Available online at www.derpharmachemica.com

Scholars Research Library

Der Pharma Chemica, 2010, 2(1): 82-90
(http://derpharmachemica.com/archive.html)

ISSN 0975-413X

Three Dimensional Quantitative Structural–Activity Relationship (3D-QSAR) Studies some 3-{4-[3-(2-aryl-phenoxy) butoxy]-phenyl} Propionic acids as novel PPAR γ/δ agonists

M. C. Sharma*, Smita Sharma[a], D. V. Kohli[b], S. C. Chaturvedi [c]

School of Pharmacy, Devi Ahilya Vishwavidyalaya, Indore (M.P), India
[a] Department of Chemistry, Yadhunath Mahavidyalya Bhind (M.P), India
[b] Department of Pharmaceutical Sciences, Dr. Hari Singh Gour University, Sagar (M.P), India
[c] Shri Arvindo Institute of Pharmacy, Ujjain Road, Indore (M.P), India

---

Abstract

The use of Three Dimensional Quantitative Structure–Activity Relationships, since its advent, has become increasingly helpful in understanding many aspects of biochemical interactions in drug research. This approach was utilized to explain the relationship of structure with biological activity of selective PPAR γ/δ agonists. The enormity of the PPAR γ/δ agonists discovery is reflected in the unprecedented speed at which research laboratories have sought to validate its clinical implications. 3D QSAR study through recently introduced k- Nearest Neighbor Molecular Field Analysis (k-NN MFA) with Step Wise (SW), Simulated Annealing (SA) and Genetic Algorithm (GA) as variable selection methods resulted in eleven statistically significant models for PPAR γ/δ agonists. These models gave a value of q^2 as high as 0.7362 for model 1 and value of pred_r^2 as high as 0.6985 for model 3.The k-NN MFA contour plots provided further understanding of the relationship between the structural features of PPAR γ/δ agonists derivatives and their activities, which should be applicable to design new, potential PPAR γ/δ agonists.

Key-words-3D QSAR, k-NN MFA, DUAL PPAR γ/δ, Diabetes.

Introduction

Peroxisome Proliferator-Activated Receptors (PPARα and PPARγ) are ligand activated nuclear hormone receptors that regulate the transcription of genes involved in carbohydrate and lipid metabolism pathways [1-4] activation of PPARα, which is predominantly expressed in adipose tissue, results in insulin-sensitizing anti-diabetic effects [5-6]. Activation of PPARγ, which is highly expressed in the liver, results in the lowering of triglycerides and the elevation of plasma HDL cholesterol levels [7-8]. In addition, both PPARα and PPARγ

selective activators have been demonstrated to suppress vessel wall inflammatory activity and reduce atherosclerosis in experimental animal models through complementary mechanisms [9-10]. Therefore, PPARα and PPARγ dual activators provide superior profile toward the control of hyperglycaemia and hyper triglyceridemia.Current trends in medicinal chemistry are to identify molecules with activity on multiple targets. Such molecules act on more than one biological target and produce a synergistic effect. This approach offers an advantage that the body has to deal only with a single molecule having multiple activities, which could eliminate a lot of complexity in terms of administering medication to the patient, the body's ability to absorb a drug, and also its side effects. Most often the biological targets involved in multiple activities are quite similar to each other and belong to the same family of receptors. An important class of compounds currently under study in the category of dual activators are members of the nuclear receptor super family viz; peroxisome proliferator- activated receptors (α, γ and ß/δ) [11]. Many topological descriptors can be used to describe organic molecular structure with QSAR aims. Resent trends in 2D, 3D QSAR have focused on the development of procedure that allows selection of optimal variables from the pool of descriptors of chemical structures i.e. ones that are most meaningful and statistically significant in terms of correlation with biological activity. This is accomplished by combining one of the stochastic search methods such as SA, GAs, or evolutionary algorithms with the correlation methods such as MLR, PLSR, or artificial neural networks [12-17]. The k-NN MFA, used for 3D QSAR analysis of the present data set adopts a k-nearest neighbour principle for generating relationships of molecular fields with the experimentally reported activity. The variables and optimal k values were chosen using three variable selection methods viz. SW, SA, and GA. Like many 3D QSAR methods, k-NN MFA requires suitable alignment of given set of molecules. This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The term descriptor is utilized in the following discussion to indicate field values at the lattice points. The optimal training and test sets are generated using the sphere exclusion algorithm. This algorithm allows the construction of training sets covering descriptor space occupied by representative points. Once the training and test sets are generated, k-NN methodology is applied to the descriptors generated over the grid [18].

