Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 17 (2007) 2156–2160
Bioorganic & Medicinal Chemistry Letters

3D-QSAR studies with the aid of molecular docking for a series of non-steroidal FXR agonists

Tao Zhang,[a] Jun-Hong Zhou,[b] Liang-Wei Shi,[b] Rui-Xin Zhu[a] and Min-Bo Chen[a,*]

[a]*Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 354 Fenglin Rd., Shanghai 200032, PR China*
[b]*School of Chemistry and Molecular Engineering Department, East China University of Science and Technology, PR China*

Received 6 November 2006; revised 21 December 2006; accepted 26 January 2007
Available online 31 January 2007

Abstract—The farnesoid x receptor (FXR) has become a potential drug target for treating cholesterol-related and bile acid-related diseases recently. In this paper, 3-dimensional quantitative structure–activity (structure–affinity and structure–efficacy) relationships are investigated for a series of non-steroidal agonists (fexaramine series) by using the comparative molecular field analysis (CoMFA), where molecular docking method (FlexX) is employed to construct molecular superimposition maps. A proposal to design some new agonists is discussed lastly.

Nuclear receptors (NRs) take part in various biological activities such as development, reproduction, differentiation, and cellular homeostasis by binding with lipophilic steroid and retinoid hormones.[1,2] The farnesoid x receptor (FXR, BAR, NR1H4[18]) was first observed to be weakly activated by farnesol, and then bile acids were identified as endogenous ligands of FXR in 1999.[3–6] FXR plays an important role in regulating bile acid synthesis and cholesterol metabolism.[7,8] Now, FXR has become an attractive target for treating cholesterol-related and bile acid-related diseases.

Many bile acid derivatives and two kinds of non-steroidal compounds have been developed to regulate FXR. Roberto Pellicciari and coworkers have synthesized and evaluated a series of body and side chain modified analogues of chenodeoxycholic acid (CDCA, Fig. 1), such as significantly active 6ECDCA (Fig. 1) and recent 'back door' derivatives.[9–11] Tomofumi Fujino et al. investigated the structure–activity relationship of other bile acids and their analogues, such as cholic acid, lithocholic acid, and alkylated CDCA.[12] Maloney et al. identified GW 4064 as a chemical tool for FXR.[13] Nicolaou et al. discovered fexaramine as a potent agonist from natural product-like libraries (Fig. 1).[14] For fexaramine series, Honorio et al. reported the studies of hologram quantitative structure–activity relationships (HQSAR), where two dimensional chemical structures and their activities are correlated.[21]

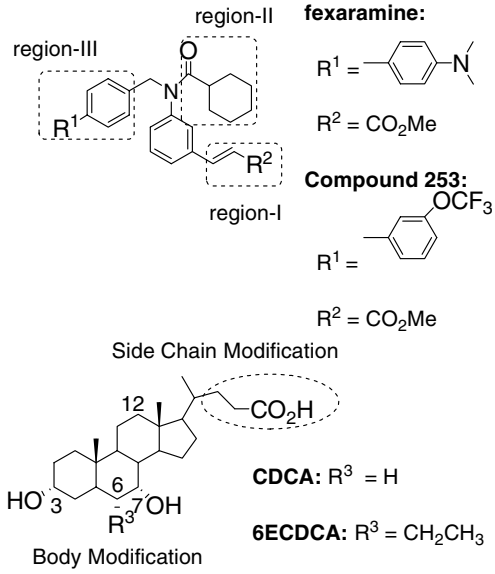

Figure 1. The chemical structures of the bile acids and fexaramine series. CDCA is modified in side chain and body (many substitutes in the sites 3, 6, 7, and 12) for pursuing more active compounds. Fexaramine is discovered in a designed product-like library (refer to *Org. Biomol. Chem.* **2003**, *1*, 908).

