

Hologram quantitative structure–activity relationships for a series of farnesoid X receptor activators

Kathia M. Honorio, Richard C. Garratt and Adriano D. Andricopulo*

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

Received 10 March 2005; revised 7 April 2005; accepted 11 April 2005

Available online 12 May 2005

Abstract—The farnesoid X receptor (FXR) is an attractive drug target for the development of novel therapeutic agents for the treatment of dyslipidemia and cholestasis. Hologram quantitative structure–activity relationship (HQSAR) studies were conducted on a series of potent FXR activators originated from natural product-like libraries. A training set containing 82 compounds served to establish the models. The best HQSAR model was generated using atoms, bonds, connections, chirality, and donor and acceptor as fragment distinction and fragment size default (4–7) with six components. 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 HQSAR model and the information obtained from HQSAR 2D contribution maps should be useful for the design of novel FXR ligands having improved potency.

© 2005 Elsevier Ltd. All rights reserved.

Nuclear receptors (NRs) have been extensively studied in recent years as they affect a wide variety of biological functions, including fatty acid metabolism, reproductive development, and detoxification of foreign substances.¹ The activity of many nuclear receptors is controlled by the binding of small, lipophilic ligands that include hormones, metabolites such as fatty acids, bile acids, oxysterols, xeno-, and endobiotics. Recognition of the importance of NRs as master regulators of genes involved in metabolic control suggests that these receptors are potential drug targets for the development of therapeutic agents for atherosclerosis, diabetes, and obesity.^{2,3} One of the nuclear receptors that has attracted medicinal chemistry interest is the farnesoid X receptor (FXR).^{4,5} First discovered as an orphan nuclear receptor, it was found subsequently to have bile salts chenodeoxycholic acid (CDCA), deoxycholic acid, and lithocholic acid (LCA) as its natural endogenous ligands.³ FXR plays a central role in regulation of bile acid synthesis, conjugation and transport, as well as lipid metabolism. Both FXR and LXR (liver X receptor) control lipoprotein metabolism, reverse cholesterol transport (RCT), and triglyceride metabolism.⁶ Because of its many attractive applications, selective FXR li-

gands have been recognized as potential drug candidates for the treatment of diseases related to cholesterol, bile acid metabolism, and homeostasis.^{7,8}

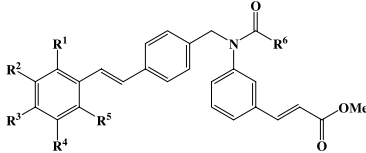
As part of our ongoing research program aimed at discovering new potent FXR activators, and in order to investigate the quantitative structure–activity relationships (QSAR) of a large series of FXR activators, we have employed the hologram QSAR (HQSAR) method to generate predictive 2D QSAR models. In this method, each molecule in the data set is broken down into a series of unique structural fragments, which are arranged to form a molecular hologram. HQSAR encodes all possible molecular fragments (linear, branched, and overlapping). It is interesting to note that the number of QSAR studies concerning the chemical structure and corresponding biological information for NRs is significant, as a result of the SAR available information. However, to the best of our knowledge, the majority of QSAR models reported in the literature is associated with the estrogen receptor (ER).⁹ This proves the importance of further QSAR studies involving other NRs such as FXR, the molecular target of our studies.

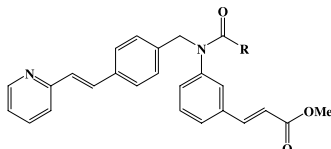
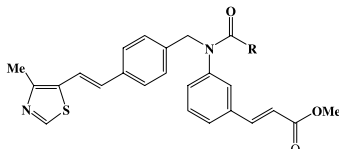
The data set of 102 compounds used for the HQSAR analyses was selected from the literature.⁸ The chemical structures and biological properties for the complete set of compounds are listed in Table 1. The EC₅₀ values

Keywords: HQSAR; Drug design; Nuclear receptors; FXR.

