Chem Biol Drug Des 2008; **72:** 257–264

Research Article

doi: 10.1111/j.1747-0285.2008.00701.x

Quantitative Structure-Activity Relationships for PPAR-γ Binding and Gene Transactivation of Tyrosine-Based Agonists Using Multivariate Statistics

Costas Giaginis[1,2], Stamatios Theocharis[2] and Anna Tsantili-Kakoulidou[1,*]

[1]Department of Pharmaceutical Chemistry, School of Pharmacy, University of Athens, Panepistimiopolis, Zografou, Athens 15771, Greece
[2]Department of Forensic Medicine and Toxicology, Medical School, University of Athens, 75 Mikras Asias Street, Athens 11527, Greece
**Corresponding author: Anna Tsantili-Kakoulidou, tsantili@pharm.uoa.gr*

Peroxisome proliferator-activated receptor-γ offers a molecular target for drugs aimed to treat type II diabetes mellitus, while its therapeutic potency against cancer disease is currently being explored in preclinical studies. Tyrosine derivatives constitute a major class of peroxisome proliferator-activated receptor-γ agonists attracting considerable research interest in drug discovery. Thus, the establishment of adequate QSAR models would serve as a guide for further molecular design. In the present study, multivariate data analysis was applied on a large set of tyrosine-based peroxisome proliferator-activated receptor-γ agonists for modelling binding affinity, expressed as pKi and gene transactivation, expressed as pEC_{50}. A pool of descriptors based on physicochemical and molecular properties as well as on specific structural characteristics was used and two PLS models with satisfactory statistics were produced for binding data. According to them, molecular weight, rotatable bonds and lipophilicity were found to exert a considerable positive influence, while excess negative and positive charge created by additional acidic or basic groups in the molecules was unfavourable. With gene transactivation data, an adequate model was obtained only for the highly active compounds if considered separately. The higher complexity incorporated in gene transactivation data was further investigated by establishing a PLS model, which improved the inter-relationship between pEC_{50} and pKi.

Key words: multivariate data analysis, peroxisome proliferator-activated receptor-γ ligands, quantitative structure-activity relationships, receptor binding, transactivation, tyrosine derivatives

Received 9 June 2008, revised 8 August 2008 and accepted for publication 12 August 2008

Peroxisome proliferator-activated receptors (PPARs) belong to the nuclear hormone receptor family. Peroxisome proliferator-activated receptors regulate gene transcription after being activated by small lipophilic ligands (1,2). To date, three different subtypes of PPARs have been identified in various species: PPAR-α, -β/δ and -γ (1,2). Peroxisome proliferator-activated receptor-γ, the most extensively studied among the three, is predominantly expressed in brown and white adipose tissues and plays a central role in the process of adipocyte differentiation and peripheral glucose utilization (3,4). It is also associated with a wide spectrum of actions beyond lipid metabolism and glucose homoeostasis, such as cell differentiation, cell cycle regulation, apoptosis, carcinogenesis, inflammation, atherosclerosis and bone metabolism (5–9). To date, it offers a molecular target mainly for drugs developed for the treatment of type II diabetes mellitus (10), while its therapeutic potency against cancer disease is currently being explored in preclinical studies (11).

A large number of synthetic ligands with potency to bind selectively to and activate the PPAR-γ have been reported. They belong mainly to five major chemical classes of acidic compounds, thiazolidinediones (TZDs), tyrosine-based (TB), indole-based, propionic-acid and phenylacetic acid derivatives. Thiazolidinediones were the first PPAR-γ agonists introduced in the market as promising oral antidiabetic agents. Tyrosine-based ligands attracted further interest as a result of the toxic side-effects reported for TZDs (12,13). A number of relevant QSAR studies have appeared in literature using CoMFA and CoMSIA, Eigenvalue and Apex-3D analyses (14–17). Generally, such studies need the use of alignments based on the optimized 3-D structure of the ligands, a procedure which is often difficult and time-consuming for a large data set. Moreover, according to X-ray structure of the PPAR-γ receptor, the entrance of the binding pocket is rather flexible allowing the accommodation of structurally diverse compounds and rendering the bioactive conformation rather ambiguous (18,19). Such models are very valuable for further prediction of ligand affinity, as well as for mapping the receptor's active site (20,21). However, they do not provide information on structural characteristics and properties related to the activity, which can be further evaluated in regard to other biological processes, including ADME, so that a more holistic approach could be achieved. In this direction, 2-D QSAR may be considered a useful tool. Recently, two

2-D QSAR studies on PPAR-γ agonists have been published based on multiple regression analysis, one study suggesting binding models for a limited number of tyrosine analogues, while the other concerns a large series of compounds belonging to different chemical classes and the derived models contain mainly topological and constitutional descriptors, which reflect the diversity of the data set (22,23). To this point, it should be noted that in the case of large data sets multiple regression analysis is associated with certain limitations, as it does not tolerate inter-relation between variables. This is a serious drawback since often the information may be found in the correlation pattern rather than in the individual measures (24). Such problems can be faced by multivariate data analysis, a powerful statistical tool, which can treat inter-related descriptors, exploiting the maximum information encoded within them and separating regularities from noise (25).

Under these considerations, we were motivated to apply multivariate data analysis in the present study for modelling PPAR-γ activity. Since a large number of PPAR-γ agonists belong to tyrosine-based derivatives, we focussed our interest on this particular class of compounds in the aim to derive models based on more general and easily interpretable parameters which could be used to guide further molecular design.

