European Journal of Medicinal Chemistry 44 (2009) 3488–3495

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Important pharmacophoric features of pan PPAR agonists: Common chemical feature analysis and virtual screening

Sandeep Sundriyal, Prasad V. Bharatam*

Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S.A.S. Nagar, Punjab 160 062, India

ARTICLE INFO

Article history:
Received 10 June 2008
Received in revised form
15 January 2009
Accepted 22 January 2009
Available online 31 January 2009

Keywords:
Pan PPAR agonists
HipHop
Pharmacophore
Docking
Virtual screening

ABSTRACT

HipHop program was used to generate a common chemical feature hypothesis for pan Peroxisome Proliferator-Activated Receptor (PPAR) agonists. The top scoring hypothesis (*hypo-1*) was found to differentiate the pan agonists (actives) from subtype-specific and dual PPAR agonists (inactives). The importance of individual features in *hypo-1* was assessed by deleting a particular feature to generate a new hypothesis and observing its discriminating ability between 'actives' and 'inactives'. Deletion of aromatic features AR-1 (*hypo-1b*), AR-2 (*hypo-1e*) and a Hydrophobic feature HYD-1 (*hypo-1c*) individually did not affect the discriminating power of the *hypo-1* significantly. However, deletion of a Hydrogen Bond Acceptor (HBA) feature (*hypo-1f*) in the hydrophobic tail group was found to be highly detrimental for the specificity of *hypo-1* leading to high hit rate of 'inactives'. Since *hypo-1* did not produce any useful hits from the database search, *hypo-1b*, *hypo-1c* and *hypo-1e* were used for virtual screening leading to the identification of new potential pan PPAR ligands. The docking studies were used to predict the binding pose of the proposed molecules in PPARγ active site.

1. Introduction

Pan Peroxisome Proliferator-Activated Receptor (PPAR) agonists are emerging as effective agents for the treatment of metabolic disorders including type 2 diabetes and obesity [1–5]. These agents act on all the three subtypes (PPARα, PPARγ and PPARδ) of PPARs in contrast to the subtype-specific or dual PPAR agonist which acts on one or two subtypes, respectively [6,7]. Ligand-based computational methods such as 3D-QSAR and pharmacophore mapping techniques have been used to understand the important chemical features of subtype-specific and dual PPAR ligands [8–14]. On the other hand, similar studies have not been reported in the case of pan PPAR agonists, except for our recent attempt to develop a CoMFA model using 'sum of activities' as a dependent parameter [15]. This may be attributed to the fact that only a few reports describe the design and synthesis of pan PPAR agonists possessing potent activity at all the three subtypes.

Pharmacophore models are important tools for virtual screening [16] for quickly searching large databases of molecules for desired compounds with therapeutic activity. In the case of PPAR, virtual screening has been utilized successfully to predict new active compounds [17–19], e.g. Scarsi et al. employed docking based virtual screening which predicted that sulfonyl ureas and glinides can act as PPARγ ligands [17]. Interestingly, these classes of molecules were already known for antidiabetic activity due to the binding to the sulfonylurea receptor. The *in silico* results were further confirmed by experimental studies.

The comparison of structures of some of the potent pan agonists suggests the existence of high degree of structural similarity in these molecules. For example, molecules **1**–**5** (Fig. 1) are composed of an acidic head group (AHG) that is further attached to hydrophobic tail group (HTG) through a linker with varying length and high flexibility. A molecule interacts with the receptor through various chemical features such as Hydrophobic (HYD), Aromatic Rings (ARs) Hydrogen Bond Acceptor–donors (HBA/D), etc. A number of such features are present within the AHG and HTG of the pan agonists (Fig. 1). It is highly probable that particular number of each feature and its 3D arrangement with respect to other features are important in determining the activity of these molecules on all the PPAR subtypes. Although, some features may be important only for acting on a particular subtype, together all the features are expected to help pan agonists to access and interact with the active sites of all the three PPAR subtypes. It would be interesting to find out which feature might be important in differentiating pan agonists from subtype-specific and dual PPAR agonists. Establishing the significance of each feature in the pan agonists would further aid in designing new such agents. However, in order to understand the importance of these features it is essential to first find out all

* Corresponding author. Tel.: +91 172 2214682; fax: +91 172 2214692.
E-mail address: pvbharatam@niper.ac.in (P.V. Bharatam).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.01.024

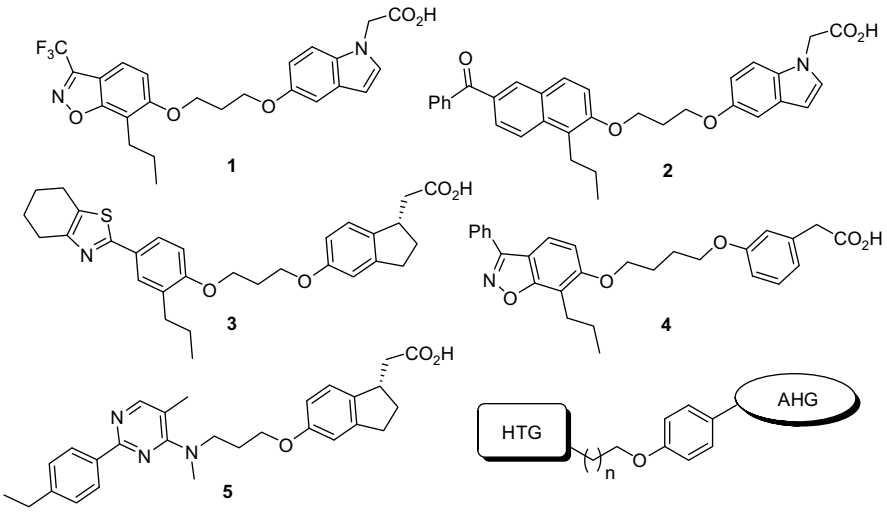

Fig. 1. Structures of various PPAR pan agonists used as training set and visualization of their general pharmacophoric features in 2D.

the commonly occurring important features in the pan agonists. It was recognized that HipHop algorithm can be used for this problem, which can identify a common 3D configuration of the chemical features shared among a set of input compounds in an automated manner [20].

