

Molecular Field Analysis and 3D-Quantitative Structure–Activity Relationship Study (MFA 3D-QSAR) Unveil Novel Features of Bile Acid Recognition at TGR5

Antonio Macchiarulo,[†] Antimo Gioiello,[†] Charles Thomas,[‡] Alberto Massarotti,[†] Roberto Nuti,[†] Emiliano Rosatelli,[†] Paola Sabbatini,[†] Kristina Schoonjans,[‡] Johan Auwerx,^{‡,§,||} and Roberto Pellicciari^{*,†}

Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, Via del Liceo 1, 06123 Perugia, Italy, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS/INSERM/ULP, 67404 Illkirch, France, Institut Clinique de la Souris, 67404 Illkirch, France, and Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Received June 10, 2008

Bile acids regulate nongenomic actions through the activation of TGR5, a membrane receptor that is G protein-coupled to the induction of adenylate cyclase. In this work, a training set of 43 bile acid derivatives is used to develop a molecular interaction field analysis (MFA) and a 3D-quantitative structure–activity relationship study (3D-QSAR) of TGR5 agonists. The predictive ability of the resulting model is evaluated using an external set of compounds with known TGR5 activity, and six bile acid derivatives whose unknown TGR5 activity is herein assessed with *in vitro* luciferase assay of cAMP formation. The results show a good predictive model and indicate a statistically relevant degree of correlation between the TGR5 activity and the molecular interaction fields produced by discrete positions of the bile acid scaffold. This information is instrumental to extend on a quantitative basis the current structure–activity relationships of bile acids as TGR5 modulators and will be fruitful to design new potent and selective agonists of the receptor.

INTRODUCTION

Known for many years for their solubilizing properties, bile acids (BAs) have been recently subjected to a paradigm shift.¹ The discoveries of signaling pathways of BAs, in particular, have upgraded these molecules from merely detergents to versatile signaling hormones endowed with endocrine functions. BAs regulate these pathways through two orders of action: as modulators of a specific nuclear hormone receptor, namely the farnesoid X receptor (FXR),^{2–4} and as ligands of a specific membrane G-protein-coupled receptor (GPCR), namely TGR5.^{5,6}

As for many other hormonal ligands, the action of BAs is referred to as nongenomic or nongenotropic whenever it is mediated by the binding to TGR5, whereas it is referred to as genomic or genotropic when it is mediated by the modulation of FXR.^{7–9}

While nongenomic actions of BAs constitute a signaling pathway to the control of immune response and energy metabolism,^{6,10–12} the genomic actions of BAs are the gateways to the control of cholesterol homeostasis, lipid, and carbohydrate metabolism.^{13–15}

These observations have fostered a considerable interest in BA-controlled signaling pathways as a source of attractive targets for the treatment of obesity, diabetes, and metabolic syndrome.¹⁶

In consequence, synthetic steroidal and nonsteroidal FXR agonists have been developed and are currently undergoing clinical testing for metabolic disorders.¹⁷ For instance, we have reported 6ECDCA (INT747, **1**, Chart 1) as the most potent steroid ligand for FXR so far available.¹⁸ 6ECDCA is actually in phase II clinical trials for the treatment of cholestasis in subjects with primary biliary cirrhosis and for the management of nonalcoholic steatohepatitis in patients with the metabolic syndrome. Crystallization studies and extensive structure-based molecular modeling approaches have allowed to gain insight into the molecular determinants of bile acid binding to the ligand binding domain (LBD) of FXR.^{19–24} These information have proven useful to the design and optimization of novel potent and selective chemical tools of bile acid genomic pathways. Conversely, the lack of structural information of TGR5 has hampered an in-depth analysis of the molecular determinants featuring the modulation of bile acid nongenomic pathways.

In this framework, while chemical intuition and biological screening of libraries of nonsteroidal compounds and natural products have been the way to the disclosure of selective TGR5 modulators (**2–4** Chart 1),^{25–28} ligand-based approaches have been the computational tools so far adopted to infer the structure–activity relationships of bile acid nongenomic actions.²⁹

Along this line of research, we have first reported that the introduction of a methyl group at the C-23(S) position of natural bile acids confers a marked selectivity for TGR5 over FXR, while the 6 α -alkyl substitution increases the potency at both receptors.²⁵ These results combine with the recently reported stereospecificity of molecular recognition of natural bile acids to TGR5.²⁸

* Corresponding author phone: +39 075 5855120; fax: +39 075 5855124; e-mail: rp@unipg.it.

