Journal of Molecular Graphics and Modelling 36 (2012) 1–9

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Development of docking-based 3D-QSAR models for PPARgamma full agonists

Laura Guasch[a], Esther Sala[a], Cristina Valls[a], Miquel Mulero[a], Gerard Pujadas[a,b], Santiago Garcia-Vallvé[a,b,*]

[a] Grup de Recerca en Nutrigenòmica, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, C/ Marcel. lí Domingo s/n, 43007 Tarragona, Catalonia, Spain

[b] Centre Tecnològic de Nutrició i Salut, TECNIO, CEICS, Camí de Valls 81-87, 43204 Reus, Catalonia, Spain

ARTICLE INFO

Article history:
Accepted 6 March 2012
Available online 14 March 2012

Keywords:
PPARgamma agonists
3D-QSAR
Antidiabetics
Drug design
Protein–ligand docking

ABSTRACT

Peroxisome proliferator-activated receptor gamma (PPARγ) has become an attractive molecular target for drugs that aim to treat diabetes mellitus type II, and its therapeutic potency against skin cancer and other skin diseases is also currently being explored. To study the relationship between the structure of several PPARγ full agonists and the trans-activation activity of PPARγ, we have performed a three-dimensional quantitative structure–activity relationship (3D-QSAR) study of tyrosine-based derivatives, based on the 3D alignment of conformations obtained by docking. Highly predictive 3D-QSAR models, with Pearson-*R* values of 0.86 and 0.90, were obtained for the transactivation activity and binding affinity of PPARγ, respectively. These models are in good agreement with the structural characteristics of the binding pocket of PPARγ and provide some structural insights for the improvement of PPARγ full agonist bioactivities.

1. Introduction

Peroxisome proliferator-activated receptors (PPAR) are fatty acid-activated transcription factors that belong to the nuclear hormone receptor family [1,2]. Three PPAR isotypes, PPARα, PPARβ/δ and PPARγ, have previously been identified. Each of these subtypes appears to be differentiated in a tissue-specific manner and plays a pivotal role in glucose and lipid homeostasis [3,4]. PPARγ constitutes a primary target for the development of drug candidates for the treatment of type II diabetes. Thiazolidinediones (TZDs) represent the first known PPARγ agonists used as oral antidiabetic agents [4,5]. In addition, several studies have suggested that oral PPARγ full agonists not only exert an antidiabetic effect but also may act as a promising therapeutic target for a broad variety of skin disorders, including inflammatory skin diseases, such as psoriasis and atopic dermatitis, melanoma and other skin malignancies [6–9]. Furthermore, PPARγ full agonists may even induce cell growth arrest, apoptosis and terminal differentiation in various human malignant tumors [7]. There are several synthetic PPARγ full agonists besides TZDs with high potency and selectivity [10–14].

Over the past decade, a number of protein structures of the PPARγ ligand-binding domain (LBD), co-crystallized with ligands or in the apo-form, have been resolved by X-ray crystallography [4,15]. The binding pocket of PPARγ is very large and has a Y-shaped form, consisting of an entrance (arm III) that branches off into two pockets [16]. Arm I is extended toward H12, and arm II is situated between helix H3 and a β-sheet [16]. Arm I is the only substantially polar cavity of the PPARγ LBD, whereas arms II and III are mainly hydrophobic. To show biological activity, only two arms need to interact with the ligand; therefore, PPARγ full agonists occupy arms I and II [17].

It is expected that the use of quantitative structure–activity relationship (QSAR) approaches could correlate the observed biological activities with structural changes of the ligands [18]. Although several QSAR models for PPARγ full agonists have been performed [19–24], some of them used a small series of ligands, analyzed only the binding affinity or the trans-activation activity or developed 2D-QSAR models. We have constructed two atom-based 3D-QSAR models, one for the binding affinity and another for the transactivation activity, that use poses obtained by docking to align the structures of a set of tyrosine-derivate PPARγ full agonists. The additional advantages of our procedure are that we take into account the structures of both the ligands and the receptor and that a comparison of the binding affinity and transactivation activity

Abbreviations: LBD, ligand binding domain; PPAR, peroxisome proliferator-activated receptor; QSAR, quantitative structure–activity relationship; SAR, structure–activity relationship; TZDs, thiazolidinediones.

* Corresponding author at: Grup de Recerca en Nutrigenòmica, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, C/ Marcel.lí Domingo s/n, 43007 Tarragona, Catalonia, Spain/Centre Tecnològic de Nutrició i Salut (CTNS), TECNIO, CEICS, Camí de Valls 81-87, 43204, Reus, Catalonia, Spain. Tel.: +34 977558778.

E-mail address: santi.garcia-vallve@urv.cat (S. Garcia-Vallvé).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2012.03.001

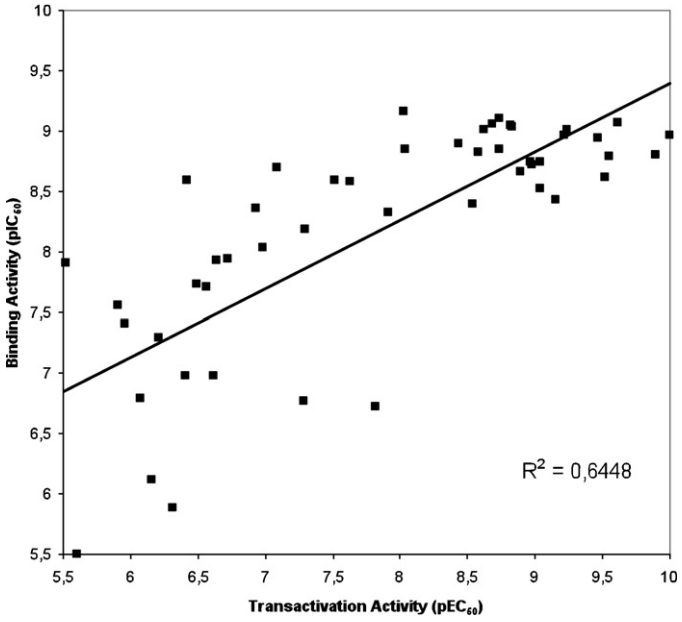

Fig. 1. Correlation between the experimental transactivation activity and experimental binding activity of the 49 tyrosine-based PPARγ full agonists used for the construction of the 3D-QSAR models.

models may provide some structural insights of the features needed by full agonists to increase the trans-activation activity of PPARγ. We have applied a similar procedure to the analysis of the structural insights of PPARγ partial agonists [17].