2. Computational methods

The HipHop module of CATALYST 4.10 software [21] installed on a Silicon Graphics Octane2 workstation with IRIX 6.5 operating system was employed to develop the common chemical features hypothesis using a set of 5 pan PPAR agonists (**1**–**5**; Fig. 1; most active on all the three PPAR subtypes). The conformation models for all the molecules were generated using 'Best' option in CATALYST that used poling algorithm for conformer selection and CHARMm force fields for minimizations [22–25]. A maximum of 250 conformations within the energy range of 20 kcal/mol of estimated global minima was demanded. The molecule **4** having potent EC_{50} values at all the three subtypes was assigned a principal value of 2 and MaxOmitFeat (MOF) value of 0 to ensure that all features of this molecule are considered for generation of the hypothesis. For other molecules of the training set, principal and MOF values were set as 1. All the default settings were maintained in the 'advanced job options' of the program. The thiazolidinedione (TZD) ring was included in the definition of Negatively Ionizable (NI) feature using 'Exclude/OR' quick tool of CATALYST. In addition, Hydrogen Bond Acceptor (HBA) feature definition was also modified to include the 'F' atom using the same procedure. Thus, the modified NI and HBA features were selected in addition to hydrophobic (HYD) and Aromatic Ring (AR) features for generation of the hypothesis. Since the acidic group makes three important H-bonds with His323, Tyr473 and His449 in the PPARγ active site [26], the minimum of one NI feature was demanded in the final hypothesis.

Final model was validated by demonstrating its ability to pick maximum number of pan agonists (actives) and rejecting maximum number of subtype-specific or dual PPAR ligands (inactives) from the respective databases. In the present work, a pan agonist was defined as the one possessing EC_{50} less than 1000 nM at all the subtypes in a standard transactivation assay. These were further sub-classified into 'highly active', 'active' and 'moderately active' based on EC_{50} values at the three subtypes. A 'highly active' molecule should have the $EC_{50} \leq 200$ nM at all the three subtypes. In case the EC_{50} against any subtype was greater than 200 nM but $\leq$500 nM, the molecule was defined as 'active'. Similarly, if EC_{50} against any subtype exceeded 500 nM but $\leq$1000 nM, the molecule was called as 'moderately active'. Therefore, within the training set, **3**–**5** could be classified as 'highly active' as they exhibited $EC_{50} \leq 200$ nM at all the three subtypes. Similarly, **1** and **2** were found to belong to 'active' class since the EC_{50} of these molecules exceeded 200 nM (but $\leq$500 nM) for PPARγ and PPARδ subtypes, respectively.

For a clear cut distinction, inactive molecule was defined as the one having less than 20% efficacy at a concentration of 10,000 nM at any subtype compared to the standard molecule. This definition is generally accepted in various literature reports describing the evaluation of PPAR ligands. Thus, a number of selective and dual ligands are included in this class. Finally, a number of literature reports were studied in order to identify 'actives' and 'inactives' classes of molecules according to the aforementioned definitions. The following selection criteria were followed in order to make databases of molecules:

(i) Only those molecules were selected for which PPAR transactivation data (EC_{50}) is reported using cell based reporter gene efficacy assay [1–5]. Thus, the molecules for which only binding affinity (in the form IC_{50} and K_i values) has been reported were not added to any database.
(ii) Only full agonists were used since the partial agonists might not share the similar interactions and binding modes as compared to the full agonists. This is evident from some of the reported crystal structures of partial agonists with PPAR [19,27–29].
(iii) The molecules tested as racemates were not used for the present study since it was not clear which enantiomer should be taken as active.
(iv) The molecules were selected from the various recent research publications (see Supplementary information for the sources). The data from the patents was not included.

The conformation models of all the molecules were generated using the same criteria as mentioned for the training set molecules and stored as spreadsheets. A total of 66 molecules were found to meet the aforementioned definition of a pan PPAR agonist which were further divided into 'highly active', 'active', and 'moderately active' categories as defined earlier. On the other hand, a large number of subtype-specific and dual ligands (inactives) have been

reported in literature, out of which a total of 122 were found to meet the aforementioned criteria. Although, efforts were made to cover all the important literature reports, possibility of missing a few cannot be discarded and thus, databases are not supposed to be exhaustive in nature. Nonetheless, these databases represent the corresponding classes of the pan agonists (actives) and subtype-specific or dual ligands (inactives) within the followed criteria.

Docking studies were carried out using the FlexX program [30] interfaced with SYBYL6.9 [31] installed on a Silicon Graphics Octane2 workstation with IRIX 6.5 operating system. In this automated docking program, the flexibility of the ligands is considered while the protein or biomolecule is considered as a rigid structure. The 3D coordinates of the active sites were taken from the X-ray crystal structures of the PPARγ (PDB code 2F4B) reported as a complex with agonist **1** [3]. The active site was defined as the area within 8.0 Å around the co-crystallized ligand, and customization was done for the His423, Tyr473 and His449 so as to represent the reported H-bonding interactions. Formal charges were assigned to all the molecules and FlexX run was submitted. The hypotheses were used to search the NCI2000 database interfaced with CATALYST and the hit molecules were transferred as a spreadsheet to the SYBYL environment in SD file format and docked using 'multiple ligand docking' protocol.

3. Results and discussion

3.1. Hypothesis generation

In general, most active and diverse set of compounds are used to generate a pharmacophore model. In the current study, these conditions were found to be satisfied by compounds **1**–**5** (Fig. 1). The HipHop program was used to find out common features among these five potent pan PPAR agonists [1–5]. Though the bioactive conformation of molecules **1** and **2** is available as X-ray complex with the PPARγ (PDB codes 2ATH and 2F4B respectively) [2,3], it was not used for the hypothesis generation because the pharmacophore model being sought is not only for the PPARγ but for all the three subtypes. Thus, the conformational models were generated for all the training set molecules.