Methods and Materials

Data set

The data set consisted of 106 tyrosine-based derivatives. Their chemical structures and numbering can be found in Table S1. According to essential pharmacophore elements, their chemical structure can be divided into three substructures, which comprise an acidic head (ah), including the tyrosine carboxylic and amine groups, linked to an aromatic centre (ac) and a large lipophilic fragment (lf) (8,17,18,26). An example is illustrated in Figure 1. The aromatic spacer in Figure 1 is common for all derivatives, while a high degree of chemodiversity is present in the ah and the lf. For most compounds (**1–95**), the substitution pattern on the tyrosine moiety in the ah contains one or two aromatic rings, while in compounds **62–64,** a third aromatic ring is present. 11 derivatives (**96–106**) bear small substituent on the tyrosine nitrogen. In derivatives **15–17,** the tyrosine nitrogen atom is replaced by carbon, sulphur or oxygen, respectively. In compounds **2** and **3,** the tyrosine carboxylic group has been converted to an amide or an ester group, respectively. Compounds **79, 80** and **88** bear a second carboxylic group in the ah. The tyrosine nitrogen possesses a considerable basic function in compounds **1, 4–6, 95–97** and **101–104**, while a non-tyrosine basic nitrogen is present in the ah of compounds **80** and **83** and in the lf of compounds **39, 43** and **44**. The presence of a basic function in these compounds suggests that they may exist as zwitterionic species at pH 7.4. The lf contains aromatic, heteroaromatic or alicyclic rings. A further variation in chemical structure concerns the presence of additional heteroatoms in both the lf and ah.

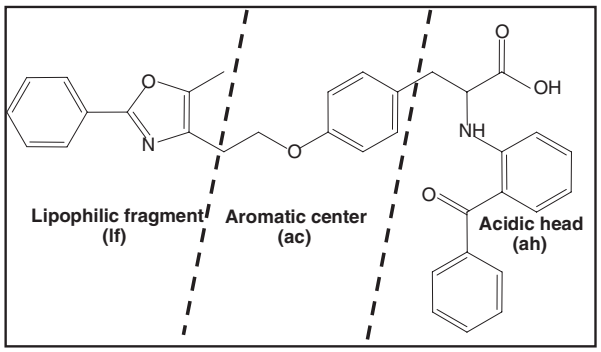

Figure 1: The structure of a representative compound (No 14, farglitazar) divided into three substructures according to essential pharmacophore elements, which comprise an acidic head (ah) linked to an aromatic centre (ac) and a large lipophilic fragment (lf).

Biological data were taken from literature (27–31). It should be noted that for the whole data set, binding affinity and gene transactivation were assessed by the same laboratory via a scintillation proximity assay (SPA) and a transient co-transfection assay, respectively (32,33). Binding data, expressed as pKi, were available for 103 compounds and gene transactivation data, expressed as pEC_{50} for 104 compounds. One hundred and one compounds possessed both pKi and pEC_{50} values. Because only the S-enantiomers of the tyrosine-based agonists bind to the receptor with high affinity, the pKi and pEC_{50} values of racemates were augmented by log 2 (i.e. 0.3) and thus their structure was treated as S-configuration (34,35).

Descriptors

The chemodiversity of the essential three substructures illustrated in Figure 1 as well as the overall structure of compounds was taken into account for the manual calculation of structural-specific constitutional descriptors. In this way, the following descriptors were calculated: the total number of rings (nRings) and number of aromatic rings ($nArC_6$), the number of heteroatoms into the rings (nNr, nOr, nSr, nFr, nHalr) and out of the rings (nNnr, nOnr, nSnr, nFnr, nHalnr) or the sum of them (nN, nO, nS, nF, nHal), as well as the number of CH_2 that link the ah with the central aromatic ring ($nCH_2(1,2)$) or the latter with the lf ($(nCH_2(2,3))$). Descriptors that take into account the presence of a tyrosine amine group attached to aromatic C_6 ring ($tyrNarC_6(ah)$) or to any ring except for aromatic C_6 (tyrNr(ah)), as well as the presence of an additional amine group in the area of the ah (non-tyrN(ah)) were also considered.

The software ADME BOXES 3.0 (Pharma-Algorithms, Toronto, Canada) was used to calculate octanol–water partition (log *P*) and distribution coefficient at pH 7.4 ($\log D_{7.4}$), hydrogen bond acceptor (HA) and donor sites (HD), the number of rotatable bonds (nRB), as well as fractions of molecular species at pH 7.4, e.g. negative (F^-), positive (F^+) and zwitterionic (Fz) fraction. Considering that zwitterionic structure may be disrupted when compounds are in contact with the cationic head in the receptor-active site, (Fz) was in a second step incorporated in (F^+) and (F^-). Alternative use of these descriptors supported the above hypothesis. The number of ionizable groups in the total structure (ablon), in the ah (ablon(ah)) and in the lf (ablon(lf)), as well as the number of acidic (alon) and basic (blon) ionizable groups in the total structure or in each substructure separately were also deduced by ADME software. The module Absolv

implemented in the same software was used to calculate Abraham's solvatochromic parameters: hydrogen bond acidity (*A*), hydrogen bond basicity (*B*), excessive refractivity (*E*), polarizability (*S*) and McGowan volume (Vmc). For lipophilicity calculation, additional software was used: CLOGP (Biobyte 4.0), CHEMDRAW ULTRA 9.0 log *P*(chem), according to Crippen fragmentation (Cambridge Soft Corporation, Cambridge, MA, USA) and tried alternatively in the analysis.

