Chemometric modelling of PPAR-α and PPAR-γ dual agonists for the treatment of type-2 diabetes

Neha Verma* and Usha Chouhan

Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal 462 003, India

Type-2 diabetes mellitus (T2DM) is an enervating and fast-growing disease characterized by hyperglycaemia. The increasing incidences of T2DM represent a public health problem. The disease is characterized by loss in sensitivity of tissues towards insulin, which can be restored by the activation of peroxisome proliferator-activated receptors (PPARs). PPARs are members of the nuclear receptor family, which function as a ligand-dependent transcription factor. The aim of the present work is to develop ligands, which can activate PPARs and are expected to lower LDL cholesterol and triglycerides, raise HDL cholesterol, and normalize hyperglycaemia. Here quantitative structure–activity relationship (QSAR) study is performed, followed by pharmacophore modelling and docking of the most active compound to the proteins PPAR-γ (PDB ID: 1FM9) and PPAR-α (PDB ID : 1K7L). Docking studies revealed the importance of hydrogen-bonding interactions for the binding of targets with the ligand. QSAR study is performed on the dataset by means of multiple linear regression and partial least squares (PLS) techniques. A good correlation is found by regression analysis between the observed and predicted activities as evident by their R^2 (0.651), Q^2 (0.649) and R^2_{pred} (0.606) for PPAR-γ, and R^2 (0.784), Q^2 (0.774) and R^2_{pred} (0.841) for PPAR-α. Subsequent analysis of the model by PLS cross-validation technique yields a similar set of coefficients. Pharmacophore studies reveal the importance of features like hydrogen bond donor, hydrogen bond acceptor and aromaticity, which contribute significantly in both models and are essential for binding of ligands to the receptor and also for their proper functioning.

Keywords: Chemometric modelling, diabetes mellitus, peroxisome proliferator-activated receptors, quantitative structure–activity relationship.
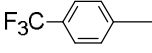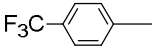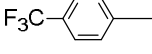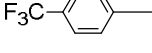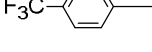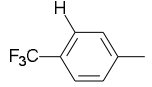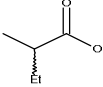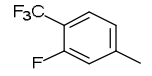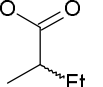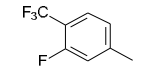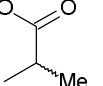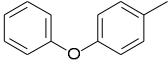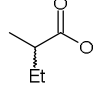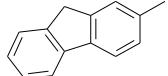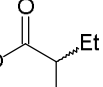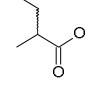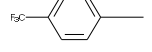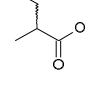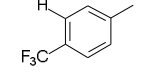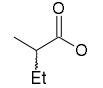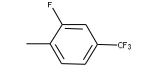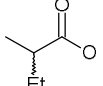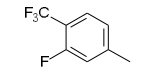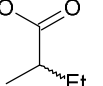
DIABETES mellitus (DM), long considered a disease of minor importance to global health, is now a major threat to human health[1]. DM is becoming the leading cause of death in developed countries, being the fourth or fifth most common non-communicable diseases worldwide[2]. It is estimated that by 2025, 300 million people will be diabetic or prediabetic[3]. Developing countries such as India have seen maximum increase in DM in the last few years. In general, people with diabetes either totally lack insulin (type-1 diabetes mellitus or T1DM) or have too little insulin or cannot use it efficiently (type-2 diabetes mellitus or T2DM); T2DM accounts for 90–95% of all diabetic patients. DM is marked by raised blood glucose levels and is a heterogeneous group of diseases[4]. The increasing incidences of T2DM and its consequences in terms of cardiovascular morbidity and mortality represent a considerable public health problem[5]. Thus, for developing new therapeutic agents for the treatment of T2DM and other metabolic syndromes, it is necessary to identify the molecular targets of the transducers critically involved in the control of glucose and lipid homeostasis. For cellular and whole-body glucose and lipid homeostasis maintenance, metabolic nuclear receptor (NR) molecules are found to be a particularly attractive target and play a key role in controlling glucose and lipid homeostasis. Among these receptors, special attention has been paid for more than a decade to the members of the peroxisome proliferator-activated receptor (PPAR) family[6].

PPARs belong to the nuclear hormone receptor family and are defined as transcriptional factors that are activated by the binding of ligands to their ligand-binding domains (LBDs)[7]. There are three subtypes of PPARs[8], namely PPAR-α, PPAR-β, and PPAR-γ, which share similar three-dimensional structure within LBDs, but display distinct tissue distribution pattern and different pharmacological profiles[9]. Thus ligands simultaneously activating all the PPARs can be strong candidates in relation to drugs and can be used to treat abnormal metabolic homeostasis. Derivatives of 3-(4-alkoxyphenyl) propanoic acid are synthetic ligands which exhibit unique PPAR agonistic activities[7]. Saturated and unsaturated fatty acids, endogenous metabolites and synthetic ligands are known to activate PPARs[10]. PPAR-α is mostly expressed in liver, kidney, skeletal, heart muscles and adrenal glands, as these tissues are involved in lipid oxidation[11]. PPAR-γ is expressed in macrophages, vascular smooth muscles and adipose tissue[12]. PPAR-γ was first determined as a key regulator for differentiation of an adipocyte, but recent molecular–biological studies have indicated that PPAR-γ activation is also linked to the expression of
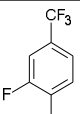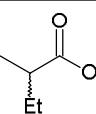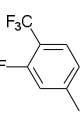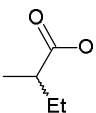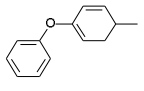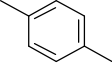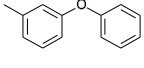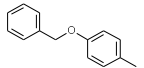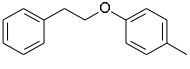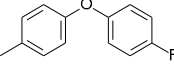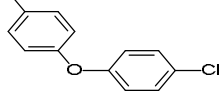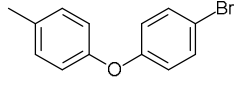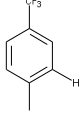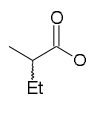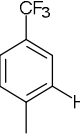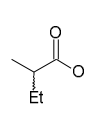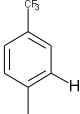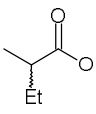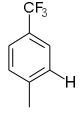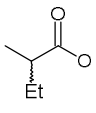
Table 1. Biological activity and structure of phenylpropanoic acid derivatives