The HipHop evaluates active molecules based on their chemical features together with their ability to adopt a conformation that allows those features to be superimposed on a common configuration. The principal value and MOF are the user defined parameters that allow complete or partial mapping of a training set molecule on to the generated hypothesis. Various hypotheses are built and ranked based on how well the training set molecules superimpose on them. The most active compound **4** (having least sum of EC_{50} values on all subtypes) was given principal = 2 and MOF = 0 to ensure that it had strongest influence on the model building phase. The HipHop with these settings produced top ten hypotheses (*hypo-1* to *hypo-10*) details of which have been specified in Table 1. The top ranked hypothesis (*hypo-1*) scored 103.46 and consisted of one each of NI and HBA features in addition to two AR and three HYD features. The next three hypotheses (*hypo-2* to *hypo-4*) also had identical types and numbers of chemical features with slightly different relative positions in space and lesser score as compared to *hypo-1*. The *hypo-1* was selected for further studies as it possessed the highest score among all the hypotheses. The mapping of training set molecule **1** to the *hypo-1* is shown in Fig. 2 as an example. The carboxyl group in indole moiety mapped to the NI feature while the 5-membered ring of indole mapped to the AR-1 feature lying near to NI. The HYD-1 feature mapped to the linker chain between AHG and HTG. The HYD-2 and HYD-3 features mapped to the *n*-Pr group and $-CF_3$ group in HTG, respectively. The benzisoxazole ring in the HTG mapped to the AR-2 (benzene ring) and HBA (N-atom) features. The 'Best Fit' values of the mappings were obtained by either forcing the training set molecules to map all the features (BestFit-0) or by allowing them to miss one feature (BestFit-1). Expectedly, highest fit value of 7.00 was obtained with the reference molecule **4** having the lowest sum of EC_{50} for all the PPAR subtypes (Table 2). All molecules of the training set mapped all the features of *hypo-1* when forced to map all the features (BestFit-0). However, when allowed to miss one feature (BestFit-1) molecules **2** and **5** missed the HYD-2 and HBA features with improved fit value compared to BestFit-0.

Table 1
Details of the top ten hypotheses generated using HipHop.

Hypothesis	Features[a]	Ranking score[b]	Direct hit (DH) and Partial hit (PH)[c]
hypo-1	NI, HBA, 2 × AR, 3 × HYD	103.460	DH: 11111 PH: 00000
hypo-2	NI, HBA, 2 × AR, 3 × HYD	100.167	DH: 01111 PH: 10000
hypo-3	NI, HBA, 2 × AR, 3 × HYD	100.167	DH: 01111 PH: 10000
hypo-4	NI, HBA, 2 × AR 3 × HYD	99.818	DH: 01111 PH: 10000
hypo-5	NI, HBA, AR, 4 × HYD	99.418	DH: 11111 PH: 00000
hypo-6	NI, AR, HBA, 4 × HYD	99.112	DH: 11111 PH: 00000
hypo-7	NI, HBA, 3 × HYD, 2 × AR	98.750	DH: 01111 PH: 10000
hypo-8	NI, HBA, 5 × HYD	98.708	DH: 11111 PH: 00000
hypo-9	NI, AR, HBA, 4 × HYD	98.431	DH: 11111 PH: 00000
hypo-10	NI, HBA, 2 × AR 3 × HYD	97.697	DH: 01111 PH: 10000

[a] AR = Aromatic Ring; NI = Negatively Ionizable; HYD = Hydrophobic; HBA = Hydrogen Bond Acceptor.
[b] Higher the ranking score, lesser the probability of chance correlation. The best hypothesis shows the highest value.
[c] DH indicates whether (1) or not (0) a training set molecule mapped every feature of the hypothesis. Similarly, PH indicates whether (1) or not (0) a molecule mapped to all but one feature in the hypothesis. The numbers (from right to left) correspond to the compounds (from top to bottom) as given in Table 2.

3.2. Hypothesis validation

In the present work the final hypothesis model was said to be valid if it could pick maximum number of pan agonists (actives) and

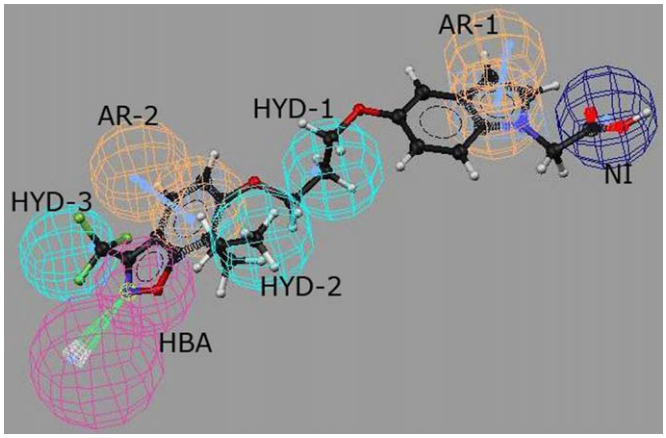

Fig. 2. The mapping of **1** to various chemical features of the *hypo-1*. The blue, orange, cyan and pink spheres represent the Negatively Ionizable (NI), Aromatic Ring (AR), Hydrophobic (HYD) and Hydrogen Bond Acceptor (HBA) features, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Biological activities and Best Fit values of training set molecules.

Compound	EC_{50}[a] (nM)			BestFit-0[b]		BestFit-1[d]		Omitted feature[e]
	PPARα	PPARγ	PPARδ	Fit value	ΔE[c] (kcal/mol)	Fit value	ΔE[c] (kcal/mol)	
1	14	230	10	6.12	17.0	6.12	17.0	–
2	8	70	500	5.22	6.8	5.58	2.1	HYD-2
3	43	12	1.9	5.47	7.5	5.47	7.5	–
4	5	6	20	7.00	15.4	7.00	15.4	–
5	200	145	10	3.46	17.0	4.40	17.1	HBA

[a] Effective concentration to produce 50% of functional activity in a standard functional assay.
[b] Best Fit value obtained using MOF = 0, i.e. molecules must map all the features of the hypothesis. The 'fit' values (with no units) show how well a molecule maps to the given hypothesis, higher the fit value better the molecule maps to the hypothesis. Fit values against the pharmacophore, calculated as in the following: Fit = Σmapped hypothesis features × $W[1 - \Sigma(\text{disp/tol})^2]$, where 'Σmapped hypothesis features' is the number of pharmacophore features that successfully superimpose corresponding chemical moieties within the fitted compound, '*W*' is the weight of the corresponding hypothesis feature spheres. This value is fixed to 1.0 in HipHop-generated models. 'disp' is the distance between the center of a particular pharmacophoric sphere and the center of the corresponding superimposed chemical moiety of the fitted compound; 'tol' is the radius of the pharmacophoric feature sphere (known as Tolerance, equals 1.6 Å by default).
[c] Value given here is the energy difference (kcal/mol) between the Best Fit conformer and the estimated global minima for the same molecule.
[d] Best Fit value obtained using MOF = 1, i.e. molecule is allowed to miss one feature. The molecules **2** and **5** show improved 'fit' values after missing a feature, with these settings.
[e] Feature not mapped to the molecule in the case of BestFit-1.

reject maximum number of subtype-specific or dual PPAR ligands (inactives) from the respective databases. However, standard definitions for selective, dual or pan PPAR agonists do not exist in terms of their activity on the different subtypes. Consequently, it is not surprising that the same molecule is sometimes referred as a dual agonist or pan agonist in literature by two different research groups. Thus, for the present work a pan PPAR agonist and subtype-specific (and dual) PPAR agonists were distinguished clearly by specifying their range of activities on all the three subtypes in a standard assay (see computational methods).