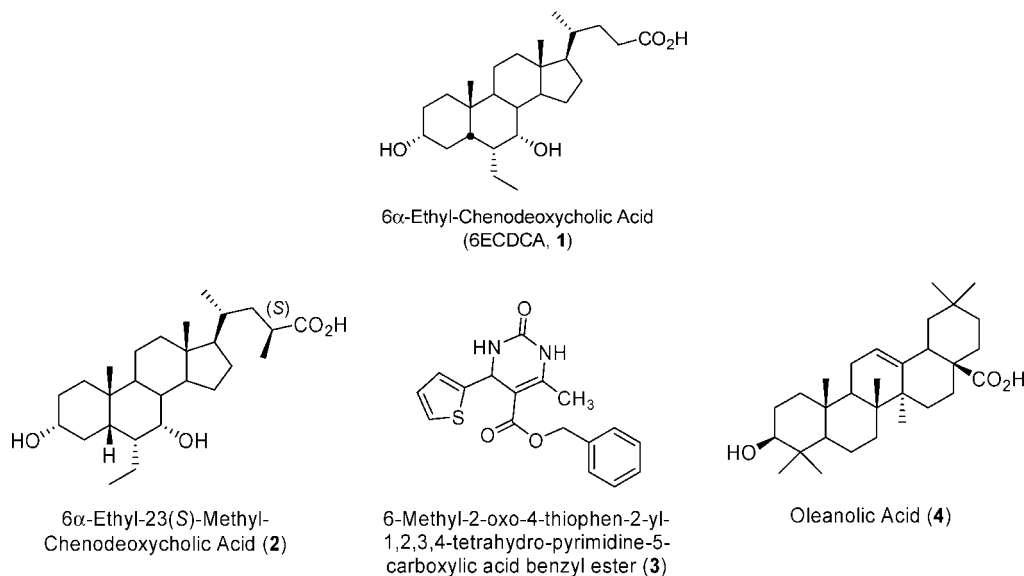
[†] Università di Perugia.

[‡] Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC).

[§] Institut Clinique de la Souris.

^{||} Ecole Polytechnique Fédérale de Lausanne.

Chart 1



More recently, the biological screening of a library of steroidal compounds has been instrumental to depict some features of the bile acid binding site in TGR5 as additionally to the aforementioned selectivity pocket lining the C-23(S) position of natural BAs.²⁹ These comprise a large polar site recognizing the acidic side chain of BAs, a hydrophobic pocket hosting the C6 and C7 positions of the BA steroid nucleus, and a narrow hydrogen bond donor site binding the 3-hydroxyl group of BAs. In order to indirectly assess the quantitative contribution that each of these pockets have on the agonistic activity of bile acids at TGR5, we herein report a molecular interaction field analysis (MFA) and a 3D-quantitative structure–activity relationship study (3D-QSAR) of 43 bile acid derivatives. The 3D-QSAR model is statistically validated using an external test set of compounds that includes six bile acid derivatives with unknown TGR5 activity that have been biologically evaluated using an *in vitro* luciferase assay of cAMP formation. These latter include four new compounds whose synthesis is herein reported. The inspection of the QSAR equation is instrumental in identifying molecular interaction fields produced by discrete positions of the bile acid scaffold that diversely affect the agonistic activity of bile acids at TGR5. Furthermore, the analysis of the coefficients of the QSAR equation reveals the quantitative contributions to the nongenomic activity that arise from the interaction of bile acids at each of the previously mapped binding pockets of TGR5.

RESULTS

Molecular Field Analysis. A structural alignment of 55 bile acid derivatives (Table 1) and six compounds with unknown TGR5 activity (Table 2) was generated using the bioactive conformation of the torsional χ -angle of the BA side chain as obtained according to the Experimental Section on the basis of the TGR5 activities of the C₂₂–C₂₃ cyclopropyl isomers of CDCA (**31**, **32**).²⁹ In particular, the better activity of 22*S*,23*S*-CCDCA (**31**) indicated a folded conformation of this angle as preferred by bile acids in binding to TGR5.