* Corresponding author. Tel.: +55 16 3373 8095; fax: +55 16 3373 9881; e-mail: aandrico@ifsc.usp.br

Table 1. Chemical structures and EC₅₀ (μM) values for a series of FXR activators

Training set molecules ^a							
							
Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	EC ₅₀
1	Cl	H	H	H	Cl	CH(CH ₃) ₂	0.195
2	H	CF ₃	H	CF ₃	H	cHex	1.470
3	H	CF ₃	H	CF ₃	H	CH(CH ₃) ₂	1.950
4	H	CF ₃	H	H	H	CH(CH ₃) ₂	0.267
5	H	CF ₃	H	H	H	NHCH(CH ₃) ₂	0.932
6	F	H	H	H	F	CH(CH ₃) ₂	0.108
7	F	H	H	H	H	cHex	0.064
8	F	H	H	H	H	NHCH(CH ₃) ₂	0.431
9	Me	H	Me	H	Me	CH(CH ₃) ₂	0.149
10	Me	H	Me	H	Me	NHCH(CH ₃) ₂	0.431
11	H	F	H	H	H	cHex	0.086
12	H	F	H	H	H	CH(CH ₃) ₂	0.071
13	H	H	F	H	H	CH(CH ₃) ₂	0.120
14	H	H	F	H	H	NHCH(CH ₃) ₂	0.348

					
Cpd	R	EC ₅₀	Cpd	R	EC ₅₀
15	cHex	0.309	17	cHex	0.227
16	CH(CH ₃) ₂	0.310	18	CH(CH ₃) ₂	0.228

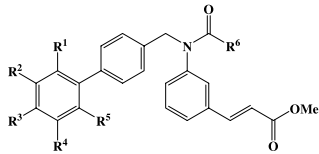
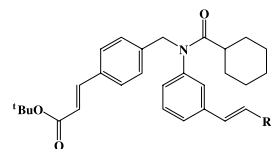
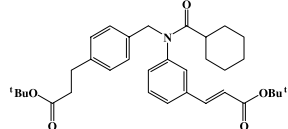
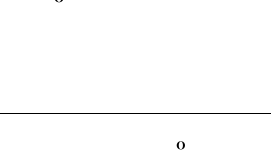
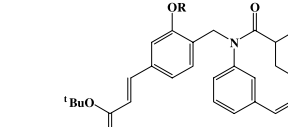
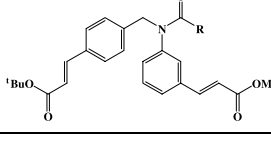
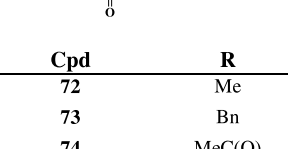
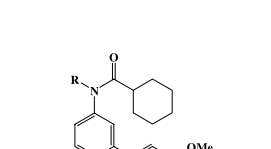
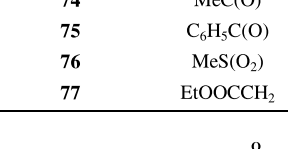
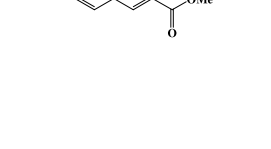
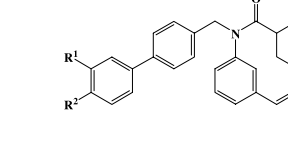
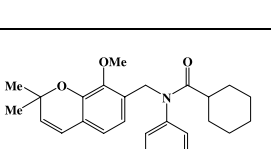
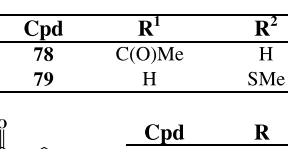
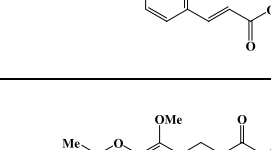
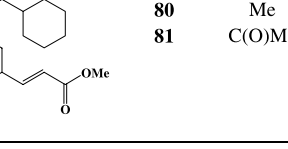
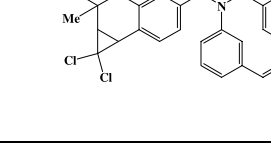
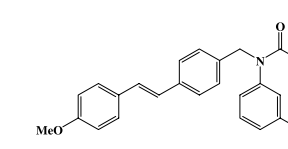
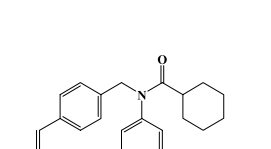
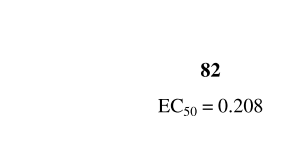
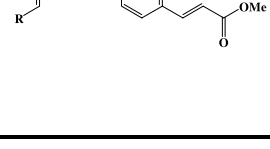