Firstly, model *hypo-1* was used to screen the databases of 'actives' (66) and 'inactives' (122). Out of the 23 'highly active' molecules, 21 (91.3%) were picked by *hypo-1* (Table 3). However, the hit percentage decreased significantly for 'active' and 'moderately active' classes of agonists with percentage hit of 64.7% and 42.3%, respectively. This showed that the model *hypo-1* is capable of selecting more number of 'highly active' class of pan agonists compared to 'active' and 'moderately active' classes. Interestingly, *hypo-1* picked only 7.4% of 'inactive' molecules showing its ability to differentiate between 'actives' and 'inactives'. The overall hit rate for the actives was found to be 65.2%, which is about 9 times more than that of inactives. Thus, the model *hypo-1* was found to discriminate not only actives and inactives but also the various subclasses of actives.

3.3. *Comparison of* hypo-1 *with PPAR active sites*

Various chemical features in the *hypo-1* (Fig. 2) were also correlated with the active site of the PPARs. The X-ray co-crystallized complexes of only two pan agonists **1** and **2** are available in literature with PDB codes 2ATH and 2F4B, respectively [2,3]. However, both of these ligands are reported only in the PPARγ active site and their precise interactions in PPARα and PPARδ active sites are not known. Nonetheless, the high number of HYD and AR features in the *hypo-1* is in agreement with the overall hydrophobic nature of the active site of the PPARs [2,3,26,32–34]. Also, an acidic group is present in many potent PPARδ [32,33] and PPARα [34] specific agonists and is known to display important H-bonds within the active sites, as is the case with PPARγ ligands [2,3,26]. The presence of an NI feature in all the potent pan agonists (represented by *hypo-1*) ensures that important H-bonding interactions are maintained in all the three subtypes of PPARs. Finally, all chemical features of *hypo-1* were found to correspond to the reported interactions of pan agonists **1** and **2** in the PPARγ active site. For example in the case of **1** (PDB code 2F4B), the –COOH group mapped to the NI feature, displayed H-bonds with His323, His449, Tyr473 and Ser289. The indole ring mapped to the AR-1 feature, formed strong hydrophobic interactions with His449, Cys285 and Phe282. The linker part (mapped to the HYD-1) was found to be close to the hydrophobic pocket of PPAR consisting of Leu300, Met334, Phe368, Val339, and Met364. The aromatic moiety in the HTG part mapped to AR-2 and also displayed the hydrophobic interactions with Cys285 and Ile341. The *n*-propyl group of the HTG part that mapped to HYD-2, provided an additional interaction with Arg288, while the $-CF_3$ group of HTG portion (mapped by HYD-3) exhibited close hydrophobic contacts with Ile281, Ile341 and Met348. However, specific role of the HBA feature, mapped by the N-atom (in the case of **1**) and carbonyl O-atom (in the case of **2**) in the HTG was not clear in these complex structures. No H-bond was observed with these atoms in the PPARγ crystal structures, yet in the case of **1** the N-atom is at a distance of 4.606 Å from the water molecule (residue 64). However, the X-ray crystal structure is just a snapshot of a more dynamic and complex ligand–receptor interaction phenomenon. Some features of a ligand may be unimportant for binding to the active site but may be essential for the initial interactions with the receptor that bring both the partners close to each other. Alternatively, some features may just assist the ligand to bind at the site of entry and help the ligand to access the active site of the macromolecular target. Notably, some PPARγ selective agonists such as rosiglitazone or pioglitazone also contain an HBA feature in the HTG but they do not possess all the features of the *hypo-1* [35]. Most likely, HBA and particular 3D arrangement of other features in the HTG of pan agonists might be helping these ligands to enter active sites of all the three PPAR subtypes. However, only speculation can be made at this stage that warrants further experimental proof.

Table 3
Database screening results with different hypotheses.

Hypotheses (deleted features)	Number of hits			
	Highly active (total = 23)	Active (total = 17)	Moderately active (total = 26)	Inactive (total = 122)
hypo-1 (none)	21 (91.3%)	11 (64.7%)	11 (42.3%)	9 (7.4%)
hypo-1a (NI)	22 (95.6%)	13 (76.5%)	21 (80.8%)	20 (16.4%)
hypo-1b (AR-1)	22 (95.6%)	13 (76.5%)	16 (61.5%)	12 (9.8%)
hypo-1c (HYD-1)	21 (91.3%)	12 (70.6%)	13 (50.0%)	11 (9.0%)
hypo-1d (HYD-2)	22 (95.6%)	14 (82.4%)	14 (53.8%)	22 (18.0%)
hypo-1e (AR-2)	21 (91.3%)	11 (64.7%)	13 (50.0%)	11 (9.0%)
hypo-1f (HBA)	22 (95.6%)	13 (76.5%)	19 (73.1%)	33 (27.0%)
hypo-1g (HYD-3)	22 (95.6%)	12 (70.6%)	14 (53.8%)	19 (15.6%)
hypo-1h (HYD-1 and AR-2)	22 (95.6%)	12 (70.6%)	15 (57.7%)	16 (13.1%)
hypo-1i (AR-1 and AR-2)	22 (95.6%)	14 (82.4%)	17 (65.4%)	18 (14.7%)
hypo-1j (HYD-1 and AR-1)	22 (95.6%)	16 (94.1%)	19 (73.1%)	22 (18.0%)

