The Application of Molecular Modelling in the Safety Assessment of Chemicals: A Case Study on Ligand-Dependent PPARγ Dysregulation

Merilin Al Sharif[1], Ivanka Tsakovska[1], Ilza Pajeva[1], Petko Alov[1], Elena Fioravanzo[2], Arianna Bassan[2], Simona Kovarich[2], Chihae Yang[3], Aleksandra Mostrag-Szlichtyng[3], Vessela Vitcheva[3], Andrew P. Worth[4], Andrea-N. Richarz[5], Mark T. D. Cronin[5]

[1] Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences, Acad. G. Bonchev Str., bl. 105, 1113 Sofia, Bulgaria

[2] S-IN Soluzioni Informatiche SRL, Via Ferrari 14, Vicenza, 36100, Italy

[3] Altamira LLC, 1455 Candlewood Drive, Columbus OH, 43235, USA

[4] Systems Toxicology Unit, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra, Varese, Italy

[5] School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, England

Abstract

The aim of this paper was to provide a proof of concept demonstrating that molecular modelling methodologies can be employed as a part of an integrated strategy to support toxicity prediction consistent with the mode of action/adverse outcome pathway (MoA/AOP) framework. To illustrate the role of molecular modelling in predictive toxicology, a case study was undertaken in which molecular modelling methodologies were employed to predict the activation of the peroxisome proliferator-activated nuclear receptor γ (PPARγ) as a potential molecular initiating event (MIE) for liver steatosis. A stepwise procedure combining different *in silico* approaches (virtual screening based on docking and pharmacophore filtering, and molecular field analysis) was developed to screen for PPARγ full agonists and to predict their transactivation activity (EC_{50}). The performance metrics of the classification model to predict PPARγ full agonists were balanced accuracy = 81%, sensitivity = 85% and specificity = 76%. The 3D QSAR model developed to predict EC_{50} of PPARγ full agonists had the following statistical parameters: $q^2_{cv} = 0.610$, $N_{opt} = 7$, $SEP_{cv} = 0.505$, $r^2_{pr} = 0.552$. To support the linkage of PPARγ agonism predictions to prosteatotic potential, molecular modelling was combined with independently performed mechanistic mining of available *in vivo* toxicity data followed by ToxPrint chemotypes analysis. The approaches investigated demonstrated a potential to predict the MIE, to facilitate the process of MoA/AOP elaboration, to increase the scientific confidence in AOP, and to become a basis for 3D chemotype development.

Graphical abstract

Highlights

Abbreviations: ADMET, absorption, distribution, metabolism, excretion and toxicity; AOP, adverse outcome pathway; BHK21 ATCC CCL10, baby hamster kidney cell line from the American Type Culture Collection; CAS, chemical abstracts service; CoMSIA, comparative molecular similarity indices analysis; COS-1 and COS-7, CV-1 in origin, with SV40 genetic material; COSMOS, Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of COSMetics to Optimise Safety; CSRML, Chemical Subgraphs and Reactions Markup Language; CV-1, CV-1 (simian - *Cercopithecus aethiops*) or normal African green monkey kidney Fibroblast Cells; DUD-E database, a database of useful decoys: enhanced; EC_{50}, effective concentration (the concentration of a drug that gives half-maximal response); FDA CFSAN's CERES, Chemical Evaluation and Risk Estimation System at the U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition; H12, helix 12; HEK293, human embryonic kidney 293 cell line; HepG2, human liver hepatocellular carcinoma cell line; Huh-7, human liver hepatocellular carcinoma cell line; IC_{50}, half maximal inhibitory concentration; LD, lipid droplet; LOO, leave-one-out cross-validation; %max, percent efficacy in relation to the maximum efficacy of a reference compound; MIE, molecular initiating event; MM, molecular modelling; MoA, mode of action; NAFLD, non-alcoholic fatty liver disease; oRepeatTox DB, oral repeated dose toxicity database; PDB, protein data bank; PLS, partial least squares analysis; PPARα, peroxisome proliferator-activated receptor α; PPARγ, peroxisome proliferator-activated receptor γ; (Q)SAR, (quantitative) structure-activity relationship; SEE, standard error of estimate; SEP_{cv}, cross-validated standard error of prediction; SEURAT-1, Safety Evaluation Ultimately Replacing Animal Testing; TG, triglycerides; VS, virtual screening.

1. Introduction

The traditional approach for the safety assessment of chemicals based on the observation of apical endpoints in animals is moving towards a new predictive paradigm based on upstream biological events that are determinants of the apical adverse outcome (OECD, 2013). This paradigm shift opens the door to a new toxicity-testing framework that evaluates biologically significant perturbations mediating key toxicity pathways by using innovative computational toxicology methods and a comprehensive array of *in vitro* tests (Krewski et al., 2010). In this context, the European FP7 Research Initiative SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing; http://www.seurat-1.eu/) adopted the mode of action/adverse outcome pathway (MoA/AOP) framework as a means to understand human adverse health effects caused by repeated exposure to chemicals and to develop integrated tools for predictive toxicology (including *in silico* and experimental *in vitro* models), toward the replacement of *in vivo* chronic toxicity testing. One of the key elements of SEURAT-1 is the application of mechanistic knowledge acquired through the definition and description of specific AOPs to develop *in silico* tools for toxicity prediction. The FP7 COSMOS Project (Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety; http://www.cosmostox.eu/), which is part of the SEURAT-1 Cluster, has explored and exploited multiple *in silico* approaches including (Quantitative) Structure-Activity Relationships ((Q)SARs), structural alerts and molecular modelling (MM). While the first two computational methodologies are now widely used in the field of toxicology and safety assessment, and even encouraged by different regulatory frameworks, the use of MM in predictive toxicology is not yet a standard procedure.

MM methodologies have been used extensively in drug design for more than 30 years to provide cost-effective virtual analysis prior to synthesis. Such approaches aim to direct drug design efforts to synthesise highly selective and potent small molecules that bind to a target biomolecule. The latter is often a key protein involved in a particular metabolic or signaling pathway that is specific to a disease condition or pathology or to the infectivity and survival of a microbial

pathogen. Therefore the array of intermolecular interactions that trigger dysregulation (inhibition or activation) of such a key biological target comprises the mechanistic basis behind the observed therapeutic effect. Since such interactions could initiate adverse effects and could play crucial role in the mechanisms of toxicity, the use of MM techniques can also be envisaged in the risk assessment framework. However, the use of MM tools, developed for drug discovery, to predict toxicity of chemicals requires that the methods are adapted and interpreted taking into account the differences (see Table 1) between the drug discovery and risk assessment frameworks (Rabinowitz et al., 2008).

Table 1. Comparative analysis of the uses of MM approaches in drug discovery and risk assessment

The present study provides a proof of concept that MM methodologies can be employed as part of an integrated strategy to support target organ toxicity prediction in the MoA/AOP framework. In particular, we present the use of MM to predict potential binding to, and potential activation of the peroxisome proliferator-activated nuclear receptor γ (PPARγ). The challenge of this study lays in the fact that *in silico* modelling of PPARγ ligands is traditionally directed to rational drug design and improvement of pharmacological over adverse effects (Al-Najjar et al., 2011; Carrieri et al., 2013; Dixit and Saxena, 2008; Liao et al., 2004; Lu et al., 2006; Rücker et al., 2006; Shah et al., 2008; Sundriyal et al., 2009; Vidović et al., 2011), while risk assessment issues are poorly addressed (Vedani et al., 2007).

PPARγ is one of the important nuclear hormone receptors that contribute to excessive accumulation of triglycerides in hepatocytes (liver steatosis or fatty liver) (Landesmann et al., 2012). Our recent review (Al Sharif et al., 2014) has compiled the existing knowledge for PPARγ and non-alcoholic fatty liver disease (NAFLD), in particular liver steatosis, to develop a prosteatotic AOP, integrating the ligand-induced activation of the PPARγ as a MIE. Within the proposed AOP, the MIE induces up-regulation of target genes for lipid transport/binding proteins, fatty acid and triglyceride synthesising enzymes, and lipid droplet-associated proteins (LD proteins). Thus, the

AOP covers the whole cascade of key molecular events and the subsequent cytological and histopathological manifestations of liver steatosis, namely increased number or size of LDs, ectopic TG deposition in hepatic instead of adipose tissue, and hepatomegaly.

Therefore the AOP developed improves the mechanistic understanding of the role of PPARγ activation as a MIE in the development of liver steatosis. In addition it underlines the key events and the data gaps for further *in vitro*/*in silico* exploration and maps the potential assays that could be proposed for development and/or modification to meet the needs of the *in vitro* safety assessment of toxicants (OECD, 2013; Patlewicz et al., 2015).

As a logical further step the present study focused on *in silico* modelling of this particular MIE as a reliable early signal for hazard identification. Its nature – ligand-receptor interactions - determined the choice of MM approaches to be applied. Taking into consideration that the prosteatotic genomic activity of PPARγ is specifically triggered by full agonists, but not by partial agonists (Chigurupati et al., 2015), our modelling strategy included an initial analysis of the available data for full agonists (e.g. binding mode, efficacy range) and subsequent development of pharmacophore-based virtual screening (VS) procedure and 3D QSAR models.

In order to strengthen the conclusions and to further confirm the prosteatotic activity of PPARγ, the results from the MM study were challenged with those of an independently applied, alternative approach, involving mechanistic data mining of available *in vivo* toxicity data, followed by ToxPrint chemotype analysis of chemical compounds associated with particular phenotypic effects (Mostrag-Szlichtyng et al., 2014). A chemotype is defined as a structural fragment encoded for connectivity and, where required, for physicochemical and electronic properties of atoms, bonds, fragments, and even a whole molecule (Yang et al., 2015).

The chemotypes approach could be considered as a ligand based screening that relies on empirical prediction of the pathological condition based on the presence of particular substructures. Thus, while the MM study allowed for mechanistically justified prediction of the first step of the AOP, namely the MIE, the chemotype analysis predicts the adverse effect directly (in particular the

histological and physiological manifestations of the pathology) disregarding the mechanistical basis of the MIE. The combinded application of both approaches proves to be helpful both for predictive purposes and in the analysis and understanding of the molecular mechanisms of the prosteatotic AOP.

2. Materials and methods

2.1. PPARγ ligands dataset used in the molecular modelling study

2.1.1. PPARγ ligands and experimental data

Structural and experimental data for 439 PPARγ ligands were collected from the Protein Data Bank (PDB) (www.rcsb.org, Berman et al., 2000) and from the literature. The dataset is publicly available at http://biomed.bas.bg/qsarmm/. The dataset includes the following experimental data; (i) binding affinity, IC_{50} (measured in *in vitro* binding assays – radioligand binding assay or fluorescence polarisation binding assay); (ii) potency, EC_{50} (measured in a cell based luciferase transcriptional reporter gene assay, evaluating the effect of the ligand-dependent PPARγ activation on the expression of its target reporter protein); (iii) relative efficacy, %max (percent response in relation to the maximum response of a reference compound in the cell based transactivation assay). Experimental data measured in different human and animal cell lines were collected. The distribution of the ligands according to the cell line and relative efficacy toward PPARγ is shown in Figure 1 and summarised in Table S.1. of the Supplementary Material.

Figure 1. PPARγ dataset: distribution of the ligands according to the cell line and their relative efficacy toward PPARγ. Numbers 1 – 7 indicate the different species and cell lines: 1 – hamster/kidney (BHK21 ATCC CCL10), 2-4 – monkey/kidney (COS-1, COS-7, CV-1, respectively), 5 – human/kidney (HEK293), 6 and 7 – human/liver (HepG2, Huh-7, respectively)

As explained in greater detail in subsequent sections, a subset of 170 PPARγ full agonists was extracted from the initial dataset since these compounds met the data requirements for

modelling purposes. The selected ligands constituted a structurally diverse dataset of PPARγ full agonists with relative efficacy ≥ 70% and/or PDB ligands with structural features fitting the recently developed PPARγ full agonists' pharmacophore (Tsakovska et al., 2014). Detailed information regarding the ligands retrieved from PDB and used for modelling is provided in Table S.2 of the Supplementary Material. The activity data were converted to micromolar concentrations and transformed to pEC_{50} (the negative logarithm of the EC_{50}) values. For a small subset of reference compounds (farglitazar, rosiglitazone and pioglitazone) that have been tested on human and animal cell lines by different research groups, preference was given to human over animal data and the mean pEC_{50} values were taken when necessary.