3-D descriptors were calculated by HYPERCHEM, v.5.0/CHEMPLUS v.1.6 software (HYPERCUBE Inc., Waterloo, ON, Canada). Energy minimization was carried out by molecular mechanics using MM + force field followed by semi-empirical AM1 method. The structures were constructed many times with different starting points and the robustness of the calculated parameters was evaluated.

Molecular size parameters, as solvent accessible and van der Waals volume (SAVol and WAVol), solvent accessible and van der Waals surface area (SASA, WASA) corresponding to the whole molecular structure, as well as refractivity (Ref), polarizability (Pol), energy parameters (E_{HOMO}, E_{LUMO} and E_{LUMO}–E_{HOMO}) and dipole moment (Dipol -X, -Y, -Z matrix) were calculated in the lowest energetic conformation. Polar surface or volume parameters (WASA-p, SASA-p, WAVol-p, SAVol-p) were calculated considering oxygen and nitrogen atoms and the attached hydrogen atoms. The complementary nonpolar surface or volume descriptors were obtained by subtraction of polar from whole molecule parameters (WASA-np, SASA-np, WAVol-np, SAVol-np). Submolecular size descriptors were also estimated taking into account each substructure separately (ah- WASA(ah), SASA(ah), WAVol(ah), SAVol(ah), ac- WASA(ac), SASA(ac), WAVol(ac), SAVol(ac), lf- WASA(lf), SASA(lf), WAVol(lf), SAVol(lf)).

In total, 120 constitutional, 2-D and 3-D descriptors were calculated. The descriptors used in the final models are summarized in Table S2.

Multivariate data analysis

Multivariate data analysis was performed using Simca-P 10.5 (Umetrics, CA, USA). Prior to analysis, data were mean cantered and scaled to unit variance. Principal component analysis (PCA) was initially performed in an *X*-matrix including pKi, pEC_{50} and the total number of descriptors. The score plot of the first two Principal Components served to explore any patterns in the data set. The corresponding loading plot was considered to explore relationships between the variables. Partial least square analysis was then applied to obtain regression models. In PLS analysis, pKi and pEC_{50} were treated as response variables *Y* separately. As regards activity inter-relationship, pEC_{50} was treated as response variable *Y* and pKi was included with the descriptors in the *X*-matrix. The derived models were validated by different procedures: First, internal validation (cross-validation) was performed using the default option of SIMCA-P, namely by developing parallel models leaving out one group of seven compounds each time. The omitted data are then considered as test set and the sum of squared differences between the measured response and the predicted value, defined as Predicting Residuals Sum of Squares (PRESS), is used to calculate cross-validated correlation coefficient Q^2: $Q^2 = 1 - PRESS/SS_Y$, SS_Y representing the variation in *Y* after mean centring and scaling. It should be noted that this cross-validation procedure provides a more severe criterion than the so-called leave-one-out approach, since by decreasing the size of CV groups the estimated Q^2 become very similar to R^2. As a second tool, permutation testing was applied based on the randomization of the response data (Y scrambling). The default option of SIMCA including 20 permutations was applied. The proposed model is considered robust if the new models on the set with randomized responses have significantly lower R^2 and Q^2 than the original model. In addition, as a more critical test, randomized response data were used to run from the beginning the model development procedure (36). Finally, external validation was performed in the final model by splitting the data set into several training and test sets. The root mean square error of prediction (RMSEP) was considered for evaluating the predictive power.

Results and Discussion

PCA overview

Initial PCA led to a model with $R^2 = 0.64$ and $Q^2 = 0.52$. Five principal components ($A = 5$) proved significant. The score plot of the first two components discriminated the compounds according to their molecular size in two main classes occupying different positions inside the Hotelling T^2 ellipse. In particular, most compounds with molecular weight (MW) exceeding the cut-off value 500 (MW > 500), proposed by Lipinski in the rule of 5 (37), lie at the right quartiles, while compounds with MW < 500 occupy the left part in the Hotelling T^2 ellipse (Figure 2A). In particular, tyrosine derivatives with small N-substituents formed a separate subcluster at the upper left quartile.

The corresponding loading plot was considered in respect to the response variables pKi, and pEC_{50}. They were found to occupy different positions on the axis of the first component, accounting for about 30% of the variance, while they were located close to each other on the axis of the second component, accounting for another 12%. For better illustration, the positions of the response variables pKi, and pEC_{50} and the descriptors included in the final models are depicted in the loading plot (Figure 2B).

QSAR analysis for binding data (pKi)

Initial PLS analysis was performed for the whole data set ($n = 103$) to screen for outliers. A two-component PLS model with rather poor statistics was derived ($R^2 = 0.60$, $Q^2 = 0.42$, RMSEE = 0.64). Analysis was repeated excluding compounds with residual higher $2 \times$ RMSEE and border cases. The new PLS model was then used to predict the activity of the border cases and those that were well predicted were re-included in the analysis. In this way, 10 compounds (numbers **1**, **41, 62, 63, 64, 79, 82, 83, 97** and **105**) were finally characterized as outliers. After exclusion of these compounds ($n = 93$), the PLS model was further refined by variable selection according to the Variable Influence to Projection (VIP) criterion, the weight (*w*) in the loadings plot and the size of the coefficients. Variables with VIP < 0.6 or low weight in the loadings plot and low coefficient were excluded. From descriptors encoding the same information, those which performed better were chosen.