Sl. no.	R_1	R_2	R_3	B	Activity experimental (EC_50) μmol PPAR-γ	PPAR-α
$_{atr}C1^{\gamma tr}$	F_3C–(phenyl)–	thiazolidinedione (TZD) ring	OMe	CH_2NHCO	800	100
$_{atr}C2^{\gamma ts}$	F_3C–(phenyl)–	$O(CH_2)_3CO_2H$	OMe	CH_2NHCO	3000	220
$_{atr}C3^{\gamma tr}$	F_3C–(phenyl)–	$CH_2CH(Et)CO_2H$	OMe	CH_2NHCO	400	4
$_{atr}C4^{\gamma tr}$	F_3C–(phenyl)–	$CH_2CH(n\text{-}Bu)CO_2H$	OMe	CH_2NHCO	2500	100
$_{atr}C5^{\gamma tr}$	F_3C–(phenyl)–	$CH_2CH(OEt)CO_2H$	OMe	CH_2NHCO	2800	160
$_{atr}C6^{\gamma ts}$	F_3C–(phenyl, H)–	$CH(Et)CO_2$– ester	OMe	CH_2NHCO	3000	7000
$_{ats}C7^{\gamma ts}$	F_3C, F–(phenyl)–	$CH(Et)CO_2$– ester	OMe	CH_2NHCO	870	7300
$_{atr}C8^{\gamma tr}$	F_3C, F–(phenyl)–	$CH(Me)CO_2$– ester	OMe	CH_2NHCO	1700	3200
$_{atr}C9^{\gamma tr}$	phenyl-O-(phenyl)–	$CH(Et)CO_2$– ester	OMe	CH_2NHCO	8200	1900
$_{atr}C10^{\gamma tr}$	fluorenyl	$CH(Et)CO_2$– ester	OMe	CH_2NHCO	2000	1500
$_{atr}C11^{\gamma tr}$	F_3C–(phenyl)–	$CH(Et)CH(CH_3)CO_2$– ester	OMe	CH_2NHCO	7000	50,000
$_{atr}C12^{\gamma tr}$	F_3C–(phenyl)–	$CH(Et)CH(CH_3)CO_2$– ester	OMe	$CONHCH_2$	3000	7000
$_{atr}C13^{\gamma tr}$	F_3C, H–(phenyl)–	$CH(Et)CO_2$– ester	OMe	$CONHCH_2$	2600	1900
$_{ats}C14^{\gamma ts}$	F, CF_3–(phenyl)–	$CH(Et)CO_2$– ester	OMe	$CONHCH_2$	2600	1900
$_{atr}C15^{\gamma ts}$	F_3C, F–(phenyl)–	$CH(Et)CO_2$– ester	OMe	$CONHCH_2$	2200	1000

(*Contd*)

Table 1. (*Contd*)

Sl. no.	R_1	R_2	R_3	B	Activity experimental (EC_{50}) μmol	
					PPAR-γ	PPAR-α
$_{atr}C16^{\gamma ts}$			OMe	$CONHCH_2$	6000	1100
$_{atr}C17^{\gamma tr}$			OMe	$CONHCH_2$	3600	6300
$_{atr}C18^{\gamma tr}$			OMe	$CONHCH_2$	1900	1000
$_{atr}C19^{\gamma tr}$			OMe	$CONHCH_2$	4900	1200
$_{ats}C20^{\gamma ts}$			OMe	$CONHCH_2$	800	850
$_{atr}C21^{\gamma tr}$			OMe	$CONHCH_2$	9600	10,000
$_{ats}C22^{\gamma ts}$			OMe	$CONHCH_2$	4400	1200
$_{atr}C23^{\gamma tr}$			OMe	$CONHCH_2$	820	880
$_{ats}C24^{\gamma tr}$			OMe	$CONHCH_2$	4600	4100
$_{atr}C25^{\gamma tr}$			OMe	$CONHCH_2$	340	760
$_{atr}C26^{\gamma ts}$			OMe	$CONHCH_2$	710	7400
$_{atr}C27^{\gamma tr}$			OEt	$CONHCH_2$	460	5200
$_{ats}C28^{\gamma tr}$			On-Pr	$CONHCH_2$	780	6000
$_{atr}C29^{\gamma tr}$			On-Bu	$CONHCH_2$	920	500

(*Contd*)

Table 1. (*Contd*)

Sl. no.	R$_1$	R$_2$	R$_3$	B	Activity experimental (EC$_{50}$) µmol	
					PPAR-γ	PPAR-α
$_{ats}$C30$^{\gamma tr}$			On-Hexyl	CONHCH$_2$	830	1500
$_{atr}$C31$^{\gamma tr}$			On-Bn	CONHCH$_2$	990	2900
$_{atr}$C32$^{\gamma tr}$			OMe	CONHCH$_2$	2600	1800
$_{atr}$C33$^{\gamma ts}$			On-Pr	CONHCH$_2$	1800	1200
$_{atr}$C34$^{\gamma tr}$			On-Bu	CONHCH$_2$	1300	7700
$_{atr}$C35$^{\gamma tr}$			OMe	CONHCH$_2$	1300	52,000
$_{ats}$C36$^{\gamma tr}$			OMe	CONHCH$_2$	2300	120,000
$_{atr}$C37$^{\gamma tr}$			OMe	CONHCH$_2$	6200	820
$_{ats}$C38$^{\gamma tr}$			OMe	CONHCH$_2$	2200	4700
$_{atr}$C39$^{\gamma tr}$			On-Bu	CONHCH$_2$	650	28,000
$_{ats}$C40$^{\gamma ts}$			–		1400	7000
$_{atr}$C41$^{\gamma tr}$			OMe	CONHCH$_2$	3600	23,000
$_{ats}$C42$^{\gamma ts}$			OMe	CONHCH$_2$	2400	2900

(*Contd*)

Table 1. (*Contd*)