2.1.2. Chemical structures preparation

The subset of 170 ligands utilised in this study comprised different homologous series retrieved from 15 literature sources as noted in Table 2. Among them, eight series of chemicals contain a PPARγ ligand with a crystal structure deposited in the PDB, one contains a PPARα ligand, and six do not contain resolved PDB structures (Table 2). The structures of the compounds within the series that include a PPARγ ligand deposited in PDB were built through modification of the PDB ligand. The structures of the other compounds were built either directly or from structurally similar PDB ligands.

Table 2. PPARγ ligands selected for modelling: research group, molecular scaffold, numbers and PDB identifiers

The stereochemistry in the ligands was fixed in accordance with the reported stereoisomery. When racemic mixtures were tested, the S stereoisomer was used for modelling as a commonly accepted active form (Rücker et al., 2006; Shah et al., 2008). The protonation state of the ligands that could not be deduced from the PDB complexes was also explored and the forms of the structures were selected corresponding to their protonation state at pH = 7.4 as calculated in

ACD/Labs Percepta suite 2015 (ACD Inc.). When two forms were reported to coexist with equal or similar protonation state percentages, both structures were considered. Finally the structures were minimised with the MMFF94s force field including electrostatics. the preparation/ modification and optimisation of chemical structures were performed using the MM platform MOE v. 2014.0901 (CCG Inc.).

2.2. Molecular modelling of PPARγ full agonists

2.2.1. Protein preparation and docking in the PPARγ binding site

The receptor's ligand binding domain (X-ray structure of PPARγ with rosiglitazone, PDB ID 1FM6) was initially prepared using the Protonate3D application in MOE. This application assigns the hydrogens following the optimal free energy proton geometry and ionisation states of titratable protein groups using the Generalized Born electrostatics model. The physiologically relevant parameters were set during the minimisation: temperature 310°K; pH = 7.4; ion concentration: 0.152 mol/L. The ligands (see structures preparation described in the section 2.1.2.) were docked into the binding site of the prepared protein structure. The London dG scoring function, without subsequent refinement, was applied to estimate free energy of binding and to score the poses of the docked ligands accordingly. The selected scoring function combines terms for van der Waals interactions, hydrogen-bonding, hydrophobic effects and deformation effects associated with conformational entropy. Coefficients for these terms have been fitted from over 400 X-ray crystal structures of protein-ligand complexes with available experimental pK_i data (MOE). The highly scored poses of each ligand with a negative value of the scoring function only were kept. They were used in the development of the VS protocol to predict PPARγ full agonists (section 3.2.1).

2.2.2. 3D QSAR

2.2.2.1. Alignment of structures and calculation of fields

The ligands were aligned based on their docking poses in the PPARγ ligand binding domain using the VS protocol developed within this study (section 3.2.1). The final conformer for each

ligand was selected from the output set of up to 10 best poses retained after its docking according to two criteria: (i) visual inspection relative to the corresponding structurally similar template - the PDB ligand used as scaffold in structure generation or the ligand UNT from 3IA6 PDB complex as one of the most active agonists with pEC_{50} = 7.886 and 103% relative efficacy and possessing structural features from among most of the structures that are typical features of the agonists (Casimiro-Garcia et al, 2009; Mahindroo et al., 2005); (ii) the value of the docking score (the smallest negative scores were preferred). The ligands extracted from the PDB complexes were maintained in their experimental bioactive conformations.

The whole set was re-aligned by the "Fit Atoms" procedure in MM software suite SYBYL-X v. 2.1 (Certara USA, Inc.) using the 4 essential pharmacophoric points of the PPARγ pharmacophore model (Tsakovska et al., 2014). The ligand UNT (3IA6 PDB complex) was chosen as a template. The set of aligned structures was subjected to 3D QSAR modelling using the CoMSIA (Comparative Molecular Similarity Indices Analysis) approach within SYBYL. For this purpose the electrostatic, steric, hydrogen bond donor, hydrogen bond acceptor, and hydrophobic fields were calculated using the default CoMSIA settings.

2.2.2.2. Model development and validation

The structures were split into training and test sets and for the training set CoMSIA 3D QSAR models were to correlate ligands' potency (pEC_{50}) with similarity indices, related to field properties of each molecule, namely the steric, electrostatic, hydrophobic and hydrogen-bonding properties. The Partial Least Squares statistical method (PLS) was used for the CoMSIA modelling (Klebe, 1998). The cross-validation analysis was performed using the Leave-One-Out (LOO) procedure to evaluate the internal predictivity of the models. The statistical characteristics of the models were described by the cross-validated correlation coefficient q^2_{cv}, the optimal number of PLS components, N_{opt}, and the cross-validated standard error of prediction, SEP_{cv}. The non-cross-validated model (characterised by the correlation coefficient, r^2, standard error of estimate, SEE, and the F-value) was obtained for the best cross-validated model with N_{opt}. The sensitivity of the

model to chance correlations was investigated by Y-randomisation test (ten randomisations) and by progressive scrambling (maximum: 20 bins, minimum: two bins and critical point: 0.85). For the purposes of the external validation the pEC_{50} values of a predefined test set of full agonists were calculated and the predictive r^2 (r^2_{pr}) was considered as a means to evaluate the model's external predictivity. Identification of outliers was performed by means of two different criteria: (i) applicability domain outliers using the "extent of extrapolation" approach (Tropsha et al., 2003; Netzeva et al., 2005) as implemented in Enalos domain leverage node (Melagraki et al., 2009) in the KNIME analytics platform (Berthold et al., 2007); (ii) response outliers, using analysis of the residuals..

2.3. Chemotype analysis

2.3.1. In Vivo Toxicity Data Mining used in the chemotype analysis

The oRepeatTox DB, part of the COSMOS database (publically available at: http://cosmosdb.cosmostox.eu) developed within the COSMOS Project was used in the chemotype analysis. It includes *in vivo* oral repeated dose toxicity data for approximately 230 cosmetics-related chemicals. The database was built through consolidation of existing databases as well as harvesting new data.

The COSMOS oRepeatTox DB chronic, subchronic and subacute (≥ 28 days) studies with rat, mouse and dog were mined for the phenotypic effects observed in liver. Mapping the morphological changes at various sites and combined phenotypic effects specifically related to steatosis, steatohepatitis or fibrosis onto the chemicals, led to the identification of 59 compounds associated with lipid deposition, liver fatty changes, cytoplasmic vacuolisation, cellular infiltration and/or inflammation in hepatocytes, ultimately leading to fibrosis (Mostrag-Szlichtyng et al., 2014). These structures were used in the subsequent chemotypes analysis.

2.3.2. ToxPrint chemotypes

The chemotype approach for chemical representation is supported by the open-source XML-based query language, Chemical Subgraphs and Reactions Markup Language (CSRML). The chemotypes coded in CSRML were applied via a software tool, ChemoTyper, jointly developed by Altamira LLC and Molecular Networks GmbH for public use under a contract from US FDA's Center for Food Safety and Nutrition (CFSAN) (available at www.chemotyper.org (Yang et al., 2015)). The ToxPrint chemotypes (developed by Altamira LLC for FDA CFSAN's CERES) were applied on the identified structures in the oRepeatTox DB (section 2.3.1). They are a set of chemical features and rules derived from various toxicity prediction models and safety assessment guidelines within FDA and other federal agencies and industries. They consist of predefined library of 729 chemotypes and are publically available at www.toxprint.org (Yang et al., 2015).

3. Results and discussion

3.1. Dataset processing

The initial dataset of 439 PPARγ full and partial agonists was passed through several filters to select the modelling dataset of 170 structures. The first step was focused on dataset refinement aiming to: (i) remove data gaps; (ii) select full agonists avoiding duplicates and data uncertainties; (iii) consider stereochemistry – with a preference for S stereoisomers when the potency of racemic mixtures was reported (see section 2.1.2).

In the selection of full agonists special attention was paid to reduction of the number of the false negative predictions. The main task was to differentiate full and partial agonists in such a way to reduce the possibility of the VS missing full agonists with prosteatotic potential. This required a cutoff within the efficacy data to differentiate between full and partial agonists, such that it is both less restrictive toward marginal efficacy and relevant to the full agonism-based MIE. To our knowledge there are three classifications of PPARγ agonists regarding their relative efficacy. According to Bruning et al., (2007), transactivation more than 80% as compared to rosiglitazone, should be considered full, less than 50% – partial, and between 50% and 80% – intermediate. In

another work (Acton et al., 2005) ligands reaching 20–60% of rosiglitazone's maximal activation are deemed partial agonists. We adopted the threshold proposed by Henke et al. (1998) who considered full agonists those compounds that elicited in average at least 70% activation of PPARγ as compared to rosiglitazone.

3.2. Molecular modelling

The modelling study of PPARγ activation as a MIE in the AOP for liver steatosis was based on a two-step strategy presented schematically in Figure 2 and discussed in detail in sections 3.2.1 and 3.2.2 below.

Figure 2. Molecular modelling scheme to study PPARγ activation: VS to predict full agonists (step 1) and 3D QSAR modelling to predict their potency (step 2)

3.2.1. Virtual Screening to predict PPARγ full agonists

A VS protocol was developed and validated to predict PPARγ full agonists using MOE (MOE, v. 2014.0901). It consists of three steps: (i) protein preparation (section 2.2.1.), (ii) docking of the ligands into the PPARγ binding site (section 2.2.1.) and (iii) filtering of the generated poses based on the recently developed pharmacophore model of PPARγ full agonists (Tsakovska et al., 2014). The last step means the final poses that did not satisfy the pharmacophore were eliminated. Since the objective of the developed protocol was high throughput virtual screening, no further refinement of the poses was performed. In addition this eliminates the possibility of poses moving away from the pharmacophore. The pharmacophore model for PPARγ full agonists (Figure 3) provides four structural features capable of hydrogen bonding and ionic interactions (F1, F2, F4 and F6) and three hydrophobic and aromatic substructures (F3, F5 and F7) (Tsakovska et al., 2014). This seven-feature pharmacophore model is rather restrictive since it is based on the bioactive conformations of the most active agonists extracted from PDB. The visual inspection of all full agonists in PDB showed that the majority of them comprise 4 or 5 pharmacophore features; thus

F1-F3, F5 and/or F4 were identified as essential pharmacophoric points (Figure 3). Therefore these pharmacophoric features were used in the proposed VS protocol.

Figure 3. Pharmacophore model of PPARγ full agonists: the essential pharmacophoric points are surrounded by a dotted line

To validate the VS protocol, docking with a filter based on the five-point pharmacophore model (F1 – F5, Figure 3) was applied to the subset of 170 PPARγ full agonists. In total 144 out of the 170 ligands were correctly predicted as full agonists (model sensitivity of 85%). Two additional validation procedures were performed on: (i) a subset of 87 PPARγ partial agonists retrieved from the initial dataset of PPARγ ligands; of which, 38 did not pass the filter and were correctly classified as not being full agonists (model specificity in relation to the partial agonists of 44%); (ii) a subset of 2527 decoys (compounds that are selected to resemble the receptor binders physicochemical properties but at the same time are topologically dissimilar to them in order to minimise the likelihood of the actual binding) randomly extracted from the full set of 25867 PPARγ decoys (each 10th structure was selected after removal of duplicates) in DUD-E database (Directory of Useful Decoys - Enhanced, http://dude.docking.org, Mysinger et al., 2012); of which 1949 were correctly classified as not being full agonists (model specificity of 77%). Of the the total number of chemicals not considered to be full agonists (2614 compounds in total), 1987 were predicted correctly, revealing an accuracy for the model of 76%. Therefore the model had balanced accuracy of 81%. Obviously, the prediction model for PPARγ full agonists has high sensitivity and can discriminate between binders and non-binders quite well. At the same time discrimination between full and partial agonists is relatively low – quite a high number of partial agonists have been classified as being full agonists according to the model. This result may reflect the fact that the full and partial agonists share the same binding pocket in the PPARγ ligand binding domain and the structural differentiation between them is not fully defined. Taking into account that the model aims to predict PPARγ full agonists as potential liver toxicants, the relatively high number of false

positive hits from the VS cannot be considered as a serious drawback.