Keywords: Farnesoid x receptor; FXR; Agonist; 3D-QSAR; Molecular docking; Bile acid; Non-steroidal agonist; Fexaramine.
* Corresponding author. Tel.: +86 21 5492 5262; fax: +86 21 6416 6128; e-mail: mbchen@mail.sioc.ac.cn

0960-894X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2007.01.079

The previous crystallographic studies have shown that 6ECDCA and fexaramine located on similar but different ligand-binding-sites in FXR ligand-binding domain. (PBD deposition code: 1osv and 1osh)[15,16] 6ECDCA cannot contact the key helix12 directly, but previous molecular dynamics studies have indicated that the ligand stabilizes the cation–π interaction between H444 and W466 that helps the helix12 to locate in active conformation.[15,17] In fexaramine–FXR complex, the terminal of region-III of fexaramine is able to contact the helix12 with W473 and F465 directly (Fig. 2).[16]

Occupancy and response are two important properties for an agonist of receptor. For FXR, the response is embodied by the specific conformational changes of FXR that are induced by the agonist to form a hydrophobic crevice (AF2 site) for recruitment of a specific coactivator protein, while the occupancy is represented by the affinity between FXR and the ligand. Fexaramine series are reported with their affinities (characterized by EC_{50} value, the concentration required for obtaining 50% of the maximum effect) and responses (characterized by the relative efficacy to the indicated compound CDCA).[14] EC_{50} values of the series vary from micromolar to nanomolar level, while the RE values (relative efficacy) change from 0.05 to 2.16. As shown in Figure 3, the higher pEC_{50} values are roughly paralleled with the better RE values. The coarse correspondence indicates that the more active is a compound, the more efficient it is. But for the high active compounds ($pEC_{50} > 6.5$), the relationship between EC_{50} and RE is blurred. This discrimination between the affinity and efficacy supplies the opportunity to design various potent modulators of FXR, from full agonist to full antagonist.

In this paper, the 3-dimensional quantitative structure–affinity relationship and the 3-dimensional quantitative structure–efficacy relationship for the fexaramine series are analyzed by using comparative molecular field analysis (CoMFA) method, where the molecular docking method (FlexX) is used to construct the molecular superimposition maps. The discrimination of mechanisms between the binding affinity and the driving efficacy is investigated by comparison of the different CoMFA results combined with the previous crystallographic results. Finally, encouraged by the activation mechanism of the FXR by 6ECDCA, we propose a new direction to design some new fexaramine-based agonists.

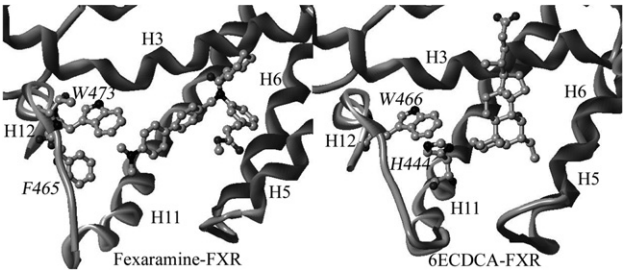

Figure 2. Comparison of two protein-ligand crystal structures: fexaramine–FXR (left, rat) and 6ECDCA-FXR (right, human). Region-III of fexaramine interacts with helix3, helix11, and helix12, inserting a pocket composed by LEU291, THR292, LEU455, PHE456, TRP473, PHE288, and TRP458 (called subpocket-C), and that region-I inserts into the gap between helix3 and helix6 (LEU291, THR292, ARG355, ILE356, ASN287, LEU352, and MET248, called subpocket-A), and that region-II locates near the disordered helix2 and the loop between helix5 and helix6 (PHE340, ILE339, MET269, and LEU352, called subpocket-B). The helix3, helix5, helix11, and helix12 are represented by ribbons, and two ligands and several key residues by ball-sticks. Other parts of the protein are not displayed here due to clarity reasons. The crystal structures were taken for RCSB Protein Data Bank, 1osv (rat) and 1osh (right), respectively.