Inactive compounds and steroid hormones from the same study²⁹ were not included herein in order to focus our

attention only on the assessment of the quantitative contributions to the modulation of TGR5 of different regions of active bile acid derivatives. Although this arbitrary choice may be arguable, we deemed it advisable in light of using the combination of the resulting QSAR model with the previously reported binary classification model,²⁹ rather than only one single model, for future activity predictions of unknown compounds.

A grid box containing 4800 points was generated as reported in the Method section to enclose all the aligned molecules (Figure 1). In order to avoid the artifacts that might be introduced using regularly spaced grids for the placement of probes and interaction energies, the Cartesian coordinates of the interaction points were randomized. For each grid point, the interaction between three probe atoms and each compound was calculated yielding three molecular interaction fields. The functional probes were the proton, the methyl, and the hydroxyl group encoding respectively the electrostatic, the steric, and the hydrogen bond interactions. Each molecular interaction field was filtered to remove the interaction points with the lowest variance. After the filtering procedure, the interaction point energies were used as descriptors to perform stepwise linear regression analyses.

Stepwise Linear Regression Analysis. A training set of 43 bile acid derivatives was defined from the initial set of aligned compounds. In particular, the training set was composed of bile acid derivatives endowed with a TGR5 efficacy higher than 95 (compounds **1**, **2**, **5**–**9**, **5a**–**8a**, **5b**, **6b**, **8b**, **10a**, **11a**, **11b**, **15**–**18**, **20**, **22**–**24**, **22a**–**24a**, **25**–**34**, **37**–**42**).

The principle of linear regression is to model a quantitative dependent variable, the EC₅₀ of the compounds at TGR5, through a linear combination of quantitative explanatory variables, namely the interaction points. The model yields an equation that is used to infer the quantitative structure–activity relationships of the compounds. While the major advantage of the method is its computational simplicity that offers the possibility to easily interpret the resulting equation, its limitation is mainly linked to the ratio of objects and independent

Table 1. TGR5 Agonist Potency (Ref 29) of Natural Bile Acids and Their Tauro- or Glyco-Conjugated Forms (A); Body-Modified Bile Acids Derivatives and Their Tauro-Conjugated Forms (B); Side-Chain-Modified Bile Acids Derivatives (C); and Body- and Side-Chain-Modified Bile Acids Derivatives (D)^a

A)

Acid-form (4-12)

Tauro-form (4a-12a)

Glyco-form (4b-12b)

Trivial Name	R ₁	R ₂	R ₃	Acid-form EC ₅₀	Acid-form Efficacy	Tauro-form (a) EC ₅₀	Tauro-form (a) Efficacy	Glyco-form (b) EC ₅₀	Glyco-form (b) Efficacy
CDCA (5)	-H	α-OH	-H	6.71	105	1.92	103	3.88	105
CA (6)	-H	α-OH	α-OH	13.6	101	4.95	104	13.6	103
LCA (7)	-H	-H	-H	0.58	101	0.29	106	0.54	92
DCA (8)	-H	-H	α-OH	1.25	105	0.79	103	1.18	105
Lago-DCA (9)	-H	-H	β-OH	4.39	114	--	--	--	--
UDCA (10)	-H	β-OH	-H	36.4	74.9	30.0	97	33.9	91
HDCA (11)	α-OH	-H	-H	31.6	79.7	24.2	107	36.7	101
Muro-CA (12)	β-OH	-H	-H	4.89	94.4	--	--	--	--

B)

Acid-form

Tauro-form (a)

Trivial Name	R ₁	R ₂	R ₃	Acid-form EC ₅₀	Acid-form Efficacy	Tauro-form (a) EC ₅₀	Tauro-form (a) Efficacy
5β-Cholanic acid (13)	-H	-H	-H	6.88	76	--	--
LCA-S (14)	α-OSO ₃ H	-H	-H	>100	0 (100μM)	47.8	92
Dehydro-LCA (15)	-O	-H	-H	0.27	106	--	--
Iso-LCA (16)	β-OH	-H	-H	1.25	99	--	--
7β-Me-LCA (17)	α-OH	-CH ₃	-H	0.076	106	--	--
7α-F-LCA (18)	α-OH	α-F	-H	0.25	99	--	--
7β-F-LCA (19)	α-OH	β-F	-H	2.29	92	--	--
3-Dehydro-CDCA (20)	=O	α-OH	-H	3.98	107	--	--
3-Deoxy-CDCA (21)	-H	α-OH	-H	14.5	81	--	--
7β-Me-CDCA (22)	-H	α-OH β-CH ₃	-H	6.18	105	2.36	112
7β-Fl-CDCA (23)	-H	α-OH β-CH ₂ CH ₃	-H	2.63	99	0.73	99
7β-Pr-CDCA (24)	-H	α-OH β-CH ₂ CH ₂ CH ₃	-H	0.78	108	0.36	104