Result and Discussion

The 3D QSAR for molecular field analysis was performed using the k-Nearest Neighbor method using software V-LIFE MDS 3.0. A set of 16 molecules was selected and divided in training (12) and test sets (4) for 3D QSAR studies. All the molecules were aligned based on template. The model was generated by kNN-MFA method which showed the cross validated squared correlation coefficient value. The model showed the better correlation with biological activity. The steric and electrostatic contribution showed the effect of substitution on biological activity. The steric and bulky groups reduced the activity, the electrostatic methyl and ethyl groups at positions R_1 and R_4 are essential for the activity. Substitution by ethyl group on position R_4 increases the activity. If the ethyl group is substituted by CF_3, the activity is increased as compared to alkyl substitutions.

Model 1

S_263 (-0.4351 to 30.0000); 23(0.9021 to 10.0000); E_151 (-0.0329 to 6.3509); E_964 (-0.7385 to -0.0814)

k Nearest Neighbor=5; n = 16; Degree of freedom = 21; q2 = 0.6371; q2_se =0.3491 ;Predr^2 = 0.5921;pred_r2se = 0.4160

Another statistically significant model 1 was obtained for PPARα and PPARγ activity through SW k-NN MFA justified by internal and external predictivity of the model as 63 % (q2=0.6371) and 37 % (pred_r2=0.5921) respectively.

Model 2
S_263 (2.581 to 30.000); E_151 (-0.0329 to 6.3509) ; E_256 (-0.5731 to 7.125) ; E_964 (-1.4721 to 0.7831)
k Nearest Neighbor= 5; n = 16; Degree of freedom = 22 ;q2 = 0.6941 ;q2_se = 0.4206 Pred_r^2 = 0.5357; pred_r2se = 0.6260

For PPARα and PPARγ, models, two were found to be statistically significant justified by the values of q2 that explained 98% internal predictivity

Model 3
S_1252 (-0.3452 to 0.5438); S_765 (-1.789 to 0.6907)
k Nearest Neighbor= 4; n = 16 ;Degree of freedom = 23 ;q2 = 0.7263 ;q2_se = 0.5791 pred_r2=0.6975; pred_r2se = 0.7433

Model 3, the best model developed through k-NN MFA had a value of q2=0.7263 and that of pred_r2=0.6975 that explained 76% of total variance (internal predictivity) and 24% predictive power for the external test set.

Model 4
S_969 (1.8742 to 2.9076); E_262 (-10.000 to -0.6902)
k Nearest Neighbor= 5 ;n = 16 ;Degree of freedom = 22 ;q2 = 0.8218 ;q2_se = 0.6211 Predr2 = 0.7941 ; pred_r2se = 0.5652

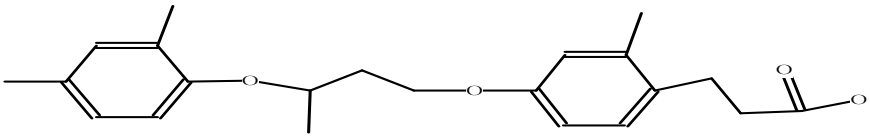
Another statistically significant model, model 4 was generated for activity against PPARα and PPARγ having a value of q2 89 % and that of pred_r2=0.7941 79 %, Plot of the k-NN MFA which shows the relative position and ranges of the corresponding important.