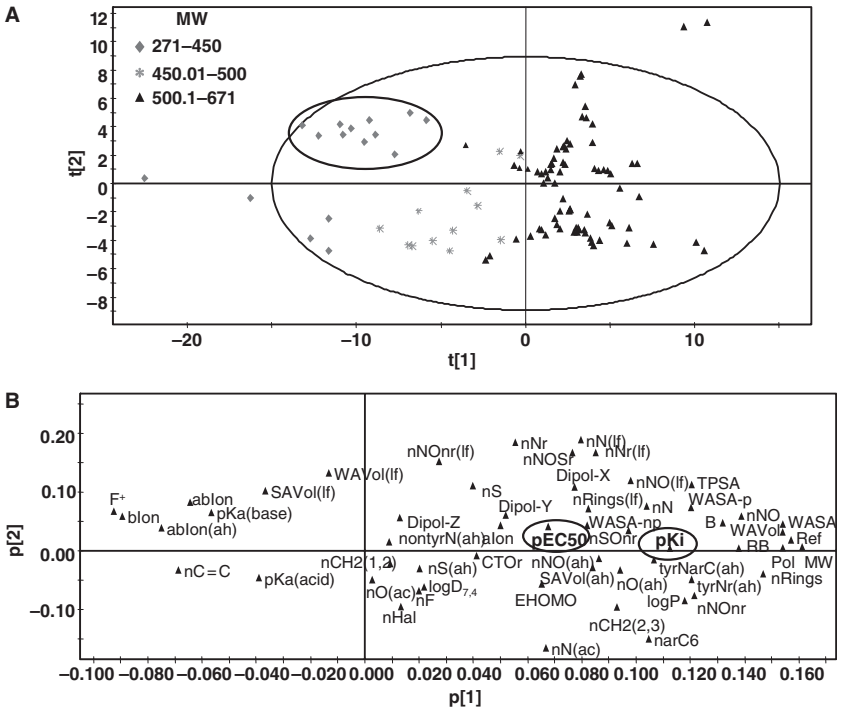

Figure 2: (A) The score plot of the first two principal components discriminated the compounds according to their MW. Encircled are depicted tyrosine derivatives with small N-substituents. (B) The loading plot of the first two principal components. Encircled are response variables pKi, and pEC_{50}. For better illustration not all the descriptors but only those included in the final models are shown.

Finally, 26 descriptors were used and a satisfactory two-component PLS model with R^2 = 0.82, Q^2 = 0.78, RMSEE = 0.43 was obtained (PLS Model A). The most important variables were found to be whole molecule descriptors related to size. In fact, MW, polarizability (Pol), refractivity (Ref), as well as WAVol and WASA exerted considerable positive influence in binding, with VIP > 1.3. Among the molecular size descriptors corresponding to individual substructure, the WAVol of the aromatic centre (WAVol(ar)) was found to be important (VIP = 1.31), also with positive impact in binding. The flexibility of the compounds, indicated by the nRB, contributed positively to the model. Protonation expressed as the positive fraction (F^+) after incorporation of zwitterionic fraction (Fz) and the number of basic ionizable groups (bIon) exhibited negative effect in binding. Hydrophobic binding was reflected in the positive contribution of lipophilicity expressed by log *P*(chem), with VIP = 1.07. Among the structural descriptors, the number of total rings (nRings) and the presence of a tyrosine amine group attached to an aromatic C6 ring (tyrNarC6(ah)) or to any ring (tyrNr(ah)) in the ah area, were the most important with a positive influence in binding. The coefficients of the descriptors for PLS Model A are presented as histogram in Figure 3A. Bars corresponding to descriptors with VIP > 1 are illustrated in black. To this point, it should be taken into consideration that among the most significant molecular properties in PLS Model A, MW, lipophilicity and rotatable bonds are included in the guidelines for druglikeness and certain cut-off values have been suggested (37,38).

PLS Model A was further validated using a permutation test through randomly re-ordering response data. R^2 and Q^2 were plotted against the correlation coefficient of the *Y* vector itself (R_Y) yielding intercepts close and below zero, respectively, indicating robustness of the model (Figure 4A). Moreover, the entire model development procedure for the 93 compounds was repeated with randomized response data. No significant model was obtained ($R^2 < 0.4$, $Q^2 < 0.3$). External validation was performed by dividing the data set into a training (n = 83) and a test set (n = 10). This procedure was repeated several times selecting the test sets representatively, in regard to the distribution of pKi values, the structure of compounds and their position in the PCA score plot. Practically, the same two-component PLS model was obtained for the training set each time and successful pKi predictions were provided for the corresponding test set. The plot of observed versus predicted pKi values for a representative training set (R^2 = 0.82, Q^2 = 0.77, RMSEE = 0.42) and the matching test set is depicted (RMSEP = 0.50) in Figure 5A.

In an attempt to limit the number of outliers, a more tolerant criterion was applied excluding only three compounds (numbers 1, 41 and 105) with residuals above 1.5 units in the initial PLS analysis. A two-component PLS model (PLS Model B) with inferior but still acceptable statistics was obtained (R^2 = 0.73, Q^2 = 0.67, RMSEE = 0.53), which however could accommodate the outliers considered for PLS Model A. Model B differed from Model A in respect to the negative impact of excess ionization in binding which seems to obscure the contribution of molecular size parameters. The negative influence of the number of ionizable groups in the acidic area (abIon(ah), VIP 1.35) and the number of the overall ionizable groups (abIon, VIP 1.22) shows that excess negative and positive charge produced by additional acidic or basic groups in the molecules is unfavourable for binding. The coefficients for Model B are presented as a histogram in Figure 3B. Bars corresponding to descriptors with VIP > 1 are illustrated in black. PLS Model B was further successfully validated using a permutation test through randomly re-ordering response data according to the default option of