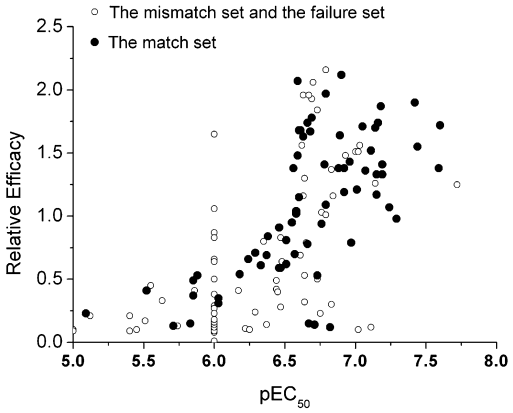

Figure 3. The comparison of pEC_{50} with relative efficacy of all compounds. The match set compounds are marked with solid circles, while the mismatch and failure sets with open circles. The details about the sets see Supporting information.

Construction of molecular superimposition maps. The 3D-QSAR method assumes that all the active molecules bind in a common manner to the same target site, so the key step to get a reasonable CoMFA result is to construct a reliable molecular superimposition map, by which the common binding manner could be reflected. As the crystal structure of fexaramine–FXR complex is available, the molecular docking method is used to explore the conformations of fexaramine series when they are binding to FXR.

As seen in Figure 4, the FlexX integrated in SYBYL[19] is able to reproduce the experimental conformation of fexaramine (rms = 1.5 angstrom). For all 149 compounds, the docking results are sorted into three sets: the match set, the mismatch set, and the failure set. The match set includes the compounds whose binding manners are similar as the experimental one, that is, the region-I is docking into subpocket-A, region-II is docking into subpocket-B, and region-III is docking into subpocket-C. As seen in Figure 3, the majority of high active ($pEC_{50} > 6.5$) and efficient ($RE > 0.5$) compounds belong to the match set, with some low active compounds and a few of inefficient compounds also included. The mismatch set includes the compounds that could be placed into the ligand-binding-site in docking computation, but their binding manners disagree with the experimental result. For example, the region-III of the molecule **253** inserts into the subpocket-A, while the region-I is docking into the subpocket-C (Fig. 4,

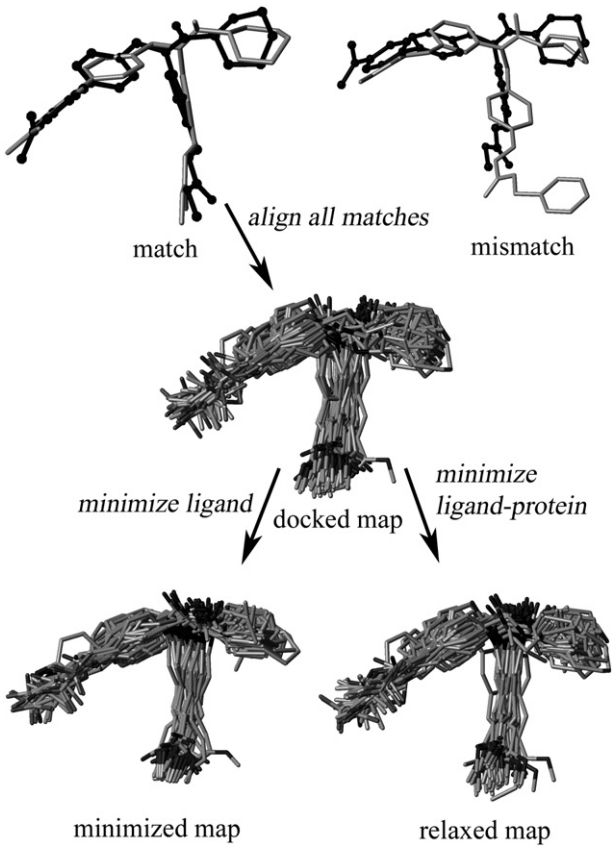

Figure 4. The comparison of fexaramine docked conformation (top left and top right, in black) with fexaramine crystallographic conformation (top left, in gray), and with the docked conformation of compound **253** (top right, in gray). Three molecular superimposition maps: the docked (center), the minimized (bottom left), and the relaxed (bottom right).

top right). The failure set includes the compounds that fail to enter ligand-binding-site in docking computation. Most of the low active and inefficient compounds fall into the two latter sets.