Materials and Methods

Data set and Molecular modelling

A dataset consisting of a series of 3-{4-[3-(2-aryl-phenoxy) butoxy]-phenyl} propionic acids derivatives acting PPAR **γ/δ** dual activators (Table 1) has been chosen to develop a dual-response QSAR model. The biological activities of the molecules have been expressed as the binding affinities measured as IC_{50} values in micro molar using recombinant PPAR **γ/δ** by [19]. For QSAR analysis, these have been converted to pIC50 (log IC_{50}) values in molar terms (Table 1). The molecular modelling was carried out on Compaq PC having Pentium IV processor and windows 98 operating system, using the software namely: Molecular Design Suite supplied by the VLife Sciences, Pune (MDS 3.5) [20]. The structures were constructed using the 2D draw application and converted to 3D structures. Energy minimization and geometry optimization was conducted using Merck Molecular Force Field (MMFF) method with Root Mean Square (RMS) gradient set to 0.01 Kcal/mol A^0 and iteration limit to 10000. Alignment of all the 16 compounds was done using template based alignment in MDS; the aligned structures were used for the study. In the template based alignment method, a

template structure was defined and used as a basis for alignment of a set of molecules. Following 3-{4-[3-(2-aryl-phenoxy) butoxy]-phenyl} propionic acids nucleus was the template used for template-based alignment, as it was common to all structures.



The alignment of molecule is shown in figure.1

Table-1 Substituted 3-{4-[3-(2-aryl-phenoxy) butoxy]-phenyl} propionic acids With IC_{50} and PIC_{50} values

S.No.	Compound	R^3	R^1	R^4	PPAR IC_{50}	log IC_{50}
1.	1mol2	Benzoyl	Me	Et	5	0.69897
2.	8 mol2	3- Thiophenyl	Me	Et	568	2.75435
3.	9 mol2	2-Thiophenyl	Me	Et	1090	3.03743
4.	10 mol2	2- Furanyl	Me	Et	940	2.97313
5.	11 mol2	2- Pyridyl	Me	Et	46	1.66276
6.	12 mol2	3- Pyridyl	Me	Et	60	1.77815
7.	13 mol2	4- Pyridyl	Me	Et	520	2.71611
8.	14 mol2	4 -Thiazolyl	Me	Et	346	2.53908
9.	15 mol2	2-Thiazolyl	Me	Et	1530	3.18469
10.	17 mol2	2- Oxazolyl	Me	Et	483	2.68395
11.	18mol2	2- Pyridyl	Me	Cl	88	1.94448
12.	19 mol2	2- Pyridyl	Et	CF_3	20	1.30103
13.	* 21 mol2	2- Pyridyl	Et	CF_3	16	1.20412
14.	* 7mol2	Phenyl	Me	Et	215	2.332438
15.	* 20mol2	2- Pyridyl	Et	Et	35	1.544068
16.	* 16mol2	4-Oxazolyl	Me	Et	979	2.990783

* Selected Test Compound

Table -2 Actual and Predicted values for model-1 and model-2 with Residual values for *Training set* Compounds 3D QSAR analysis

S. No.	Molecules	Actual activity	Predicted model	Residual value
1.	1mol2	0.69897	1.30294	-0.60397
2.	8mol2	2.75435	2.88018	-0.12583
3.	9mol2	3.03743	2.80411	0.23332
4.	10mol2	2.97313	3.16774	-0.19461
5.	11mol2	1.66276	1.5466	0.11616
6.	12mol2	1.77815	1.80576	-0.02761
7.	13mol2	2.71611	1.64432	1.07179
8.	14mol2	2.53908	2.64115	-0.10207
9.	15mol2	3.18469	2.96391	0.22078
10.	17mol2	2.68395	2.86027	-0.17632
11.	18mol2	1.94448	2.37214	-0.42766
12.	19mol2	1.30103	0.714547	0.586483

Table 3 - Actual Activity, Predicted Activity and Residual values of *test set* Compounds