3.2.2. 3D QSAR modelling to predict pEC$_{50}$ of PPARγ full agonists

A further logical step in the evaluation of the PPARγ full agonists is the quantitative prediction of their transactivation activity. The ligand-induced *in vitro* transactivation (expressed as potency, EC$_{50}$) was chosen as a relevant dependent variable capable to reflect the agonistic activity of the compounds studied. The EC$_{50}$ covers the complex cascade of receptor binding and activation followed by the downstream molecular events triggering gene expression, therefore it is an appropriate *in vitro* experimental model of the link between the MIE (PPARγ activation) and the earliest downstream key event − increased synthesis of target proteins. Although complex in its nature and thus challenging to be modelled, EC$_{50}$ may reflect, in a more complete manner, the mechanism behind the particular pathology (Rücker et al., 2006; Sundriyal et al., 2009). Therefore, a 3D QSAR model to predict pEC$_{50}$ values of full agonists was developed. The whole 3D QSAR modelling process is presented in the multistep workflow (Figure 4) and described below.

Figure 4. The 3D QSAR modelling workflow to predict the potency of full PPARγ agonists

Dataset processing and structure alignment (1st and 2nd steps, Figure 4)

The dataset of full agonists to be subjected to 3D QSAR modelling (1^{-st} step) was selected as described in section 3.1. The final set of 170 compounds from 6 research groups' publications included structures and potency data measured in human (77 ligands) or animal (93 ligands) cell lines. In the 2nd step all ligands were aligned according to the procedures described in Section 2.2.2.1 with a 4-feature pharmacophore used as a filter of the generated docking poses. Preliminary 3D QSAR analysis was performed on the whole dataset and 48 outliers were removed based on criteria defined in section 2.2.2.2.

Model generation and validation (3rd and 4th steps, Figure 4)

Since the preliminary CoMSIA 3D QSAR analyses on separated human and animal data

indicated similar results, the final analysis was performed on a combined data set. In the final data set, nearly 40% of the structures have been tested on human cell lines.

After exclusion of outliers, the remaining dataset was split into training (n=83) and test (n=39) sets (3rd step). The training set was assembled to include structures from all selected research groups as well as to cover a broad structural variety and a wide range of activities (pEC_{50} = 5.4 − 9.1). The remaining compounds comprised the test set of similar structural variability and pEC_{50} range (pEC_{50} = 5.5 − 8.1). The relatively high number of the test compounds (about half of the training set) ensures a robust validation of the derived model.

Based on the LOO cross-validation procedure the best model was selected that included three fields (electrostatic, hydrogen bond acceptor and hydrophobic) and had the following statistical parameters: q^2_{cv} = 0.610, N_{opt} = 7, SEP_{cv} = 0.505. The statistical parameters are comparable with other published models predicting transactivation activity of PPARγ full agonists. However, the training set considered in this study is the largest of any published, comprising structurally diverse compounds, covering as much as possible the available structural data in PDB and the literature to ensure a broader applicability domain of the model.

A further Y-randomisation procedure was applied to assess the probability of generating a good model by chance. This procedure (also known as scrambling) is based on comparison of the statistical performance of the original model to models built on randomly permuted responses among the original descriptors pool. If all QSAR models obtained in the Y-randomisation test have high q^2_{cv} values, it implies that the original QSAR model is not acceptable as a predictive tool. In the present study, a Y-randomisation test was performed ten times with a low average q^2_{cv} = -0.114 and high SEP_{cv} = 0.824 thus indicating the proposed CoMSIA model is acceptable. To investigate the stability of the model further, progressive scrambling was applied. This procedure analyses the sensitivity of the developed model to small systemic perturbations of the response variable. It is particularly useful for large redundant datasets where the q^2_{cv} obtained from LOO cross-validation may give a false sense of confidence, because a "near-by" molecule, with very similar descriptor

values, to each of the omitted molecules is likely to remain in the training data (SYBYL-X, 2013). The statistical parameters resulting from the applied progressive scrambling to the CoMSIA, together with a brief description of the descriptors are presented in the Table 3. When assessing the results, two general considerations have to be taken into account: (i) since the introduced noise makes the parameter Q^2 quite conservative, a value of Q^2 as low as 0.35 signifies that the original, unperturbed model is robust (ii) the effective slope is the critical statistic, therefore stable models have slopes near unity (SYBYL-X, 2013). Comparison between these reference values and the results of the progressive scrambling further confirmed the stability of the developed model (Table 3).

Table 3. Progressive scrambling of the CoMSIA model

The model was also validated on the external test set of 39 structures (4[-th] step). Good predictive power was obtained for the model ($r^2_{pr} = 0.552$, comparable to q^2_{cv}) demonstrating the stability of the predictions. The plot of predicted pEC$_{50}$ values obtained by the optimal non-cross-validation 3D QSAR model versus observed pEC$_{50}$ values for the training and the test set compounds is given in Figure 5.

Figure 5. Predicted (pEC$_{50}$ predicted) vs. observed pEC$_{50}$ (pEC$_{50}$ observed) values for training (83) and test (39) set compounds. Regression statistics: r^2 – determination coefficient; SEE – standard error of estimate, $F_{(1, 120)}$ – F-ratio between explained and unexplained variance for the given number of degrees of freedom at 95% level of significance.

The fractional contributions of the CoMSIA electrostatic, hydrogen bond acceptor and hydrophobic fields related to the differences in the transactivation activity were 0.293, 0.346 and 0.360, respectively. These results indicate that the model is not dominated by any of the three fields and they explain similar portions (approximately one third each) of the variation in the pEC$_{50}$ data.

While the significance of the electrostatic field has been already emphasised by other

authors, the hydrogen bond acceptor and hydrophobic effects have not been explicitly discussed in relation to variations in pEC_{50} in the 3D QSAR models published so far (Shah et al., 2008; Sundriyal et al., 2009). The parity between the three fields can be explained by their role in the agonist interactions: each field has its own contribution and complements the others. The hydrogen bond acceptor field, together with the electrostatic fields, contribute mostly to the ligand-receptor interactions, while the hydrophobic effects stabilise the occupancy of the ligand binding domain of PPARγ to guarantee the optimal orientation and distances of the ligand to the key amino acid residues within the pocket. This indirectly mediates the specific donor-acceptor interactions between the receptor activation helix H12 and the electronegative substructures of the full agonists and is a prerequisite for the electrostatic effects over the whole interface area. Thus, the stabilisation of the receptor in its active agonist conformation by the ligand binding can be explained by complex molecular interactions that are additive in their nature and mutually benefit each other.

3.3. Linking PPARγ full agonism prediction to observed *in vivo* phenotypic effects associated with chronic liver disease development: Case study using COSMOS oRepeatTox DB

Predicting full agonistic activity of PPARγ ligands by MM approaches involved investigating key protein-ligand interactions. However, increasing the scientific confidence in an AOP requires coupling a comprehensive understanding of the nature of the interaction between the chemical and the biological system, with mechanistic understanding of the biological response (OECD, 2013). In this particular case it requires going beyond prediction of the MIE and exploring its relation to later key events, at superior levels of organisation (tissue, organ, organism), that are phenotypic effects specific to the pathology (Patlewicz et a., 2015). Thus a two-step procedure, combining chemotype analysis and MM was proposed to predict respectively: (i) prosteatotic compounds; (ii) those compounds identified as being prosteatotic which potentially act through binding and activation of PPARgamma.

This procedure was applied on the COSMOS oRepeatTox DB. First pathologically justified, ontology-based data mining of the available *in vivo* toxicity data from COSMOS oRepeatTox DB

was performed independently of the MM. The chemicals associated with liver steatosis, steatohepatitis and fibrosis phenotypic effects were identified and subjected to ToxPrint chemotypes structural analysis (section 2.3.2). The results were challenged with the MM predictions. This research is described in detail below.

3.3.1 ToxPrint Chemotype Analysis

The structural analysis of the 59 compounds identified through data mining of the COSMOS oRepeatTox DB (see section 2.3.1) was performed in terms of ToxPrint chemotypes. It included matching the substructural fragments present in the chemicals associated with liver steatosis/steatohepatitis/fibrosis with the predefined library of ToxPrint chemotypes. Chemical categories identified for liver steatosis/steatohepatitis/fibrosis included alcohols, diols, glycol ethers, aminophenols, aromatic amines, aromatic halides, polychlorinated short alkanes and halogenated amines. Figure 6a presents the result of ToxPrint chemotype analysis performed with the ChemoTyper for the case study compound, piperonyl butoxide (Mostrag-Szlichtyng et al., 2014).

3.3.2 Integrating MM approach with ToxPrint Chemotype analysis

Structural analysis revealed that the identified set of chemicals contains potential PPARγ agonists (e.g.: piperonyl butoxide), i.e. compounds with rigid hydrophobic substructural fragments and flexible aliphatic chains. In order to verify the hypothesis arising from the *in vivo* toxicity data mining/ToxPrint chemotype analysis, the VS procedure developed (section 3.2.1.) was applied to the subset of 59 compounds with liver phenotypic effects (Mostrag-Szlichtyng et al., 2014). The VS involved docking query structures in the binding pocket of PPARγ and filtering them with the pharmacophore model of full agonists to retrieve only the potential full agonists. Piperonyl butoxide was retrieved as a hit in the VS procedure, i.e. predicted as a PPARγ full agonist (Figure 6b). Therefore PPARγ activation by piperonyl butoxide binding may be the MIE triggering further downstream events and leading to the adverse outcome effect outlined.

Figure 6. Integrated application of ToxPrint chemotypes and the pharmacophore based VS

procedure to retrieve the prosteatotic compound piperonyl butoxide as potential PPARγ full agonist. 6a. ChemoTyper structural analysis of piperonyl butoxide: matching the substructural fragments present in query chemical (left-hand side) with the predefined library of ToxPrint chemotypes (right-hand side). 6b. Matching of the piperonyl butoxide's structure to the PPARγ pharmacophore model: F1 (Don/Acc), F3 (Hyd/Aro), F4 (Don/Acc) and F5 (Hyd/Aro) essential pharmacophoric features

Therefore the application of ToxPrint chemotypes analysis to the selected dataset of 59 structures provided opportunities to identify structures with particular phenotypic effects to the liver. Further, the application of the VS procedure developed predicted one of them (piperonyl butoxide) as a potential PPARγ full agonist.

In general the integrated application of different approaches (*in vivo* toxicity data mining/ToxPrint chemotype analysis/molecular modelling) enables the identification of chemicals (1) associated with liver steatosis/steatohepatitis/fibrosis phenotypic effects; (2) containing alerting ToxPrint chemotypes; (3) predicted to be PPARγ full agonists. For the considered case of piperonyl butoxide these would respectively include: (1) effects observed in 1-year study with dogs, namely relative liver weight increase, hepatocytes hypertrophy, clinical chemistry (alkaline phosphatase and cholesterol) changes and effects observed in 90-days study with mice: absolute liver weight increase, hypertrophy and cellular infiltration of hepatocytes (http://cosmosdb.cosmostox.eu); (2) alerting substructural features, for example the glycol side chain (Fig. 6A); (3) structural features defining full PPARγ agonists (Fig. 6B). The case of piperonyl butoxide shows that pathology relevant mining of *in vivo* toxicity data combined with structural analysis and the results of the MM study complement each other within the developed AOP framework.