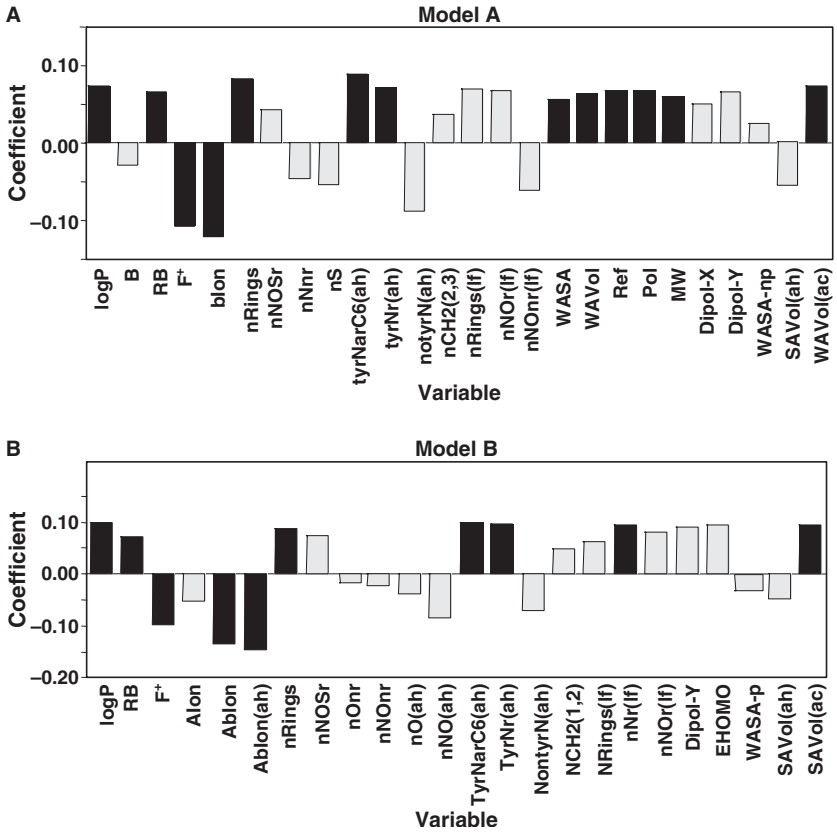

Figure 3: The coefficients of the descriptors used in: (A) PLS Model A, (B) PLS Model B. Bars corresponding to descriptors with VIP > 1 are illustrated in black.

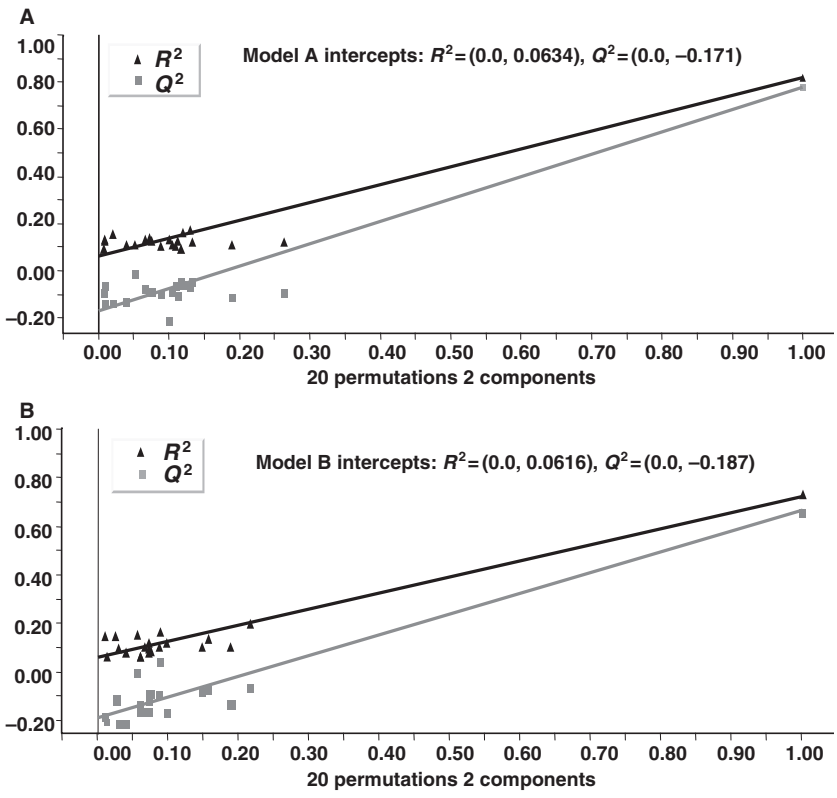

Figure 4: Validation plot (based on 20 randomization cycles) for (A) Model A and (B) Model B. R_Y refers to the correlation coefficient of the *Y* vector (pKi) itself.