2. Materials and methods

2.1. Datasets

A dataset of 49 tyrosine-based compounds with measured pK_i (*i.e.*, binding affinity) and pEC_{50} (*i.e.*, transactivation activity) values obtained from the same laboratory [10–12] was used to generate two 3D-QSAR models (see Supporting Information Figure 1). The chemical structures of these 49 compounds are unequivocally known (*i.e.*, there are either no chiral atoms in their structure or the chirality of the molecules is defined), their pEC50 and pK_i values span six and five orders of magnitude, respectively, and each order of magnitude is represented by several compounds. Of the 49 molecules, 25 were randomly assigned to the training set, whereas the remaining 24 molecules were assigned to the test set. An additional set of 45 tyrosine-derivative compounds [10–12], 6 thiazolidinediones [10] and 68 indanyacetic acid derivates (for which only pEC_{50} values were available) [14] were used as an external validation set (see Supporting Information Figures 2–4). Because the measured pK_i and pEC_{50} values for the set of 45 tyrosine-derivative compounds were for a racemate solution of these compounds, not for the enantiomerical pure compounds, we added 0.3 (*i.e.*, log10 2) to all measured pK_i and pEC_{50} values of these compounds. This is equivalent to assume that the concentration of a racemate required to obtain a certain effect is twice the concentration of the corresponding active enantiomer [24]. All compound structures were built with ChemDraw Ultra v11.0 (CambridgeSoft Corporation, Cambrigde, MA, USA; http://www.cambridgesoft.com), and their 3D structures were further minimized with the LigPrep v2.4 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com), using the OPLS_2005 force field at pH 7.0 and the rest of the parameter values by default.

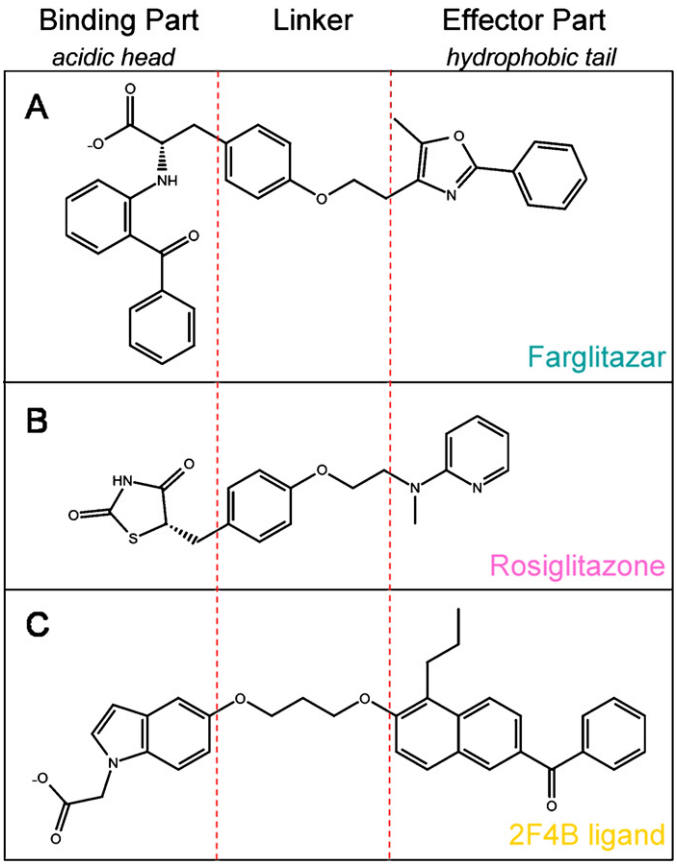

Fig. 2. Schematic representation of the common parts of PPARγ full agonists.

2.2. Molecular alignments

The most crucial step for a 3D-QSAR construction model is the alignment of the molecules. We chose a structure-based docking strategy that was carried out using the poses predicted by docking using the Glide v5.6 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). All tyrosine-based PPARγ full agonists were docked within the binding site of the 1FM9 structure. Meanwhile the 6 thiazolidinediones and the 68 indanyacetic acid derivates used as an external set were docked within the binding sites of the 1FM6 and 2F4B structures, respectively. The binding site was defined using the *Receptor Grid Generation* panel with the default options. Standard-precision (SP) docking was selected for screening the ligands. We selected the flexible docking mode, in which the Glide program generates conformations internally during the docking process. We did not request any constraint for docking. Each docking run generated at most twenty poses per ligand that survived the post-docking minimization process. The GlideScore was used as a function of fitness. The best scoring pose was selected for each ligand and used as an input structure for the subsequent 3D-QSAR analysis.

2.3. Generation of the 3D-QSAR models

The selected conformations of the ligands, obtained with the previously described alignment protocol, were used for the generation of two 3D-QSAR models, one for pIC50 and the other for pEC_{50}. The Phase v3.2 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) was utilized for constructing the 3D-QSAR models. These models can be atom-based (where all of the atoms of each ligand are taken into account) or

Table 1
Statistics of the best 3D-QSAR models for analyzing the transactivation activity of PPARγ (pEC_{50} model) and the binding affinity (pIC_{50} model) derived from a 50% randomly selected training set. To avoid an over-fitting effect, two factor models (marked in bold) were chosen.

Model	# factors	SD	R^2	F	P	Stability	RMSE	Q^2	Pearson-R
	1	0.9170	0.6288	39.0	2.28e−06	0.726	0.91	0.5882	0.8381
	2	**0.4747**	**0.9049**	**104.6**	**5.78e−12**	**0.240**	**0.78**	**0.6966**	**0.8625**
pEC50	3	0.2687	0.9709	233.6	2.77e−16	0.161	0.84	0.6479	0.8383
	4	0.1722	0.9886	434.3	3.97e−19	0.082	0.90	0.5950	0.8042
	5	0.1126	0.9954	818.0	1.73e−21	0.063	0.91	0.5895	0.7982
	1	0.5578	0.6289	39.0	2.28e−06	0.723	0.74	0.5270	0.9078
	2	**0.2609**	**0.9223**	**130.6**	**6.2e−13**	**0.164**	**0.65**	**0.6385**	**0.9035**
pIC50	3	0.1159	0.9854	471.8	2.03e−19	0.070	0.66	0.6262	0.8998
	4	0.0684	0.9951	1023.8	8.05e−23	0.082	0.69	0.5847	0.8834
	5	0.0414	0.9983	2249.6	1.2e−25	0.086	0.71	0.5679	0.8721

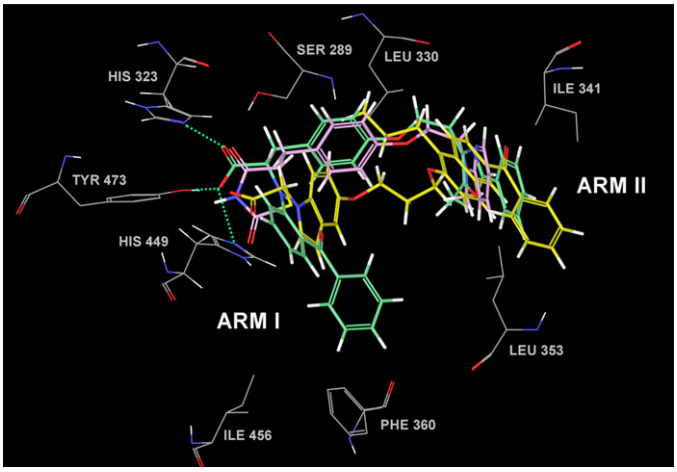

Fig. 3. The main interactions between PPARγ full agonists (farglitazar colored in green, rosiglitazone colored in pink and 2F4B ligand colored in yellow) and the LBD of PPARγ. The acidic head of the ligands is involved in a hydrogen bond network with the residues of arm I (*i.e.*, Ser289, His323, His449 and Tyr473), whereas the remaining residues form hydrophobic interactions. Hydrogen bonds of the carboxylic acid from farglitazar are shown by dashed green lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