Sl. no.	R_1	R_2	R_3	B	Activity experimental (EC_{50}) μmol	
					PPAR-γ	PPAR-α
$_{\alpha ts}$C43$^{\gamma tr}$			OMe	CONHCH$_2$	1500	14,000
$_{\alpha tr}$C44$^{\gamma tr}$			OMe	CONHCH$_2$	1000	82,000
$_{\alpha tr}$C45$^{\gamma tr}$			OMe	CH$_2$NHCO	1000	100,000
$_{\alpha tr}$C46$^{\gamma tr}$			OMe	CONHCH$_2$	8600	100

γtr, Training set ligand for PPAR-γ; γts, Test set ligand for PPAR-γ; αtr, Training set ligand for PPAR-α; αts, Test set ligand for PPAR-α.
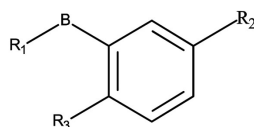
Figure 1. Common scaffold of phenylpropanoic acid derivatives.

many important genes such as TNF-α, leptin and adiponectin genes that affect energy metabolism[13]. Development of indigenously engineered remedies and cost-effective treatment become important as costs of synthetic drugs have increased now. Chemometric modelling is part of the enormous field of cheminformatic technology that has become popular among the pharmaceutical industries and researchers to designate the drug targets by means of computational tools. Structure–activity relationship (SAR), which is a statistical technique capable of analysing screening datasets and deriving predictive models of biologically interested activity, pharmacophore mapping and molecular docking are already proved vital chemometric techniques to optimize the drug candidates[14]. In the present study, these chemometric techniques are used to identify important features necessary for a compound to behave as an activator of PPAR-α and PPAR-γ receptors. PPAR-β/δ transactivation activity is not examined, since the lead compound shows very weak PPAR-β/δ transactivation activity.

Materials and methods

In the present work a set of 46 phenylpropanoic acid derivatives (Table 1) have been considered for quantitative structure activity relationship (QSAR)[15] study to design potential lead compounds for the treatment of non-insulin dependent diabetes mellitus (NIDDM)[6,16–19]. Figure 1 shows a common scaffold of phenylpropanoic acid

derivatives. QSAR is the study of the quantitative relationship between experimental activity of a set of compounds and their physico-chemical properties using statistical methods. Structure generation and minimization (using Chem3D ultra)[19], descriptor generation (using MOE)[20], and QSAR studies by means of multiple linear regressions (MLR)[21] and partial least squares (PLS)[21] analysis using SPSS[22] have been performed on the dataset. The interaction of the targets (PDB ID: 1FM9 and 1K7L)[23] with the most active compounds of the dataset has been observed via docking studies. A number of models have been generated and the best one selected based on high R^2 (correlation coefficient) and Q^2 (cross-validated correlation coefficient) along with low se (standard error of estimation) and sp (standard error of prediction), and good prediction of test set compound.

QSAR

This is a mathematical relationship between biological activity of a molecule and its chemical properties or chemical structure, and is a widespread approach for predicting biological activities in drug design. It is based on the assumption that the changes in molecular features of compounds can be correlated with variations in their physico-chemical/structural properties.

$$y = f(x), \tag{1}$$

where y is the biological activity and x is the chemical property or structural property.

The effective concentration data on phenylpropanoic acid derivatives (Table 1) are taken from the literature[6,16–19]. The dataset is divided into a training set for generating QSAR models and a test set for validating the quality of the models. The dataset is divided randomly keeping in

mind that biological activity of all compounds in the test set lies within maximum and minimum value range of biological activity of the training set compounds. The training set is considered for statistical analysis using MLR and PLS for model building methods[21]. QSAR models have been generated using the negative logarithm of half maximal effective concentration (pEC$_{50}$) values as the dependent variable and values of descriptors as independent variables. The independent variables (descriptors) are calculated from MOE and care is taken that the descriptors used for model generation do not have inter-correlation.

Docking studies

In order to check the binding interaction of the compound with PPAR-γ and PPAR-α receptors, docking is performed. The molecular docking tool, MOE[20], is used to study binding modes of the most active compound to the receptor molecule. Fine 3D structure with a resolution of 2.65 Å of nuclear receptors is retrieved from the Protein Data Bank (PDB)[23]. Hydrogen bonding interactions are the basic properties required for interaction between PPARs and phenylpropanoic acid receptors[24]. Docking algorithm is able to generate a large number of possible structures. Using this computational method, the preferred orientation of one molecule to a second when bound to each other to form a stable complex can be predicted[26].

Pharmacophore analysis

Pharmacophores are a group of features that are significant in a set of molecules. A pharmacophore indicates important groups necessary for the binding of ligands to the receptor binding pocket. The most common features are the presence of hydrophobic, aromatic ring, hydrogen bond acceptor and donor groups. The pharmacophore model can be used for virtual screening of ligands as well as for the *de novo* design of ligands to create completely novel candidate structures that conform to the requirements of a given pharmacophore. Pharmacophores are basically of three types, i.e. structure-based, ligand-based and both receptor- and ligand-based in which structure of the receptor, structure of the ligand and structure of both receptor and ligand is required respectively. For a pharmacophore model preparation, several molecules are aligned to find common features among a set of molecules which can be used for pharmacophore searching.

Validation

This is an important step of any SAR model to evaluate predictivity and robustness. In this study, QSAR models

are validated both internally and externally. The best models are validated internally using leave-one out (LOO) cross-validation method[26] followed by modified R^2 (r^2m) prediction[27]. During LOO cross-validation procedure, one compound is deleted randomly from the training set and the model is regenerated using the rest of the compound in each cycle; the new model generated is used to predict the activity of the deleted compound. Better predictive ability of the model is explained by high Q^2 (>0.5) and low se (<0.5) values. r^2m can be defined as the measure of the degree of deviation of the predicted activity from the observed ones, and the model may be considered if r^2m >0.5. External validation is performed by test set prediction which proves the true predictivity of the model that is judged best by statistical parameters such as R^2_{pred} (threshold value >0.5) and sp (threshold value <0.5)[28]. To judge the predictive ability of the model, r^2m (test) value is calculated. To validate the binding models for docking studies, protein ligand binding energy is considered within the range −5 to −15 kcal/mol (ref. 29).

Results and discussion

QSAR study

Phenylpropanoic acid derivatives, a potent PPAR-γ and PPAR-α activator and active at micro-molar level (Table 1), are considered in the present work for molecular modelling. The biological activity, expressed as pEC$_{50}$, is used as the dependent variable for modelling. A number of models have been prepared and the correlation coefficient is predicted using the MLR and PLS analysis method. Figure 2 shows a correlation plot between the experimental and predicted values of training and test sets for PPAR-γ and PPAR-α agonists.