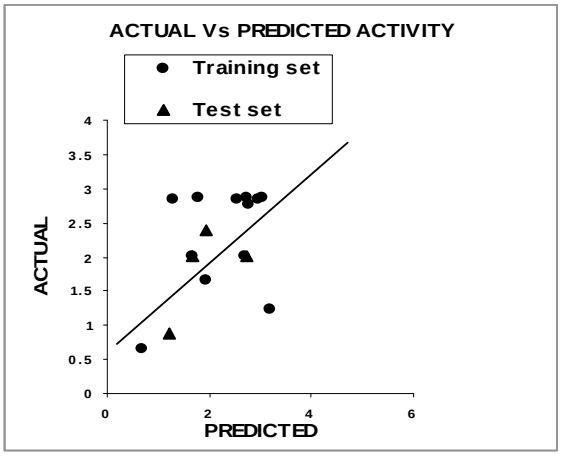
S. No.	Molecules	Actual activity	Predicted model	Residual value
1.	21mol2	1.20412	1.64768	-0.44356
2.	7mol2	2.332438	2.31224	0.020198
3.	20mol2	1.544068	1.35126	0.192808
4.	16mol2	2.990783	2.74123	0.249553

Table-4 Selected Descriptors list using Generated Models

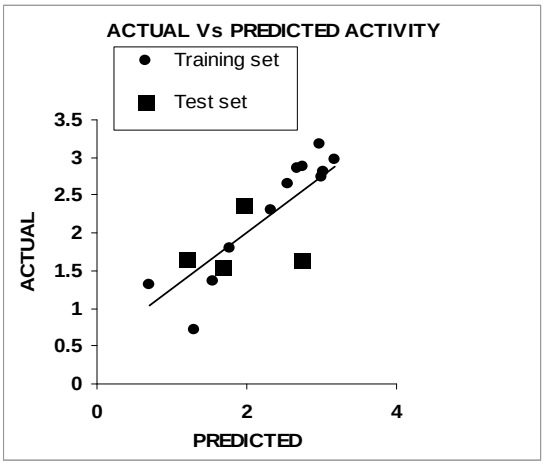
S.No	E_151	E_262	E_256	E_964	S_263	S_765	S_969	S_1252
1	-0.6567	-0.0503	-0.06234	-0.07263	0.44088	1.38526	4.38549	0.64375
2	-0.0367	-0.0169	-0.03408	-0.03444	0.55717	1.28250	3.06462	0.80289
3	-0.2458	-0.0094	-0.02765	-0.0335	0.25044	0.78778	3.94302	1.00822
4	-0.4954	-0.0295	-0.04694	-0.05248	0.35934	0.86284	4.49181	0.98329
5	-0.2894	-0.0096	-0.02967	-0.03132	0.49431	1.26850	2.71637	0.72219
6	-0.4928	-0.0572	-0.04939	-0.06361	0.35494	1.32161	2.98633	0.58047
7	-0.1091	0.0195	-0.13171	-0.12818	0.44423	1.17249	1.51944	0.88284
8	-0.2003	0.0134	-0.02407	-0.0187	0.55180	1.18664	4.04545	0.70536
9	-0.5959	-0.0442	-0.05667	-0.06559	0.54389	1.46701	4.40631	0.69588
10	0.0590	-0.0258	0.012139	0.006988	0.26011	1.59464	2.98761	0.41411
11	-0.0197	-0.0288	-0.01755	-0.02662	0.30763	0.62858	4.67969	0.88627
12	-0.0246	-0.0013	-0.02773	-0.02682	0.39123	1.27606	4.70028	0.68648
13	-0.0582	-0.0601	-0.05639	-0.07047	0.26475	1.09756	2.52539	0.35704
14	-0.0886	-0.0323	-0.08055	-0.08934	0.41369	0.87093	3.63106	1.00149
15	0.0031	-0.0123	0.006059	0.006477	0.49731	2.62949	1.75036	0.47314
16	-0.0158	-0.0383	-0.01177	-0.02201	0.45613	1.04476	0.68546	0.45757

The position of each atom is important for kNN-MFA because the descriptors were calculated based on the 3D space grid. Thus, the method to determine the conformation of each molecule and the way to align molecules together are two sensitive input parameters to build reasonable model [21] in the present study two deferent alignment rules were adopted.