pharmacophore-based (where only the features involved in matching a pharmacophore are used) [25]. The choice of which type of model depends largely on whether or not the training set molecules are sufficiently rigid and congeneric [25]. If the structures contain a relatively small number of rotatable bonds and some common structural framework, then an atom-based model may work quite well [25]. However, if the molecules are highly flexible or if they exhibit significant chemical diversity, a pharmacophore-based model may be more appropriate [25]. Pharmacophore-based models assume that the activity is explained entirely by the pharmacophore model itself, and therefore cannot predict activities where other features are important to activity, such as steric clashes. This requires consideration of the entire molecular structure, *i.e.*, an atom-based QSAR [25]. We chose the atom-based option because the molecules used to construct the models are tyrosine-derivative compounds that contain a common structural framework. In atom-based QSAR, a molecule is treated as a set of overlapping van der Waals spheres. To encode the basic characteristics of the local chemical structure, each atom (and hence each sphere) is placed into one of six categories according to a simple set of rules: (a) hydrogens attached to polar atoms are classified as hydrogen bond donors (D); (b) carbons, halogens and C–H hydrogens are classified as hydrophobic/non-polar (H); (c) atoms with an explicit negative ionic charge are classified as negative ionic (N); (d) atoms with an explicit positive ionic charge are classified as positive ionic (P); (e) non-ionic nitrogens and oxygens are classified as electron-withdrawing (W); and (f) all other types of atoms are classified as miscellaneous (X) [25]. The QSAR model partitions the space occupied by the ligands into a cubic grid. Any structural component can occupy part of one or more cubes. The allowed range of cube size range between 0.5 and 2.0 Å. We chose 1 Å because is the default value and it is an intermediate value between this range. The independent variables in the regression are given by the binary-valued occupancies ("bits") of the cubes (by structural components), while the dependent variables are the transactivation activity or the binding affinity. The regression is performed by constructing a series of models with an increasing number of partial least square (PLS) factors. The accuracy of the models increases when the number of PLS factors increases, until over-fitting begins to occur.

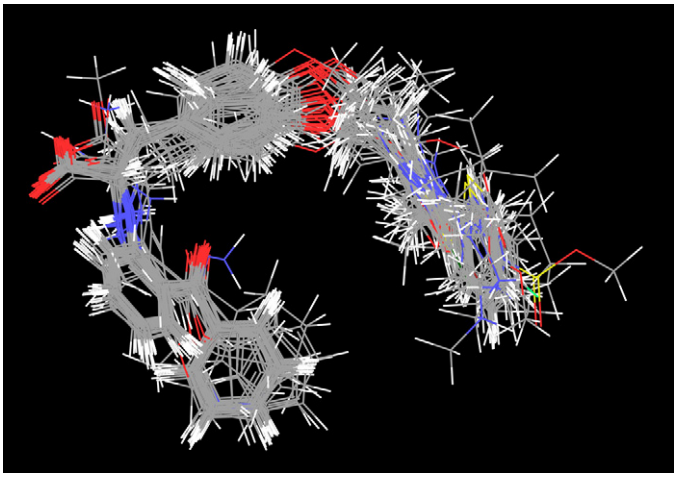

Fig. 4. Structural alignment of the selected docking poses of the 49 tyrosine-based PPARγ full agonists used to develop the 3D-QSAR models.

2.4. Statistical validations of the QSAR models

The performance of the QSAR models was evaluated by measuring the accuracy of the predictions. The statistical parameters that were used to evaluate the predictions for the training set were as follows (see the phase user manual for a complete description): (a) the coefficient of determination (R^2); (b) the standard deviation of regression (SD); (c) the F statistic that measures the overall significance of the model. Large values of F indicate a more statistically significant regression; (d) the significance level (P) of F when treated as a ratio of Chi-squared distributions. This value measures the probability that the correlation could occur by chance. Smaller values indicate a greater degree of confidence. A P value of 0.05 means F is significant at the 95% level; and (e) a *stability* value that has a maximum value of 1 and measures

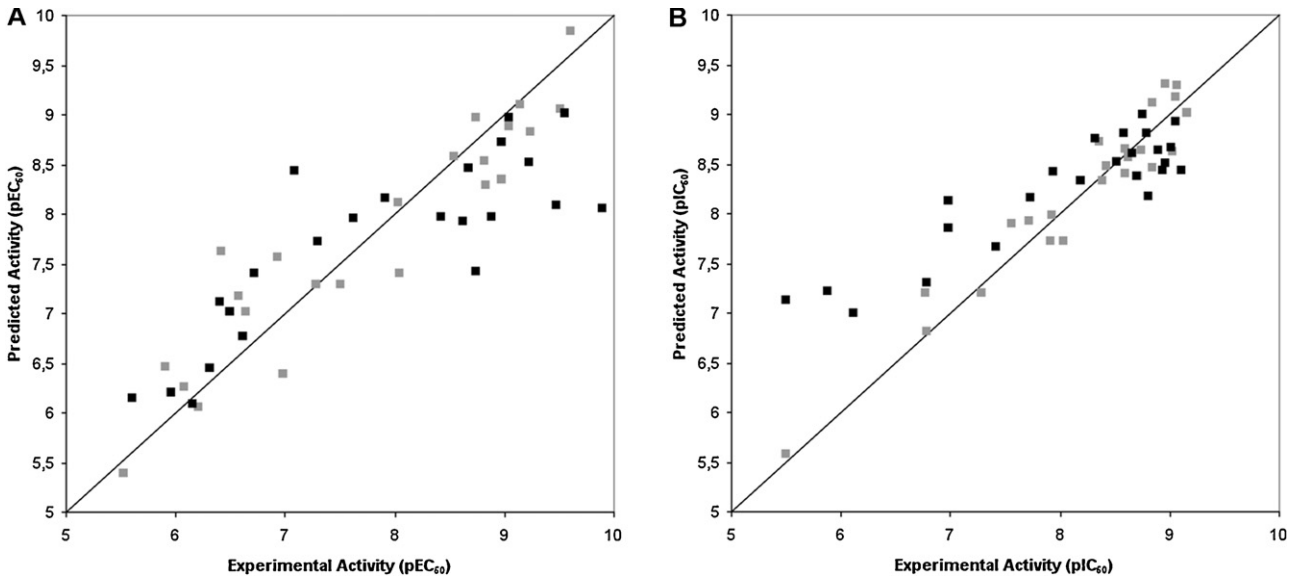

Fig. 5. Scatter plots of the (A) pEC_{50} and (B) pIC_{50} models applied to the training set (colored in gray) and the test set (colored in black).

the stability of the model predictions to changes in the training set composition. The stability value is defined as the R^2 value between the leave-10%-out predictions and the predictions from the model built on the full training set. The parameters used to evaluate the predictions for the test set were as follows: (a) the Q^2 (*i.e.*, the equivalent of the R^2 for the test set); (b) the root-mean-square error (RMSE); and (c) the Pearson correlation coefficient (R) between the predicted and observed activity for the test set.

