Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 921–927
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators

Káthia M. Honório [a,b], Richard C. Garratt [a], Igor Polikarpov [a], Adriano D. Andricopulo [a,*]

[a] *Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-Carlense, 400, 13560-970 São Carlos, SP, Brazil*
[b] *Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Av. Arlindo Bettio, 1000, 03828-000 São Paulo, SP, Brazil*

Received 10 August 2006; received in revised form 8 September 2006; accepted 10 September 2006
Available online 14 September 2006

Abstract

Three-dimensional quantitative structure–activity relationships (3D QSAR) were performed for a series of farsenoid X receptor activators using comparative molecular field analysis (CoMFA). A training set containing 77 compounds served to establish the models. The best statistical results among all models were obtained with region focusing weighted by a S.D. × coefficient values of 0.8 and a grid spacing of 1.0 ($r^2 = 0.963$, SEE = 0.097; $q^2 = 0.742$, SEP = 0.255). The model was used to predict the potency of 20 test set compounds that were not included in the training set, and the predicted values were in good agreement with the experimental results. The final CoMFA model along with the information obtained from 3D contour maps should be useful for the design of novel FXR ligands having improved potency.

Keywords: Nuclear receptors; FXR; QSAR; CoMFA

1. Introduction

The farnesoid X receptor (FXR) belongs to a large family (48 members in humans) of transcription factors called nuclear receptors [1]. Nuclear receptors share several structural features, including an amino-terminal, highly conserved DNA binding domain (DBD) and a COOH-terminal ligand binding domain (LBD). Several nuclear receptors have act as important receptors or sensors of small molecules of clinical relevance in pathological processes [2,3].

Recent studies have shown that FXR can bind and be activated by a variety of endogenous bile acids, indicating the key role played by FXR in a variety of physiological processes and pathological conditions related to the regulation of bile acid and lipid metabolism [4].

FXR can also be activated by a number of nonsteroidal and steroidal compounds not structurally related to bile acids [5,6]. The most potent endogenous FXR activator, chenodeoxycholic acid (CDCA), is shown in Fig. 1. Other FXR activators, including the synthetic steroid agonist 6-ECDCA [7], and the nonsteroidal agonists GW4064 [8] and fexaramine are also shown in Fig. 1.

The characterization of the physiological role of FXR has suggested several important therapeutic applications associated with its selective modulation. For example, several bile acid analogs have been developed as hypocholesterolemic agents using a combination of structure- and ligand-based approaches, as well as other important therapeutic applications resulting from lipid, cholesterol and bile acid abnormalities [9,10]. In particular, quantitative structure–activity relationships (QSAR) have been employed as a useful strategy for the design of new small molecule drug candidates, ranging from enzyme inhibitors to nuclear receptor ligands [11–14].

As part of our ongoing research program aimed at designing new selective FXR activators, and in order to investigated the three-dimensional quantitative structure–activity relationships (3D QSAR) of a large series of FXR activators, we have employed the comparative molecular field analysis (CoMFA) method to generate 3D QSAR models possessing substantial predictive power. In this method, it is possible to predict the

* Corresponding author. Tel.: +55 16 3373 8095; fax: +55 16 3373 9881.
E-mail address: aandrico@if.sc.usp.br (A.D. Andricopulo).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.09.003