The combined use of the chemotypes and pharmacophore based approach was further exploited as a basis for development of 3D chemotypes for liver steatosis. It includes: (i) coding the essential pharmacophore points as particular structural features extracted from the PPARγ full

agonists dataset; (ii) determining the distances between the essential pharmacophoric points; (iii) based on (i) and (ii), coding the disconnected graphs with the 3D distances. At this stage the steps (i) and (ii) have been developed (Table 4).

Table 4. Distances (Å) between the essential pharmacophoric points within the PPARγ full agonists

4. Conclusions

In this paper a case-study involving the combined use of different MM methodologies (docking, pharmacophore, 3D QSAR) is presented to screen chemicals based on their potential ability to bind and activate PPARγ. The VS procedure developed showed good discrimination between binders and non-binders and a high sensitivity in the prediction of binders that are full agonists. This demonstrates the feasibility of the approach for screening chemicals for hepatotoxic potential with the aim of minimising false negative predictions. A 3D QSAR model was developed based on structurally diverse dataset of full agonists as extracted from the collected PPARγ ligands dataset. The model successfully predicts the complex effect of transactivation activity which is associated with a number of downstream prosteatotic molecular events. The results of the MM were combined with independently applied mechanistic mining of *in vivo* toxicity data followed by ToxPrint chemotypes analysis, to provide insights into the molecular mechanisms associated with the particular AOP. The combined application of the described approaches is able to facilitate the process of MoA/AOP analysis. In addition it is a basis for 3D chemotypes development, which includes information about the spatial structural features, crucial for the ligand-receptor interactions. The definition of chemotypes predictive of PPARγ-mediated adverse effects is also of broader toxicological significance since the MIE has been implicated in a range of adverse effects, including developmental and reproductive toxicity (see putative AOPs in the OECD AOP Knowledge Base; www.aopkb.org).

The integrated approach developed within this study could be used for an *in silico* screening of hepatic PPARγ agonists that can function as steatosis inducers. Particularly the MM approaches are

primarily useful for mechanistic (MIE) elucidation, whereas chemotypes are more amenable to toxicological screening, thus complementing each other. Their combined application provides the basis both for prioritising compounds potentially of major concern for liver toxicity and / or grouping chemicals potentially sharing a common AOP (e.g., from PPARγ activation to liver steatosis) with a view to supporting read-across of toxicological properties. The integration of such multistage *in silico* prediction within the AOP framework exemplifies the global effort toward development of robust and mechanistically justified alternatives to animal testing and thus brings us a step closer to a new generation of hazard identification strategies.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgments

Supplementary data

Supplementary files associated with this article can be found at….

References

Acton, J.J., Black, R.M., Jones, A.B., Moller, D.E., Colwell, L., Doebber, T.W., MacNaul, K.L., Berger, J., Wood, H.B., 2005. Benzoyl 2-methyl indoles as selective PPARγ modulators. Bioorg. Med. Chem. Lett. 15, 357–362. doi:10.1016/j.bmcl.2004.10.068

ACD/Labs Percepta suite 2015; Advanced Chemistry development, Inc., http://www.acdlabs.com/products/percepta/

Adverse Outcome Pathway Knowledge Base (AOP-KB), https://aopkb.org/ (Last accessed in July, 2015)

Al-Najjar, B.O., Wahab, H.A., Tengku Muhammad, T.S., Shu-Chien, A.C., Ahmad Noruddin, N.A., Taha, M.O., 2011. Discovery of new nanomolar peroxisome proliferator-activated receptor γ activators via elaborate ligand-based modeling. Eur. J. Med. Chem. 46, 2513–2529. doi:10.1016/j.ejmech.2011.03.040

Al Sharif, M., Alov, P., Vitcheva, V., Pajeva, I., Tsakovska, I., 2014. Modes-of-Action Related to Repeated Dose Toxicity: Tissue-Specific Biological Roles of PPARγ Ligand-Dependent

Dysregulation in Nonalcoholic Fatty Liver Disease. PPAR Res. 2014, 1–13. doi:10.1155/2014/432647

Bénardeau, A., Benz, J., Binggeli, A., Blum, D., Boehringer, M., Grether, U., Hilpert, H., Kuhn, B., Märki, H.P., Meyer, M., Püntener, K., Raab, S., Ruf, A., Schlatter, D., Mohr, P., 2009. Aleglitazar, a new, potent, and balanced dual PPARα/γ agonist for the treatment of type II diabetes. Bioorg. Med. Chem. Lett. 19, 2468–2473. doi:10.1016/j.bmcl.2009.03.036

Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E., 2000. The Protein Data Bank. Nuc. Acids Res. 28, 235–242. doi: 10.1093/nar/28.1.235

Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007); Springer, 2007

Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W., Griffin, P.R., 2007. Partial Agonists Activate PPARγ Using a Helix 12 Independent Mechanism. Structure 15, 1258–1271. doi:10.1016/j.str.2007.07.014

Carrieri, A., Giudici, M., Parente, M., De Rosas, M., Piemontese, L., Fracchiolla, G., Laghezza, A., Tortorella, P., Carbonara, G., Lavecchia, A., Gilardi, F., Crestani, M., Loiodice, F., 2013. Molecular determinants for nuclear receptors selectivity: Chemometric analysis, dockings and site-directed mutagenesis of dual peroxisome proliferator-activated receptors α/γ agonists. Eur. J. Med. Chem. 63, 321–332. doi:10.1016/j.ejmech.2013.02.015

Casimiro-Garcia, A., Bigge, C.F., Davis, J.A., Padalino, T., Pulaski, J., Ohren, J.F., McConnell, P., Kane, C.D., Royer, L.J., Stevens, K.A., Auerbach, B., Collard, W., McGregor, C., Song, K., 2009. Synthesis and evaluation of novel α-heteroaryl-phenylpropanoic acid derivatives as PPARα/γ dual agonists. Bioorg. Med. Chem. 17, 7113–7125. doi:10.1016/j.bmc.2009.09.001

Casimiro-Garcia, A., Bigge, C.F., Davis, J.A., Padalino, T., Pulaski, J., Ohren, J.F., McConnell, P., Kane, C.D., Royer, L.J., Stevens, K.A., Auerbach, B.J., Collard, W.T., McGregor, C., Fakhoury, S.A., Schaum, R.P., Zhou, H., 2008. Effects of modifications of the linker in a series of phenylpropanoic acid derivatives: Synthesis, evaluation as PPARα/γ dual agonists, and X-ray crystallographic studies. Bioorg. Med. Chem. 16, 4883–4907. doi:10.1016/j.bmc.2008.03.043

Chigurupati, S., Dhanaraj, S.A., Balakumar, P., 2015. A step ahead of PPARγ full agonists to PPARγ partial agonists: Therapeutic perspectives in the management of diabetic insulin resistance. Eur. J. Pharmacol. 755, 50–57. doi:10.1016/j.ejphar.2015.02.043

Cronet, P., Petersen, J.F., Folmer, R., Blomberg, N., Sjöblom, K., Karlsson, U., Lindstedt, E.L., Bamberg, K., 2001. Structure of the PPARalpha and -gamma ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. Structure 9, 699-706. doi:10.1016/S0969-2126(01)00634-7

Devasthale, P.V., Chen, S., Jeon, Y., Qu, F., Ryono, D.E., Wang, W., Zhang, H., Cheng, L., Farrelly, D., Golla, R., Grover, G., Ma, Z., Moore, L., Seethala, R., Sun, W., Doweyko, A.M., Chandrasena, G., Sleph, P., Hariharan, N., Cheng, P.T.W., 2007. Discovery of tertiary

aminoacids as dual PPARα/γ agonists-I. Bioorg. Med. Chem. Lett. 17, 2312–2316. doi:10.1016/j.bmcl.2007.01.060

Dixit, A., Saxena, A.K., 2008. QSAR analysis of PPAR-γ agonists as anti-diabetic agents. Eur. J. Med. Chem. 43, 73–80. doi:10.1016/j.ejmech.2007.03.004

Ebdrup, S., Pettersson, I., Rasmussen, H.B., Deussen, H.-J., Frost Jensen, A., Mortensen, S.B., Fleckner, J., Pridal, L., Nygaard, L., Sauerberg, P., 2003. Synthesis and Biological and Structural Characterization of the Dual-Acting Peroxisome Proliferator-Activated Receptor α/γ Agonist Ragaglitazar. J. Med. Chem. 46, 1306–1317. doi:10.1021/jm021027r

Gampe, R.T., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.V., Willson, T.M., Xu, H.E., 2000. Asymmetry in the PPARγ/RXRα Crystal Structure Reveals the Molecular Basis of Heterodimerization among Nuclear Receptors. Mol. Cell 5, 545–555. doi:10.1016/S1097-2765(00)80448-7

Grether, U., Bénardeau, A., Benz, J., Binggeli, A., Blum, D., Hilpert, H., Kuhn, B., Märki, H.P., Meyer, M., Mohr, P., Püntener, K., Raab, S., Ruf, A., Schlatter, D., 2009. Design and Biological Evaluation of Novel, Balanced Dual PPARα/γ Agonists. ChemMedChem 4, 951–956. doi:10.1002/cmdc.200800425

Guasch, L., Sala, E., Valls, C., Mulero, M., Pujadas, G., Garcia-Vallvé, S., 2012. Development of docking-based 3D-QSAR models for PPARgamma full agonists. J. Mol. Graphics Modell. 36, 1–9. doi:10.1016/j.jmgm.2012.03.001

Henke, B.R., Blanchard, S.G., Brackeen, M.F., Brown, K.K., Cobb, J.E., Collins, J.L., Harrington, W.W., Hashim, M.A., Hull-Ryde, E.A., Kaldor, I., Kliewer, S.A., Lake, D.H., Leesnitzer, L.M., Lehmann, J.M., Lenhard, J.M., Orband-Miller, L.A., Miller, J.F., Mook, R.A., Noble, S.A., Oliver, W., Parks, D.J., Plunket, K.D., Szewczyk, J.R., Willson, T.M., 1998. N-(2-Benzoylphenyl)-L-tyrosine PPARγ Agonists. 1. Discovery of a Novel Series of Potent Antihyperglycemic and Antihyperlipidemic Agents. J. Med. Chem. 41, 5020–5036. doi:10.1021/jm9804127

http://cosmosdb.cosmostox.eu (Last accessed in July, 2015)

http://www.cosmostox.eu/ (Last accessed in July, 2015)

http://www.seurat-1.eu/ (Last accessed in July, 2015)

Klebe, G., 1998. Comparative Molecular Similarity Indices Analysis: CoMSIA, in: Kubinyi, H, Folkers, G., Martin, Y., (Eds.) 3D QSAR in drug design. Recent Advances, KLUWER/ESCOM: Dordrecht, Vol 3, pp. 87-104.

Krewski D., Acosta D. Jr., Andersen M., Anderson H., Bailar J.C. 3rd, Boekelheide K., Brent R., Charnley G., Cheung V.G., Green S. Jr, Kelsey K.T., Kerkvliet N.I., Li A.A., McCray L., Meyer O., Patterson R.D., Pennie W., Scala R.A., Solomon G.M., Stephens M., Yager J., Zeise L., 2010. Toxicity testing in the 21st century: a vision and a strategy. J. Toxicol. Environ. Health. B. Crit. Rev. 13, 51-138. doi: 10.1080/10937404.2010.483176

Kuhn, B., Hilpert, H., Benz, J., Binggeli, A., Grether, U., Humm, R., Märki, H.P., Meyer, M., Mohr, P., 2006. Structure-based design of indole propionic acids as novel PPARα/γ co-agonists.

Bioorg. Med. Chem. Lett. 16, 4016–4020. doi:10.1016/j.bmcl.2006.05.007

Kuwabara, N., Oyama, T., Tomioka, D., Ohashi, M., Yanagisawa, J., Shimizu, T., Miyachi, H., 2012. Peroxisome Proliferator-Activated Receptors (PPARs) Have Multiple Binding Points That Accommodate Ligands in Various Conformations: Phenylpropanoic Acid-Type PPAR Ligands Bind to PPAR in Different Conformations, Depending on the Subtype. J. Med. Chem. 55, 893–902. doi:10.1021/jm2014293

Landesmann, B., Goumenou, M., Munn, S., Whelan, M., Institute for Health and Consumer Protection, 2012. Description of prototype modes-of-action related to repeated dose toxicity. Publications Office, Luxembourg.