Based on docking results, the match set compounds are identified as the legal compounds that have similar binding manners to FXR as the fexaramine. The compounds of mismatch and failure sets are ignored in next 3D-QSAR studies, because their binding conformations cannot satisfy the basic requirement of the 3D-QSAR method in docking computations. Frankly, some useful information hidden in parts of the compounds of the mismatch and failure sets might be improperly abandoned due to fault in docking computation. However, this docking based identification promises the validation for the following CoMFA analysis to the best of the docking's ability.

Three molecular superimposition maps are established: docked, minimized, and relaxed maps. The docked map is constructed by aligning the docking conformations directly. The poor statistic indices, crossvalidated q^2 of 0.301 with optimal principal components of only 1 in pEC_{50} PLS computation, indicate that the docked map does not completely reflect the general binding manner and needs some refinement. Two optimized superimposition maps are obtained after the molecular docking results are refined differently. The minimized map is obtained by aligning the molecules after the docked ligands are energy-minimized with fixed protein in Dock program, where the electrostatic interaction between the ligand (Gasteiger–Hückel charge) and protein (Amber 99 FF) is considered. The relaxed map is obtained by aligning the molecules after the ligand–protein complexes are energy-optimized. The relaxation manipulation takes account of partial flexibility of the protein and removes the close contacts between the ligand and protein. As seen in Figure 4, the two optimized maps are more compact than the docked map, while the statistic indices are improved, especially for the crossvalidated q^2 from 0.301 to 0.507 in pEC_{50} PLS (Table 1).

The CoMFA results. The 58 compounds in the relaxed match set are randomly selected as the training set for the CoMFA studies, while the rest of the 10 compounds are used to evaluate their predictabilities. Both the pEC_{50} and RE values are correlated with standard CoMFA steric and electrostatic fields. As seen in Figure 5, the models deserve good predictabilities. To get more information from the original data, two CoMFA models for all 68 compounds are built for pEC_{50} and RE, respectively. Both models deserve good and reasonable statistic indices (Table 1, Fig. 5 bottom).

Comparison of the CoMFA contour plots with crystal structure provides useful information about how the structural changes of the ligands affect their affinities and efficacies. The relative contribution fraction data show that the electrostatic interaction rather than steric interaction dominates the pEC_{50} and RE (Fig. 6).

The similarity between the two CoMFA contour plots of steric field (top left and bottom left, Fig. 6) indicates that the steric factor affects the affinity and efficacy in a similar manner. The contribution shows that the bulk group near subpocket-B favors the affinity and efficacy, while the bulk group near W458 and M454 disfavors them.

Table 1. The statistic indices of three superimposition maps in PLS

Map		Docked[a]	Minimized[a]	Relaxed[a]	Relaxed[b]
q^2	pEC_{50}	0.301	0.397	0.507	0.408
	RE	0.545	0.445	0.677	0.605
n	pEC_{50}	1	6	6	6
	RE	6	6	6	6
R^2	pEC_{50}	0.522	0.948	0.962	0.963
	RE	0.979	0.949	0.978	0.979
SE	pEC_{50}	0.340	0.117	0.100	0.101
	RE	0.084	0.131	0.086	0.086
F	pEC_{50}	72	184	218	219
	RE	471	190	376	386

q^2, square of crossvalidated correlation coefficient; n, number of principal components; R^2, square of non-crossvalidated correlation coefficient, and SE, standard error.

[a] All 68 molecules are used partial least in squares analysis.

[b] Just the training set (58 molecules) is used in partial least-squares analysis (PLS).