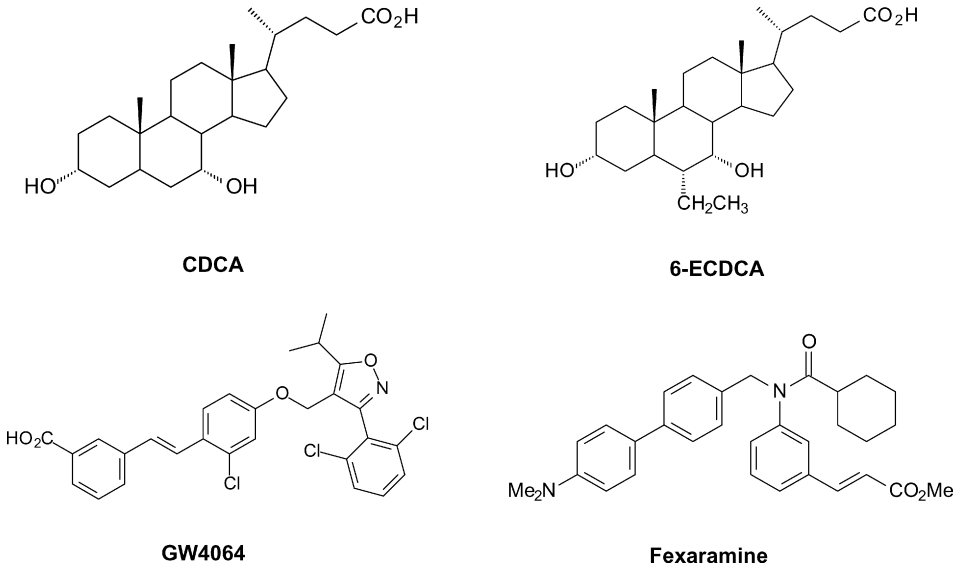


Fig. 1. Endogenous and synthetic ligands of FXR.

biological activity of molecules and represent the relationships between molecular properties (steric and electrostatic) and biological activity in the form of contour maps.

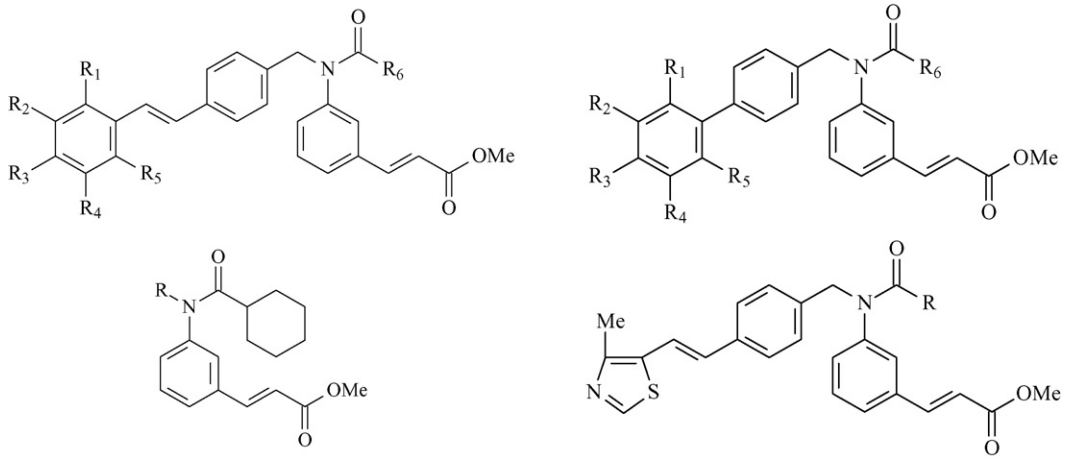
The vast majority of QSAR studies found in the literature for the large chemical structure and corresponding biological information for nuclear receptors is associated with the estrogen receptor (ER) [15]. To the best of our knowledge, no 3D QSAR investigation for FXR modulators has been reported to date. This proves the importance of QSAR studies involving the molecular target of our studies.

2. Methodology

2.1. Computational approach and data set

The QSAR modeling analyses, calculations, and visualizations for CoMFA were performed using the SYBYL 7.2 package (Tripos Inc., St. Louis, USA) running on Red Hat Enterprise Linux workstations. Docking protocol as implemented in GOLD 2.1 (Cambridge Crystallographic Data Centre, Cambridge, UK) was employed to search the possible binding conformations of activators into the FXR active site. The data set used for the 3D QSAR CoMFA analyses contains 97 FXR activators, which were selected from the literature [16]. Table 1 shows the general chemical structures of the series of nonsteroidal FXR activators employed in this work. The complete version of Table 1, including the chemical structures and corresponding biological data, is provided as supplementary information, as well as the chemical structures in SMILES format for the complete data set. The 3D structures of the data set members were constructed using CONCORD and standard geometric parameters of molecular modeling software package SYBYL 7.2. Each single optimized conformation of each molecule in the data set was energetically minimized employing the atom-centered partial charge AM1-ESP calculations implemented in MOPAC 6.0.