Li, J., Kennedy, L.J., Shi, Y., Tao, S., Ye, X.Y., Chen, S.Y., Wang, Y., Hernández, A.S., Wang, W., Devasthale, P.V., Chen, S., Lai, Z., Zhang, H., Wu, S., Smirk, R.A., Bolton, S.A., Ryono, D.E., Zhang, H., Lim, N.K., Chen, B.C., Locke, K.T., O'Malley, K.M., Zhang, L., Srivastava, R.A., Miao, B., Meyers, D.S., Monshizadegan, H., Search, D., Grimm, D., Zhang, R., Harrity, T., Kunselman, L.K., Cap, M., Kadiyala, P., Hosagrahara, V., Zhang, L., Xu, C., Li, Y.X., Muckelbauer, J.K., Chang, C., An, Y., Krystek, S.R., Blanar, M.A., Zahler, R., Mukherjee, R., Cheng, P.T., Tino, J.A., 2010. Discovery of an oxybenzylglycine based peroxisome proliferator activated receptor alpha selective agonist 2-((3-((2-(4-chlorophenyl)-5-methyloxazol-4-yl)methoxy)benzyl)(methoxycarbonyl)amino)acetic acid (BMS-687453). J. Med. Chem. 53, 2854-2864. doi:10.1021/jm9016812

Liao, C., Xie, A., Zhou, J., Shi, L., Li, Z., Lu, X.P., 2004. 3D QSAR studies on peroxisome proliferator-activated receptor gamma agonists using CoMFA and CoMSIA. J. Mol. Model. 10, 165-177. doi: 10.1007/s00894-003-0175-4

Lin, C.H., Peng, Y.H., Coumar, M.S., Chittimalla, S.K., Liao, C.C., Lyn, P.C., Huang, C.C., Lien, T.W., Lin, W.H., Hsu, J.T., Cheng, J.H., Chen, X., Wu, J.S., Chao, Y.S., Lee, H.J., Juo, C.G, Wu, S.Y., Hsieh, H.P., 2009. Design and structural analysis of novel pharmacophores for potent and selective peroxisome proliferator-activated receptor gamma agonists. J Med Chem. 52, 2618-2622. doi: 10.1021/jm801594x

Lu, Y., Guo, Z., Guo, Y., Feng, J., Chu, F., 2006. Design, synthesis, and evaluation of 2-alkoxydihydrocinnamates as PPAR agonists. Bioorg. Med. Chem. Lett. 16, 915–919. doi:10.1016/j.bmcl.2005.10.104

Mahindroo, N., Huang, C.-F., Peng, Y.-H., Wang, C.-C., Liao, C.-C., Lien, T.-W., Chittimalla, S.K., Huang, W.-J., Chai, C.-H., Prakash, E., Chen, C.-P., Hsu, T.-A., Peng, C.-H., Lu, I.-L., Lee, L.-H., Chang, Y.-W., Chen, W.-C., Chou, Y.-C., Chen, C.-T., Goparaju, C.M.V., Chen, Y.-S., Lan, S.-J., Yu, M.-C., Chen, X., Chao, Y.-S., Wu, S.-Y., Hsieh, H.-P., 2005. Novel Indole-Based Peroxisome Proliferator-Activated Receptor Agonists: Design, SAR, Structural Biology, and Biological Activities. J. Med. Chem. 48, 8194–8208. doi:10.1021/jm0506930

Mahindroo, N., Wang, C.-C., Liao, C.-C., Huang, C.-F., Lu, I.-L., Lien, T.-W., Peng, Y.-H., Huang, W.-J., Lin, Y.-T., Hsu, M.-C., Lin, C.-H., Tsai, C.-H., Hsu, J.T.-A., Chen, X., Lyu, P.-C., Chao, Y.-S., Wu, S.-Y., Hsieh, H.-P., 2006a. Indol-1-yl Acetic Acids as Peroxisome Proliferator-

Activated Receptor Agonists: Design, Synthesis, Structural Biology, and Molecular Docking Studies. J. Med. Chem. 49, 1212–1216. doi:10.1021/jm0510373

Mahindroo, N., Peng, Y.-H., Lin, C.-H., Tan, U.-K., Prakash, E., Lien, T.-W., Lu, I.-L., Lee, H.-J., Hsu, J.T.-A., Chen, X., Liao, C.-C., Lyu, P.-C., Chao, Y.-S., Wu, S.-Y., Hsieh, H.-P., 2006b. Structural Basis for the Structure−Activity Relationships of Peroxisome Proliferator-Activated Receptor Agonists. J. Med. Chem. 49, 6421–6424. doi:10.1021/jm060663c

Melagraki, G., Afantitis, A., Sarimveis, H., Koutentis, P.A., Kollias, G., Igglessi-Markopoulou, O., 2009. Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. Mol. Diversity 13, 301–311. doi:10.1007/s11030-009-9115-2

MOE (Molecular Operating Environment), version 2014.091; Chemical Computing Group Inc., 2015, http://www.chemcomp.com.

Mostrag-Szlichtyng, A.S., Vitcheva, V., Nelms, M.D., Alov, P., Tsakovska, I., Enoch, S.J., Worth, A.P., Cronin, M.T., Yang, C., 2014. Data Mining Approach to Formulate Alerting Chemotypes for Liver Steatosis/Steatohepatitis/Fibrosis. Poster presented at SOT 53rd Annual Meeting, 24–27 March 2014, Phoenix, Arizona, USA. Abstract 2254. Available at: http://www.molecularnetworks.de/files/presentations/sot2014/liverChemotypes_aMostrag_sot2014_abstract_2254.pdf (Last accessed in July, 2015)

Mueller, J.J., Schupp, M., Unger, T., Kintscher, U., Heinemann, U. Binding Diversity of Pioglitazone by Peroxisome Proliferator-Activated Receptor-Gamma. doi:10.2210/pdb2xkw/pdb

Mysinger, M.M., Carchia, M., Irwin, J.J., Shoichet, B.K., 2012. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 55, 6582-6594. doi: 10.1021/jm300687e

Netzeva, T.I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M.T., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D., Schultz, T., Stanton, D.W., van de Sandt, J.J., Tong, W., Veith, G., Yang, C., 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. Altern Lab Anim. 33, 155-173

OECD, Guidance Document on Developing and Assessing Adverse Outcome Pathways, Series on Testing and Assessment No. 184, OECD, Paris, France, 2013.

Ohashi, M., Oyama, T., Nakagome, I., Satoh, M., Nishio, Y., Nobusada, H., Hirono, S., Morikawa, K., Hashimoto, Y., Miyachi, H., 2011. Design, Synthesis, and Structural Analysis of Phenylpropanoic Acid-Type PPARγ-Selective Agonists: Discovery of Reversed Stereochemistry−Activity Relationship. J. Med. Chem. 54, 331–341. doi:10.1021/jm101233f

Ohashi, M., Oyama, T., Putranto, E.W., Waku, T., Nobusada, H., Kataoka, K., Matsuno, K., Yashiro, M., Morikawa, K., Huh, N., Miyachi, H., 2013. Design and synthesis of a series of α-benzyl phenylpropanoic acid-type peroxisome proliferator-activated receptor (PPAR) gamma partial agonists with improved aqueous solubility. Bioorg. Med. Chem. 21, 2319–2332.

doi:10.1016/j.bmc.2013.02.003

Otake, K., Azukizawa, S., Fukui, M., Shibabayashi, M., Kamemoto, H., Miike, T., Kunishiro, K., Kasai, M., Shirahase, H., 2011a. A Novel Series of (S)-2, 7-Substituted-1, 2, 3, 4-tetrahydroisoquinoline-3-carboxylic Acids: Peroxisome Proliferator-Activated Receptor α/γ Dual Agonists with Protein-Tyrosine Phosphatase 1B Inhibitory Activity. Chem. Pharm. Bull. 59, 1233–1242. doi.org/10.1248/cpb.59.1233

Otake, K., Azukizawa, S., Takahashi, K., Fukui, M., Shibabayashi, M., Kamemoto, H., Kasai, M., Shirahase, H., 2011b. 2-Acyl-tetrahydroisoquinoline-3-carboxylic acids: lead compounds with triple actions, peroxisome proliferator-activated receptor α/γ agonist and protein-tyrosine phosphatase 1B inhibitory activities. Chem Pharm Bull (Tokyo) 59, 876-879. doi: 10.1248/cpb.59.876

Otake, K., Azukizawa, S., Fukui, M., Kunishiro, K., Kamemoto, H., Kanda, M., Miike, T., Kasai, M., Shirahase, H., 2012. Novel (S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acids: Peroxisome proliferator-activated receptor γ selective agonists with protein-tyrosine phosphatase 1B inhibition. Bioorg. Med. Chem. 20, 1060–1075. doi:10.1016/j.bmc.2011.11.035

Patlewicz G., Simon T.W., Rowlands J.C., Budinsky R.A., Becker R.A., 2015. Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Regul. Toxicol. Pharmacol. 71, 463-477. doi: 10.1016/j.yrtph.2015.02.011

Rabinowitz J.R., Goldsmith M.R., Little S.B., Pasquinelli M.A., 2008. Computational molecular modeling for evaluating the toxicity of environmental chemicals: prioritizing bioassay requirements. Environ. Health Perspect. 116, 573-577. doi: 10.1289/ehp.11077

Rücker, C., Scarsi, M., Meringer, M., 2006. 2D QSAR of PPARgamma agonist binding and transactivation. Bioorg. Med. Chem. 14, 5178-5195. doi:10.1016/j.bmc.2006.04.005

Sauerberg, P., Pettersson, I., Jeppesen, L., Bury, P.S., Mogensen, J.P., Wassermann, K., Brand, C.L., Sturis, J., Wöldike, H.F., Fleckner, J., Andersen, A.-S.T., Mortensen, S.B., Svensson, L.A., Rasmussen, H.B., Lehmann, S.V., Polivka, Z., Sindelar, K., Panajotova, V., Ynddal, L., Wulff, E.M., 2002. Novel Tricyclic-α-alkyloxyphenylpropionic Acids: Dual PPARα/γ Agonists with Hypolipidemic and Antidiabetic Activity. J. Med. Chem. 45, 789–804. doi:10.1021/jm010964g

Sauerberg, P., Bury, P.S., Mogensen, J.P., Deussen, H.-J., Pettersson, I., Fleckner, J., Nehlin, J., Frederiksen, K.S., Albrektsen, T., Din, N., Svensson, L.A., Ynddal, L., Wulff, E.M., Jeppesen, L., 2003. Large Dimeric Ligands with Favorable Pharmacokinetic Properties and Peroxisome Proliferator-Activated Receptor Agonist Activity in Vitro and in Vivo. J. Med. Chem. 46, 4883–4894. doi:10.1021/jm0309046

Sauerberg, P., Mogensen, J.P., Jeppesen, L., Svensson, L.A., Fleckner, J., Nehlin, J., Wulff, E.M., Pettersson, I., 2005. Structure–activity relationships of dimeric PPAR agonists. Bioorg. Med. Chem. Lett. 15, 1497–1500. doi:10.1016/j.bmcl.2004.12.084

Shah, P., Mittal, A., Bharatam, P.V., 2008. CoMFA analysis of dual/multiple PPAR activators. Eur. J. Med. Chem. 43, 2784–2791. doi:10.1016/j.ejmech.2008.01.017

Sundriyal, S., Bharatam, P.V., 2009. "Sum of activities" as dependent parameter: A new CoMFA-

based approach for the design of pan PPAR agonists. Eur. J. Med. Chem. 44, 42–53. doi:10.1016/j.ejmech.2008.03.014

SYBYL-X, version 2.1, Tripos International, 2013, https://www.certara.com/

Tropsha, A., Gramatica, P., Gombar, V., 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb. Sci. 2, 69–77. doi: 10.1002/qsar.200390007

Tsakovska, I., Al Sharif, M., Alov, P., Diukendjieva, A., Fioravanzo, E., Cronin, M., Pajeva, I., 2014. Molecular Modelling Study of the PPARγ Receptor in Relation to the Mode of Action/Adverse Outcome Pathway Framework for Liver Steatosis. Int. J. Mol. Sci. 15, 7651–7666. doi:10.3390/ijms15057651

Vedani, A., Descloux, A.-V., Spreafico, M., Ernst, B., 2007. Predicting the toxic potential of drugs and chemicals in silico: A model for the peroxisome proliferator-activated receptor γ (PPAR γ). Toxicol. Lett. 173, 17–23. doi:10.1016/j.toxlet.2007.06.011

Vidović, D., Busby, S.A., Griffin, P.R., Schürer, S.C., 2011. A Combined Ligand- and Structure-Based Virtual Screening Protocol Identifies Submicromolar PPARγ Partial Agonists. ChemMedChem 6, 94–103. doi:10.1002/cmdc.201000428

Xu, H.E., Lambert, M.H., Montana, V.G., Plunket, K.D., Moore, L.B., Collins, J.L., Oplinger, J.A., Kliewer, S.A., Gampe, R.T., McKee, D.D., others, 2001. Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. Proceedings of the National Academy of Sciences 98, 13919–13924.