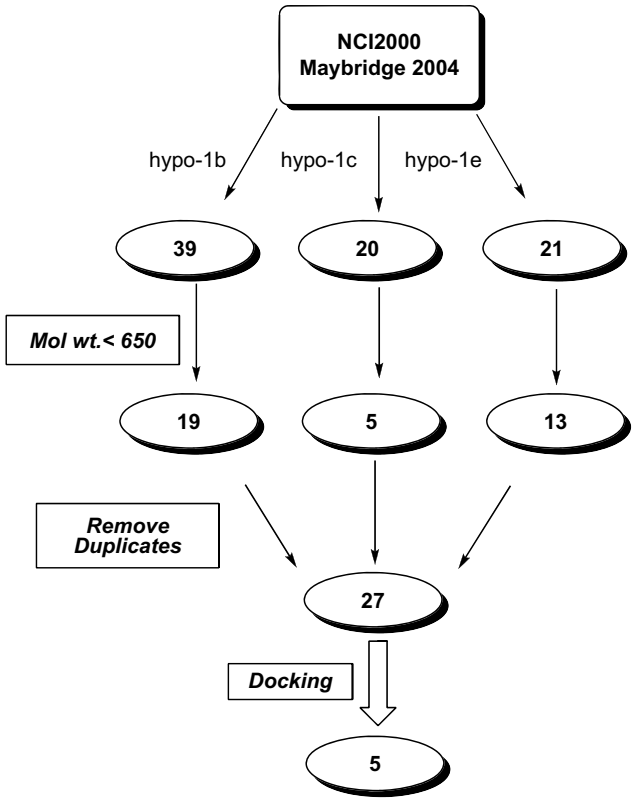

Fig. 3. Virtual screening strategy using *hypo-1b*, *hypo-1c* and *hypo-1e*. The numbers in ovals represent the hit molecules obtained after each step of virtual screening.

3.4. Assessment of individual features of hypo-1

Objective of the present work was to identify the important pharmacophoric features of pan PPAR agonists. The model *hypo-1* was found to represent the common features among pan PPAR agonists and was able to differentiate between the 'actives' and 'inactives' (as defined earlier). Nevertheless, it is possible that all the features may not be contributing significantly towards the pan agonistic activity. In the similar manner, some of the features might be more important for the activity as compared to others. It was reasoned that the importance of a chemical feature in the *hypo-1* can be obtained by deleting it from the *hypo-1* and then observing the hit rates for 'actives' and 'inactives' with the modified hypothesis. Thus, modified versions of the *hypo-1* were generated by deleting various chemical features from it one at a time while retaining the relative positions of others. For example, NI feature was deleted from *hypo-1* to give *hypo-1a* and AR-1 feature was deleted from *hypo-1* to give *hypo-1b* and so on (Table 3). These modifications of the *hypo-1* resulted in seven new hypotheses (*hypo-1a* to *hypo-1g*), which were used to screen the databases of actives and inactives and the results were compared with the *hypo-1* (Table 3).

As evident from Table 3, none of the modified hypotheses was better than the *hypo-1* in discriminating the actives and inactives. However, models *hypo-1b*, *hypo-1c* and *hypo-1e* were found to be comparable to the *hypo-1* for the overall hit rate of actives and inactives. This showed that the AR-1, HYD-1 and AR-2 may not be important features for discriminating the pan agonists from the subtype-specific or dual PPAR agonists. When two of the AR-1, HYD-1 or AR-2 features were deleted at a time (*hypo-1h* to *hypo-1j*), the discriminating ability of the corresponding hypotheses decreased significantly compared to *hypo-1*. This demonstrates that

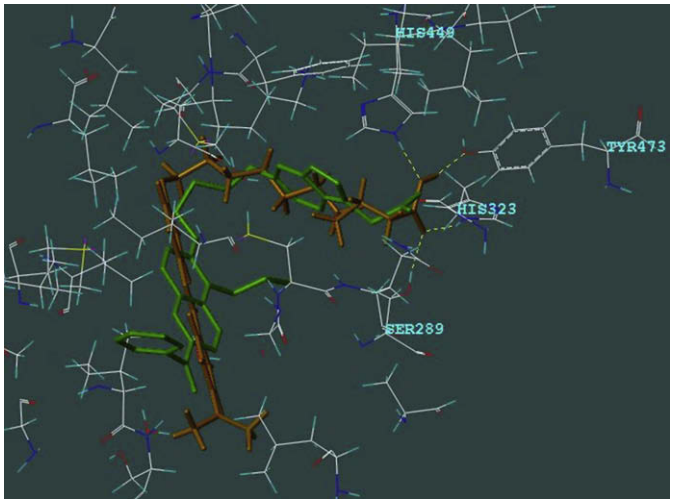

Fig. 4. Docked conformation of **6** (orange colored) together with the co-crystallized ligand (**2**, green colored) (PDB entry 2F4B) and important amino acid residues of the PPARγ active site. The ligands (**6** and **2**) are shown in capped stick model while amino acids are shown in sticks. The H-bonds are displayed using dashed yellow lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

although these features are not important individually, at least 2 of them should be present to maintain the specificity of *hypo-1* for the pan agonists. On the other hand, deletion of other features such as NI, HYD-2, HBA and HYD-3 considerably reduced the discriminating ability of the *hypo-1*. Most notably, deletion of the HBA feature led to the highest hit rate (27.0%) of inactive compounds. This was even higher than *hypo-1h*, *hypo-1i* and *hypo-1j* where two of the features were deleted simultaneously. Thus, it is clear that presence of an HBA feature in the HTG of these ligands is important for the highest discriminating ability of the *hypo-1*. However, as

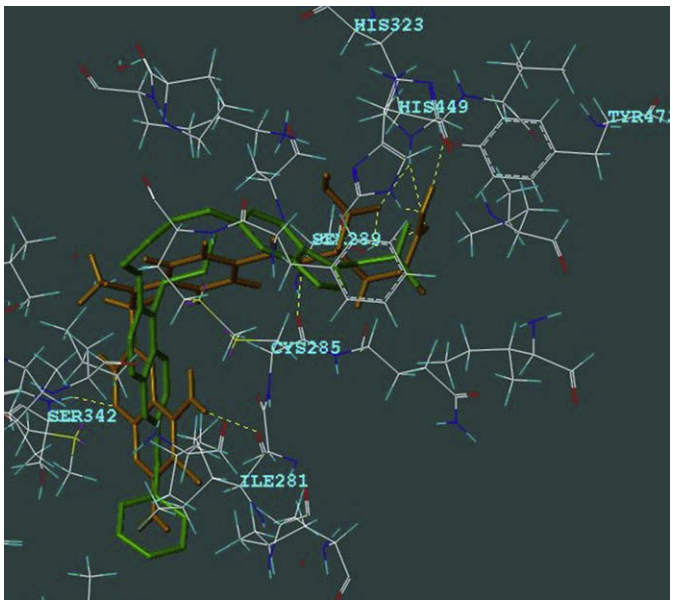

Fig. 5. Docked conformation of **7** (orange colored) together with the co-crystallized ligand (**2**, green colored) (PDB entry 2F4B) and important amino acid residues of the PPARγ active site. The ligands (**7** and **2**) are shown in capped stick model while amino acids are shown in sticks. The H-bonds are displayed using dashed yellow lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

described earlier, the exact interaction of this feature in the PPARγ active site could not be ascertained with the available structural information.