Table 1
Representative chemical structures of the series of nonsteroidal FXR activators employed in the QSAR studies



2.2. Protein structure

The preparation of the 3D protein structure employed in the CoMFA analyses involved four major steps: (i) selection and analysis of a high resolution X-ray crystal structure; (ii) removal of the bound ligand from the active site of the crystallographic structure; (iii) elimination of the water molecules from the structure and addiction of hydrogen atoms; and (iv) analysis of protonation state and correct tautomers of aminoacid residues found in the active site of protein. The X-ray crystallographic data for FXR in complex with the small molecule agonist fexaramine used in docking simulations and QSAR simulations were retrieved from the Protein Data Bank (PDB code: 1OSH, with resolution of 1.8 Å) [17]. In the process of structure preparation, fexaramine was removed from the complex and hydrogen atoms were added using the program Pymol (DeLano Scientific, San Carlos, USA).

2.3. Molecular alignment and docking studies

Molecular alignment is one of the most important steps in 3D QSAR studies. A variety of useful approaches have been described in the literature for this purpose [18]. The 3D alignment approach used in this work is based on molecular conformations obtained from the docking program GOLD. The active site for the docking simulations was defined considering a radius of 6 Å around the central carbon atom of the ligand (fexaramine) found in the X-ray structure. Default parameters and GOLDscore function were employed in all docking runs, and only the best ranked conformation was considered for 3D QSAR studies. Fig. 2 shows the alignment obtained from docking simulations for the complete set of FXR activators.

2.4. CoMFA models

To better understand and explore the contributions of electrostatic and steric fields in the binding of the data set inhibitors to FXR, and to build predictive 3D QSAR models, CoMFA studies were performed based on the molecular alignment described. CoMFA calculates steric and electrostatic properties according to Lennard–Jones and Coulomb potentials, respectively [19]. The aligned 77 training set molecules (supplementary information) were placed in a 3D grid box of 2.0 Å in the x, y, and z directions, and the grid region was automatically generated by the CoMFA routine to encompass all molecules with an extension of 4.0 Å in each direction. CoMFA steric and electrostatic fields were generated at each grid point with Tripos force field using an sp^3 carbon atom probe carrying a +1 net charge. CoMFA region focusing method was applied to increase the resolution of CoMFA models. The default value of 30 kcal mol^{-1} was set as the maximum steric and electrostatic energy cutoff. Minimum-sigma (column filtering) was set to be 2.0 kcal mol^{-1} to improve the signal-to-noise ratio by omitting those lattice points where energy variation is below the threshold. All models were investigated using full cross-validated (q^2), partial least squares (PLS), leave-one-out (LOO), and leave many-out (LMO) methods with CoMFA standard options for scaling of variables. Progressive scrambling method was applied to determine the sensitivity of the QSAR models to chance correlations.

3. Results and discussion

3D QSAR CoMFA models were derived for a series of 97 FXF activators. The *in vitro* EC_{50} values employed in this work were measured under the same experimental conditions [16], a fundamental requirement for QSAR studies [12–14]. The EC_{50} values were then converted into the corresponding pEC_{50} values ($-\log EC_{50}$) and used as dependent variables in the 3D QSAR investigations. The pEC_{50} values span approximately one and a half orders of magnitude and property values are acceptably distributed across the range of values. Thus, the data set is appropriate for the purposes of QSAR model development.

