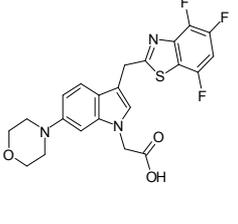
16	5-phenoxy-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole- N-acetic acid		7.522	7.628	-0.108	7.604	-0.084
17	7-fluoro-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.154	8.001	0.149	8.199	-0.049
18	7-bromo-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N- acetic acid		7.853	7.71	0.140	7.852	-0.002
19	7-chloro-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N -acetic acid		8.040	8.196	-0.156	8.007	0.033
20	3-[6-Fluorbenzothiazole-2-yl] methyl-indole-N-acetic Acid		5.982	6.493	-0.513	6.232	-0.252
21*	3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N-2-propionic acid		6.8	7.799	0.99	6.797	0.002
22*	3-(4-5,7-trifluorobenzothiazol-2-yl) methyl-indole-N-3-propionic acid		7.77	8.019	0.249	7.802	-0.032
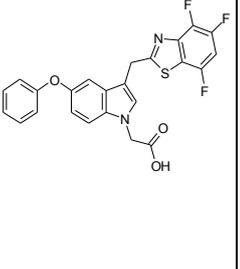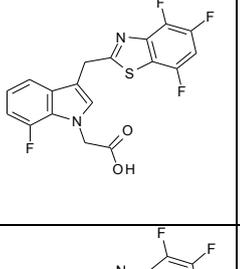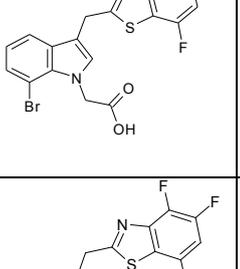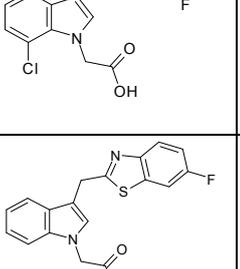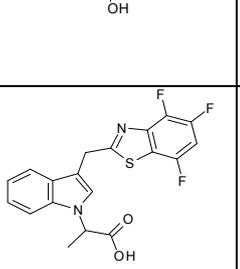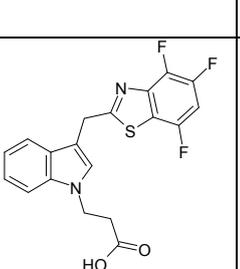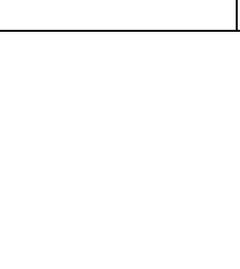
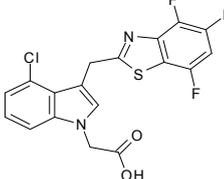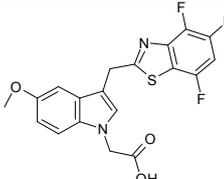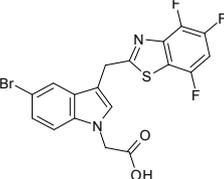|							
23*	6-Bromo-3-(5-trifluoromethylbenzothiazol-2-yl) methyl-indole- N-acetic acid		7.283	7.651	0.367	7.263	0.020
24	6-Methoxy-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N-acetic acid		8.301	8.414	-0.114	8.407	-0.107
25	4-Chloro-3-(4,5,7-trifluorobenzothiazol-2yl) methyl-indole -N-acetic acid		7.958	7.755	0.205	8.042	-0.062
26	5-Methoxy-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole-N-acetic acid		8.113	8.117	-0.007	8.13	-0.02
27	5-Bromo-3-(4,5,7-trifluorobenzothiazol-2-yl) methyl-indole -N-acetic acid		7.886	7.832	0.058	7.944	-0.054
28	3-(6-chlorobenzothiazol-2-yl) methyl-indole-N-acetic acid		6.180	6.302	-0.122	5.96	0.22
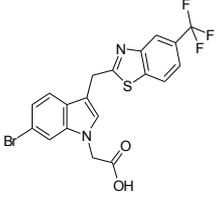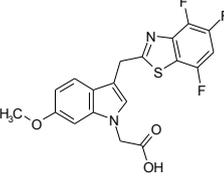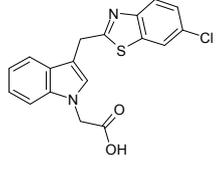
Table 2. Statistical results of CoMFA and CoMSIA models

Index	CoMFA	CoMSIA
q^2	0.850	0.856
r^2	0.983	0.977
CV	0.858	0.855
SEE	0.100	0.120
F-value	171.950	99.708

Conclusion

In this study, 3-D CoMFA and CoMSIA QSAR analyses were used to predict the anti diabetic activity of a set of indolealkanoic acid derivatives. The QSAR models gave good statistical results in terms of q^2 and r^2 values. The CoMFA model provided significant correlation of steric and electrostatic fields with biological activity values. The effects of steric, electrostatic, hydrophobic and hydrogen-bond donor fields around the aligned molecules on biological activity were clarified by analyzing the CoMSIA contour maps. The information obtained in this study provides a methodology for predicting the affinity of related indolealkanoic acid compounds for guiding structural design of novel yet potent inhibitors of diabetes mellitus. The strategy of combining conformations and alignment from the FlexX with CoMFA and CoMSIA produces natural and reasonable elucidation of activation from a 3D-QSAR calculation.