PPAR-γ

MLR analysis: The best univariate model shows TPSA which is a polar surface area ($n_{Tr} = 34$, $R^2 = 0.182$) as the important feature contributing to the activity. Bivariate model is obtained with TPSA and SMR_VSA3 which is subdivided surface area ($R^2 = 0.322$) descriptors. The rsynth (feasibility of chemical structure) descriptors dominate in explaining the variation in activity as evidenced by the final QSAR equation.

$$pEC_{50} = 9.487(\pm 0.986) - 0.067(\pm 0.010)$$
$$\times\ TPSA + 0.011(\pm 0.002) \times SMR\_VSA3$$
$$- 1.457(\pm 0.390) \times rsynth + 5.970(\pm 2.108)$$
$$\times\ vsurf\_WP7 - 0.085(\pm 0.031) \times SMR\_VSA4.$$

(Model 1)

The values of independent variables used in Model 1 are not inter-correlated ($R < 0.5$; Table 2). The statistical

Table 2. Correlation between descriptors used for PPAR-γ model generation

Descriptors	TPSA	SMR_VSA3	rsynth	vsurf_WP7	SMR_VSA4
TPSA	1				
SMR_VSA3	0.171	1			
rsynth	−0.345*	0.271	1		
Vsurf_WP7	0.232	−0.012	−0.107	1	
SMR_VSA4	−0.359*	0.270	0.120	−0.053	1
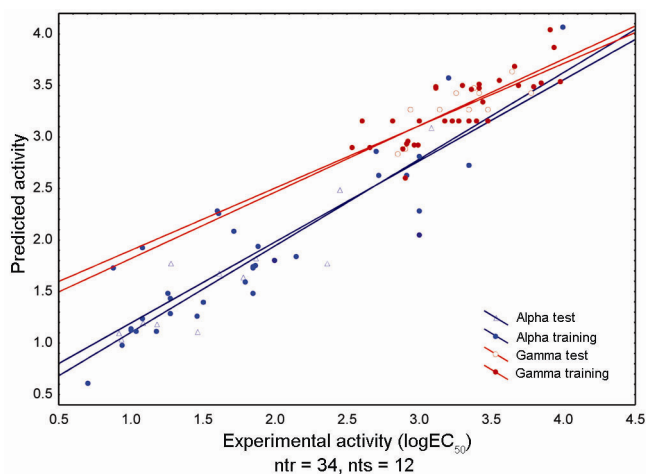
Figure 2. Correlation plot between experimental and predicted activities of training and test set molecules of PPAR-γ and PPAR-α agonists.

quality of the model is: $n_{Tr} = 34$, $R^2 = 0.651$, $Q^2 = 0.649$, se $= 0.251$, $r^2m = 0.638$, $n_{Ts} = 12$, $R^2_{pred} = 0.606$, sp $= 0.231$ and r^2m(test) $= 0.592$.

PLS analysis: For the training set PLS analysis gave a model with correlation coefficient 0.784, which is further validated by evaluating the test set model using standard expression.

$$pEC_{50} = 9.487 - 0.067 \times TPSA + 0.011$$

$$\times \ SMR\_VSA3 + 5.970 \times vsurf\_WP7$$

$$- 0.085 \times SMR\_VSA4 - 1.457 \times rsynth. \quad \text{(Model 2)}$$

The model obtained from the MLR studies reveals the importance of TPSA (polar surface area), rsynth (feasibility of the chemical structure), SMR_VSA3 and SMR_VSA4 (subdivided surface areas) and vsurf_Wp7 (polar volume) descriptors for the interaction of ligand and receptor. PLS studies (model 2) justify that the polar surface area, subdivided surface area, feasibility of chemical structure and polar volume are crucial for the binding interaction of ligand and receptor. In the model, the negative value of the coefficient for polar surface area suggests that increase in polar surface area is responsible for an increase in the activity of the molecule. The positive coefficient for the polar volume indicates that an increase

in the polar volume positively affects the activity of the ligands. The coefficient for the van der Waals surface area is also negative, which implies that increase in the surface area of ligands enhances the activity, and hence the interaction between ligand and receptor also increases. Furthermore, the smaller van der Waals surface area (SMR_VSA3) and polar volume of ligand have a positive impact on the effective concentration, as suggested by the positive value of the coefficient. Table 3 shows the values of the descriptors.

PPAR-α

MLR analysis: The best univariate model shows E_ang (number of angle bend potential energy) ($R^2 = 0.276$) as the important feature contributing to the activity. Bivariate model is obtained with E_ang and PEOE_VSA_FPOS which is fractional van der Walls surface area ($R^2 = 0.514$) descriptors.

$$pEC_{50} = 16.821(\pm 2.108) + 0.024(\pm 0.008)$$

$$\times E\_ang - 16.195(\pm 3.993) \times PEOE\_VSA\_FPOS$$

$$+ 20.438(\pm 3.458) \times BCUT\_SMR\_1$$

$$+ 24.772(\pm 6.620) \times GCUT\_PEOE\_2$$

$$- 0.693(\pm 0.191) \times opr\_leadlike. \quad \text{(Model 3)}$$

The values of independent variables used in model 3 are not intercorrelated ($R < 0.5$; Table 4). The statistical quality of the model is: $n_{Tr} = 34$, $R^2 = 0.784$, $Q^2 = 0.774$, se $= 0.426$, $r^2m = 0.782$, $n_{Ts} = 12$, $R^2_{pred} = 0.841$, sp $= 0.364$ and r^2m(test) $= 0.592$. Table 5 provides the predicted activity for training and test set compounds obtained from the model.

PLS analysis: The model reveals the importance of number of angle bend potential energy (E_ang), fractional positive van der Waals surface area (PEOE_VSA_FPOS), atomic contribution to molar refractivity (BCUT_SMR_1), atom count and bond count descriptor (opr_leadlike) and partial charge (GCUT_PEOE_2) descriptors, which are also crucial in the MLR analysis.