The generation of consistent models is dependent on the creation of appropriate training and test sets. From the original data set of 97 FXR activators, 77 compounds (**1–77**, supplementary information) were selected as members of the training set for model construction, and the other 20 compounds (**78–97**, supplementary information) as members of the test set for external model validation, in the ratio of 3.85:1. A statistical cluster analysis was carried out with Tsar 3D (Accelrys, San

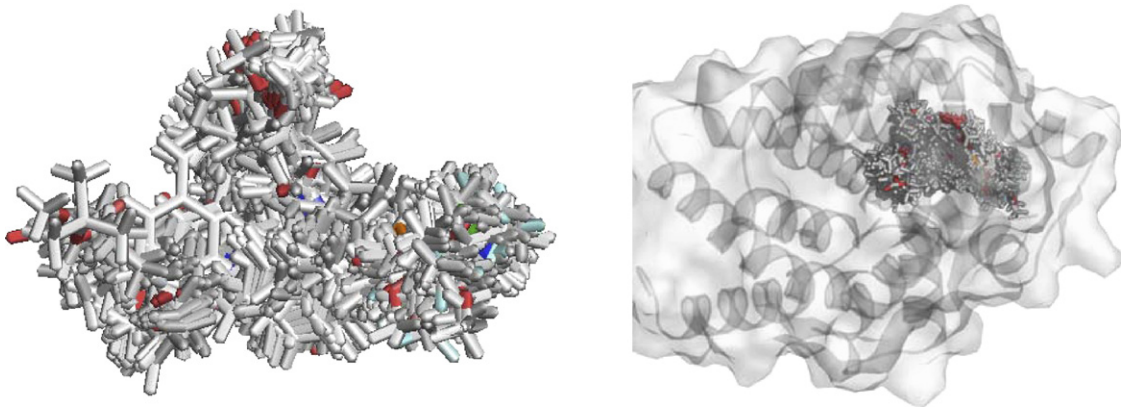


Fig. 2. Three-dimensional data set alignment for conformations generated by GOLD 2.1.

Table 2
CoMFA statistical results for the aligned data set

	Region focusing												
	No	$w = 0.3$			$w = 0.5$			$w = 0.8$			$w = 1.2$		
	–[a]	0.5[a]	1.0[a]	1.5[a]	0.5[a]	1.0[a]	1.5[a]	0.5[a]	1.0[a]	1.5[a]	0.5[a]	1.0[a]	1.5[a]
q^2	0.259	0.348	0.592	0.197	0.443	0.692	0.190	0.550	0.742	0.179	0.626	0.732	0.167
SEP	0.420	0.395	0.323	0.438	0.384	0.281	0.437	0.339	0.255	0.439	0.307	0.262	0.443
N	3	3	8	3	8	8	2	8	7	2	7	8	2

Region focusing was weighted by standard deviation coefficient values (w); cross-validated correlation coefficient (q^2); cross-validated standard error (SEP); optimal number of components (N).
[a] Grid spacing.

Diego, USA) using the complete linkage clustering method with no standardization to ensure that structurally diverse molecules possessing activities of wide range were included in both sets. The PLS method was used for all 3D analyses. CoMFA descriptors were used as independent variables in the PLS regression analyses to derive QSAR models. The predictive ability of the generated models was assessed by their q^2 values.

Employing the molecular alignment obtained from GOLD (Fig. 2) for the training set FXR activators, several PLS analyses were carried out, and the main statistical results obtained are presented in Table 2.

As it can be seen, an initial analysis without applying the option ''region focusing'' (an advanced method of noise reduction) produced a non-significant cross-validated correlation coefficient q^2 of 0.259. Thus, we have applied the region focusing that was weighted by S.D. × coefficient values ranging from 0.3 to 1.2, with a grid spacing ranging from 0.5 to 1.5. From the data shown in Table 2, the best statistical results among all models were obtained with region focusing weighted by a S.D. × coefficient values of 0.8 and a grid spacing of 1.0 ($r^2 = 0.963$, SEE = 0.097, $q^2 = 0.742$, SEP = 0.255, and seven components). The region focusing not only increased q^2 values during the process of model generation, but also resulted in the refinement of 3D contour maps. In order to check the reliability of the PLS models generated in our studies to small systematic perturbations of the response variable, we have used the progressive scrambling method [20], which is a nonparametric approach that does not disturb the underlying covariance structure of the data, being used to determine the sensitivity of a QSAR model to chance correlations. The scrambling experiments to ascertain the validity of the model lead to models with an average q^2 of 0.450, associated with higher values of calculated crossvalidated standard error (cSDEP), which is modeled as a function of the correlation coefficient between the true values of the dependent variables (Y) and the perturbed values of the dependent variables (Y'). The low q^2 values obtained for the scrambling experiments confirm the robustness of our model (results not shown).