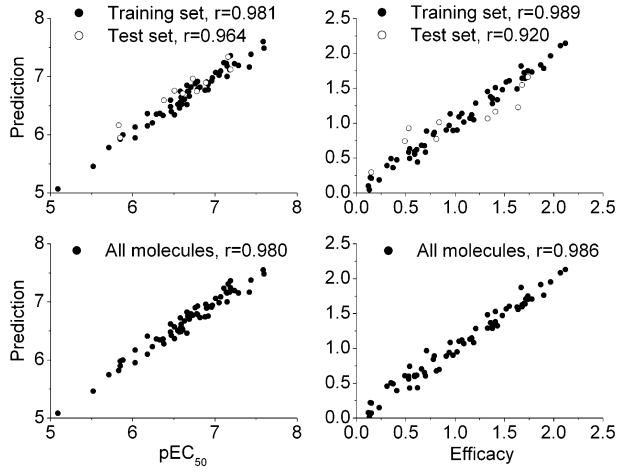

Figure 5. Comparison of the affinity and the efficacy to their corresponding prediction by CoMFA models. The top left (for pEC_{50}) and top right (for RE) prediction values are based on the CoMFA models with 58 compounds in training set, and the bottom left (for pEC_{50}) and bottom right (for RE) ones on the models with all 68 compounds.

Through comparison of the electrostatic contour plots of pEC_{50} CoMFA model with that of the RE CoMFA model, it is indicated that the electrostatic factor has a similar effect for the affinity and efficacy in most regions, but rather different in one region near W473. In these models, it is indicated that appearance of negative group near M454 and W458 favors the affinity and efficacy. There is a large red region among W473, F465, and L455 in the RE CoMFA model, but no counterpart in the pEC_{50} CoMFA model correspondingly. It indicates that some positive group on the proper position of region III might be helpful to improve efficacy, but maybe no effect for affinity. We consider that it represents the difference in mechanism between the binding affinity and the driving efficacy.

In both CoMFA models, the electrostatic count plots near the carbonyl oxygen look confusing. It is considered that the confusion in this region originates from the molecular superimposition map, where the chemical compositions in this region are exactly the same (acylamine) but the only difference is in their orientation (Fig. 4). The positional inconsistency of the acylamine part is obviously due to existence of different groups on the three regions of compounds. No significant conclusion should be drawn from it.

On being reminded by the referee, we found a similar work published very recently by Honorio et al.[22] following their former report.[21] Our CoMFA for EC_{50} is consistent with the counterpart in Honorio's report.

Proposal for new agonists. For the activation of FXR by fexaramine series, the CoMFA studies show that some small bulk groups near F465 and W473 favor the affinity and efficacy, and that some positive groups favor the