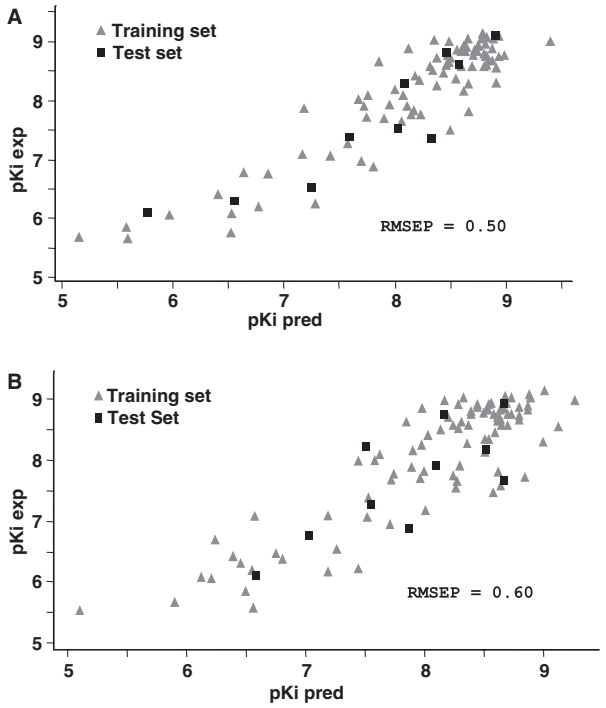

Figure 5: Plot of the observed versus predicted pKi values for a representative training set and the matching test set for (A) Model A and (B) Model B.

Simca (Figure 4B) and by re-running the model development procedure for the 100 compounds, as well as dividing the data set into several training (n = 90) and test sets (n = 10), as previously described for Model A. In the training and test sets, the compounds considered as outliers in Model A were representatively included. Practically, the same two components PLS model was obtained for each training set and successful pKi predictions were provided for the corresponding test set. The plot of observed versus predicted pKi values for a representative training set (R^2 = 0.74, Q^2 = 0.67, RMSEE = 0.52) and the matching test set (RMSEP = 0.60) is illustrated in Figure 5B.

Observed and predicted pKi values generated by the two proposed models A and B are presented in Table S3. Nevertheless, it should be noted that the predominance of the unfavourable descriptors in Model B renders it less suitable to serve as a guide.

QSAR analysis for gene transactivation data (pEC_{50})

Using gene transactivation data, pEC_{50} poor PLS models were derived for the whole data set ($R^2 < 0.5$, $Q^2 < 0.4$), due to the higher complexity incorporated in these experimental data, as discussed also in the work of Rücker *et al.* (22). However, considering the highly active compounds separately (n = 24, pEC_{50} range: 8.5–10), a satisfactory two-component model with 24 descriptors was obtained with the following statistics: R^2 = 0.85, Q^2 = 0.78, RMSEE = 0.17. For these compounds, pKi values vary between 8.4 and 9.1 and there is no correlation with pEC_{50} indicating that, when high transactivation is expressed, binding may not be the decisive step in the overall process. This is also reflected in the descriptors included in the PLS model for the highly active compounds. Thus, instead of molecular parameters like bulk properties or rotatable bonds, which played a major role in Model A for pKi, specific structural descriptors were found to be more important. Lipophilicity expressed as log P of the neutral species was also not significant. In contrast, the distribution coefficient log $D_{7.4}$ had a strong positive impact (VIP > 1). Thus, the lipophilicity term may be considered in this case to reflect rather cell and/or nucleus permeability and not hydrophobic binding. The histogram of the coefficients of the variables included in the model is presented in Figure 6. Bars corresponding to descriptors with VIP > 1 are depicted in black. The model was successfully validated through permutation and by dividing the data into training (n = 20, R^2 = 0.84, Q^2 = 0.78, RMSEE = 0.17) and test set (n = 4, RMSEP = 0.18) (Figures not shown). Observed and predicted pEC_{50} using the PLS model for highly active compounds are available in Table S4.

Activity inter-relation

Although no correlation was found between pEC_{50} and pKi values for the highly active compounds, for the entire data set there was an activity inter-relation with R^2 = 0.58. Considering that binding is one step among others for transactivation process, the inter-relation between pEC_{50} and pKi was further investigated by PLS analysis incorporating the latter in the X-matrix. Using 12 descriptors, next

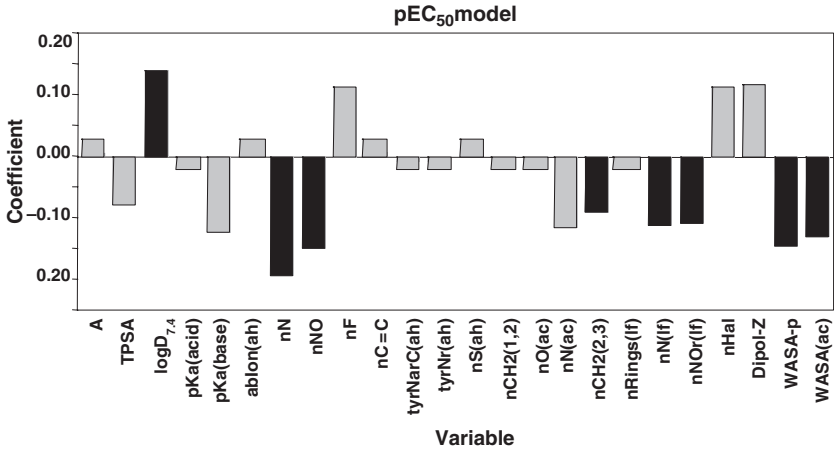

Figure 6: The coefficients of the descriptors used in the pEC_{50} model for compounds with high expression of gene transactivation. Bars corresponding to descriptors with VIP > 1 are illustrated in black.