The predictive ability of the most significant CoMFA model derived using the 77 training set molecules was assessed by predicting pEC_{50} values for 20 test set molecules (compounds **78–97**, supplementary information), which were not included in the training set. The external validation process can be considered the most valuable validation method for a predictive QSAR model, as the test set compounds are completely excluded during the training of the model. The results obtained in the external validation for the 20 test set compounds are listed in Table 3, and the graphic results for the experimental versus predicted activities of both training set and test set are displayed in Fig. 3.

The good agreement between experimental and predicted values for the 20 test set compounds indicates the reliability of the constructed CoMFA model (see Table 3 and Fig. 3). The predicted pEC_{50} values for the test set compounds fall close to the experimental values, not deviating by more than 0.41 log units. No outliers were detected in this series of FXR activators. From the low residual values, it can be seen that the CoMFA model obtained is highly reliable and can be used to predict the biological activity of novel compounds within this structural class.

Another important role of a CoMFA model, besides predicting the biological property of interest for untested

Table 3
Experimental and predicted activities (pEC_{50}) with residual values for the test set compounds employed in CoMFA analyses

Test set compounds	pEC_{50}		
	Experimental	Predicted	Residual[a]
78	6.61	6.41	0.20
79	6.66	6.47	0.19
80	6.00	5.89	0.11
81	6.48	6.60	−0.12
82	6.63	6.69	−0.06
83	6.00	6.06	−0.06
84	6.69	6.68	0.01
85	6.82	6.89	−0.07
86	6.79	7.05	−0.26
87	6.37	6.30	0.07
88	6.73	6.76	−0.03
89	6.51	6.92	−0.41
90	7.14	7.02	0.12
91	7.05	7.15	−0.10
92	7.01	7.27	−0.26
93	6.93	7.29	−0.36
94	6.35	6.13	0.22
95	6.89	6.78	0.11
96	6.58	6.46	0.12
97	6.79	6.55	0.24

[a] The difference between experimental and predicted values.

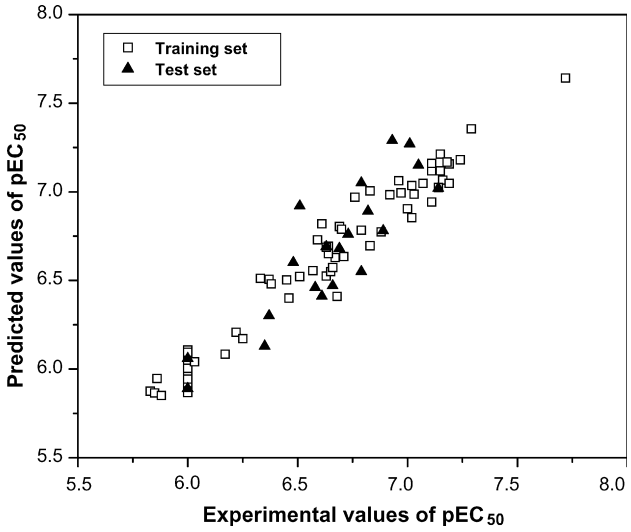


Fig. 3. Plot of predicted vs. experimental values of pEC_{50} for the 97 FXR activators (training and test sets).