Figure-2 Graph of observed v/s predicted activities of statistically significant models obtained through kNN MFA



Model 1



Model 2

Figure 3- Alignment of 3D structures of the series of 3-{4-[3-(2-aryl-phenoxy) butoxy]-phenyl} propionic acids with steric and electrostatic involvement in molecule 11

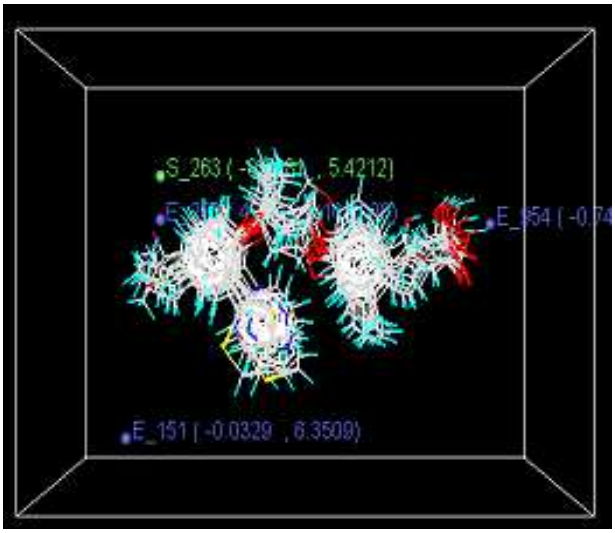
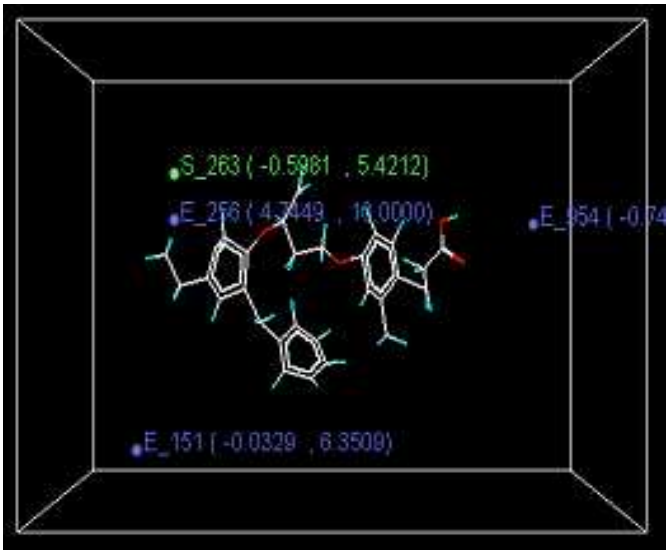


Figure 4 - Best Model Steric and electrostatic involvement in molecule 11



Calculation of field descriptor values

For calculation of field descriptor values, both electrostatic and steric field type with cut offs 10.0 and 30.0 Kcal/mol respectively were selected and charge type was selected as Gasteiger – Marsili. Dielectric constant was set to 1.0 considering the distance dependent dielectric function. Probe setting was carbon atom with charge 1.0 and grid.

k-Nearest Neighbor (kNN) Method:

The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows [22]

(1) Calculate distances between an unknown object (*u*) and all the objects in the training set;
(2) Select *k* objects from the training set most similar to object *u*, according to the calculated distances.
(3) Classify object *u* with the group to which the majority of the *k* objects belongs. An optimal *k* value is selected by optimization through the classification of a test set of samples or by leave-one out cross-validation. The variables and optimal *k* values were chosen using different variable selection methods as described below.

kNN-MFA 3D-QSAR Models:
To derive the kNN-MFA descriptor fields, a 3D cubic lattice grid in *x, y* and *z* directions, was created to encompass the aligned molecules. kNN-MFA descriptors were calculated using an sp3 carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric field energies and electrostatic fields with the distance dependant dielectric at each lattice point. The steric and electrostatic energy values were truncated at a default value of 30 kcal/mol.

kNN-MFA with Simulated Annealing
Simulated annealing (SA) is the simulation of a physical process, 'annealing', which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature). During this process, the system samples possible configurations distributed according to the Batsman distribution so that at equilibrium, low energy states are the most populated.