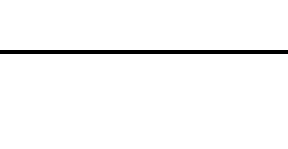

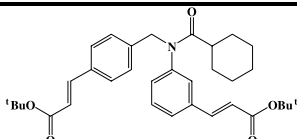
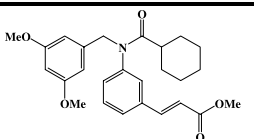
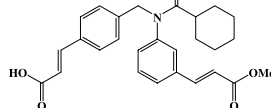
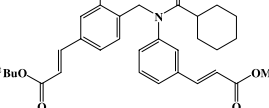
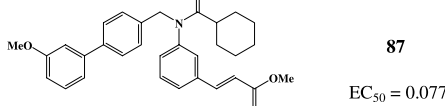
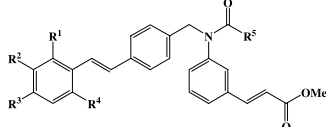
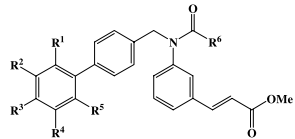
							
Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ⁶	EC ₅₀
19	H	F	F	H	H	cHex	0.072
20	H	H	SMe	H	H	CH(CH ₃) ₂	0.051
21	H	OMe	H	H	H	cHex	0.101
22	H	OMe	H	H	H	NHCH(CH ₃) ₂	1.370
23	H	OEt	H	H	H	cHex	0.147
24	H	OEt	H	H	H	CH(CH ₃) ₂	0.173
25	H	H	OMe	H	H	cHex	0.089
26	H	H	OMe	H	H	CH(CH ₃) ₂	0.097
27	H	Cl	H	H	H	cHex	0.094
28	H	Cl	H	H	H	CH(CH ₃) ₂	0.077
29	H	H	Me	H	H	CH(CH ₃) ₂	0.118
30	H	H	Me	H	H	NHCH(CH ₃) ₂	0.449
31	H	Me	H	H	H	cHex	0.109
32	H	Me	H	H	H	CH(CH ₃) ₂	0.163
33	OMe	H	H	Cl	H	CH(CH ₃) ₂	0.226
34	H	-OCH ₂ O-	H	H	H	cHex	0.038
35	H	-OCH ₂ O-	H	H	H	CH(CH ₃) ₂	0.019
36	H	-OCH ₂ O-	H	H	H	NHCH(CH ₃) ₂	0.096
37	H	Cl	F	H	H	cHex	0.066
38	H	H	OCF ₃	H	H	CH(CH ₃) ₂	0.219
39	H	OCF ₃	H	H	H	CH(CH ₃) ₂	0.247
40	OMe	H	H	H	OMe	cHex	0.077
41	OMe	H	H	H	OMe	CH(CH ₃) ₂	0.095
42	H	H	NMe ₂	H	H	cHex	0.025
43	H	H	NMe ₂	H	H	CH(CH ₃) ₂	0.057
44	H	H	NMe ₂	H	H	NHCH(CH ₃) ₂	0.162
45	H	H	t-Bu	H	H	cHex	0.132

Table 1 (continued)