Table 3. Values of descriptors used in QSAR studies

Sl. no.	TPSA	rsynth	SMR_VSA3	SMR_VSA4	Vsurf_WP7	E_ang	PEOE_VSA_FPOS	BCUT_SMR1	GCUT_PEOE_2
C1	84.5	1	20.76	0	0	13.73	0.48	−0.39	0.07
C2	75.63	1	17.99	0	0	34.92	0.52	−0.39	0.08
C3	75.63	1	17.99	0	0	37.38	0.49	−0.39	0.05
C4	75.63	1	17.99	0	0	28.33	0.49	−0.39	0.05
C5	84.86	0.5	24.45	0	0	69.95	0.54	−0.36	0.08
C6	75.63	1	17.99	0	0	14.84	0.49	−0.39	0.05
C7	75.63	1	17.99	0	0	287.53	0.50	−0.39	0.04
C8	75.63	1	17.99	0	0	13.70	0.50	−0.39	0.04
C9	75.63	1	97.53	0	0	15.63	0.51	−0.39	0.05
C10	84.86	1	104.83	0	0	20.66	0.51	−0.48	0.08
C11	75.63	1	100.71	6.37	0	38.96	0.50	−0.40	0.06
C12	75.63	1	17.99	0	0	14.45	0.49	−0.39	0.05
C13	75.63	0.76	14.80	0	0	13.05	0.50	−0.40	0.05
C14	75.63	0.76	17.99	0	0	12.98	0.50	−0.40	0.05
C15	75.63	1	17.99	0	0	13.50	0.50	−0.39	0.05
C16	75.63	0.77	17.99	0	0	10.91	0.50	−0.40	0.04
C17	75.63	0.73	17.99	0	0	17.95	0.50	−0.40	0.05
C18	75.63	1	17.99	0	0	14.24	0.50	−0.39	0.04
C19	75.63	0.77	17.99	0	0	16.13	0.50	−0.40	0.04
C20	84.86	1	22.10	0	0	19.08	0.51	−0.45	0.06
C21	75.63	0.74	17.99	0	0	14.31	0.53	−0.35	0.04
C22	84.86	0.77	20.04	0	0	21.95	0.52	−0.33	0.03
C23	84.86	0.79	25.29	0	0	17.83	0.51	−0.47	0.06
C24	84.86	0.79	25.29	0	0.13	17.33	0.51	−0.43	0.06
C25	75.63	0.78	14.80	6.37	0	13.24	0.49	−0.41	0.07
C26	75.63	0.31	14.80	6.37	0	14.21	0.49	−0.41	0.07
C27	84.86	0.79	23.23	0	0	19.30	0.52	−0.41	0.06
C28	84.86	0.8	23.23	0	0	18.98	0.52	−0.40	0.06
C29	84.86	0.79	25.29	0	0	17.59	0.51	−0.47	0.05
C30	84.86	0.79	28.79	0	0	17.53	0.49	−0.47	0.05
C31	84.86	0.79	25.29	0	0	17.85	0.47	−0.47	0.05
C32	75.63	0.76	17.99	0	0	13.19	0.50	−0.40	0.05
C33	75.63	0.77	17.99	0	0	8.53	0.47	−0.39	0.05
C34	75.63	0.77	17.99	0	0	9.01	0.47	−0.39	0.05
C35	75.63	0.78	17.99	0	0	9.80	0.47	−0.39	0.05
C36	75.63	0.79	17.99	0	0	11.42	0.47	−0.39	0.05
C37	75.63	0.8	21.17	0	0	32.14	0.44	−0.41	0.08
C38	75.63	1	17.99	0	0	14.47	0.50	−0.39	0.04
C39	75.63	1	17.99	0	0	8.75	0.48	−0.39	0.04
C40	75.63	1	17.99	0	0	13.89	0.48	−0.39	0.04
C41	75.63	0.73	17.99	0	0	13.47	0.50	−0.40	0.05
C42	75.63	0.71	17.99	0	0	13.99	0.50	−0.40	0.05
C43	75.63	1	17.99	0	0	15.09	0.50	−0.39	0.04
C44	75.63	1	17.99	0	0	14.31	0.50	−0.39	0.04
C45	75.63	1	17.99	0	0	25.02	0.48	−0.39	0.04
C46	59.42	1	15.73	4.12	0	15.19	0.46	−0.34	0.00

$$pEC_{50} = 16.128 + 0.693 \times \text{opr\_leadlike}$$

$$+ 0.024 \times \text{E\_ang} - 16.194 \times \text{PEOE\_VSA\_FPOS}$$

$$+ 20.438 \times \text{BCUT\_SMR\_1} + 20.772$$

$$\times \text{GCUT\_PEOE\_2}. \quad \text{(Model 4)}$$

The model obtained from MLR studies reveals the importance of E_ang descriptors for the interaction of ligand and receptor. In the model, the positive value of coefficient for E_ang suggests that angle bend potential energy is responsible for an increase in the activity of the molecules. The negative coefficient for PEOE_VSA_FPOS indicates that an increase in the fractional positive van der Waals surface area positively affects the activity of the ligands. Furthermore, the negative value of coefficient for opr_leadlike descriptor indicates that higher values of atom and bond count will enhance the activity of the molecules while positive values of coefficient of BCUT_SMR_1 and GCUT_PEOE_2 indicate that molar refractivity of ligands and partial charge have a positive impact on the effective concentration of molecule respectively. The model obtained from the PLS studies (model 4)

Table 4. Correlation between descriptors used for PPAR-α model generation

Descriptors	E_ang	PEOE_VSA_FPOS	BCUT_SMR_1	GCUT_PEOE_2	opr_leadlike
E_ang	1				
PEOE_VSA_FPOS	0.317	1			
BCUT_SMR_1	0.106	−0.016	1		
GCUT_PEOE_2	0.460**	0.227	−0.455**	1	
opr_leadlike	−0.038	0.251	0.419*	−0.002	1