3.5. Virtual screening and docking

Another objective of the present work was to identify novel scaffolds for the design of new pan PPAR ligands. Hence, *hypo-1* was employed as a query for searching NCI2000 (238,819 molecules) and Maybridge2004 (59,652 molecules) databases containing known small organic compounds. However, only 9 hits were obtained which were found to be of high molecular weight and were thus, discarded. This showed that although *hypo-1* is highly efficient in distinguishing highly active PPAR pan agonists and inactives, it is too restrictive to find novel hits through virtual screening. Probably, the restrictive nature of *hypo-1* was due to the large number of chemical features representing it. Thus, the models *hypo-1b*, *hypo-1c* and *hypo-1e* were used as queries for the virtual screening since these are comparable to the *hypo-1* in picking actives and inactives. Also, these hypotheses contain one feature less than the *hypo-1* and are likely to produce more number of hits. Indeed queries *hypo-1b*, *hypo-1c* and *hypo-1e* produced a total of 39, 20 and 21 hits, respectively, from NCI2000 and Maybridge2004 databases (Fig. 3). However, a few of the hits were found to be of high molecular weight and were discarded from the list. The cut off value of 650 g/mol was chosen based on the highest molecular weight reported for the PPAR agonists with good pharmacokinetic profiles [36]. Furthermore, a few hits were found to be common for the three hypotheses and thus were included only once.

Finally, a total of 27 molecules (see Supporting information for structures) were docked in the PPARγ active site (PDB code 2F4B) [3] for the prediction of their binding strength and pose. Out of the 27 molecules, 5 were (**6**–**10**, Fig. 6) found to have similar binding pose and better (or comparable) docking score as compared to the reference molecule **2** (co-crystallized ligand). The mapping of **6**–**10** to various features of *hypo-1* is depicted in Fig. 6. The predicted binding conformation of the two of the hits (**6** and **7**) is given in Figs. 4 and 5. The total FlexX score for the co-crystallized ligand **2** was found to be −23.3 kcal/mol with root mean square (RMS) deviation of 1.48 Å between the docked and the bioactive conformations. Similarly, the FlexX score for **6** and **7** was found to be −20.0 kcal/mol and −33.9 kcal/mol, respectively. The molecule **6** possessed a long aliphatic carbon chain and resembles the fatty acids which are the known natural ligands of PPARγ [7]. The free –COOH group of **6** displayed the H-bonding interactions with His323, Tyr473, Ser289 and His449 while the other structural features of the molecule exhibited hydrophobic interactions with the active site residues similar to the co-crystallized ligand (Fig. 4).

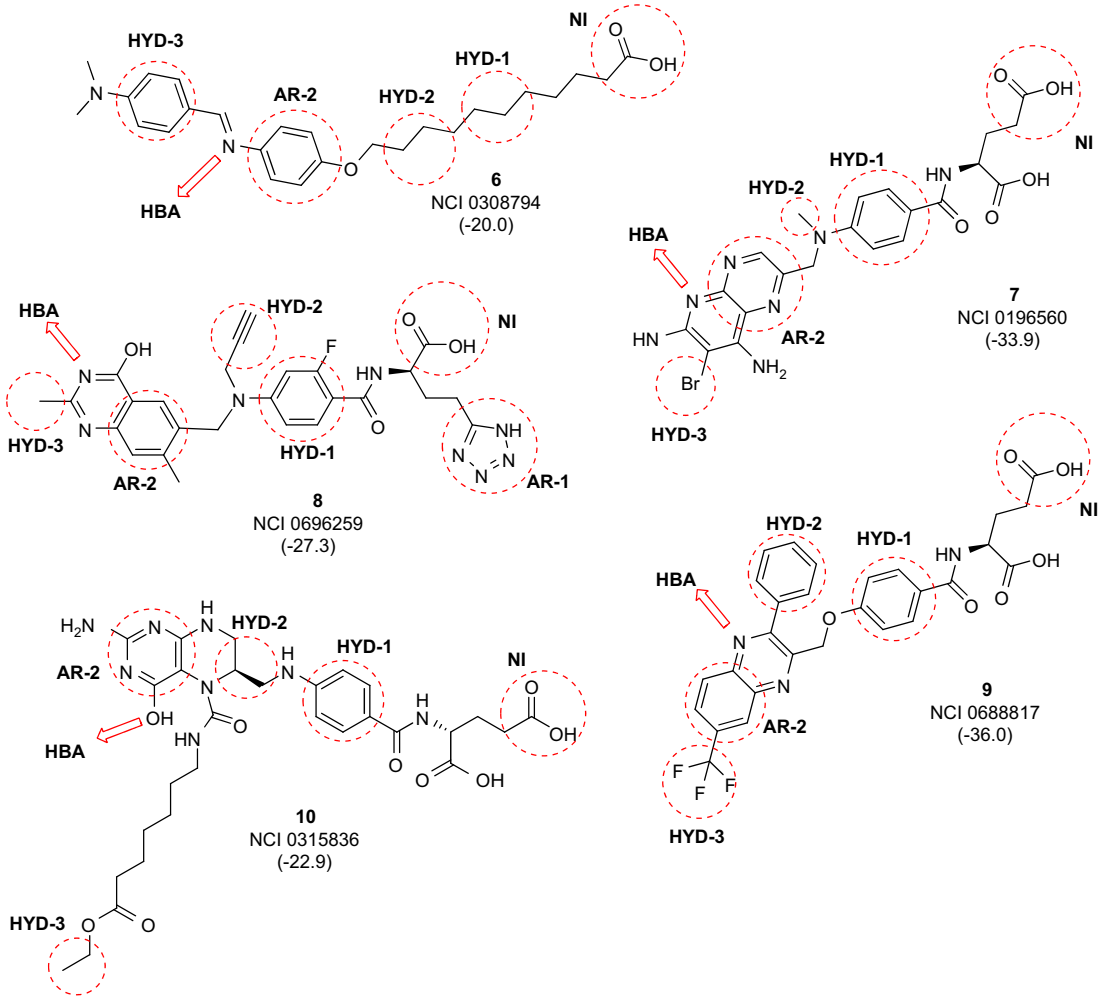

Fig. 6. Structures of the final hits obtained through virtual screening (Using Fast Flexible Search) using *hypo-1b*, *hypo-1c* and *hypo-1e*. Substructures of hit molecules mapping to the various features of *hypo-1* are encircled (except HBA, indicated with the arrow) and labeled. The docking scores (in kcal/mol) in the PPARγ active site are given in the parentheses. Refer to Supporting information for structures of the remaining hits.