	Cpd	R	EC₅₀	 <p>71 EC₅₀ > 1.0</p>
	46	CONH ₂	> 1.0	
	47	CH ₂ OMe	0.243	
	48	CH ₂ OEt	0.220	
	49	phenyl	0.236	
	50	2-furyl	0.205	
	51	isopropylamino	0.212	
	52	benzylamino	> 1.0	
	53	H	> 1.0	
	54	methyl	> 1.0	
	55	benzyl	> 1.0	
	56	2-naphthyl	0.680	
	57	2-bromobenzyl	> 1.0	
	58	3-bromobenzyl	> 1.0	
	59	4-bromobenzyl	> 1.0	
	60	4-tert-butylbenzyl	> 1.0	
	61	3-methoxybenzyl	> 1.0	
	62			
	63			
	EC ₅₀ = 0.358			
	64	COOMe	> 1.0	
	65	COOEt	> 1.0	
	66	COOBn	> 1.0	
	67	CONMe ₂	> 1.0	
	68	CONH ^t Bu	> 1.0	
	69	CH ₂ OMe	0.233	
	70	CH ₂ OEt	0.198	
				
				
				<p>82 EC₅₀ = 0.208</p>

(continued on next page)

Table 1 (continued)

Test set molecules ^a							
			83 EC ₅₀ > 1.0		 84 EC ₅₀ = 0.606		
			85 EC ₅₀ > 1.0		 86 EC ₅₀ > 1.0		
			87 EC ₅₀ = 0.077				
							
Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	EC ₅₀	
88	Cl	H	H	Cl	cHex	0.150	
89	H	Cl	H	H	cHex	0.165	
90	H	Cl	H	H	CH(CH ₃) ₂	0.164	
91	F	H	H	H	CH(CH ₃) ₂	0.070	
92	H	H	H	H	CH(CH ₃) ₂	0.065	
93	H	F	H	H	NHCH(CH ₃) ₂	0.467	
94	H	H	F	H	cHex	0.185	
							
Cpd	R ¹	R ²	R ³	R ⁴	R ⁵	R ₆	EC ₅₀
95	H	OMe	H	H	H	CH(CH ₃) ₂	0.072
96	H	Cl	H	H	H	NHCH(CH ₃) ₂	1.400
97	H	Me	H	H	H	NHCH(CH ₃) ₂	1.330
98	OMe	H	H	Cl	H	cHex	0.233
99	H	Cl	F	H	H	CH(CH ₃) ₂	0.129
100	H	H	OCF ₃	H	H	cHex	0.264
101	H	OCF ₃	H	H	H	cHex	0.420
102	OMe	H	H	H	OMe	NHCH(CH ₃) ₂	0.561

^a All values of EC₅₀ are reported in μM.

vary from 0.019 to 1.950 μM (Table 1) and were measured under the same experimental conditions,⁸ a fundamental requirement for successful QSAR studies. The QSAR modeling analyses, calculations, and visualizations were performed using the SYBYL 6.9.2 package (Tripos Inc., St. Louis, USA) running on Red Hat Linux 7.3 workstations. The EC₅₀ values were converted to the corresponding pEC₅₀ (−logEC₅₀) and used as dependent variables in the QSAR investigations. The pEC₅₀ values span approximately two orders of magnitude and property values are acceptably distributed across the range of values. From the original data set of 102 FXR activators, 82 compounds (1–82, Table 1) were selected as members of the training set for model construction, and the other 20 compounds (83–102, Table 1) as members of the test set for external model validation. Training and test sets were selected such that structurally diverse molecules possessing activities of a wide range were included in both sets. Thus, the data set is suitable for QSAR model development.