Table 5. Predicted and observed activities of training and test sets

Sl. no. (dataset)	Observed activity (logEC$_{50}$)		Predicted activity			
			PPAR-γ		PPAR-α	
	PPAR-γ	PPAR-α	MLR	PLS	MLR	PLS
C1	2.90	3	2.6	2.6	2.28	2.28
C2	3.48	3.34	3.27	3.27	2.73	2.73
C3	2.60	1.60	3.16	3.16	2.28	2.28
C4	3.40	3	3.16	3.16	2.05	2.05
C5	3.45	3.20	3.34	3.34	3.57	3.57
C6	3.48	1.85	3.27	3.27	1.74	1.74
C7	2.94	1.86	3.27	3.27	1.83	1.83
C8	3.23	1.51	3.16	3.16	1.4	1.4
C9	3.91	1.28	4.04	4.04	1.43	1.43
C10	3.30	1.18	3.5	3.5	1.12	1.12
C11	3.85	2.70	3.53	3.53	2.86	2.86
C12	3.48	1.85	3.16	3.16	1.73	1.73
C13	3.41	1.28	3.48	3.48	1.29	1.29
C14	3.41	1.28	3.43	3.43	1.78	1.78
C15	3.34	1	3.27	3.27	1.13	1.13
C16	3.78	1.04	3.43	3.43	1.11	1.11
C17	3.56	1.80	3.55	3.55	1.59	1.59
C18	3.28	1	3.16	3.16	1.14	1.14
C19	3.69	1.08	3.5	3.5	1.24	1.24
C20	2.90	0.93	2.89	2.89	1.04	1.04
C21	3.98	2	3.54	3.54	1.8	1.8
C22	3.64	1.08	3.63	3.63	1.2	1.2
C23	2.91	0.94	2.93	2.93	0.98	0.98
C24	3.66	1.61	3.68	3.68	1.68	1.68
C25	2.53	0.88	2.9	2.9	1.73	1.73
C26	2.85	1.87	2.84	2.84	1.75	1.75
C27	2.66	1.72	2.9	2.9	2.09	2.09
C28	2.89	1.78	2.89	2.89	1.64	1.64
C29	2.96	0.70	2.92	2.92	0.61	0.61
C30	2.92	1.18	2.96	2.96	1.19	1.19
C31	3.00	1.46	2.92	2.92	1.26	1.26
C32	3.41	1.26	3.51	3.51	1.48	1.48
C33	3.26	1.08	3.43	3.43	1.92	1.92
C34	3.11	1.89	3.49	3.49	1.94	1.94
C35	3.11	2.72	3.48	3.48	2.63	2.63
C36	3.36	3.08	3.46	3.46	3.09	3.09
C37	3.79	4	3.49	3.49	4.06	4.06
C38	3.34	0.91	3.16	3.16	1.1	1.1
C39	2.81	1.61	3.16	3.16	2.26	2.26
C40	3.15	2.45	3.27	3.27	2.49	2.49
C41	3.56	1.85	3.55	3.55	1.48	1.48
C42	3.38	2.36	3.47	3.47	1.78	1.78
C43	3.18	1.46	3.16	3.16	1.11	1.11
C44	3.00	2.15	3.16	3.16	1.84	1.84
C45	3.00	2.91	3.16	3.16	2.63	2.63
C46	3.93	3	3.87	3.87	2.81	2.81

reveals the importance of all the above descriptors for interaction of ligand and receptor. From the PLS and MLR studies of the dataset similar results are obtained, which justify that smaller van der Waals surface area and atom and bond count favour the activity of molecules while increase in angle bend potential energy, molar refractivity and partial charge can contribute in enhancing the activity.

Validation of QSAR models

Internal validation: The activity of the training compounds is predicted using LOO cross-validation method in QSAR studies of both α- and γ-subtypes. Q^2 is found to be 0.649 and 0.774 for model 1 and model 3 respectively, whereas SE 0.251 and 0.426 respectively. The r^2m value for the respective models is 0.638 and 0.782. The statistical results (Q^2 and $r^2m > 0.5$) of the both studies show that the selected models are robust.

External/test set validation: The activity of the test compounds is predicted in QSAR studies. Correlation (R) between observed and estimated activities of test compounds is 0.777 and 0.917 for model 1 and model 3 respectively. In QSAR study, R^2_{pred} is 0.606 with sp = 0.231, and it is 0.841 with sp = 0.364 for model 1 and model 3 respectively. For better determination of the predictive abilities of the models, the value of r^2m (test) is also calculated. To determine whether the predicted activity values are close to the corresponding observed ones the value of r^2m (test) is predicted, as the high value of R^2_{pred} may not always indicate a low residual between the observed and predicted activity data. In QSAR study, the r^2m (test) value is found to be 0.592 and 0.837 for model 1 and model 3 respectively. It is observed that all models in the present study show high R^2_{pred} (>0.5) and r^2m (test) (>0.5) values, which explains the superiority of these models.

Docking studies

PPAR-γ: To further observe the interactions, docking study is performed and the most active compound is docked for its binding interactions with the active site of (PDB 1FM9) protein. Hydrogen bond (HB) is the most widely used parameter for evaluation of docking results, as it is an influential parameter in the activity of the drug compound. The number of HB interactions is observed via docking. The docked conformation of compounds reveals that the compound interacts with the binding pocket residues (GLN286, ASP243, LYS232, HIS449, LYS232, GLU295 etc.) of targeted proteins through several favourable interactions, including HB donor and acceptor with residue GLU295 and LYS232 respectively (Figure 3). The N atom present in the ligand behaves as a HB donor and interacts with the GLU295 residue, while the

LYS232 residue interacts with terminal hydroxyl group which behaves like a HB acceptor. The HB feature is also observed in the QSAR studies.

PPAR-α: To further support the QSAR results, the most active compound is docked for its binding interactions in the active site of (PDB 1K7L) protein. Group R is at the entrance of the pocket and is surrounded by THR279, SER280, LEU321, PHE273, LYS257, GLU282 residues of targeted proteins (Figure 4). The presence of a hydrogen bonding feature is found crucial for the interaction, which is also observed in the QSAR studies. Residue LYS257 and GLU282 are found to show favourable interaction, including HB donor and HB acceptor with the most active conformer. The GLU282 residue in receptor binding pocket interacts with the terminal hydroxyl group of most active conformer while the LYS257 residue
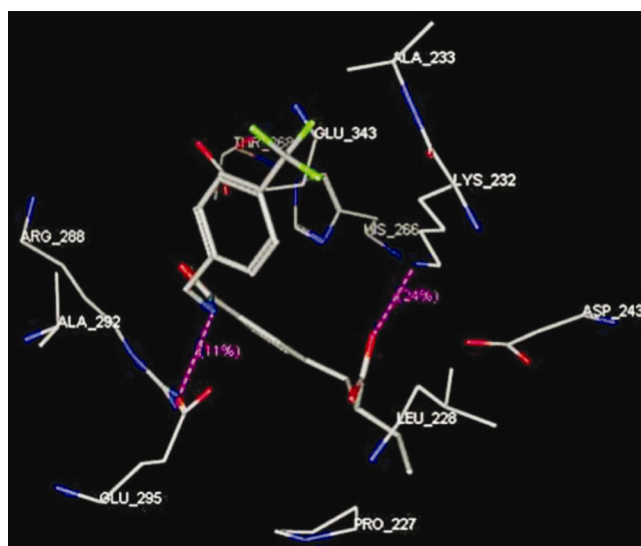