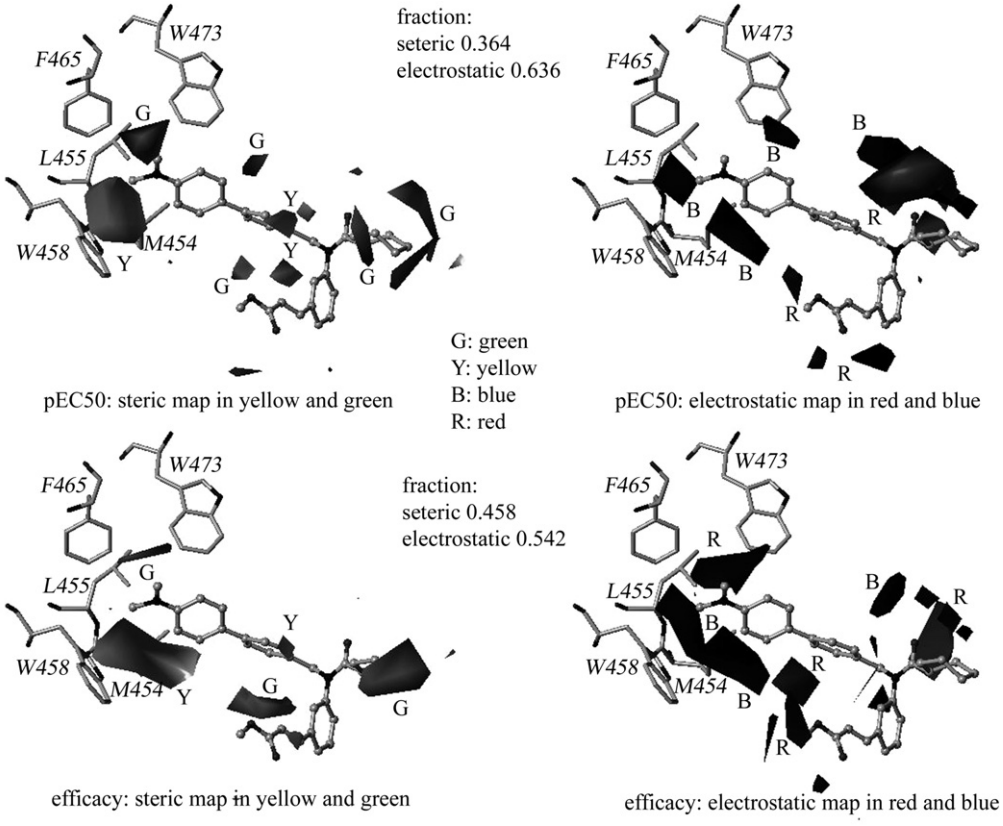

Figure 6. The graphic view of the CoMFA results. The contours of the steric map are shown in yellow and green, and those of the electrostatic map are shown in red and blue. Greater values are correlated with: more bulk near green; less bulk near yellow; more positive charge near blue, and more negative charge near red. The relative contribution fraction data are indicated.

efficacy (Fig. 6). A potential candidate satisfying both the conditions above is hydroxyl group on proper position of the region-III. As mentioned previously, the cation–π interaction between H444 and W466 plays a key role in stabilizing the helix12, when the 6ECDCA activates the FXR (Fig. 2). We notice that the terminal of the region-III of fexaramine can directly contact W473 that belongs to the helix12. Tryptophan is often involved in the cation–π interaction in bio-macromolecules as the π parter.[20] And the space among W473, F465, and L455 is big enough to accommodate a hydroxyl group (Fig. 6). Therefore, based on the previous studies and 3D-QSAR results, we propose that some agonists can be discovered when the proper modification (hydroxyl substitute is suggested) on the region-III of fexaramine is carried out, if the modification is able to lead to a aproper cation–π interaction between W473 and agonist. This potential cation–π interaction of protein–ligand might improve both the affinity and the efficacy, especially for the latter, because the cation–π interaction may stabilize the helix12 in active conformation directly.

Methods and materials. The chemical structures and bioevaluation data of the fexaramine series were taken from the reference.[14] The molecular docking program FlexX integrated in Sybyl 7.2[19] was used in default setting, except that the radius of the ligand-binding site was set to 10 angstroms rather than the default 6.5 angstroms. For each ligand, the docked conformation with the best score was selected to the next molecular alignment manipulation. The program Dock integrated in Sybyl was used to get the optimized conformations of the ligands for the minimized and relaxed maps. In Dock, FXR are parameterized by Amber99 force field, while the small molecules are charged by Gasteiger–Hückel method.

In 3D-QSAR studies, QSAR module integrated in Sybyl was employed. The standard steric and electrostatic fields were set up for all molecular superimposition maps. To confirm the number of optimal principal components, the cross-validation PLS computation with leave-one-out option was performed. With the number of optimal principal components, the non-crossvalidation PLS computation was carried out to get final CoMFA results.

Supplementary data

The chemical structures of the fexaramine series see the supporting information available on http://ees.elsevier.com/bmcl/ or refer to *Org. Biomol. Chem.* **2003**, *1*, 908. The docking results see the supporting information. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2007.01.079.