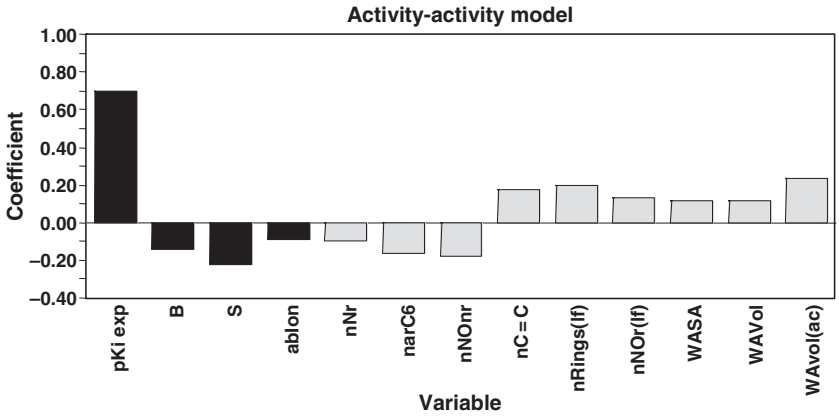

Figure 7: The coefficients of the descriptors used in the activity–activity (pEC_{50}–pKi) model. Bars corresponding to descriptors with VIP > 1 are illustrated in black.

to pKi, the relationship was improved by 18% of the variance and a four-component PLS model was obtained with the following statistics: $n = 101$, $R^2 = 0.76$, $Q^2 = 0.67$, RMSEE = 0.61. Among the most important descriptors, hydrogen bond basicity (*B*), polarizability (*S*), presence of acidic and basic groups exerted a negative effect, while molecular volume parameters showed excess positive contribution. Coefficients of the descriptors are presented as a histogram in Figure 7. The model was successfully validated through permutation and by dividing the data into training ($n = 91$, $R^2 = 0.76$, $Q^2 = 0.67$, RMSEE = 0.61) and test set ($n = 10$, RMSEP = 0.63) (Figures not shown). Observed and predicted pEC_{50} derived from activity–activity PLS model are available in Table S5.

Conclusions and Future Directions

In the present study, we demonstrated the efficiency of multivariate data analysis as a valuable tool for modelling the activity of PPAR-γ ligands. Principal component analysis allowed a fast overview of the data set by discriminating the compounds according to their molecular size. Partial least square analysis produced satisfactory and robust PLS models for binding of PPAR-γ agonists to their receptor, while at the same time providing a clear informative illustration of the contributing molecular properties and structural descriptors. Among them, MW, rotatable bonds and log *P* were found to exert a considerable positive influence in binding. Excess bulk was found to be important also for transactivation following binding as suggested by the activity–activity PLS model. This information is important and can be incorporated in a more holistic approach for further design of tyrosine-based PPAR-γ agonists, if considered that for the above properties cut-off values have been proposed in guidelines for druglikeness (37,38).

References

1. Desvergne B., Wahli W. (1999) Peroxisome proliferator-activated receptors: nuclear control of metabolism. Endocr Rev;20:649–688.
2. Kliewer S.A., Forman B.M., Blumberg B., Ong E.S., Borgmeyer U., Mangelsdorf D.J., Umesono K., Evans R.M. (1994) Differential expression and activation of a family of murine peroxisome proliferator-activated receptors. Proc Natl Acad Sci USA;91:7355–7359.
3. Nolte R.T., Wisely G.B., Westin S., Cobb J.E., Lambert M.H., Kurokawa R., Rosenfeld M.G., Willson T.M., Glass C.K., Milburn M.V. (1998) Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. Nature;395:137–143.
4. Braissant O., Foufelle F., Scotto C., Dauça M., Wahli W. (1996) Differential expression of peroxisome proliferator-activated receptors (PPARs): tissue distribution of PPAR-alpha, -beta, and -gamma in the adult rat. Endocrinology;137:354–366.
5. Theocharis S., Margeli A., Viehl P., Kouraklis G. (2004) Peroxisome proliferator-activated receptor-gamma ligands as cell-cycle modulators. Cancer Treat Rev;30:545–554.
6. Hamerman D. (2005) Osteoporosis and atherosclerosis: biological linkages and the emergence of dual-purpose therapies. QJM;98:467–484.
7. Giaginis C., Tsantili-Kakoulidou A., Theocharis S. (2007) Peroxisome proliferator-activated receptor-gamma ligands as bone turnover modulators. Expert Opin Investig Drugs;16:195–207.
8. Willson T.M., Brown P.J., Sternbach D.D., Henke B.R. (2000) The PPARs: from orphan receptors to drug discovery. J Med Chem;43:527–550.
9. Fiévet C., Fruchart J.-C., Staels B. (2006) PPARalpha and PPAR-gamma dual agonists for the treatment of type 2 diabetes and the metabolic syndrome. Curr Opin Pharmacol;6:606–614.
10. Wysowski D.K., Armstrong G., Governale L. (2003) Rapid increase in the use of oral antidiabetic drugs in the United States, 1990–2001. Diabetes Care;26:1852–1855.
11. Fenner M.H., Elstner E. (2005) Peroxisome proliferator-activated receptor-gamma ligands for the treatment of breast cancer. Expert Opin Investig Drugs;6:557–568.
12. Patel C., Wyne K.L., McGuire D.K. (2005) Thiazolidinediones, peripheral oedema and congestive heart failure: what is the evidence? Diab Vasc Dis Res;2:61–66.
13. Smith M.T. (2003) Mechanisms of troglitazone hepatotoxicity. Chem Res Toxicol;16:679–687.
14. Rathi L., Kashaw S.K., Dixit A., Pandey G., Saxena A.K. (2004) Pharmacophore identification and 3D-QSAR studies in N-(2-benzoyl phenyl)-L-tyrosines as PPAR gamma agonists. Bioorg Med Chem;12:63–69.