Figure 3. Interaction between receptor pocket of PPAR-γ and most active ligand.
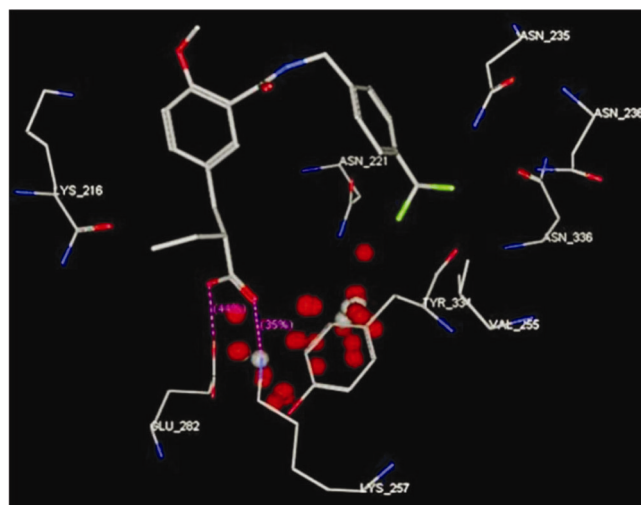

Figure 4. Interaction between receptor pocket of PPAR-α and most active ligand.

shows HB interactions with the oxygen atom of carboxylic group of conformer.

The docking study is performed on the receptor molecules with the most active compound for both PPAR-α and PPAR-γ reveals the importance of HB donor and acceptor character which is also observed in the QSAR studies, and hence signifying the validity of the model.

Validation of docking studies

The docking analysis between the receptor molecules PPAR-α and PPAR-γ and the most active compound gives binding energy of –9.36 and –9.66 kcal/mol respectively, for the best docking pose. This is further compared with the binding energy for the interaction of available phenylpropanoic acid-based drug, ragaglitazar (zinc ID: ZINC01481830)[30] with the α and γ subtypes, which is found to be –9.48 and –9.76 kcal/mol respectively. The obtained binding energy for the interaction falls within the threshold range (between –5 and –15 kcal/mol), which implies that the binding models for both subtypes are acceptable. Also, the above comparative analysis implies that the binding energy of interaction for the most active ligand with the respective receptors is close to that of the available drug and hence verifies the validity of the binding model.

Pharmacophore development

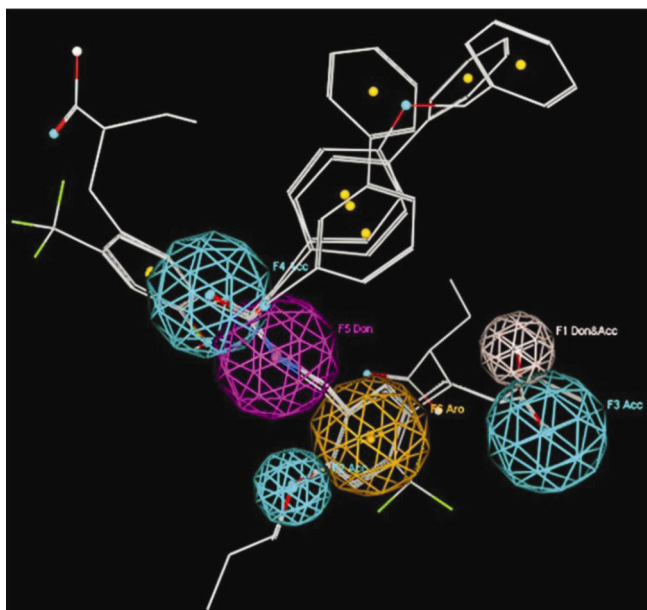Using the Pharmacophore Query Editor tool in MOE, the Pharmacophore model has been developed. This

Figure 5. Pharmacophore model of PPAR-γ full agonists (pink: F1: Don & Acc, cyan : F2, F3, F4 : Acc, magenta : F5: Don, yellow: F6: Aro).

application generates a pharmacophore query by a computerized representation of the binding interactions with a particular active site using a hypothesized pharmacophore. A pharmacophore query in MOE is a set of features created typically from ligand annotation points. The pharmacophore model is developed by alligning the five most active agonists (Figure 5). It outlines six important pharmacophore features that are observed in the most active agonists: four polar atoms and functional groups capable of performing HB (F1, F2, F3, F4 and F5), and aromatic structural elements (F6). The hydrophobic/aromatic features stabilize the positions of the hydrophilic ones; the terminal F1 donor and acceptor feature can form HB with the oxygen atom, which is also confirmed by docking interactions, and F6 features are directed inside and contribute additionally to the stabilization of the ligand pose into the pocket. F2, F3 and F4 are the acceptor features contributing to the hydrogen bond interaction as evidenced by docking interactions of LYS257 residue with F2 feature. Similar functions can be assigned to the F5 feature being the donor group of interacting with the GLU295 residue, either directly or through water molecules. Thus the hydroxyl group is found to be crucial for the binding of ligand to the receptor, as evident from the pharmacophore and docking studies.

Conclusion

From the MLR and PLS analysis studies of the compounds, we have obtained descriptors which contribute significantly to both subtypes. Features like polar surface area, van der Waals surface area and polar volume are important for better prediction of activity and binding interaction of ligands with the PPAR-γ receptor. Furthermore, features like angle bend potential energy, refractivity of ligands, partial charge and fractional positive van der Waals surface area are crucial for binding of ligands to the PPAR-α receptor. Docking studies between receptor and most active compound supports the fact that hydrogen bonding interaction is important for binding interaction of ligands to the receptors, which is confirmed by the pharmacophore studies. Thus, the hydroxyl group is found to be significant for the binding of ligand to the receptor, as justified by the pharmacophore modelling and docking studies.