C)

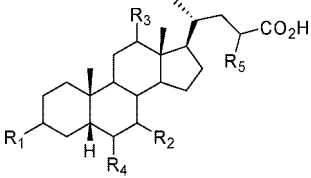
Trivial Name	R ₁	R ₂	R ₃	EC ₅₀	Efficacy
LCA Me Ester (25)	-H	-H	CO ₂ CH ₃	0.59	102
Nor-LCA (26)	-H	-H	CO ₂ H	0.77	102
CDCA Me Ester (27)	α-OH	-H	CO ₂ CH ₃	4.39	114
Nor-CDCA (28)	α-OH	-H	CO ₂ H	10.4	102
CDC-OH (29)	α-OH	-H	OH	0.12	103
CDC-Sul (30)	α-OH	-H	SO ₃ H	0.44	103
22S,23S-CCDCA (31)	α-OH	-H	CO ₂ H	1.33	110
22S,23R-CCDCA (32)	α-OH	-H	CO ₂ H	2.91	102
C-OH (33)	α-OH	α-OH	OH	0.87	103
C-Sul (34)	α-OH	α-OH	SO ₃ H	1.00	103
Nor-UDCA (35)	β-OH	-H	CO ₂ H	47.2	79
UDC-OH (36)	β-OH	-H	OH	2.18	79
UDC-Sul (37)	β-OH	-H	SO ₃ H	5.02	109

D)

Trivial Name	R ₁	R ₂	R ₃	R ₄	R ₅	EC ₅₀	Efficacy
23(S)-Me-CA (38)	α-OH	α-OH	α-OH	-H	-CH ₃	4.39	105
23(S)-Me-CDCA (39)	α-OH	α-OH	-H	-H	-CH ₃	3.58	110
23(R)-Me-CDCA (40)	α-OH	α-OH	-H	-H	-CH ₃	25.5	100
6α-Me-23(S)-Me-CDCA (41)	α-OH	α-OH	-H	α-CH ₃	-CH ₃	0.14	105
6α-Me-23(S)-Fl-CDCA (42)	α-OH	α-OH	-H	α-CH ₂ CH ₃	-CH ₃	0.10	102

^a Data represent average values of at least three independent experiments of CRE-driven luciferase reporter assays in TGR5-transfected CHO cells. Units are μM for EC₅₀ and % of 10 μM LCA value for efficacy.

Table 2. TGR5 Agonist Potency of Novel Body-Modified and Side-Chain-Modified Bile Acids^a

							
trivial name	R ₁	R ₂	R ₃	R ₄	R ₅	EC ₅₀	efficacy
CDCA (5)	α-OH	α-OH	-H	-H	-H	6.71	105
6α-Me-CDCA (43)	α-OH	α-OH	-H	α-CH ₃	-H	1.90	165
CA (6)	α-OH	α-OH	α-OH	-H	-H	13.6	101
6α-Me-23(S)-Me-CA (44)	α-OH	α-OH	α-OH	α-CH ₃	-CH ₃	0.80	76
6α-Me-23(R)-Me-CA (45)	α-OH	α-OH	α-OH	α-CH ₃	-CH ₃	12.4	81
Muro-CA (12)	α-OH	-H	-H	β-OH	-H	4.89	94.4
7αF-Muro-CA (46)	α-OH	α-F	-H	β-OH	-H	6.99	77
LCA (7)	α-OH	-H	-H	-H	-H	0.58	101
23(S)-Me-LCA (47)	α-OH	-H	-H	-H	-CH ₃	0.56	102
23(R)-Me-LCA (48)	α-OH	-H	-H	-H	-CH ₃	3.96	105