kNN-MFA with Stepwise (SW) Variable Selection
This method employs a stepwise variable selection procedure combined with kNN to optimize the number of nearest neighbours (k) and the selection of variables from the original pool as described in simulated annealing.

kNN-MFA with Genetic Algorithm
Genetic algorithms (GA) first described by Holland mimic natural evolution and selection. In biological systems, genetic information that determines the individuality of an organism is stored in chromosomes. Chromosomes are replicated and passed onto the next generation with selection criteria depending on fitness. The 3D QSAR for molecular field analysis was performed using the k Nearest Neighbour method using software V-LIFE MDS 3.5.

PLS analysis:
The partial least squares method (PLS), [23-25] was used to derive a linear relationship and cross-validation was performed using leave-one out method, [26-27] to check consistency and predictive ness. Models were generated by using three significant statistical methods, namely, partial least square analysis, multiple regressions, and principle component analysis. The cross-validation analysis was performed using the leave-one-out method. In the selected equations, the cross-correlation limit was set at 0.5, the number of variables at 10, and the term selection criteria at r^2. An F value was specified to evaluate the significance of a variable. The higher the F value, the more stringent was the significance level: F test ''in'' as 4 and F test ''out'' as 3.99. The variance cut off was set at 0, and scaling was auto scaling in which the number of random iterations was set at 100.The following statistical parameters were considered for comparison of the generated QSAR models: correlation coefficient (r), squared correlation coefficient (r2), predictive r^2 for external test set (pred r^2) for external validation, and Fischer's (F).The predicted r^2 (pred_r^2) value was calculated using Eq. 1, where yi and yˆi are the actual and predicted activities of the i^{th} molecule in the test set,

respectively, and ymean is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The pred_r^2 value indicates the predictive power of the current model for the external test set as follows

$$pred\_r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \qquad (1)$$

Internal validation was carried out using leave-one-out (q^2, LOO) method. For calculating q^2, each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The q^2 was calculated using the equation which describes the internal stability of a model:

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2} \qquad (2)$$

Where y_i, and $\hat{y}_i$ are the actual and predicted activity of the $_i$th molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

Randomization Test.

To evaluate the statistical significance of the QSAR model for an actual data set, we have employed a one-tail hypothesis testing. The robustness of the QSAR models for experimental training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging biological activities of the training set molecules. The significance of the models hence obtained was derived based on calculated Z_{score} [28-29].

Evaluation of the QSAR Models.

The QSAR models were evaluated using following statistical measures: *n*, number of observations (molecules); Vn, number of descriptors; *k*, number of nearest neighbours; *q*2, cross validated *r*2 (by the leave-one-out method); pred_r2, predicted *r*2 for the external test set; Zscore, the Z score calculated by *q*2 in the randomization test; best_ran_q2, the highest *q*2 value in the randomization test; and R, the statistical significance parameter obtained by the randomization test.

Conclusion

The steric and bulky groups reduced the activity, the electrostatic methyl and ethyl groups at positions R_1 and R_4 are essential for the activity. Substitution by ethyl group on position R_4 increases the activity. If the ethyl group is substituted by CF_3, the activity is increased as compared to alkyl substitutions. For the best 3D-QSAR model, the cross-validated squared correlation coefficient (Q^2) was 0.8218 and predicted R^2 value was found to be 0.7941. Models have given significant information to build a strategy to improve the biological activity of the compounds. Substituted methyl and ethyl groups at R^1 are essential for the biological activity. The CF_3 group at R^4 gives compounds with better biological activity than the ethyl substituents. The electrostatic contribution of 2-Pyridyl and 3-Pyridyl groups at R^3 led to compounds with good selectivity over PPARα and potent PPAR γ/δ affinity and functional activity.

Acknowledgement

The authors are thankful to Mr. Amit Badi Technical scientist Vlife Science Technologies Pvt. Ltd for Cooperating and given valuable suggestions and Vlife Science Technologies Pvt. Ltd, 1, Akshay 50, Anand Park, Aundh, Pune, India to provide trial version of software.