The hit molecules **7**–**10** belonged to the N-benzoylated glutamic acids with two acidic functions in the AHG part of the molecule that displayed conserved H-bonds with the His323, Tyr473, Ser289 and His449 residues. Additionally, molecule **7** was found to exhibit H-bonds with the backbone 'NH' of Ser342, carbonyls of Ile281 and Cys285 (Fig. 5) which might be responsible for the higher score. The scaffold of the hit molecules is different from the previously known PPAR pan agonists and can be further optimized to design novel series of compounds belonging to this class.

The next step was to predict the binding strength and pose of hit molecules **6** and **7** in the active site of PPARα and PPARδ. For this purpose, the crystal structures of these subtypes in complex with a potent ligand were sought in the protein data bank. For PPARα, the structure 1K7L (resolution = 2.50 Å) [34] was selected for the docking studies. Unfortunately, when the co-crystallized (GW409544) ligand was docked in the active site, FlexX program could not reproduce its binding pose. When the active site size was selected as 6.5 Å around the bound ligand, the RMS value for the docked and the bioactive pose was found to be more than 12.0 Å. The RMS deviation was not affected much by changing the active site size to 7.0 Å and 8.0 Å. Thus, docking of hit molecules **6** and **7** was not performed due to the absence of suitable docking score for the reference ligand for comparison. For PPARδ, the structures 1Y0S (resolution = 2.65 Å) [33] and 1GWX (resolution = 2.50 Å) [32] were selected from the protein data bank. However, in this case also FlexX was unable to predict the binding pose of the co-crystallized ligands. The RMS values were found to be high ranging from 6.0 to 10.0 Å for the different sizes of active site in both the cases. Consequently, docking of the hit molecules **6** and **7** could not be performed in PPARδ, either.

Thus, FlexX performed poorly in the cases of PPARα and PPARδ in comparison to PPARγ, in reproducing the bound conformation of the respective co-crystallized structures. Hence, binding pose of **6** and **7** could not be evaluated in PPARα and PPARδ subtypes.

4. Conclusions and future prospects

We have attempted to generate a common chemical feature pharmacophore for pan PPAR agonists using HipHop program. The chemical features in the top scoring hypothesis (*hypo-1*) were found to match with the known interactions of the pan agonists in the PPARγ active site. In addition, *hypo-1* was also found to discriminate among the pan (actives) and subtype-specific or dual PPAR agonists (inactives) with reasonably good accuracy. The modified versions *(hypo-1a* to *hypo-1g*) of *hypo-1* were developed by deleting all the features of *hypo-1* one at a time. The ability of the *hypo-1a* to *hypo-1g* to differentiate 'actives' from 'inactives' was measured and compared with the *hypo-1*. The presence of an HBA feature in the HTG portion of pan agonists was found to be most vital in discriminating actives from inactives. On the other hand, *hypo-1b*, *hypo-1c* and *hypo-1e* generated by deleting AR-1, HYD-1 and AR-2 features of *hypo-1*, respectively, were found to be comparable to the parent hypothesis in discriminating 'actives' and 'inactives'. The *hypo-1* was found to be highly restrictive in picking up the novel hits from the NCI and Maybridge databases and thus, hypotheses *hypo-1b*, *hypo-1c* and *hypo-1e* were used for the purpose of virtual screening. The virtual screening results finally provided a total of 27 new hits which were docked in the active site of PPARγ. The 5 molecules were found to have comparable or better score than the reference molecule. The hit molecules exhibited important H-bonding interactions with His323, Tyr473, Ser289 and His449, which are conserved in the case of various PPARγ agonists.

The above pharmacophore mapping approach helped in understanding the importance of various chemical features present in the pan PPAR agonists. The different hypotheses such as *hypo-1*, *hypo-1b*, *hypo-1c* and *hypo-1e* can be used for virtual screening depending on the specificity required in the final results. The validation of these hypotheses can further be tested in future when a large number of potent PPAR pan agonists become available. It would be interesting to design experiments to understand the importance of HBA in the HTG part of these ligands. Additionally, the methodology adopted in this work will be useful in understanding the structural requirements and design of other ligands acting on multiple targets.

Acknowledgements

Sandeep Sundriyal gratefully acknowledges the Department of Science and Technology (DST), New Delhi, India, for providing senior research fellowship. Authors acknowledge reviewers for their valuable suggestions.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2009.01.024.