Ye, X.-Y., Chen, S., Zhang, H., Locke, K.T., O'Malley, K., Zhang, L., Srivastava, R., Miao, B., Meyers, D., Monshizadegan, H., Search, D., Grimm, D., Zhang, R., Lippy, J., Twamley, C., Muckelbauer, J.K., Chang, C., An, Y., Hosagrahara, V., Zhang, L., Yang, T.-J., Mukherjee, R., Cheng, P.T.W., Tino, J.A., 2010. Synthesis and structure–activity relationships of 2-aryl-4-oxazolylmethoxy benzylglycines and 2-aryl-4-thiazolylmethoxy benzylglycines as novel, potent PPARα selective activators- PPARα and PPARγ selectivity modulation. Bioorg. Med. Chem. Lett. 20, 2933–2937. doi:10.1016/j.bmcl.2010.03.019
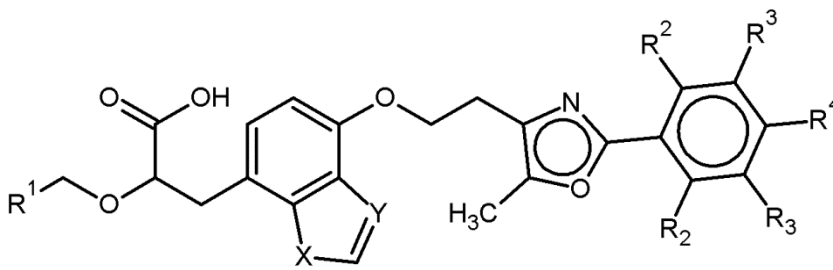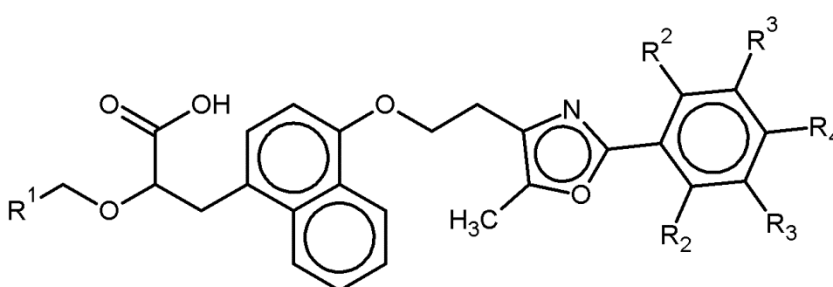
Zhang, H., Ryono, D.E., Devasthale, P., Wang, W., O'Malley, K., Farrelly, D., Gu, L., Harrity, T., Cap, M., Chu, C., Locke, K., Zhang, L., Lippy, J., Kunselman, L., Morgan, N., Flynn, N., Moore, L., Hosagrahara, V., Zhang, L., Kadiyala, P., Xu, C., Doweyko, A.M., Bell, A., Chang, C., Muckelbauer, J., Zahler, R., Hariharan, N., Cheng, P.T.W., 2009. Design, synthesis and structure–activity relationships of azole acids as novel, potent dual PPAR α/γ agonists. Bioorg. Med. Chem. Lett. 19, 1451–1456. doi:10.1016/j.bmcl.2009.01.030

Yang, C., Tarkhov, A., Marusczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C.H., Schwöbel, J., Terfloth, L., Arvidson, K., Richard, A., Worth, A., Rathman, J., 2015, New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J. Chem. Inf. Model. 2015, 55, 510-528. doi: 10.1021/ci500667v

Table 1. Comparative analysis of the uses of MM approaches in drug discovery and risk assessment

	DRUG DISCOVERY	**RISK ASSESSMENT**
GOALS	Explore the chemical space in order to extract a molecule with a given activity and desired properties	Evaluate the possible risk of an adverse effect initiated by a specific molecule in biological systems (humans and environment) under defined conditions of exposure
CHEMICAL SPACE	Drugs have prescribed chemical properties to ensure strong interactions with a specific target and specific absorption, distribution, metabolism, excretion and toxicity (ADMET) profiles	Chemicals span a large chemical space; "undesirable" property space from an ADMET perspective; elicit effects from both weak and strong interactions with targets
MAIN TASKS	- Hit identification; Lead generation/ optimisation; ADMET optimisation; - Drug candidates screening: identification of the most potent chemicals; reducing the number of false positives (i.e. chemicals incorrectly predicted to have the desired therapeutic properties)	- Support existing data; Priority setting & data gap filling; Mechanistic information - Screening of chemicals: reducing the number of false negatives (i.e. chemicals incorrectly predicted to be non-toxic)

Table 2. PPARγ ligands selected for modelling: research group, molecular scaffold, numbers and PDB identifiers