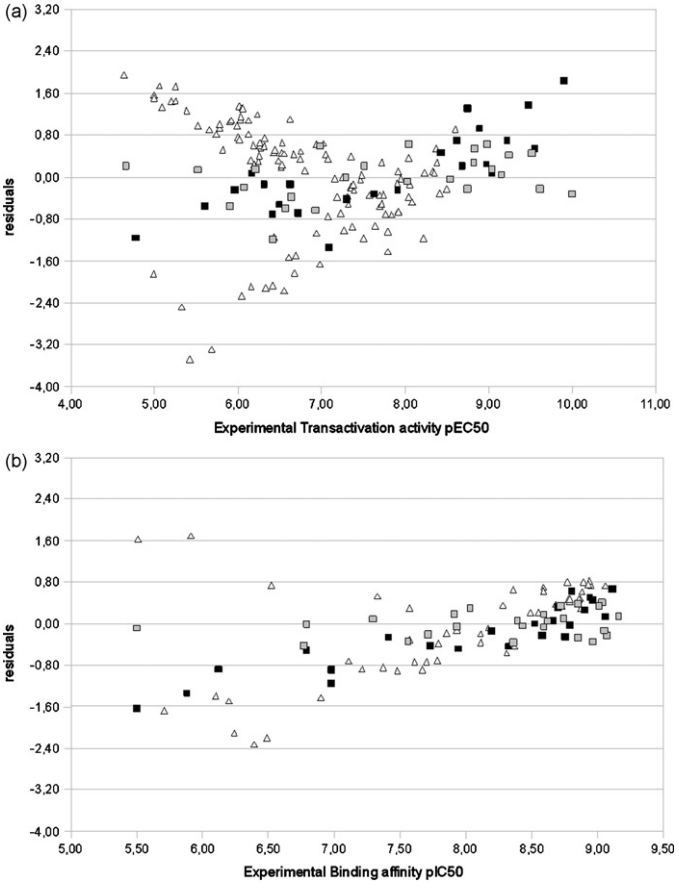

Fig. 6. Plot of residuals of the (A) pEC_{50} and (B) pIC_{50} models. Training set, test set and external test set values are represented by gray squares, black squares and white triangles, respectively. Residuals lower than 0.8 are considered good predictions, residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk in Tables 2–5) and residual higher than 1.6 are considered poor predictions (marked with two asterisks in Tables 2–5).

3. Results and discussion

3.1. Datasets

Although different SAR studies of PPARγ full agonists have been reported, we only selected the studies that contain compounds with a wide range of both transactivation and binding activity values. The dataset that we selected contained 49 tyrosine-based PPARγ full agonists [10–12], and Fig. 1 shows the correlation ($R^2 = 0.6448$) between the experimental transactivation activity (pEC50) and the experimental binding activity (pIC50) of these 49 tyrosine-based PPARγ full agonists. A relationship between the two variables is expected, because when the binding affinity increases, the AF-2 domain will be better stabilized, thus increasing the transactivation activity of PPARγ. However, some compounds, such as the compounds **3**, **4**, **6** and **21_2**, are slightly away from the line of best fit of Fig. 1. Because not all the molecular interactions between PPARγ agonists and PPARγ contribute to the stabilization of the AF-2 domain and to the trans-activation activity, there are different profiles of binding, such as those presented by partial agonists [17]. For this reason, it is interesting to study in more detail the structural differences between the features of PPARγ full agonists used for binding and those needed to activate the trans-activation activity of PPARγ. To do so, two 3D-QSAR models, one for the binding affinity and the other for the trans-activation activity were developed and compared.

3.2. Molecular alignment

A structure-based docking strategy was adopted for aligning the 49 tyrosine-based PPARγ full agonists analyzed. This set of PPARγ full agonists follows a common structural pattern (see Fig. 2) that consists of a three-module structure, comprising an acidic head linked to an aromatic center and a hydrophobic tail. The 49 molecules were docked into the crystal structure of PPARγ, and the highest scoring pose was selected for each of the

Table 2
Experimental and predicted values of the transactivation activity (pEC_{50}) and binding affinity (pIC_{50}) of the 49 tyrosine-based molecules used for the construction of the 3D-QSAR models. Residuals lower than 0.8 are considered good predictions; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered poor predictions (marked with two asterisks).

Ligands	Sets	Transactivation activity			Binding affinity		
		Experimental activity	Predicted activity	Residual	Experimental activity	Predicted activity	Residual
2	Training	6.64	7.02	−0.38	7.93	7.99	−0.06
15	Training	6.21	6.06	0.15	7.29	7.2	0.09
18	Training	8.04	7.41	0.63	8.85	8.47	0.38
30	Training	6.07	6.26	−0.19	6.79	6.81	−0.02
13_2	Training	8.98	8.35	0.63	8.72	8.38	0.34
14_2	Training	9.61	9.83	−0.22	9.07	9.29	−0.22
15_2	Training	8.82	8.54	0.28	9.05	9.18	−0.13
16_2	Training	8.74	8.97	−0.23	8.85	9.12	−0.27
18_2	Training	5.91	6.46	−0.55	7.56	7.9	−0.34
19_2	Training	5.52	5.39	0.13	7.91	7.72	0.19
20_2	Training	7.51	7.29	0.22	8.59	8.41	0.18
21_2	Training	7.29	7.29	0	6.77	7.2	−0.43
23_2	Training	6.93	7.56	−0.63	8.36	8.72	−0.36
36_3	Training	8.54	8.58	−0.04	8.39	8.33	0.06
4_2	Training	10	10.33	−0.33	8.96	9.31	−0.35
49_2	Training	6.98	6.39	0.59	8.03	7.73	0.3
5_2	Training	6.42	7.62	−1.2*	8.59	8.65	−0.06
58_3	Training	8.83	8.29	0.54	9.03	8.62	0.41
59_3	Training	9.04	8.89	0.15	8.74	8.64	0.1
63_3	Training	9.15	9.11	0.04	8.43	8.48	−0.05
65_3	Training	9.52	9.06	0.46	8.62	8.57	0.05
66_3	Training	9.24	8.82	0.42	9.01	8.66	0.35
7_2	Training	8.03	8.11	−0.08	9.16	9.02	0.14
70_2	Training	6.57	7.17	−0.6	7.71	7.92	−0.21
ent-2	Training	4.66	4.44	0.22	5.5	5.58	−0.08
3	Test	6.31	6.45	−0.14	5.88	7.22	−1.34*
4	Test	6.16	6.09	0.07	6.12	7	−0.88*
6	Test	5.6	6.15	−0.55	5.5	7.14	−1.64**
9	Test	4.78	5.94	−1.16*	6.79	7.3	−0.51
16	Test	7.3	7.73	−0.43	8.19	8.33	−0.14
20	Test	9.47	8.09	1.38*	8.94	8.43	0.51
10_2	Test	7.91	8.17	−0.26	8.32	8.76	−0.44
11_2	Test	9.9	8.06	1.84**	8.8	8.17	0.63
12_2	Test	9.22	8.53	0.69	8.96	8.51	0.45
17_2	Test	8.68	8.46	0.22	9.06	8.92	0.14
22_2	Test	8.74	7.43	1.31*	9.11	8.44	0.67
24_2	Test	8.89	7.97	0.92*	8.66	8.6	0.06
25_2	Test	8.62	7.93	0.69	9.01	8.66	0.35
26_2	Test	7.63	7.95	−0.32	8.58	8.81	−0.23
28_2	Test	6.62	6.76	−0.14	6.98	7.86	−0.88*
29_2	Test	5.96	6.21	−0.25	7.41	7.67	−0.26
56_3	Test	9.55	9.01	0.54	8.79	8.81	−0.02
57_2	Test	6.49	7.01	−0.52	7.73	8.16	−0.43
6_2	Test	7.09	8.43	−1.34*	8.7	8.38	0.32
64_3	Test	9.04	8.97	0.07	8.52	8.52	0
76_2	Test	6.72	7.4	−0.68	7.94	8.42	−0.48
8_2	Test	8.97	8.73	0.24	8.75	9	−0.25
9_2	Test	8.43	7.96	0.47	8.9	8.63	0.27
ent-18	Test	6.41	7.11	−0.7	6.98	8.13	−1.15*