molecules, is to provide hints about what molecular regions are directly related to biological activity. In CoMFA, visualization of the steric and electrostatic interaction fields is essential for further SAR studies and the design of novel molecules with improved properties, using an association of medicinal chemistry knowledge with expertise in synthetic chemistry. Fig. 4 displays the PLS S.D. × coefficient contour maps obtained from CoMFA for steric and electrostatic fields for the most potent activator of the data set (compound **33**, $pEC_{50} = 7.72$). In CoMFA contour maps, unfavorable steric regions are represented in yellow and favorable steric regions in green, while red contours represent regions where electronegative substituents may increase the biological activity, and the blue contours indicate regions where electropositive groups would contribute to enhance the activity. According to the CoMFA/PLS analysis, the steric and electrostatic field properties contribute in a 45/55 ratio to the total variance, respectively. In Fig. 4a, the favorable steric contours show that there is still room for bulkier groups at the methyl groups region of the alkyl amide. It is also possible to see in Fig. 4a two unfavorable steric contours. The most significant information obtained is related to the cyclic acetal group, suggesting that addiction of bulky substituents in this region may decrease the biological activity. Fig. 4b shows several blue contours surrounding the aromatic rings, indicating that electropositive substituents would generate compounds with increased biological activity. The small red contour surrounding the carbonyl oxygen at the methyl acrylate moiety suggests that electronegative substituents would enhance the biological activity.

The understanding of the molecular determinants of ligand binding is critical in drug design. In cases where the crystal structure of the target protein in complex with a ligand is available, structure-based molecular docking can be used to investigate the detailed intermolecular interactions. The most favorable conformation of the compound **33** was generated in the FXR active site by the docking program GOLD, as shown in

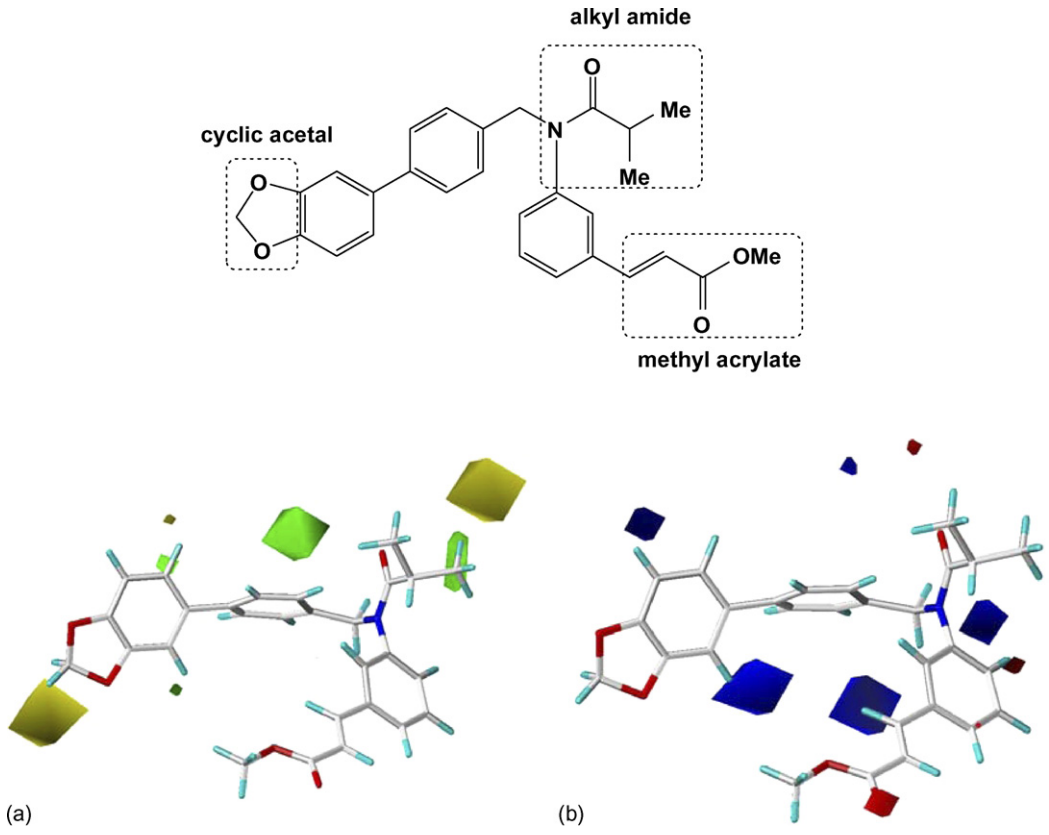


Fig. 4. (a) Steric and (b) electrostatic CoMFA contour maps for the most potent FXR activator of the series (**33**, $pEC_{50} = 7.72$).