Research group *	Scaffold used in the source paper	Ligands (number)	PDB complex code	PDB ligand code	Comment
			TEMPLATES FOR STRUCTURE GENERATION		
	DATA SOURCE		PDB complex code	PDB ligand code	Comment
1ᵃ		10	3G9E	RO7	
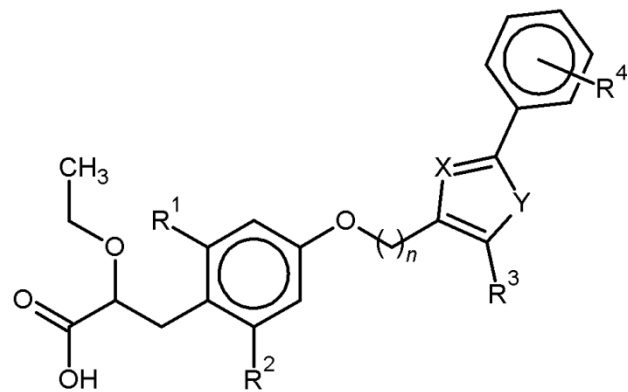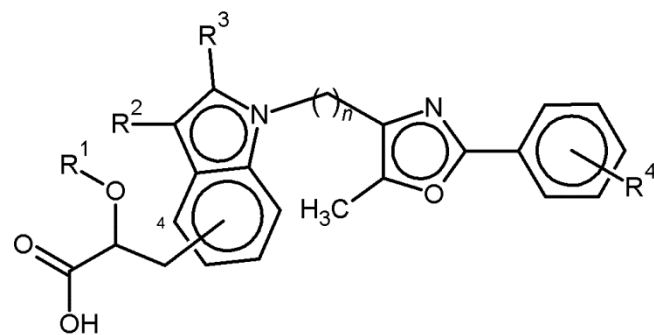
1[b]		12	3FEJ	CTM
1[c]		17	2GTK	208
2[a]		12	2Q8S	L92
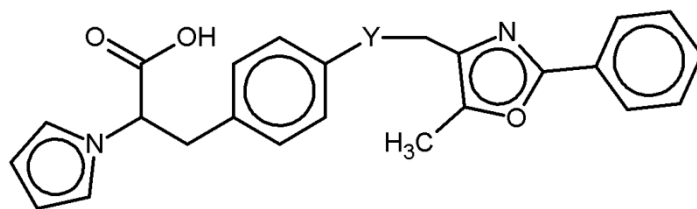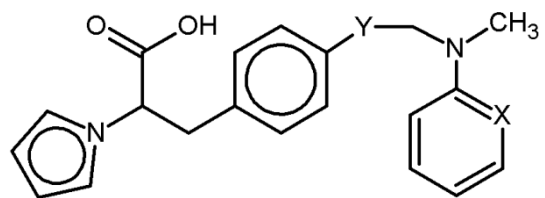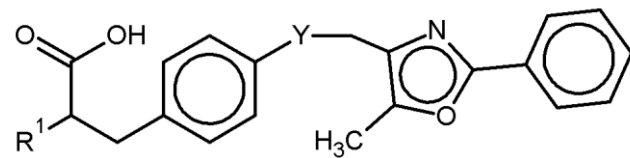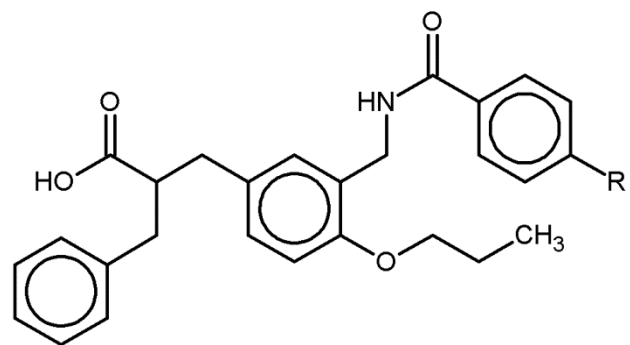
2[b]		3	3IA6	UNT
3		10	3VSO	EK1
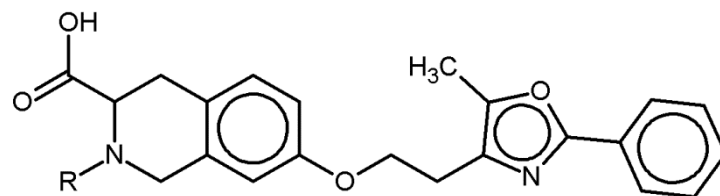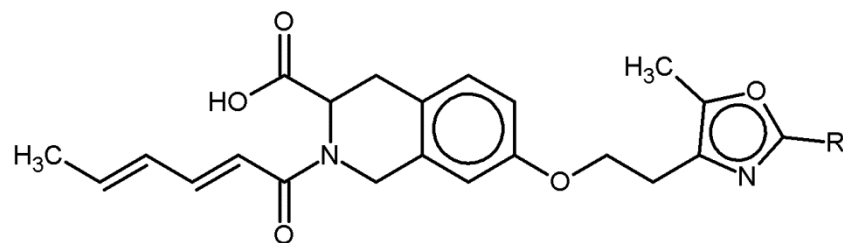
4[a]		10	no	no
4[b]		9	no	no
4[c]		25	no	no
5[a]		13	1KNU	YPA
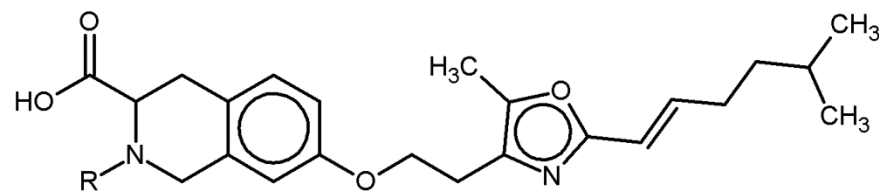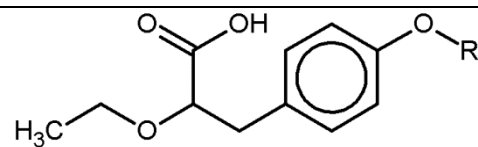
5[b]		2	no	no	1KNU/ YPA used as a template
5[c]		3	no	no	1KNU/ YPA used as a template
6[a]		12	no	no	1FM9/570 used as a template
6[b]		11	3BC5	ZAA	
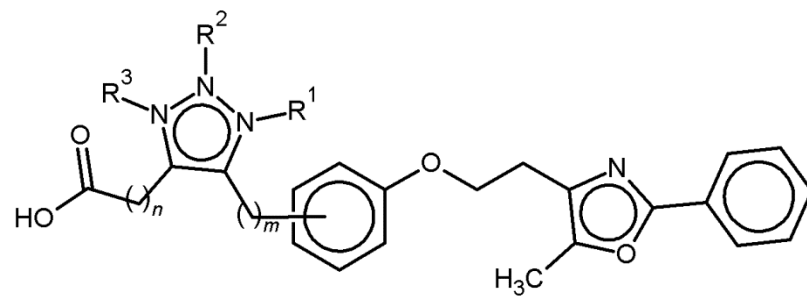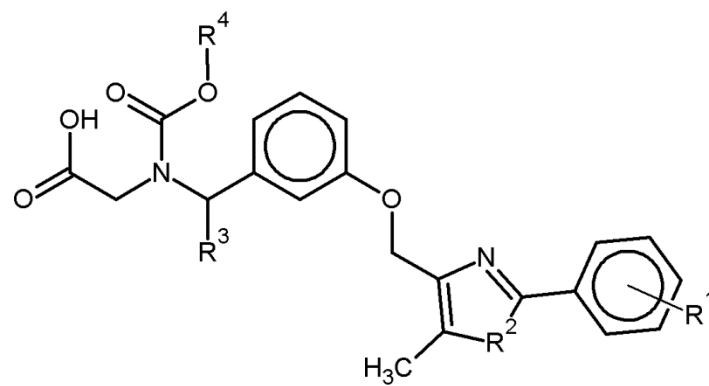
| 6^c | | 9 | 3KDU | NKS | NKS used only as a template however not included in the modelling dataset since 3KDU is a complex of PPARα |
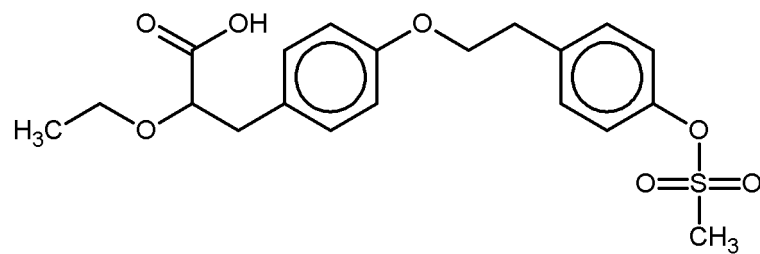
7   1   1I7I   AZ2
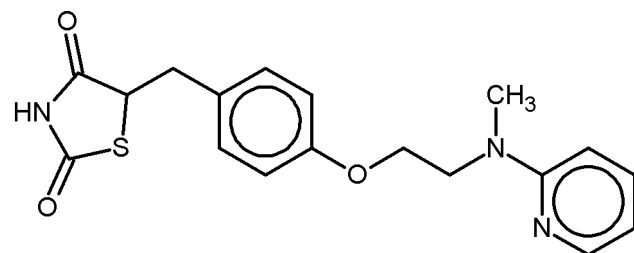
8   1   1FM6   BRL
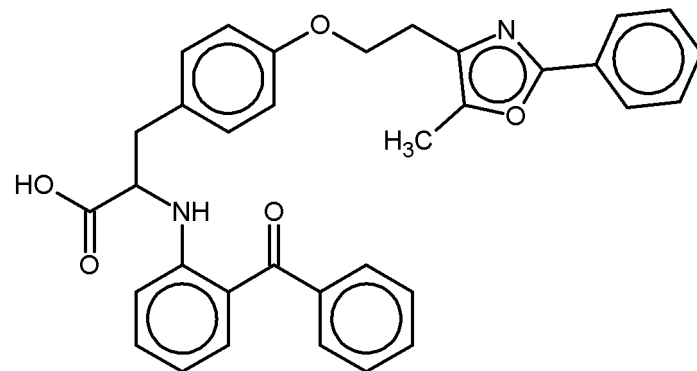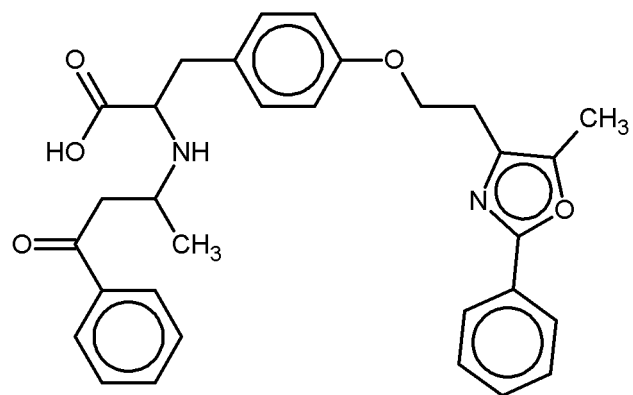
		1	1FM9	570
9		1	1K74	544
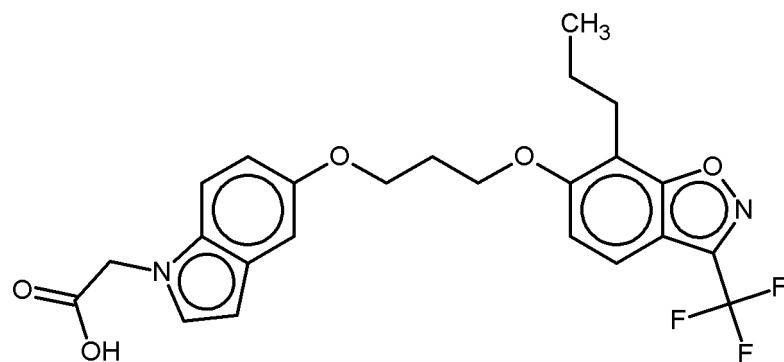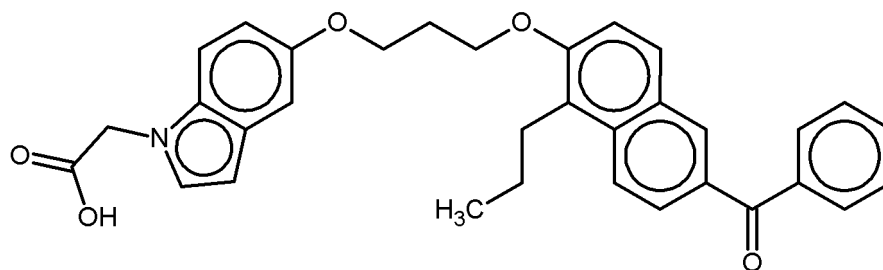
| 10 | | 1 | 2ATH | 3EA |

| 11 | | 1 | 2F4B | EHA |
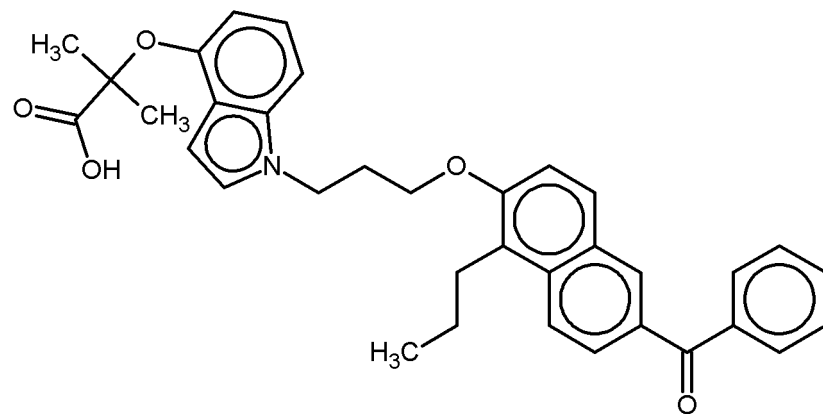
12   1   2HWR   DRD
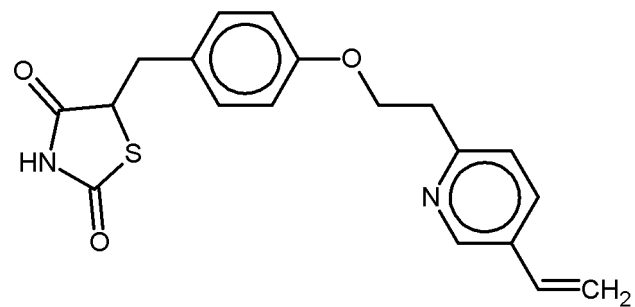
13   1   2XKW   PIB
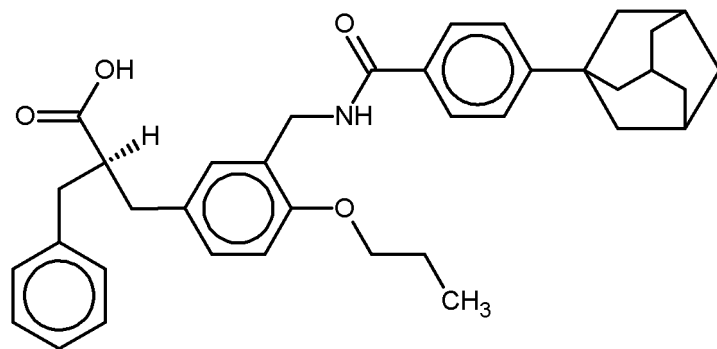
1 3AN3 M7S