molecules. These selected poses are predicted to be the most stable conformation of each molecule for binding to the PPARγ active site. All of the selected poses of the 49 analyzed molecules were visually inspected to demonstrate that they were able to establish the molecular interactions that can establish other PPARγ full agonists. These interactions include several hydrogen bonds with residues Ser289, His323, His449 and Tyr473 from arm I of the LBD of PPARγ, and several hydrophobic interactions with residues Phe360 and Ile456 from arm I and Ile281, Leu330, Ile341, Leu353 and Met364 from arm II of the LBD of PPARγ [17] (Fig. 3). Fig. 4 represents the final alignment of this dataset. The chemical similarity between the compounds analyzed, their excellent alignment and the evidence that they can interact with the LBD of PPARγ similarly to other PPARγ full agonists, ensure that a 3D-QSAR model can be obtained from all the selected compounds.

3.3. 3D-QSAR models

The aligned compounds were used to generate two 3D-QSAR models, one for analyzing the binding affinity between the ligands and PPARγ (namely the pIC_{50} model) and the other for analyzing the transactivation activity of PPARγ (namely the pEC_{50} model). Table 1 and Fig. 5 show the statistic of the constructed 3D-QSAR models. In both models, to avoid an over-fitting effect, two PLS factors were chosen. If we had chosen three or more PLS factors, the R^2 value would be greater, but the Q^2 and Pearson-R values would be lower (see Table 1). This would mean that the model performs very well with the training set, but not with the test set. The Pearson correlation coefficient of the pEC_{50} model was 0.8625 with an R^2 of 0.9049 for the training set and a Q^2 of 0.6966 for the test set. For the pIC_{50} model, the Pearson correlation coefficient

Table 3
Experimental and predicted values of the transactivation activity (pEC_{50}) and binding affinity (pIC_{50}) of 45 tyrosine-derivate compounds used as an external set to validate the models. In order to render racemate data comparable to pure S enantiomer data, we have added 0.3 (=log10 2) to all experimental pK_i and pEC50 values of racemates. Residuals lower than 0.8 are considered good prediction; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered poor predictions (marked with two asterisks).

Ligands	Transactivation activity			Binding affinity		
	Experimental activity	Predicted activity	Residual	Experimental activity	Predicted activity	Residual
4	5.71	7.39	−1.68**			
6	6.10	7.50	−1.40*	4.99	6.85	−1.86**
8	6.20	7.69	−1.49*			
11	8.28	7.92	0.36	8.04	7.67	0.37
15	7.85	8.03	−0.18	7.08	7.83	−0.75
16	7.78	8.49	−0.71	6.16	8.25	−2.09**
17	7.79	8.17	−0.38	6.33	8.45	−2.12**
18	7.37	8.22	−0.85*	6.61	8.14	−1.53*
19	8.59	7.89	0.71	8.60	7.68	0.92
20	8.78	8.31	0.47	7.84	8.56	−0.72
21	7.21	8.09	−0.88*	6.04	8.32	−2.28**
22	8.73	8.39	0.34	7.27	8.29	−1.02*
53	8.87	8.38	0.49	7.70	8.27	−0.57
54	8.88	8.26	0.62	8.23	8.14	0.09
55	8.59	7.96	0.63	7.92	7.81	0.11
56	8.95	8.21	0.74	8.33	8.22	0.11
57	8.87	8.58	0.29	8.08	8.56	−0.48
58	8.85	8.47	0.38	8.37	7.83	0.54
59	9.06	8.34	0.72	7.72	8.23	−0.51
60	8.94	8.11	0.83*	7.23	7.92	−0.69
61	8.55	8.34	0.21	7.79	8.82	−1.03*
62	7.57	7.88	−0.31	6.43	7.58	−1.15*
63	7.48	8.39	−0.91*	6.42	8.49	−2.07**
64	7.61	8.34	−0.73	7.92	8.57	−0.65
65	8.49	8.28	0.21	8.04	8.43	−0.39
67	7.70	8.43	−0.73	8.41	8.73	−0.32
68	8.86	8.39	0.47	7.64	8.57	−0.93*
69	8.93	8.17	0.76	7.50	8.67	−1.17*
70	8.79	8.36	0.43	7.37	8.33	−0.96*
71	8.17	8.25	−0.08	6.94	8.01	−1.07*
72	8.36	7.70	0.66	7.35	7.53	−0.18
73	7.93	8.05	−0.12	7.72	7.44	0.28
74	8.89	8.09	0.80	8.38	8.10	0.28
75	8.31	8.88	−0.57	8.22	9.40	−1.18*
76	7.11	7.83	−0.72	5.33	7.81	−2.48**
77	7.67	8.56	−0.89*	5.69	8.98	−3.29**
78	8.68	8.30	0.38	8.06	8.20	−0.14
79	6.49	8.70	−2.21**	5.42	8.89	−3.29**
80	8.11	8.31	−0.20	6.69	8.19	−1.50*
81	8.77	7.97	0.80	7.91	8.02	−0.11
82	6.39	8.72	−2.33**	6.98	8.65	−1.67**
83	6.24	8.35	−2.11	6.68	8.51	−1.83**
84	8.79	8.32	0.47	8.35	8.27	0.08
89	8.11	8.48	−0.37	7.79	9.22	−1.43*
90	6.90	8.32	−1.42*	6.55	8.72	−2.17**

was 0.9035 with an R^2 of 0.9223 for the training set and a Q^2 of 0.6385 for the test set. Table 2 shows the accuracy of the two models in predicting the transactivation activity and binding affinity of the training and test sets. The predictions of different activities have been classified according to the following residual scale (*i.e.*, residual is computed as the difference between the experimental activity and the estimated activity): residuals lower than 0.8 are considered good predictions; residuals between 0.8 and 1.6 are considered weak predictions and residuals higher than 1.6 are considered poor predictions. The predictions were good for the majority of the training molecules of both models. The predictions for the molecules from the test set are also reasonably good and validate the use of both models. Only 5 (*i.e.*, the compounds **5_2**, **9**, **20**, **22_2**, **24_2** and **6_2**) and 4 (*i.e.*, the compounds **3**, **4**, **28_2**

Table 4
Experimental and predicted values of the transactivation activity (pEC_{50}) and binding affinity (pIC_{50}) of the 6 thiazolidinediones used as an external set to validate the models. Residuals lower than 0.8 are considered good prediction; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered poor predictions (marked with two asterisks).