References

[1] S. Kersten, B. Desvergne, W. Wahli, *Nature.*, **2000**, 405, 421-424.
[2] S. Kliewer, E. Xu, M. Lambert, T.Willson, *Recent Prog. Horm. Res.*, **2001**, 56, 239-263.
[3] J. Berger, D. Moller, *Annu. Rev. Med.*, **2002**, 53, 409–435.
[4] H.Bays,L.Mandarino,R.DeFronzo, *J.Clin. Endocrinol. Metab,* **2004,** 89, 478.
[5] J. Olefsky, *J. Clin. Invest.*, **2000,**106, 467 – 472.
[6] H. Lebovitz, *Diabetes Metab. Res. Rev.*, **2002**, 18, 23-29.
[7]S.Robins,D.Collins,J.Wittes,V.Papademetriou,P.Deedwania,E.Schaefer,J.McNamara,M.Kashyap,J.Hershman,L.Wexler,H. Rubins,*J.Am.Med.Assoc.*,**2001**,285,1585 – 1591.
[8] Diabetes Atherosclerosis Intervention Study Investigators, *The Lancet.*, **2001,**357, 905-910.
[9] K. Buchan, D. Hassall, *Med. Res. Rev.*, **2000,** 20,350-366.
[10] C.Duval, G.Chinetti, F.Trottein, J. Fruchart, B. Staels, *Trends Mol. Med.*, **2002,** *8*, 422-430.
[11] T. M. Willson, P. J. Brown, D. D. Sternbach, B. R. Henke, *J.Med.Chem.*, **2000,** 43, 527-541.
[12]J.MSutter,S.L.Dixon,P.C.Jurs,*J.Chem.Inf.Comput.Sci.*,**1995**,35,77-84.
[13] D. Rogers, A.J.Hopfinger, *J.Chem. Inf. Comput. Sci.*, **1994**, 34, 854-866.
[14]H.Kubinyi,*Quant.Struct.Act.Relat.*,**1994**,13,285-294.
[15] H. Kubinyi, *Quant. Struct.Act. Relat.*, **1994,**13,393-401.
[16] B. T. Luke, *J. Chem. Inf. Comput. Sci.*, **1994,** 34,1279-1287.
[17] S.S.So, M.Karplus, *J. Med. Chem.*, **1996,** 39, 1521-1530.
[18] S. Ajmani, K. Jadhav, S.A Kulkarni, *J.Chem. Inf. Model.*, **2006**, 46, 24-31.
[19] I. C. Gonzalez, J. Lamar, F.Iradier, *Bio. Med. Chem. Lett.*, **2007**, 17, 1052-1055.
[20] V-Life MDS 3.5 Vlife Science Technologies Pvt. Ltd, 1 akshay residency, plot 50, Anand Park, aundh pune 411007.
[21] R. D. Cramer, D.E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.*,**1988**, 110,5967.
[22] M.A.Sharaf, D.L.Illman, B. R. Kowalski, *Chemometrics.*, Wiley, New York, **1986**.
[23]S.Wold, C. Albano,W.J. Dunn, U. Edlund, K. Esbenson, P. Gelad, S. Hellberg, W. Lindberg, M. Sjostrom, In; *Chemometrics*, Int. Ed. Kowalski, B., Reidel Dortrecht, The Netherlands, 17, **1984**.
[24] W.J.Dunn, S.Wold, U. Edlund, S.Hellberg, J.Gasteiger, *Quant. Struct. Act. Relat.Chem.Bio.*, **1984**, *3*, 31.
[25] P. Geladi, *J. Chemom.*, **1998,** 2, 231.
[26] S. Wold, *Technometrics.*, **1978,** 4,397.
[27] R.D.Cramer,J.D.Bunce,D.E.Patterson,*Quant. Struct. Act. Relat.*, **1998,** 7, 18.
[28] M.A.Sharaf, D. L. Illman, B. R. Kowalski, Chemometrics, Wiley, New York, **1986.**
[29] J.Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, **1975**.