14

1 3AN4 M7R

15   1   3GBK   2PQ

16   1   3VJI   J53
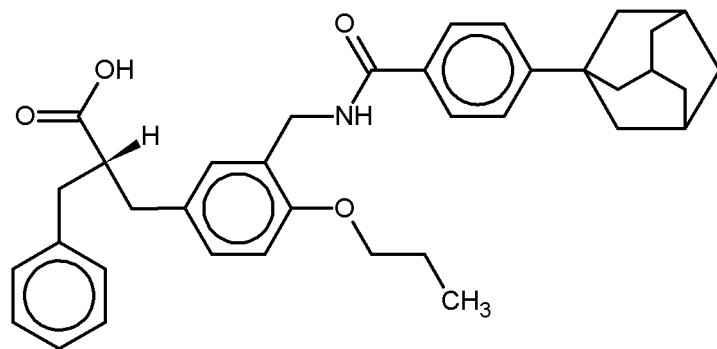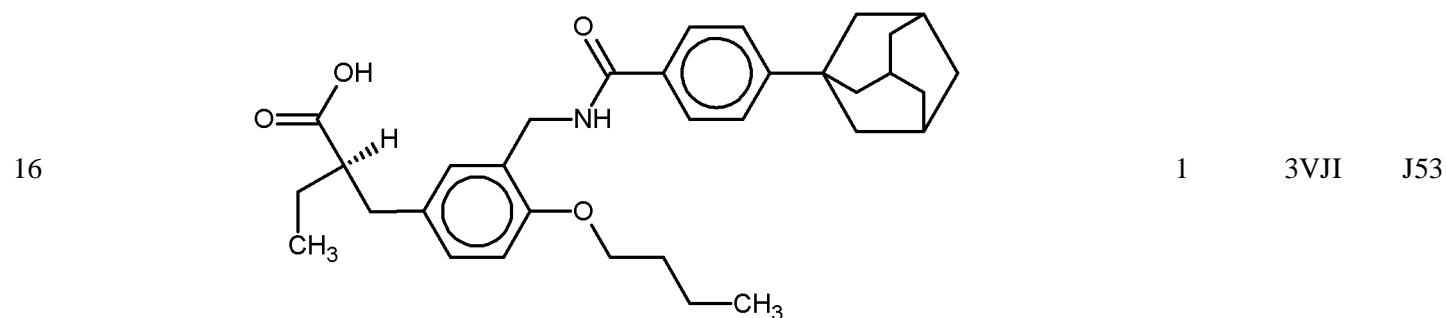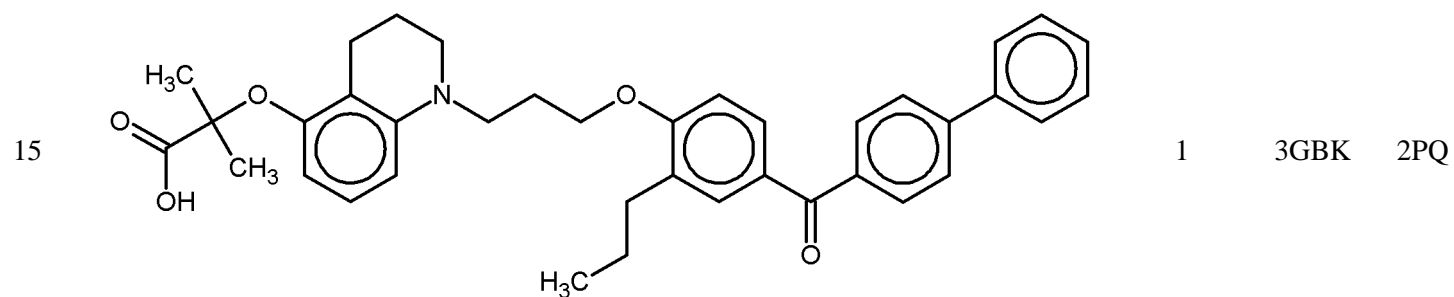
* 1[a] – Bènardeau et al.; 2009; 1[b] – Grether et al., 2009; 1[c] – Kuhn et al., 2006; 2[a] – Casimiro-Garcia et al., 2008; 2[b] – Casimiro-Garcia et al., 2009; 3 – Ohashi et al., 2013; 4[a] – Otake et al., 2011a; 4[b] – Otake et al., 2011b; 4[c] – Otake et al., 2012; 5[a] – Sauerberg et al., 2002; 5[b] – Sauerberg et al., 2003; 5[c] – Sauerberg et al., 2005; 6[a] – Devasthale et al., 2007, 6[b] – Zhang et al., 2009 and 6[c] – Ye et al., 2010., 7 – Cronet et al., 2001; 8 – Gampe et al., 2000; 9 – Xu et al., 2001; 10 – Mahindroo et al., 2005; 11 – Mahindroo et al., 2006a; 12 – Mahindroo et al., 2006b; 13 – **DOI**: 10.2210/pdb2xkw/pdb; 14 – Ohashi et al., 2011; 15 –Lin et al., 2009; 16 – Kuwabara et al., 2012. Indices a, b, and c correspond to different papers of one and the same research group designated by a number.

Table 3. Statistics of the CoMSIA model's progressive scrambling

Parameter	Description	Calculated value
Q^2	The predictivity of the model after potential effects of redundancy have been removed, that is, the expected value of q^2 at the specified critical point for $r^2_{yy'}$ (the correlation of the scrambled responses with the unperturbed data)	0.437
cSDEP	The estimated cross-validated standard error at the specified critical point	0.598
dq/dr	The slope of q^2 - the cross-validated correlation coefficient evaluated at the specified critical point with respect to the correlation of the original dependent variables versus the perturbed dependent variables	1.06

Table 4. Distances (Å) between the essential pharmacophic points within the PPARγ full agonists

Feature	F1-F2	F1-F3	F1-F5	F2-F3	F2-F5	F3-F5
Average, Å	2.76	6.4	13.1	5.8	13.1	9.3
min÷max, Å	1.9÷3.4	4.9÷9.2	11.2÷15.5	4.4÷7.3	10.8÷15.4	7.1÷11.7
Feature	F1-F2	F1-F3	F1-F5	F2-F3	F2-F5	F3-F5
Average, Å	2.76	6.4	13.1	5.8	13.1	9.3
min÷max, Å	1.9÷3.4	4.9÷9.2	11.2÷15.5	4.4÷7.3	10.8÷15.4	7.1÷11.7

Figure 1. PPARγ dataset: distribution of the ligands according to the cell line and their relative efficacy toward PPARγ. Numbers 1 − 7 indicate the different species and cell lines: 1 − hamster/kidney (BHK21 ATCC CCL10), 2-4 − monkey/kidney (COS-1, COS-7, CV-1, respectively), 5 − human/kidney (HEK293), 6 and 7 − human/liver (HepG2, Huh-7, respectively)

Figure 2. Molecular modelling scheme to study PPARγ activation: VS to predict full agonists (step 1) and 3D QSAR modelling to predict their potency (step 2)

Figure 3. Pharmacophore model of PPARγ full agonists: the essential pharmacophoric points are surrounded by a dotted line

Figure 4. The 3D QSAR modelling workflow to predict the potency of full PPARγ agonists

Figure 5. Predicted (pEC$_{50}$ predicted) vs. observed pEC$_{50}$ (pEC$_{50}$ observed) values for training (83) and test (39) set compounds. Regression statistics: r^2 − determination coefficient; SEE − standard error of estimate, F (1, 120) − F-ratio between explained and unexplained variance for the given number of degrees of freedom at 95% level of significance.

Figure 6. Integrated application of ToxPrint chemotypes and the pharmacophore based VS procedure to retrieve the prosteatotic compound piperonyl butoxide as potential PPARγ full agonist. 6a. ChemoTyper structural analysis of piperonyl butoxide: matching the substructural fragments present in query chemical (left-hand side) with the predefined library of ToxPrint chemotypes (right-hand side). 6b. Matching of the piperonyl butoxide's structure to the PPARγ pharmacophore model: F1 (Don/Acc), F3 (Hyd/Aro), F4 (Don/Acc) and F5 (Hyd/Aro) essential pharmacophoric features
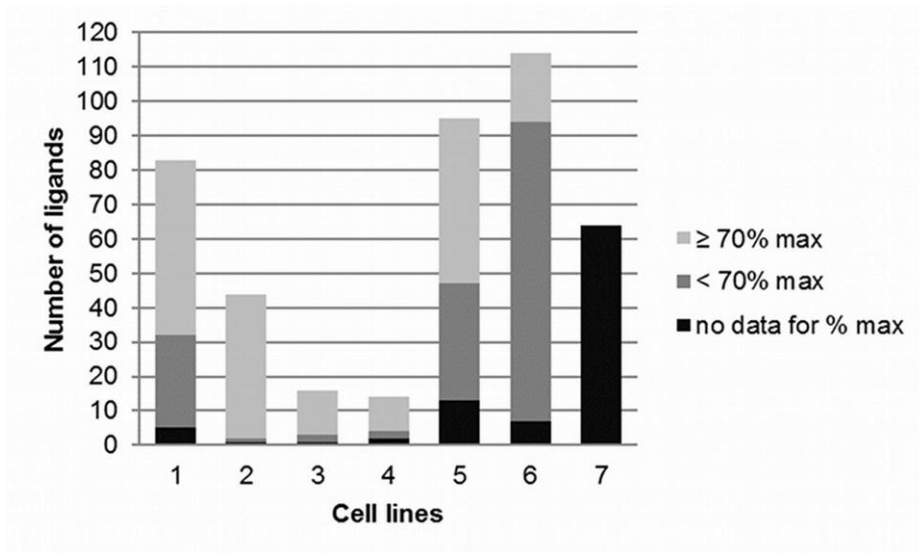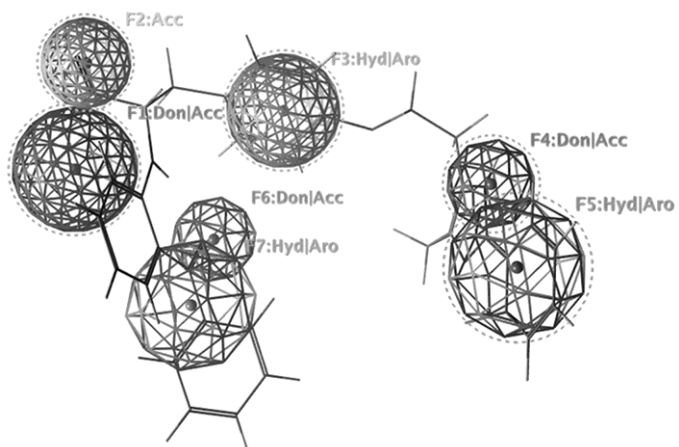
Figure 1

Figure 2

Virtual Screening to predict PPARγ full agonists
- docking + pharmacophore filter
- hits identification

step 1

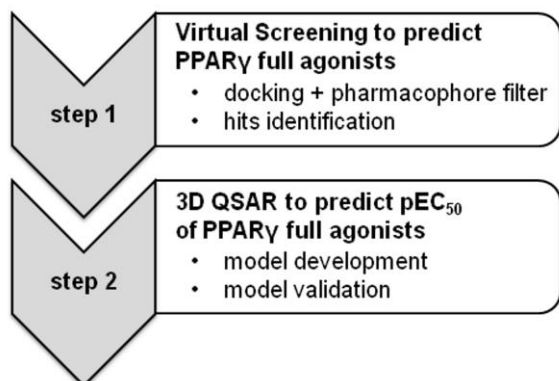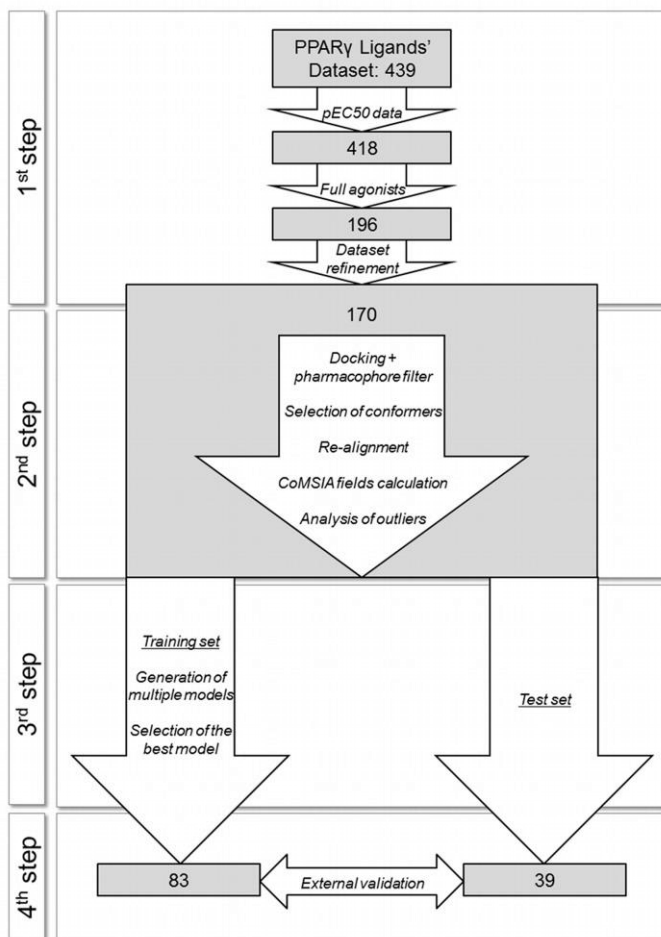**3D QSAR to predict pEC$_{50}$ of PPARγ full agonists**
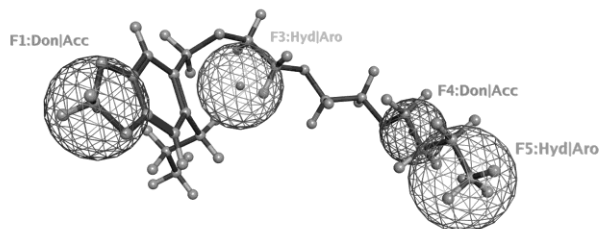- model development
- model validation

step 2

Figure 3

Figure 4

Figure 5

6a

6b

Figure 6