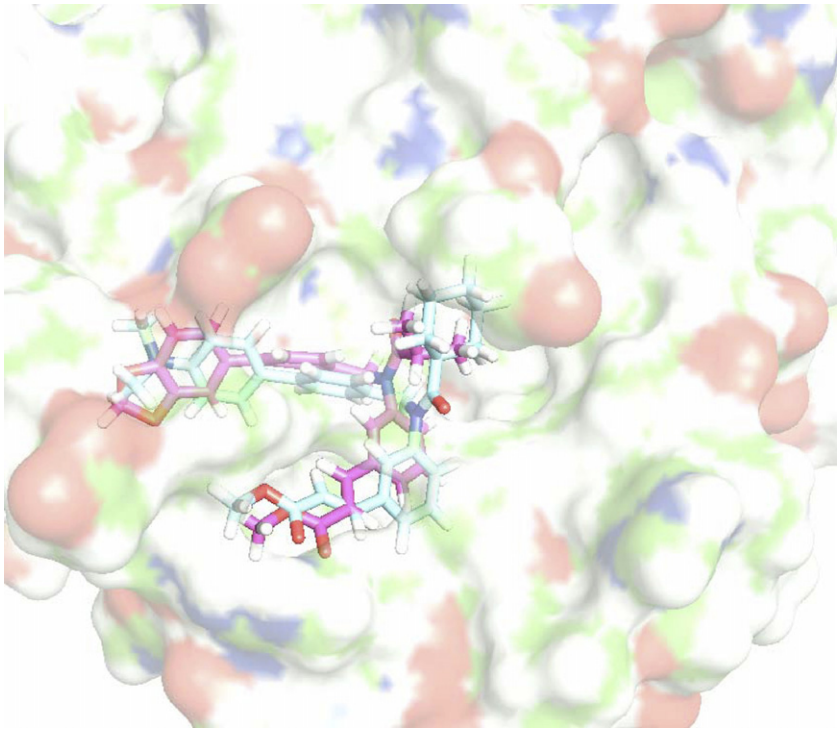


Fig. 5. Representation of the binding mode of the most potent activator of the series (**33**, colored in pink) into the crystal structure of FXR bound to fexaramine (in green).

Fig. 5 (in pink). The crystallographic binding mode of fexaramine (in green) in the FXR active site (PDB code: 1OSH) is also shown in Fig. 5. As it can be seen, all the main interactions between FXR and fexaramine are also found in the docked compound used as template for the CoMFA studies. In particular, intermolecular interactions that stabilized the hexyl group of fexaramine, the biaryl rings, as well as the methyl ester and dimethyl amine moieties. This group of protein–small molecule interactions indicates that the 3D aligned molecules that produced statically significant CoMFA models are also in good agreement with the crystallographic data.

4. Conclusions

The 3D CoMFA model described herein possesses good internal and external consistency. The good correlation between experimental and predicted pEC_{50} values for the test compounds further proved the predicted power of the constructed QSAR model. CoMFA contour maps are important tools in medicinal chemistry, as they show regions in 3D space where modifications of steric and electrostatic fields are strongly correlate with concomitant changes in biological activity. Moreover, the 3D QSAR model generated is compatible with the 3D protein environment in the FXR binding site, suggesting possible synergism between QSAR and structure-based technologies. The QSAR model should be useful for the design of new structurally related activators of FXR having improved potency.

Acknowledgment

The authors would like to thank financial support from FAPESP, The State of São Paulo Research Foundation, Brazil.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2006.09.003.