15. Hyun K.H., Lee D.Y., Lee B.S., Kim S.K. (2004) Receptor-based 3D QSAR studies on PPARγ agonists using CoMFA and CoMSIA approaches. QSAR Comb Sci;23:637–649.
16. Liao C., Xie A., Shi L., Zhou J., Lu X. (2004) Eigenvalue analysis of peroxisome proliferator-activated receptor gamma agonists. J Chem Inf Comput Sci;44:230–238.
17. 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.
18. Chen G., Zheng S., Luo X., Shen J., Zhu W., Liu H., Gui C., Zhang J., Zheng M., Puah C.M., Chen K., Jiang H. (2005) Focused combinatorial library design based on structural diversity, druglikeness and binding affinity score. J Comb Chem;7:398–406.
19. Liao C., Liu B., Shi L., Zhou J., Lu X.-P. (2005) Construction of a virtual combinatorial library using SMILES strings to discover potential structure-diverse PPAR modulators. Eur J Med Chem;40:632–640.
20. 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 gamma (PPAR gamma). Toxicol Lett;173:17–23.
21. Vedani A., Dobler M., Lill M.A. (2005) In silico prediction of harmful effects triggered by drugs and chemicals. Toxicol Appl Pharmacol;207:S398–S407.
22. Rücker C., Scarsi M., Meringer M. (2006) 2D QSAR of PPAR-gamma agonist binding and transactivation. Bioorg Med Chem;14:5178–5195.
23. Dixit A., Saxena A.K. (2008) QSAR analysis of PPAR-gamma agonists as anti-diabetic agents. Eur J Med Chem;43:73–78.
24. Eriksson L., Johansson E., Kettaneh-Wold N., Wold S. (2001) Multi-and Megavariate Data Analysis, Principles and Applications. Umea, Sweden: Umetrics AB.
25. Eriksson L., Johansson E. (1996) Multivariate design and modelling in QSAR. Chemometr Intell Lab Syst;34:1–19.
26. Kuhn B., Hilpert H., Benz J., Binggeli A., Grether U., Humm R., Marki H.P., Meyer M., Mohr P. (2006) Structure-based design of indole propionic acids as novel PPARalpha/gamma co-agonists. Bioorg Med Chem Lett;16:4016–4020.
27. Henke B.R., Blanchard S.G., Brackeen M.F., Brown K.K., Cobb J.E., Collins J.L., Harrington W.W. Jr, *et al.* (1998) N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 1. Discovery of a novel series of potent antihyperglycemic and antihyperlipidemic agents. J Med Chem;41:5020–5036.
28. Collins J.L., Blanchard S.G., Boswell G.E., Charifson P.S., Cobb J.E., Henke B.R., Hull-Ryde E.A., *et al.* (1998) N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 2. Structure-activity relationship and optimization of the phenyl alkyl ether moiety. J Med Chem;41:5037–5054.
29. Cobb J.E., Blanchard S.G., Boswell E.G., Brown K.K., Charifson P.S., Cooper J.P., Collins J.L., *et al.* (1998) N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 3. Structure-activity relationship and optimization of the N-aryl substituent. J Med Chem;41:5055–5069.
30. 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., Moore J.T., Willson T.M (2001) Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. Proc Natl Acad Sci USA;98:13919–13924.
31. Liu K.G., Lambert M.H., Ayscue A.H., Henke B.R., Leesnitzer L.M., Oliver W.R., Plunket K.D., Xu H.E., Sternbach D.D., Willson T.M. (2001) Synthesis and biological activity of L-tyrosine-based PPARgamma agonists with reduced molecular weight. Bioorg Med Chem Lett;11:3111–3113.
32. Nichols J.S., Parks D.J., Consler T.G., Blanchard S.G. (1998) Development of a scintillation proximity assay for peroxisome proliferator-activated receptor gamma ligand binding domain. Anal Biochem;257:112–119.
33. Lehmann J.M., Moore L.B., Smith-Oliver T.A., Wilkinson W.O., Willson T.M., Kliewer S.A. (1995) An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). J Biol Chem;270:12953–12956.
34. Parks D.J., Tomkinson N.C.O., Villeneuve M.S., Blanchard S.G., Willson T.M. (1998) Differential activity of rosiglitazone enantiomers at PPAR gamma. Bioorg Med Chem Lett;8:3657–3658.
35. Haigh D., Allen G., Birrell H.C., Buckel D.R., Cantello B.C.C., Eggleston D.S., Haltiwanger R.C., Holder J.C., Lister C.A., Pinto I.L., Rami H.K., Sime J.T., Smith S.A., Sweeney J.D. (1999) Non-thiazolidinedione antihyperglycaemic agents. Part 3: the effects of stereochemistry on the potency of alpha-methoxy-beta-phenylpropanoic acids. Bioorg Med Chem;7:821–830.
36. Rücker C., Rücker G., Meringer M. (2007) γ-Randomization and its variants in QSPR/QSAR. J Chem Inf Model;47:2345–2357.
37. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev;23:3–25.
38. Veber D.F., Johnson S.R., Cheng H.Y., Smith B.R., Ward K.W., Kopple K.D. (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem;45:2615–2623.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Structures of the tyrosine-based PPAR-γ ligands.

Table S2. Descriptors used in the final models.

Table S3. Experimental pKi data taken from the literature and pKi data predicted from the model A and B.

Table S4. Experimental pEC_{50} data taken from the literature and pEC_{50} data predicted from the PLS model derived for the highly active compounds.

Table S5. Experimental pKi and pEC_{50} data taken from the literature and pEC_{50} data predicted from the PLS activity-activity model.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.