Ligands	Transactivation activity			Binding affinity		
	Experimental activity	Predicted activity	Residual	Experimental activity	Predicted activity	Residual
2_ciglitazone	4.64	6.58	1.94**	5.51	7.14	1.63**
2_pioglitazone	6.23	7.42	1.19*	5.91	7.60	1.69**
2_troglitazone	6.27	6.82	0.55	6.52	7.26	0.74
2_rosiglitazone	7.05	7.48	0.43	7.33	7.86	0.53
2_BRL48482	7.95	7.92	−0.03	7.57	7.87	0.30
2_AD7057	8.5	8.27	−0.23	8.37	7.93	−0.44

Table 5
Experimental and predicted values of the transactivation activity (pEC_{50}) of the 68 indanyacetic acid derivates used as an external set to validate the transactivation model. Residuals lower than 0.8 are considered good predictions; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered poor predictions (marked with two asterisks).

Ligands	Transactivation activity			Ligands	Transactivation activity		
	Experimental activity	Predicted activity	Residual		Experimental activity	Predicted activity	Residual
10_17a	5.00	6.56	1.56**	10_34u	6.32	7.06	0.74
10_17j	5.00	6.51	1.51**	10_34b	6.43	6.89	0.46
10_17w	5.06	6.80	1.74**	10_34an	6.44	6.96	0.52
10_17i	5.10	6.43	1.33*	10_34i	6.49	6.80	0.31
10_17b	5.21	6.65	1.44**	10_17n	6.52	6.71	0.19
10_17v	5.25	6.98	1.73**	10_29g	6.52	6.98	0.46
10_17c	5.25	6.71	1.45**	10_34ae	6.52	6.75	0.23
10_29b	5.39	6.65	1.26*	10_34o	6.53	7.18	0.65
10_17e	5.52	6.50	0.98*	10_34j	6.55	7.00	0.44
10_17l	5.66	6.56	0.90*	10_34aa	6.62	7.72	1.10*
10_29i	5.74	6.56	0.82*	10_34t	6.67	7.10	0.43
10_17q	5.78	6.71	0.93*	10_34s	6.74	7.09	0.35
10_17r	5.78	6.79	1.01*	10_34am	6.75	7.24	0.49
10_17m	5.82	6.34	0.52	10_17x	6.80	6.93	0.13
10_17t	5.91	6.96	1.05*	10_34n	6.94	7.56	0.62
10_34l	5.92	7.00	1.08*	10_34m	7.03	7.66	0.64
10_17d	5.99	6.97	0.98*	10_34q	7.08	7.42	0.34
10_34ab	6.00	6.76	0.76	10_34af	7.16	7.13	−0.03
10_29f	6.01	6.73	0.71	10_34f	7.19	6.80	−0.39
10_34c	6.02	7.37	1.35*	10_34h	7.26	7.25	−0.01
10_17f	6.03	7.11	1.08*	10_34w	7.31	6.94	−0.37
10_34ai	6.04	7.19	1.15*	10_34a	7.32	6.80	−0.52
10_34ac	6.05	7.37	1.32*	10_34aj	7.35	7.16	−0.19
10_34d	6.06	7.37	1.31*	10_34e	7.35	7.13	−0.22
10_34ag	6.12	6.94	0.81*	10_34ak	7.36	7.74	0.39
10_29c	6.13	7.21	1.08*	10_34r	7.38	7.24	−0.14
10_17s	6.15	6.48	0.32	10_34ah	7.39	7.13	−0.26
10_29a	6.19	6.79	0.61	10_34al	7.47	7.41	−0.05
10_17g	6.20	6.43	0.24	10_34k	7.48	7.52	0.04
10_17o	6.24	6.54	0.30	10_34ad	7.57	7.22	−0.35
10_17u	6.26	6.67	0.41	10_34x	7.70	7.35	−0.35
10_29d	6.26	6.85	0.59	10_34g	7.74	7.41	−0.34
10_29h	6.26	6.90	0.64	10_34v	7.77	7.06	−0.71*
10_29e	6.31	6.90	0.58	10_34p	7.92	7.26	−0.66

and **ent-18**) out of 22 compounds have a weak prediction and only one compound for each model (*i.e.*, compound **11_2** for the transactivation model and compound **6** for the binding model) displayed a poor prediction. The poor predicted compounds are those with the lowest experimental values of pIC_{50} and those with the lowest and highest experimental values of pEC_{50} (see Fig. 6). This is most likely due to the fact that compounds with extreme experimental values are not well represented in the training set.

For a practical assessment of the study, the predictability of the 3D-QSAR models was evaluated using three external test sets of 45 tyrosine-derivative compounds [10–12], 6 thiazolidinediones [10] and 68 indanyacetic acid derivates [14]. The results of the predictions of the three external sets are shown in Tables 3–5. 54.7%, 33.3% and 12% of the compounds of the three external test sets were classified, respectively, as good, weak and poor predictions in the trans-activation 3D-QSAR model. 72.5%, 19.6% and 7.8% of the compounds of were classified, respectively, as good, weak and poor predictions in the binding 3D-QSAR model. In general, the predictions are good, although the models over-predict the values of the molecules that have the lowest experimental values (see Fig. 6). This is most likely due to the fact that compounds with extreme experimental values are not well represented in the training set. Other than this limitation, our models can reasonably predict the order of the activity, *i.e.*, the predicted activity of the molecules with the highest experimental values of pEC_{50} and pIC_{50} tend to be lower than the predicted activity of molecules with lower experimental values of pEC_{50} and pIC_{50}. This point is relevant because when the 3D-QSAR model is applied to the results of a virtual screening, it is more important to know which compounds have the highest activity values rather than knowing the exact activity value for each compound. Therefore, these results are very encouraging in view of future applications of the study aimed at (a) the priorization of analogues of active ligands resulting from virtual screening and (b) the optimization of known PPARγ full agonists.

Fig. 6 shows the representation of the 3D-QSAR models. In these types of figures, the cubes that represent the model are displayed and colored according to the sign of their coefficient values. Blue and red cubes are used for positive and negative coefficients, respectively, and indicate regions that increase or decrease the analyzed parameter. One of the advantages of using these representations is that the position of the cubes in the 3D-QSAR model can be compared with the positions of the amino acid residues in the active site. This might provide insight into which functional groups are desirable or undesirable at certain positions of a molecule. Fig. 7A shows the favorable and unfavorable regions for the transactivation activity. The favorable region for transactivation is mainly located in regions that interact with arm II of the LBD of PPARγ. The contribution of this region is mainly hydrophobic with some contribution of electron-withdraw (results not shown). This region corresponds to the effector part of PPARγ full agonists and is the most variable region in the molecules analyzed (see Fig. 4). Although this region is far away from Tyr473 and H12, it explains the differences in affinity and potency between very similar glitazones; *i.e.*, the larger the effector module, more hydrophobic interactions will occur with arm II of the LBD of PPARγ, thus stabilizing better the PPARγ–ligand complex. Fig. 6C and E display the cubes of the 3D-QSAR model grid that are occupied by two of the compounds analyzed. These representations show which parts of the ligand have a positive or negative contribution to the parameter analyzed, which is, in this case, the transactivation activity. Fig. 7C shows the **4_2** compound,