1. Zimmet, P., Globalization, coca-colonization and the chronic disease epidemic: can the dooms day scenario be averted. *J. Intern. Med.*, 2001, **247**, 301–310.
2. Amos, A., McCarty, D. and Zimmet, P., The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabetic Med.*, 1987, **14**, S1–S85.
3. King, H., Aubert, R. and Herman, W., Global burden of diabetes. 1995–2025. Prevalence, numerical estimates and projections. *Diabetes Care*, 1998, **21**, 1414–1431.
4. WHO, Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classifications of

diabetes mellitus. Department of Non-communicable Disease Surveillance, World Health Organization, Geneva, 1999; http://whqlibdoc.who.int/hq/1999/who_ncd_ncs_99.2.pdf

5. Alberti, K. and Zimmet, P., Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic. Med.*, 1998, **7**, 539–553.

6. Makishima *et al.*, Design, synthesis, and evaluation of a novel series of α-substituted phenylpropanoic acid derivatives as human peroxisome proliferator-activated receptor (PPAR) α/δ dual agonists for the treatment of metabolic syndrome. *Bioorg. Med. Chem.*, 2006, **14**, 8405–8414.

7. Oyama, T. *et al.*, Adaptability and selectivity of human peroxisome proliferator-activated receptor (PPAR) pan agonists revealed from crystal structures. *Acta Crystallogr. D*, 2009, **65**, 786–795.

8. Klug, J., Nuclear Receptor Nomenclature Committee. *Cell*, 1999, **97**, 161–163.

9. Willson, T., Brown, P., Sternbach, D. and Henke, B. R., The PPARs: from orphan receptors to drug discovery. *J. Med. Chem.*, 2000, **43**, 527–550.

10. Banner, C., Gottlicher, M., Widmark, E., Sjovall, J., Rafter, J. and Gustafsson, J., A systematic analytical chemistry/cell assay approach to isolate activators of orphan nuclear receptors from biological extracts: characterization of peroxisome proliferator-activated receptor activators in plasma. *Lipid Res.*, 1993, **34**, 1583–1593.

11. Mukherjee, R., Jow, L., Noonan, D. and McDonnell, D. P., PPARA antibody: peroxisome proliferator-activated receptor alpha, mRNA [monoclonal antibody] steroid. *Biochem. Mol. Biol.*, 1994, **51**, 157–166.

12. Okuno, A. *et al.*, Troglitazone increases the number of small adipocytes without the change of white adipose tissue mass in obese Zucker rats. *J. Clin. Invest.*, 1988, **101**, 1354–1361.

13. Okuno, A. *et al.*, Drug discovery: practices, processes, and perspectives. *J. Clin. Invest.*, 1998, **101**, 1354.

14. Esposito, E., Hopfinger, A. and Madura, J., Methods for applying the quantitative structure–activity relationship paradigm. In *Chemoinformatics Concepts, Methods and Tools for Drug Discovery* (ed. Bajorath J.), Humana Press Inc., Totowa, NJ, pp. 131–213.

15. Bohm, M., Sturzebecher, J. and Klebe, G., Three-dimensional quantitative structure–activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin and factor-Xa. *J. Med. Chem.*, 1999, **42**, 458–477.

16. Miyachi, H. *et al.*, Design, synthesis and evaluation of substituted phenylpropanoic acid derivatives as peroxisome proliferator-activated receptor (PPAR) activators: novel human PPAR-γ selective activators. *Bioorg. Med. Chem. Lett.*, 2002, **12**, 77–80.

17. Ban, S. *et al.*, Structure-based design, synthesis, and nonalcoholic steatohepatitis (NASH) – preventive effect of phenylpropanoic acid peroxisome proliferator-activated receptor (PPAR) α-selective agonists. *Bioorg. Med. Chem.*, 2011, **19**, 3183–3191.

18. Kasuga, J. *et al.*, Design, synthesis, and evaluation of potent, structurally novel peroxisome proliferator-activated receptor (PPAR) δ-selective agonists. *Bioorg. Med. Chem.*, 2007, **15**, 5177–5190.

19. Chemdraw ultra package version 7.0 http://www.cambridgesoft.com/.

20. MOE, Molecular operating environment (MOE). Chemical Computing Group Inc, Montreal, Canada, Trial version, 2013.

21. Jain, S., Ghate, M., Bhadoriya, K., Bari, S., Chaudhari, A. and Borse, J., 2D, 3D-QSAR and docking studies of 1,2,3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. *Org. Med. Chem. Lett.*, 2012, **2**, 22.

22. IBM, IBM SPSS Statistics for Windows, Version 20.0, IBM Corp, NY, released 2011.

23. Protein Data Bank; http://www.rcsb.org/pdb/home/home.do

24. Shen, X., Jiang, H., Luo, X., Chen, K., Shen, J., Cheng, F. and Xu, X., Interaction models of a series of oxadiazole-substituted – isopropoxy phenylpropanoic acids against PPAR and PPAR: molecular modeling and comparative molecular similarity indices analysis studies. *Protein Pept. Lett.*, 2009, **16**(2), 150–162.

25. Julio, C. *et al.*, Study of differences in the VEGFR2 inhibitory activities between semaxanib and SU5205 using 3D-QSAR, docking, and molecular dynamics simulations. *J. Mol. Graph. Model.*, 2011, **32**, 39–48.

26. Kubinyi, H., Hamprecht, F. A. and Mietzner, T., Three-dimensional quantitative similarity–activity relationships (3D QSiAR) from SEAL similarity matrices. *J. Med. Chem.*, 1998, **41**, 2553–2564.

27. Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N. and Kabir, H., Comparative studies on some metrics for external validation of QSPR models. *J. Chem. Inf. Model.*, 2012, **52**, 396–408.

28. Roy, P. P., Paul, S., Mitra, I. and Roy, K., On two novel parameters for validation of predictive QSAR models. *Molecules*, 2009, **14**, 1660–1701.

29. Mujwar, S. and Pardasani, K. R., Prediction of riboswitch as a potential drug target for infectious diseases: an In silico case study of anthrax. *J. Med. Imag. Health Inf.*, 2015, **5**, 7–16.

30. Zinc database, University of California, San Francisco, http://zinc.docking.org/ (accessed on 1 March 2016).