In HQSAR, each molecule is hashed to a molecular fingerprint, which encodes the frequency of occurrence of various molecular fragment types using a pre-defined set of rules. To construct a molecular hologram, this molecular fingerprint is cut into strings at a fixed interval as specified by a hologram length (HL) parameter. The strings are then aligned and the sum of each column constitutes the individual component of the molecular hologram of a particular length. The progress of incorporating information about each fragment, and each of its constituent sub-fragments, implicitly encodes 3D structural information (such as hybridization and chirality). It is possible that when the molecular holograms produce a QSAR model, the most important fragments for the biological activity are hashed to the same bin site in the hologram, without discrimination. This phenomenon known as fragment collision can lead to poor models with little predictive power. To avoid bad collision, a HQSAR run requires selecting values for parameters that specify the size of the hologram that is to

be used, as well as the size and type of fragment substructures that are to be encoded.^{10,11} Thus, final HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction. In our studies, holograms were generated using the standard parameters implemented in SYBYL 6.9.2. The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). In order to assess the process of hologram generation, several combinations of these parameters were considered using the fragment size default (4–7) as follows: AB, ABC, ABCH, ABCHCh, ABCHChDA, ABH, ABCCh, ABDA, ABCDA, ABHDA, ABCChDA, ABCHDA, and ABHChDA. The HQSAR analysis was performed by screening the 12 default series of hologram length values ranging from 53 to 401 bins. The patterns of fragment counts from the training set compounds were then related to the measured biological activity. All models generated in our studies were investigated using full cross-validated q^2 (Partial Least Squares (PLS) Leave-One-Out (LOO) method. The predictive ability of the models was assessed by their q^2 values. The statistical results from the PLS analyses for the 82 training set compounds using several fragment distinction combinations are presented in Table 2.

As it can be seen in Table 2, the best statistical results among all models were obtained for model 11 ($q^2 = 0.841$, and $r^2 = 0.931$). This model was derived using atoms, bonds, connections, chirality, and donor and acceptor, with 6 being the optimum number of PLS components. The results also indicate that the use of other fragment distinctions into molecular hologram decreases the quality of the models as measured by statistical parameters in Table 2. It is worth noting that our results confirm previous studies, which indicate that when donor and acceptor are being used in the fragment generation, the hydrogen atoms should not be considered.¹² This can be explained by the drastic increase in the number of fragments generated when both of these

Table 3. HQSAR analysis for the influence of various fragment sizes on the key statistical parameters using the best fragment distinction (atoms, bonds, connections, chirality, donor, and acceptor)

Fragment size	q^2	SEP	r^2	SEE	HL	N
2–5	0.802	0.221	0.923	0.138	307	6
3–6	0.822	0.209	0.926	0.135	353	6
4–7	0.841	0.198	0.931	0.130	353	6
5–8	0.803	0.220	0.902	0.155	199	6
6–9	0.800	0.222	0.901	0.157	199	6
7–10	0.789	0.228	0.888	0.166	61	6

options are considered in the model construction. However, due to the intrinsic nature of different data sets, it is important to test various combinations of fragments in order to obtain the best final HQSAR model.

The influence of different fragment sizes in the statistical parameters was further investigated for the best HQSAR model generated previously (model 11, Table 2). Fragment size parameters control the minimum and maximum length of fragments to be included in the hologram fingerprint. The HQSAR results for the different fragment sizes used are summarized in Table 3. The results show that the variation of the fragment size did not provide any improvement in the HQSAR model. The fragment size default (4–7) led to the better statistical results in comparison with other fragment sizes (Table 3).

As the structure encoded within a 2D fingerprint is directly related to biological activity of molecules within the training set, the HQSAR model is able to predict the activity of new related molecules from its fingerprint. The predictive power of the best HQSAR model derived using the 82 training set molecules was assessed by predicting pEC₅₀ values for 20 test set molecules (compounds 83–102, Table 1), which were not included in the training set. The external validation process can be considered the most valuable validation method, as these compounds are completely excluded during the training of the model. The results are listed in Table 4,