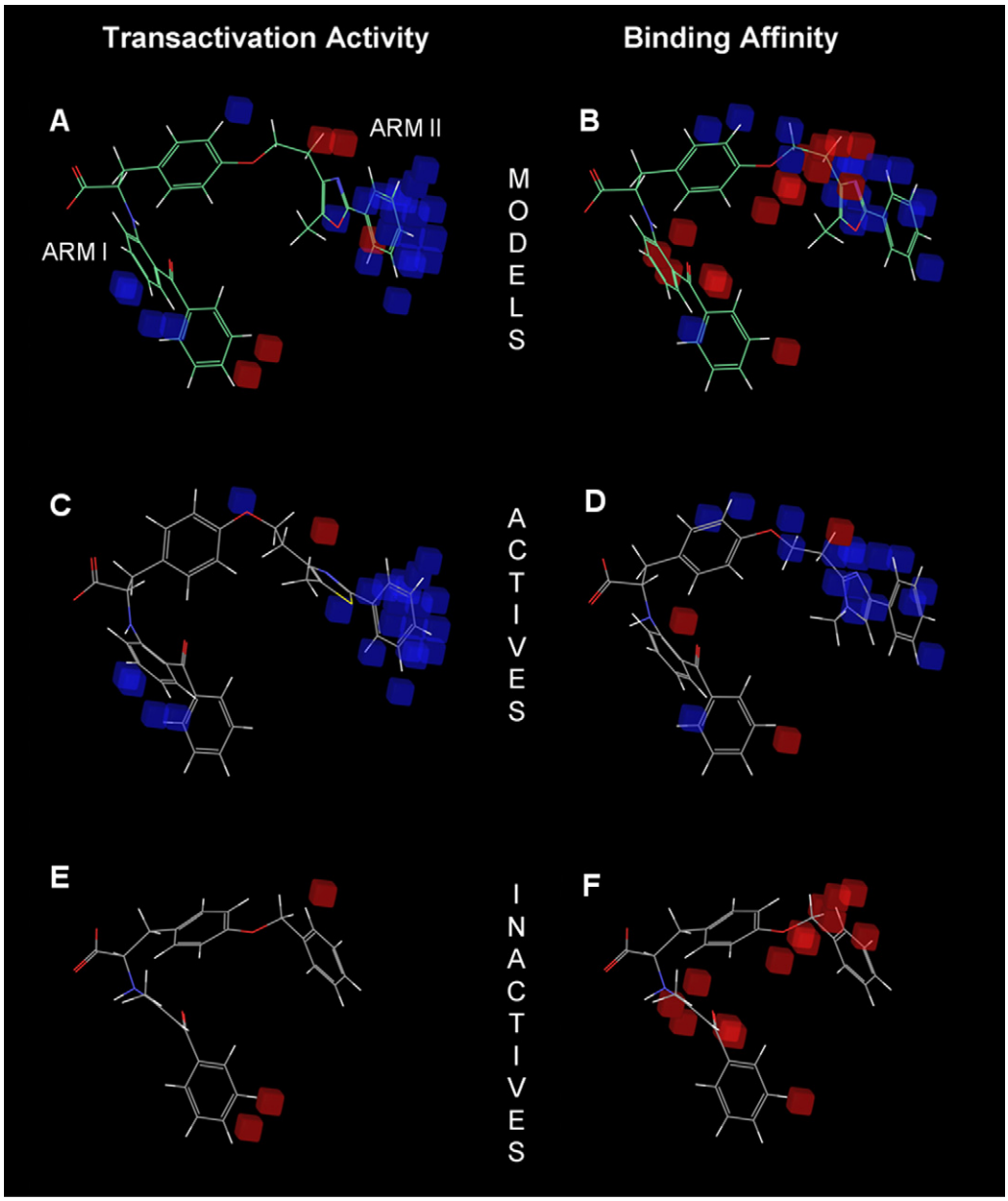

Fig. 7. Representation of the pEC_{50} and the (B) pIC_{50} models. Blue and red cubes indicate, respectively, regions that are favorable and unfavorable for the transactivation activity or binding affinity to the LBD of PPARγ. Panels (C) and (E) show the cubic volume elements that are occupied by one of the compounds of the series with the highest (*i.e.*, **4_2**) and the lowest (*i.e.*, **ent-2**) transactivation activity, respectively. Panels (D) and (F) show the cubic volume elements that are occupied by one of the compounds of the series with the highest (*i.e.*, **7_2**) and the lowest (*i.e.*, **ent-2**) binding affinity, respectively. All panels are presented in the same relative orientation to allow for an easier comparison.

which has one of the highest transactivation activities in this series (pEC_{50} = 10). Our 3D-QSAR model explains the high transactivation activity of this compound because it provides a heterocyclic and a phenyl ring at the effector module. In comparison, the compound from our ligand dataset with the lowest transactivation activity (*i.e.*, **ent-2** with pEC_{50} = 4.66 in Fig. 7E) only interact with arm II through one phenyl ring. Moreover, this compound has a (R)-configuration, and it is known that compounds with an (S)-configuration, derived from naturally occurring L-tyrosine, are more active as PPARγ full agonists [10].

Fig. 7B, D and F shows a representation of the binding model. In this model, the binding affinity of PPARγ is the analyzed variable. Fig. 7B shows the favorable and unfavorable regions for binding. The favorable region is again located at arm II, and the main contribution to the binding affinity of PPARγ is also caused by hydrophobic interactions (results not shown). The unfavorable regions are located at arm I and at the beginning of arm II. Regarding the first unfavorable region, it occupies the place of one internal hydrogen bond in the 2-aminobenzophenone moiety. Therefore, the presence of some hydrophobic residues at this position will not allow the formation of this internal interaction. The second unfavorable region is due to steric clashes with residues Arg288, Leu330 and Ile341. Fig. 6D and F show the 3D-QSAR models represented only by the cubic volume elements that are occupied by one of the compounds with the highest (*i.e.*, the **7_2** compound with a pIC_{50} of 9.16) and lowest (*i.e.*, the **ent-2** compound with a pIC_{50} of 5.5) binding affinities, respectively. The **7_2** compound fits the blue areas of the 3D-QSAR model perfectly, as it can interact with the receptor through the favorable regions at arm II (see Fig. 7D). However, the **ent-2** compound does not occupy all of arm II (see Fig. 7F), and it is

not able to make the favorable interactions shown as blue cubes in Fig. 7.

4. Conclusion

In this work, a structure-based docking strategy was used for aligning the molecules, and highly predictive 3D-QSAR transactivation and binding models were developed for PPARγ full agonists. These models match well with the known features of the different parts of the PPARγ binding site and show that the binding portion of the PPARγ agonists is essential for the transactivation activity of PPARγ. The hydrophobic interactions between the effector portion and the receptor are also important for increasing the transactivation activity of PPARγ.

Acknowledgements

This manuscript was edited for English language fluency by American Journal Experts. This study was supported by grants AGL2008-00387/ALI and AGL2011-25831/ALI from the "Ministerio de Educación y Ciencia" of the Spanish Government and the ACC1Ó (TECCT11-1-0012) program (Generalitat de Catalunya). The authors wish to thank the "Servei de Disseny de Fàrmacs" (Pharmaceutical Design Service) of the Center for Scientific and Academic Services of Catalonia (CESCA) for providing access to the Schrödinger software.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2012.03.001.