References

1. Ghacha, R., Sinha, A.K. and Karkar, A.M., HbA$_{1C}$ and serum fructosamine as markers of the chronic glycemic state in type 2 diabetic hemodialysis patients, Dial Transplant, 2001, 30, 214–217.
2. Liew, C.F. and Cheah J.S., Hereditary spherocytosis, a pitfall in the assessment of glycaemic control, Singapore Med. J., 2003, 44, 94–97.
3. Dunstan, D.W., Daly, R.M., Owen, N., Jolley, D., De Courten, M., Shaw, J. and Zimmet, P. High-intensity resistance training improves glycemic control in older patients with type 2 diabetes, Diabetes Care, 2002, 25, 1729–1736.
4. Genuth, S., Alberti, K.G., Bennett P., Buse, J., Defronzo, R., Kahn, R., kitzmiller, J., Knowler, W.C., Lebovitz, H., Lernmark, A., Nathan, D., Palmer, J., Rizza, R., Saudek, C., Shaw, J., Steffes, M., Stern, M., Tuomilehto, J. and Zimmet, P., Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Follow-up report on the diagnosis of diabetes mellitus, Diabetes Care, 2003, 26, 3160–3167.
5. Alberti, G., Zimmet, P., Shaw, J., Bloomgarden, Z., Kaufman, F. and Silink M., Type 2 diabetes in the young: the evolving epidemic: the international diabetes federation consensus workshop, Diabetes Care, 2004, 27, 1798–1811.
6. Brown, S.A., Effects of educational interventions in diabetes care: a meta-analysis of findings, Nurs. Res., 1988, 37, 223–230.
7. Arnold, J.G. and McGowan, H.J., Delay in diagnosis of diabetes mellitus dueto inaccurate u se of hemoglobin A$_{1C}$ levels, J. Am. Board Fam. Med., 2007, 20, 93-96.
8. Karki, R.G. and Kulkarni, V.M., Three-dimensional Quantitative Structure-Activity Relationship (3D-QSAR) of 3-Aryloxazolidin-2-one Antibacterials, Bioorg. Med. Chem., 2001, 9, 3153-3160.
9. Puntambekar, D., Giridhar, R. and Yadav, M.R., 3D-QSAR studies of farnesyltranferase inhibitors: A comparative molecular field analysis approach, Bioorg. Med. Chem. Lett., 2006, 16, 1821-1827.
10. Cramer, R.D., Patterson, D.E. and Bunce J.D., Comparative molecular field analysis (CoMFA): Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc., 1988, 110, 5959-5967.
11. Sybyl 6.7, Tripos Associates, 1699 South Hanley Road, St. Louis, USA, mo 63144.
12. Clark, M.C., Cramer III, R.D. and van Opden Bosch, N., Validation of the General Purpose Tripos 5.2 Force Field, J. Comput. Chem., 1989, 10, 982–1012.
13. Gasteiger, J. and Marsilli, M., Iterative partial equalization of orbital electro negativity- a rapid

access to atomic charges, Tetrahedron, 1980, 36, 3219-3225.

14. Bhongade, B.A. and Gadad, A.K., 3D-QSAR CoMFA/CoMSIA studies on Urokinase plasminogen activator (uPA) inhiubitors: A strategic design in novel anticancer agents, Bioorg. Med. Chem., 2004, 12, 2797-2805.

15. Klebe, G., The use of composite crystal field environments in molecular recognition and the de novo design of protein ligands, J. Mol. Biol., 1994, 237, 212-235.

16. Geladi, P. and Kowalski, B., Partial least-squares regression: A tutorial, Analytica Chimica Acta, 1996, 185, 1-17.

17. Wold, S., Ruhe, A., Wold, H. and Dunn III, W.J., The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, J. Sci. Stat. Comput., 1984, 5, 735-743.

18. Kramer, B., Metz, G., Rarey, M. and Lengauer, T., Ligand docking and screening with FlexX, Med. Chem. Res., 1999, 9, 463-478.

19. Gilbert, K.M., Skawinski, W.J., Misra, M., Paris, K.A., Naik, N.H., Buono, R.A., Deutsch, H.M. and Venanzi, C.A., Conformational analysis of methylphenidate: comparison of molecular orbital and molecular mechanics methods, J. Comput. Aided Mol. Des., 2004, 18, 719-738.

*****