References

[1] N.Y. Kalaany, D.J. Mangelsdorf, LXRs and FXR: the Yin and Yang of cholesterol and fat metabolism, Ann. Rev. Physiol. 68 (2006) 159–191.
[2] H. Gronemeyer, J.A. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily, Nat. Rev. Drug Discov. 3 (2004) 950–964.
[3] K.H. Pearce, M.A. Iannone, C.A. Simmons, J.G. Gray, Discovery of novel nuclear receptor modulating ligands: an for peptide interaction integral role profiling, Drug Discov. Today 9 (2004) 741–751.
[4] R. Pellicciari, A. Gioiello, G. Costantino, Potential therapeutic applications of farnesoid X receptor (FXR) modulators, Expert Opin. Ther. Pat. 16 (2006) 333–341.
[5] D. Duran-Sandoval, B. Cariou, J.C. Fruchart, B. Staels, Potential regulatory role of the farnesoid X receptor in the metabolic syndrome, Biochimie 87 (2005) 93–98.
[6] F. Kuipers, T. Claudel, E. Sturm, B. Staels, The farnesoid X receptor (FXR) as modulator of bile acid metabolism, Rev. Endocr. Metab. Disorders 5 (2004) 319–326.
[7] R. Pellicciari, S. Fiorucci, E. Camaioni, C. Clerici, G. Costantino, P.R. Maloney, A. Morelli, D.J. Parks, T.M. Willson, 6Rethyl-chenodeoxycholic acid (6-ECDCA), a potent and selective FXR agonist endowed with anticholestatic activity, J. Med. Chem. 45 (2002) 3569–3572.
[8] P.R. Maloney, D.J. Parks, C.D. Haffner, A.M. Fivush, G. Chandra, K.D. Plunket, K.L. Creech, L.B. Moore, J.G. Wilson, M.C. Lewis, S.A. Jones, T.M. Willson, Identification of a chemical tool for the orphan nuclear receptor FXR, J. Med. Chem. 43 (2000) 2971–2974.
[9] R. Pellicciari, G. Costantino, S. Fiorucci, Farnesoid X receptor: from structure to potential clinical applications, J. Med. Chem. 48 (2005) 5383–5403.
[10] T. Claudel, B. Staels, F. Kuipers, The farnesoid X receptor: a molecular link between bile acid and lipid and glucose metabolism, Arterioscler. Thromb. Vasc. Biol. 25 (2005) 2020–2031.

[11] A.D. Andricopulo, C.A. Montanari, Structure–activity relationships for the design of small-molecule inhibitors, Mini-Rev. Med. Chem. 5 (2005) 585–593.
[12] K.M. Honorio, R.C. Garratt, I. Polikarpov, A.D. Andricopulo, Hologram QSAR studies on farnesoid X receptor activators, Lett. Drug Des. Discov. 3 (2006) 261–267.
[13] K.M. Honorio, R.C. Garratt, A.D. Andricopulo, Hologram quantitative structure–activity relationships for a series of farnesoid X receptor activators, Bioorg. Med. Chem. Lett. 15 (2005) 3119–3125.
[14] M.S. Castilho, M.P. Postigo, C.B.V. de Paula, et al., Two- and three-dimensional quantitative structure–activity relationships for a series of purine nucleoside phosphorylase inhibitors, Bioorg. Med. Chem. 14 (2006) 516–527.
[15] H. Fang, W. Tong, W.J. Welsh, D.M. Sheehan, QSAR models in receptor-mediated effects: the nuclear receptor superfamily, J. Mol. Struct. (Theochem) 622 (2003) 113–125.
[16] K.C. Nicolaou, R.M. Evans, A.J. Roecker, R. Hughes, M. Downes, J.A. Pfefferkorn, Discovery and optimization of non-steroidal FXR agonists from natural product-like libraries, Org. Biomol. Chem. 1 (2003) 908–920.
[17] M. Downes, M.A. Verdecia, A.J. Roecker, R. Hughes, J.B. Hogenesch, H.R. Kast-Woelbern, M.E. Bowman, J.L. Ferrer, A.M. Anisfeld, P.A. Edwards, J.M. Rosenfeld, J.G. Alvarez, J.P. Noel, K.C. Nicolaou, R.M. Evans, A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR, Mol. Cell 11 (2003) 1079–1092.
[18] A.M. Doweyko, 3D-QSAR illusions, J. Comput. Aid. Mol. Des. 18 (2004) 587–596.
[19] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular-field analysis (COMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[20] R.D. Clark, P.C. Foz, Statistical variation in progressive scrambling, J. Comput. Aid. Mol. Des. 18 (2004) 563–576.