References and notes

1. Gronemeyer, H.; Gustafsson, J.-Å.; Laudet, V. *Nat. Rev. Drug Disc.* **2004**, *3*, 950.
2. Francis, G. A.; Fayard, E.; Picard, F.; Auwerx, J. *Annu. Rev. Physiol.* **2003**, *65*, 261.
3. Pellicciari, R.; Costantino, G.; Fiorucci, S. *J. Med. Chem.* **2005**, *48*, 5383.
4. Wang, H.; Chen, J.; Hollister, K.; Sowers, L. C.; Forman, B. M. *Mol. Cell.* **1999**, *3*, 543.
5. Park, D. J.; Blanchard, S. G.; Bledsoe, R. K.; Chandra, G.; Consler, T. G.; Kliewer, S. A.; Stimmel, J. B.; Willson, T. M.; Zavacki, A. M.; Moore, D. D.; Lehmann, J. M. *Science* **1999**, *284*, 1356.
6. Makishima, M.; Okamoto, A. Y.; Repa, J. J.; Tu, H.; Learned, R. M.; Luk, A.; Hull, M. V.; Lustig, K. D.; Mangelsdorf, D. J.; Shan, B. *Science* **1999**, *284*, 1362.
7. Goodwin, B.; Jones, S. A.; Price, R. R.; Watson, M. A.; McKee, D. D.; Moore, L. B.; Galardi, C.; Wilson, J. G.; Lewis, M. C.; Roth, M. E.; Maloney, P. R.; Willson, T. M.; Kliewer, S. A. *Mol. Cell.* **2000**, *6*, 517.
8. Lu, T. T.; Makishima, M.; Repa, J. J.; Schoonjans, K.; Kerr, T. A.; Auwerx, J.; Mangelsdorf, D. J. *Mol. Cell.* **2000**, *6*, 507.
9. Pellicciari, R.; Fiorucci, S.; Camaioni, E.; Clerici, C.; Costantino, G.; Maloney, P. R.; Morelli, A.; Parks, D. J.; Willson, T. M. *J. Med. Chem.* **2002**, *45*, 3569.
10. Pellicciari, R.; Costantino, G.; Camaioni, E.; Sadeghpour, B. M.; Entrena, A.; Willson, T. M.; Fiorucci, S.; Clerici, C.; Gioiello, A. *J. Med. Chem.* **2004**, *47*, 4559.
11. Pellicciari, R.; Gioiello, A.; Costantino, G.; Sadeghpour, B. M.; Rizzo, G.; Meyer, U.; Parks, D. J.; Entrena-Guadix, A.; Fiorucci, S. *J. Med. Chem.* **2006**, *49*, 4208.
12. Fujino, T.; Une, M.; Imanaka, T.; Inoue, K.; Nishimaki-Mogami, T. *J. Lipid Res.* **2004**, *45*, 132.
13. Maloney, P. R.; Parks, D. J.; Haffner, C. D.; Fivush, A. M.; Chandra, G.; Plunket, K. D.; Creech, K. L.; Moore, L. B.; Wilson, J. G.; Lewis, M. C. *J. Med. Chem.* **2000**, *43*, 2971.
14. Nicolaou, K. C.; Evans, R. M.; Roecker, A. J.; Hughes, R.; Downes, M.; Pfefferkorn, J. A. *Org. Biomol. Chem.* **2003**, *1*, 1079.
15. Mi, L. Z.; Devarakonda, S.; Harp, J. M.; Hand, Q.; Pellicciari, R.; Willson, T. M.; Khorasanizadeh, S.; Rastinejad, F. *Mol. Cell.* **2003**, *11*, 1093.
16. Downes, M.; Verdecia, M. A.; Roecker, A. J.; Hughes, R.; Hogenesc, J. B.; Kast-Woelbern, H. R.; Bowman, M. E.; Ferrer, J. L.; Anisfeld, A. M.; Edwards, P. A.; Rosenfeld, J. M.; Alvarez, J. G.; Noel, J. P.; Nicolaou, K. C.; Evans, R. M. *Mol. Cell.* **2003**, *11*, 1079.
17. Costantino, G.; Entrena-Guadix, A.; Macchiarulo, A.; Gioiello, A.; Pellicciari, R. *J. Med. Chem.* **2005**, *48*, 3251.
18. Nuclear Receptors Nomenclature Committee, *Cell*, **1999**, *97*, 161.
19. SYBYL, Molecular Modelling Software, Tripos Inc., 1699 S. Hanley Road, St. Louis, MO 63944, USA.
20. Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303.
21. Honorio, K. M.; Garratt, R. C.; Andricopulo, A. D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119.
22. Honorio, K.M.; Garratt, R.C.; Polikarpov, I.; Andricopulo, A.D. *J. Mol. Graphics Modell.* **2006**, in press. Available online 14 September 2006.