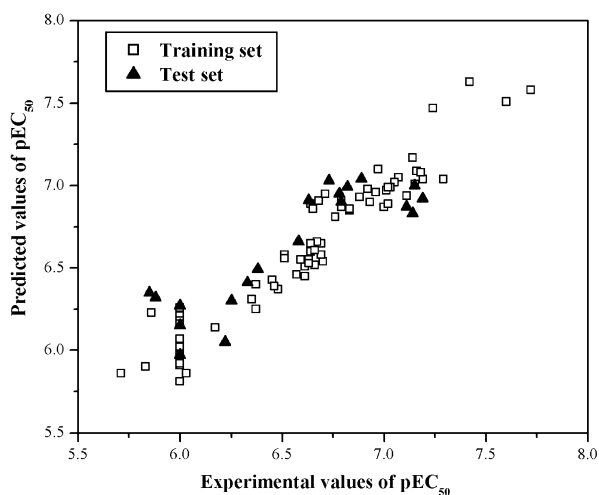
Table 2. Results of HQSAR analyses for various fragment distinctions on the key statistical parameters using fragment size default (4–7)

Model	Fragment distinction	q^2	SEP	r^2	SEE	HL	N
1	A/B	0.775	0.236	0.897	0.159	307	6
2	A/B/C	0.779	0.233	0.912	0.147	307	6
3	A/B/C/H	0.704	0.270	0.860	0.186	199	6
4	A/B/C/H/Ch	0.700	0.272	0.866	0.182	307	6
5	A/B/C/H/Ch/DA	0.691	0.276	0.872	0.178	307	6
6	A/B/H	0.681	0.281	0.844	0.196	353	6
7	A/B/C/Ch	0.807	0.219	0.923	0.138	307	6
8	A/B/DA	0.774	0.236	0.869	0.180	59	6
9	A/B/C/DA	0.813	0.215	0.924	0.137	401	6
10	A/B/H/DA	0.709	0.268	0.847	0.194	53	6
11	A/B/C/Ch/DA	0.841	0.198	0.931	0.130	353	6
12	A/B/C/H/DA	0.671	0.285	0.833	0.203	83	6
13	A/B/H/Ch/DA	0.718	0.264	0.843	0.197	53	6

q^2 , cross-validated correlation coefficient; SEP, cross-validated standard error; r^2 , noncross-validated correlation coefficient; SEE, noncross-validated standard error; HL, hologram length; N , optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

Table 4. Experimental and predicted activities (pEC_{50}) with residual values for the test set compounds

Test set compounds	pEC_{50}		
	Experimental	Predicted	Residual
83	6.00	6.15	0.15
84	6.22	6.05	-0.17
85	6.00	6.27	0.27
86	6.00	5.97	-0.03
87	7.11	6.87	-0.24
88	6.82	6.99	0.17
89	6.78	6.95	0.17
90	6.79	6.90	0.11
91	7.15	7.00	-0.15
92	7.19	6.92	-0.27
93	6.33	6.41	0.08
94	6.73	7.01	0.28
95	7.14	6.85	-0.29
96	5.85	6.35	0.50
97	5.88	6.32	0.44
98	6.63	6.91	0.28
99	6.89	7.04	0.15
100	6.58	6.66	0.08
101	6.38	6.49	0.11
102	6.25	6.30	0.05

**Figure 1.** Plot of predicted versus experimental values of pEC_{50} for the 102 FXR activators (training and test sets).

and the graphic results for the experimental versus predicted activities of both training set and test set are displayed in Figure 1.

The good agreement between experimental and predicted values for the test set compounds indicates the

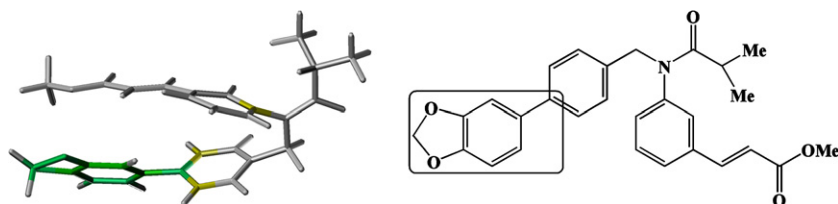
reliability of the constructed HQSAR model ($r^2 = 0.904$, $\text{SEE} = 0.158$). From the low residual values, it can be seen that the HQSAR model obtained is highly reliable and can be used to predict the biological activity of novel compounds within this structural class. The predicted values fall close to the experimental pEC_{50} values, deviating by less than 0.3 log units. The only exceptions are compounds **96** and **97**, the two weakest FXR activators in the test set, for which the predicted values are more substantially in error (0.50 and 0.44 log units, respectively).