References

[1] A. Yessoufou, W. Wahli, Multifaceted roles of peroxisome proliferator-activated receptors (PPARs) at the cellular and whole organism levels, Swiss Med. Wkly. 140 (2010) w13071.
[2] S. Garcia-Vallve, J. Palau, Nuclear receptors, nuclear-receptor factors, and nuclear-receptor-like orphans form a large paralog cluster in Homo sapiens, Mol. Biol. Evol. 15 (1998) 665–682.
[3] J. Berger, D.E. Moller, The mechanisms of action of PPARs, Annu. Rev. Med. 53 (2002) 409–435.
[4] T.M. Willson, M.H. Lambert, S.A. Kliewer, Peroxisome proliferator-activated receptor gamma and metabolic disease, Annu. Rev. Biochem. 70 (2001) 341–367.
[5] J.M. Lehmann, L.B. Moore, T.A. Smith-Oliver, W.O. Wilkison, T.M. Willson, S.A. Kliewer, An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma), J. Biol. Chem. 270 (1995) 12953–12956.
[6] L. Michalik, W. Wahli, Peroxisome proliferator-activated receptors (PPARs) in skin health, repair and disease, Biochim. Biophys. Acta 1771 (2007) 991–998.
[7] P. Sertznig, M. Seifert, W. Tilgen, J. Reichrath, Peroxisome proliferator-activated receptors (PPARs) and the human skin: importance of PPARs in skin physiology and dermatologic diseases, Am. J. Clin. Dermatol. 9 (2008) 15–31.
[8] R. Behshad, K.D. Cooper, N.J. Korman, A retrospective case series review of the peroxisome proliferator-activated receptor ligand rosiglitazone in the treatment of atopic dermatitis, Arch. Dermatol. 144 (2008) 84–88.
[9] D. Schadendorf, Peroxisome proliferator-activating receptors: a new way to treat melanoma? J. Invest. Dermatol. 129 (2009) 1061–1063.
[10] B.R. Henke, S.G. Blanchard, M.F. Brackeen, K.K. Brown, J.E. Cobb, J.L. Collins, W.W. Harrington, M.A. Hashim, E.A. Hull-Ryde, I. Kaldor, S.A. Kliewer, D.H. Lake, L.M. Leesnitzer, J.M. Lehmann, J.M. Lenhard, L.A. Orband-Miller, J.F. Miller, R.A. Mook, S.A. Noble, W. Oliver, D.J. Parks, K.D. Plunket, J.R. Szewczyk, T.M. Willson, N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 1. Discovery of a novel series of potent antihyperglycemic and antihyperlipidemic agents, J. Med. Chem. 41 (1998) 5020–5036.
[11] J.L. Collins, S.G. Blanchard, G.E. Boswell, P.S. Charifson, J.E. Cobb, B.R. Henke, E.A. Hull-Ryde, W.M. Kazmierski, D.H. Lake, L.M. Leesnitzer, J. Lehmann, J.M. Lenhard, L.A. Orband-Miller, Y. Gray-Nunez, D.J. Parks, K.D. Plunkett, W.Q. Tong, N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 2. Structure–activity relationship and optimization of the phenyl alkyl ether moiety, J. Med. Chem. 41 (1998) 5037–5054.
[12] J.E. Cobb, S.G. Blanchard, E.G. Boswell, K.K. Brown, P.S. Charifson, J.P. Cooper, J.L. Collins, M. Dezube, B.R. Henke, E.A. Hull-Ryde, D.H. Lake, J.M. Lenhard, W. Oliver, J. Oplinger, M. Pentti, D.J. Parks, K.D. Plunket, W.Q. Tong, N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 3. Structure–activity relationship and optimization of the N-aryl substituent, J. Med. Chem. 41 (1998) 5055–5069.
[13] S. Azukizawa, M. Kasai, K. Takahashi, T. Miike, K. Kunishiro, M. Kanda, C. Mukai, H. Shirahase, Synthesis and biological evaluation of (S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acids: a novel series of PPAR gamma agonists, Chem. Pharm. Bull. (Tokyo) 56 (2008) 335–345.
[14] J. Rudolph, L. Chen, D. Majumdar, W.H. Bullock, M. Burns, T. Claus, F.E. Dela Cruz, M. Daly, F.J. Ehrgott, J.S. Johnson, J.N. Livingston, R.W. Schoenleber, J. Shapiro, L. Yang, M. Tsutsumi, X. Ma, Indanylacetic acid derivatives carrying 4-thiazolyl-phenoxy tail groups, a new class of potent PPAR alpha/gamma/delta pan agonists: synthesis, structure–activity relationship, and in vivo efficacy, J. Med. Chem. 50 (2007) 984–1000.
[15] R.T. Gampe, V.G. Montana, M.H. Lambert, A.B. Miller, R.K. Bledsoe, M.V. Milburn, S.A. Kliewer, T.M. Willson, H.E. Xu, Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors, Mol. Cell 5 (2000) 545–555.
[16] V. Zoete, A. Grosdidier, O. Michielin, Peroxisome proliferator-activated receptor structures: ligand specificity, molecular switch and interactions with regulators, Biochim. Biophys. Acta 1771 (2007) 915–925.
[17] L. Guasch, E. Sala, C. Valls, M. Blay, M. Mulero, L. Arola, G. Pujadas, S. Garcia-Vallve, Structural insights for the design of new PPARgamma partial agonists with high binding affinity and low transactivation activity, J. Comput. Aided Mol. Des. 25 (2011) 717–728.
[18] C. Giaginis, S. Theocharis, T.K. Anna, A QSAR Study on indole-based PPARg agonists in respect to receptor binding and gene transactivation data, QSAR Comb. Sci. 28 (2009) 802–805.
[19] L. Rathi, S.K. Kashaw, A. Dixit, G. Pandey, A.K. Saxena, Pharmacophore identification and 3D-QSAR studies in N-(2-benzoyl phenyl)-L-tyrosines as PPAR gamma agonists, Bioorg. Med. Chem. 12 (2004) 63–69.
[20] C. Liao, A. Xie, L. Shi, J. Zhou, X. Lu, Eigenvalue analysis of peroxisome proliferator-activated receptor gamma agonists, J. Chem. Inf. Comput. Sci. 44 (2004) 230–238.
[21] K.H. Hyun, D.Y. Lee, B. Lee, C.K. Kim, Receptor-based 3D QSAR studies on PPAR-gamma agonists using CoMFA and CoMSIA approaches, QSAR Comb. Sci. 23 (2004) 637–649.
[22] Y. Kurogi, Three-dimensional quantitative structure–activity relationships (3D-QSAR) of antidiabetic thiazolidinediones, Drug Des. Discov. 16 (1999) 109–118.
[23] M. Kapoor, M. McCann, S. Liu, K. Huh, C.P. Denton, D.J. Abraham, A. Leask, Loss of peroxisome proliferator-activated receptor gamma in mouse fibroblasts results in increased susceptibility to bleomycin-induced skin fibrosis, Arthritis Rheum. 60 (2009) 2822–2829.
[24] C. Rücker, M. Scarsi, M. Meringer, 2D QSAR of PPARgamma agonist binding and transactivation, Bioorg. Med. Chem. 14 (2006) 5178–5195.
[25] S.L. Dixon, A.M. Smondyrev, E.H. Knoll, S.N. Rao, D.E. Shaw, R.A. Friesner, PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results, J. Comput. Aided Mol. Des. 20 (2006) 647–671.