References

[1] A.D. Adams, W. Yuen, Z. Hu, C. Santini, A.B. Jones, K.L. MacNaul, J.P. Berger, T.W. Doebber, D.E. Moller, Bioorg. Med. Chem. Lett. 13 (2003) 931–935.
[2] N. Mahindroo, C.-F. Huang, Y.-H. Peng, C.-C. Wang, C.-C. Liao, T.-W. Lien, S.K. Chittimalla, W.-J. Huang, C.-H. Chai, E. Prakash, C.-P. Chen, T.-A. Hsu, C.-H. Peng, I.-L. Lu, L.-H. Lee, Y.W. Chang, W.-C. Chen, Y.-C. Chou, C.-T. Chen, C.M.V. Goparaju, Y.-S. Chen, S.-J. Lan, M.-C. Yu, X. Chen, Y.-S. Chao, S.-Y. Wu, H.-P. Hsieh, J. Med. Chem. 48 (2005) 8194–8208.
[3] N. Mahindroo, C.-C. Wang, C.-C. Liao, C.-F. Huang, I.L. Lu, T.W. Lien, Y.-H. Peng, W.-J. Huang, Y.-T. Lin, T.-A. Hsu, C.-H. Lin, C.-H. Tsai, J.T.-A. Hsu, X. Chen, P.-C. Lyu, Y.-S. Chao, S.-Y. Wu, H.-P. Hsieh, J. Med. Chem. 49 (2006) 1212–1216.
[4] J. Rudolph, L. Chen, D. Majumdar, W.H. Bullock, M. Burns, T. Claus, F.E.D. Cruz, M. Daly, F.J. Ehrgott, J.S. Johnson, J.N. Livingston, R.W. Schoenleber, J. Shapiro, L. Yang, M. Tsutsumi, X. Ma, J. Med. Chem. 50 (2007) 984–1000.
[5] L.-D. Cantin, S. Liang, H. Ogutu, C.I. Iwuagwu, K. Boakye, W.H. Bullock, M. Burns, R. Clark, T. Claus, F.E. delaCruz, M. Daly, F.J. Ehrgott, J.S. Johnson, C. Keiper, J.N. Livingston, R.W. Schoenleber, J. Shapiro, C. Town, L. Yang, M. Tsutsumi, M. Xin, Bioorg. Med. Chem. Lett. 17 (2007) 1056–1061.
[6] B.R. Henke, J. Med. Chem. 47 (2004) 4118–4127.
[7] T.M. Willson, P.J. Brown, D.D. Sternbach, B.R. Henke, J. Med. Chem. 43 (2000) 527–541.
[8] P. Markt, D. Schuster, J. Kirchmair, C. Laggner, T. Langer, J. Comput. Aided Mol. Des. 21 (2007) 575–590.
[9] S. Khanna, M.E. Sobhia, P.V. Bharatam, J. Med. Chem. 48 (2005) 3015–3025.
[10] C. Liao, A. Xie, L. Shi, J. Zhou, X. Lu, J. Chem. Inf. Comput. Sci. 44 (2004) 230–238.
[11] C. Liao, A. Xie, L. Shi, J. Zhou, L. Shi, Z. Li, X.-P. Lu, J. Mol. Model. 10 (2004) 165–177.
[12] C. Rücker, M. Scarsi, M. Meringer, Bioorg. Med. Chem. 14 (2006) 5178–5195.
[13] S.S. Kulkarni, L.K. Gediya, V.M. Kulkarni, Bioorg. Med. Chem. 7 (1999) 1475–1485.
[14] S. Khanna, R. Bahal, P.V. Bharatam, in: S.P. Gupta (Ed.), QSAR and Molecular Modeling Studies in Heterocyclic Drugs I, Topics in Heterocyclic Chemistry, vol. 3, Springer-Verlag, Berlin Heidelberg, 2006, pp. 149–180.
[15] S. Sundriyal, P.V. Bharatam, Eur. J. Med. Chem. 44 (2009) 42–53.
[16] A.D. Andricopulo, R.V.C. Guido, G. Oliva, Curr. Med. Chem. 15 (2008) 37–46.
[17] M. Scarsi, M. Podvinec, A. Roth, H. Hug, S. Kersten, H. Albrecht, T. Schwede, U.A. Meyer, C. Rücker, Mol. Pharmacol. 71 (2007) 398–406.
[18] S. Derksen, O. Rau, P. Schneider, M. Schubert-Zsilavecz, G. Schneider, ChemMedChem 1 (2006) 1346–1350.
[19] I.-L. Lu, C.-F. Huang, Y.-H. Peng, Y.-T. Lin, H.-P. Hsieh, C.-T. Chen, T.-W. Lien, H.-J. Lee, N. Mahindroo, E. Prakash, A. Yueh, H.-Y. Chen, C.M.V. Goparaju, X. Chen, C.-C. Liao, Y.-S. Chao, J.T.-A. Hsu, S.Y. Wu, J. Med. Chem. 49 (2006) 2703–2712.
[20] D. Barnum, J. Greene, A. Smellie, P. Sprague, J. Chem. Inf. Comput. Sci. 36 (1996) 563–571.
[21] Catalyst Program, Version 4.10, Accelrys Inc., San Diego, CA, USA.
[22] A. Smellie, S.L. Teig, P. Towbin, J. Comput. Chem. 16 (1995) 171–187.
[23] A. Smellie, S.D. Kahn, S.L. Teig, J. Chem. Inf. Comput. Sci. 35 (1995) 285–294.
[24] A. Smellie, S.D. Kahn, S.L. Teig, J. Chem. Inf. Comput. Sci. 35 (1995) 295–304.
[25] B.R. Brooks, R.E. Brucolleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187–217.
[26] R.T. Nolte, G.B. Wisely, S. Westin, J.E. Cobb, M.H. Lambert, R. Kurokawa, M.G. Rosenfeld, T.M. Willson, C.K. Glass, M.V. Milburn, Nature 395 (1998) 137–143.

[27] E. Burgermeister, A. Schnoebelen, A. Flament, J. Benz, M. Stihle, B. Gsell, A. Rufer, A. Ruf, B. Kuhn, H.P. Marki, J. Mizrahi, E. Sebokova, E. Niesor, M. Meyer, Mol. Endocrinol. 20 (2006) 809–830.
[28] C.R. Hopkins, S.V. O'Neil, M.C. Laufersweiler, Y. Wang, M. Pokrosss, M. Mekel, A. Evdokimov, W. Walter, M. Kontoyianni, M.E. Petrey, G. Sabatakos, T.M. Roesgen, E. Richardson, T.P. Demuth Jr., Bioorg. Med. Chem. Lett. 16 (2006) 5659–5663.
[29] T. Ostberg, S. Svensson, G. Selen, J. Uppenberg, M. Thor, M. Sundbom, M. Sydow-Backman, A.L. Gustavsson, L. Jendeberg, J. Biol. Chem. 279 (2004) 41124–41130.
[30] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[31] SYBYL6.9, Tripos Inc., 1699, South Hanley Road, St. Louis, MO 631444, USA.
[32] H.E. Xu, M.H. Lambert, V.G. Montana, D.J. Parks, S.G. Blanchard, P.J. Brown, D.D. Sternbach, J.M. Lehmann, G.B. Wisely, T.M. Willson, S.A. Kliewer, M.V. Milburn, Mol. Cells 3 (1999) 397–403.
[33] I. Takada, R.T. Yu, H.E. Xu, M.H. Lambert, V.G. Montana, S.A. Kliewer, R.M. Evans, K. Umesono, Mol. Endocrinol. 14 (2000) 733–740.
[34] H.E. Xu, M.H. Lambert, V.G. Montana, K.D. Plunket, L.B. Moore, J.L. Collins, J.A. Oplinger, S.A. Kliewer, R.T. Gampe Jr., D.D. McKee, J.T. Moore, T.M. Willson, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 13919–13924.
[35] B.B. Lohray, V. Bhushan, Curr. Med. Chem. 11 (2004) 2467–2503.
[36] P. Sauerberg, P.S. Bury, J.P. Mogensen, H.-J. Deussen, I. Pettersson, J. Fleckner, J. Nehlin, K.S. Frederiksen, T. Albrektsen, N. Din, L.A. Svensson, L. Ynddal, E.M. Wulff, L. Jeppesen, J. Med. Chem. 46 (2003) 4883–4894.