An important role of a QSAR model, besides predicting the activities of untested molecules, is to provide hints about what molecular fragments are directly related to biological activity. This information, combined with knowledge of synthetic chemistry, could lead to the synthesis of new molecules with improved properties. In HQSAR, it is possible to visualize the individual contribution to activity of each atom in a given molecule of the data set through the generation of contribution maps. The HQSAR module implemented in SYBYL 6.9.2 uses a color code in order to discriminate the main atomic contributions to activity. The colors at the red end of the spectrum (red, red orange, and orange) reflect poor contributions, whereas colors at the green end (yellow, green blue, and green) reflect favorable contributions. Atoms with intermediate contributions are colored white. Figure 2 displays the individual atomic contributions for the most potent compound (**35**, $\text{pEC}_{50} = 7.72$) of the data set. In particular, it was found that one fragment of the molecular structure, represented by the 1,3-benzodioxol moiety, is suggested to be strongly related to the biological activity of this compound. Moreover, in any molecule, regions with intermediate or poor contributions can be identified as potential targets for molecular modification and further SAR studies.

In sum, the final HQSAR model described herein possesses both good internal and external consistency, which indicate, that this model is statistically robust with good correlative and predictive power. This HQSAR model and the information obtained from 2D contribution maps should be useful for the design of new potent FXR modulators that are structurally related to the training set compounds.

Acknowledgment

The authors acknowledge financial support from FAPESP (The State of São Paulo Research Foundation, Brazil).

**Figure 2.** Individual atomic contributions for the activity of the most potent FXR activator of the series.

References and notes

1. Khan, S. A.; Vanden Heuvel, J. P. *J. Nutr. Biochem.* **2003**, *14*, 554.
2. Francis, G. A.; Fayard, E.; Picard, F.; Auwerx, J. *Annu. Rev. Physiol.* **2003**, *65*, 261.
3. Redinger, R. N. *J. Lab. Clin. Med.* **2003**, *142*, 7.
4. Forman, B. M.; Goode, E.; Chen, J.; Oro, A. E.; Bradley, D. J.; Perlmann, T.; Noonan, D. J.; Burka, L. T.; McMorris, T.; Lamph, W. W.; Evans, R. M.; Weinberger, C. *Cell* **1995**, *81*, 687.
5. Claudel, T.; Sturm, E.; Kuipers, F.; Staels, B. *Exp. Opin. Invest. Drugs* **2004**, *13*, 1135.
6. Chiang, J. Y. L. *J. Hepatol.* **2004**, *40*, 539.
7. Mi, L.; Devarakonda, S.; Harp, J. M.; Han, Q.; Pellicciari, R.; Willson, T. M.; Khorasanizadeh, S.; Rastinejad, F. *Mol. Cell* **2003**, *11*, 1093.
8. Nicolaou, K. C.; Evans, R. M.; Roecker, A. J.; Hughes, R.; Downes, M.; Pfefferkorn, J. A. *Org. Biomol. Chem.* **2003**, *1*, 908.
9. Fang, H.; Tong, W.; Welsh, W. J.; Sheehan, D. M. *J. Mol. Struct. (Theochem)* **2003**, *622*, 113.
10. Chen, Da.; Yin, C.; Wang, X.; Wang, L. *Chemosphere* **2004**, *57*, 1739.
11. So, S.; Karplus, M. *J. Comput. Aided Mol. Des.* **1999**, *13*, 243.
12. HQSAR™ Manual, SYBYL 6.9.2, Tripos Inc., St. Louis, MO, 2003.