^a Data represent average values of at least three independent experiments of CRE-driven luciferase reporter assays in TGR5-transfected CHO cells. Units are μM for EC₅₀ and % of 10 μM LCA value for efficacy.

variables (Obs/Vi) that, as a rule of thumb, leads to a large risk in chance correlation if below a value of five.³⁰

A stepwise selection procedure was adopted to identify the interaction points highly correlated to the dependent variable

(see the Method section for details). This procedure was carried out using a diverse combination of the proton (H), methyl (M), and hydroxyl (O) interaction fields of the training set which led to generate the seven models reported in Table 3.

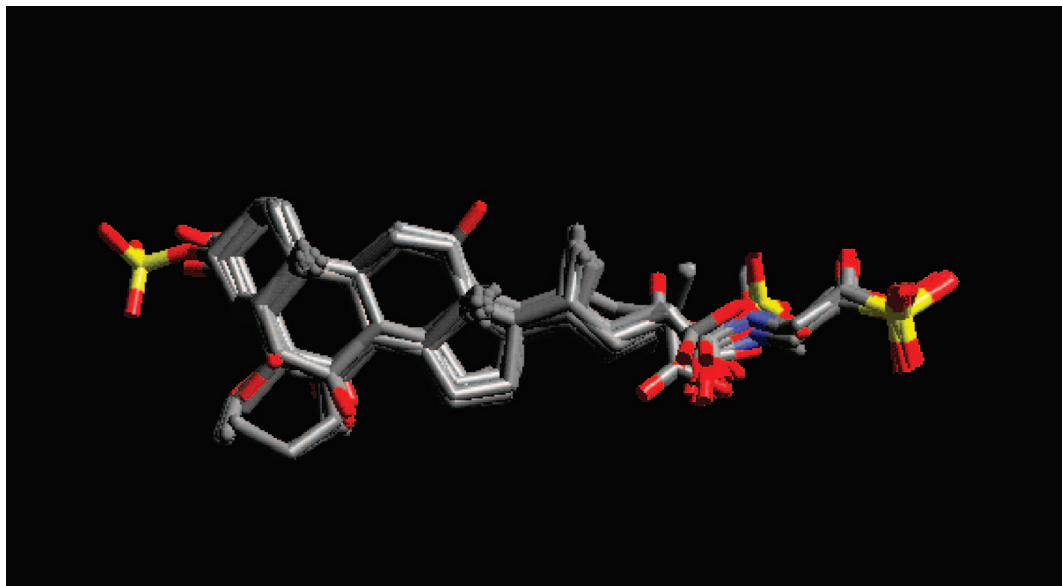


Figure 1. The grid box containing 4800 interaction points (red dots) and enclosing the alignment of the bile acids used in this study.

Table 3. Stepwise Linear Regression Analysis^a

model	R ²	R _{XV} ²	no. of variables	selected variable	RMSE ^b
a) proton	0.53	0.47	5	H25/H30/H49/H68/H73	0.526
b) methyl	0.35	0.30	3	M48/M50/M53	0.606
c) hydroxyl	0.34	0.29	3	O15/O37/O58	0.608
d) proton/methyl	0.75	0.69	9	H30/H47/H51/H53/H68/M32/M60/M69/M76	0.404
e) proton/hydroxyl	0.57	0.50	6	H30/H49/H68/H71/O16/O37	0.512
f) methyl/hydroxyl	0.58	0.50	6	M60/M69/O15/O20/O37/O61	0.508
g) proton/methyl/hydroxyl	0.65	0.58	7	H30/H49/H60/H68/M31/M50/O37	0.466

^a 3D-QSAR models (a-g) generated using a diverse combination of the proton (H), methyl (M), and hydroxyl (O) interaction fields. ^b Root mean square of prediction errors.

Table 4. Parameters of Model “g_{opt}”

source ^a	value ^b	standard error ^c	t ^d	Pr > t ^e	lower bound (95%) ^f	upper bound (95%) ^f
intercept	−0.512	0.413	−1.241	0.224	−1.353	0.328
H30	0.025	0.005	4.735	<0.0001	0.014	0.036
H49	−0.036	0.010	−3.689	0.001	−0.056	−0.016
H60	0.028	0.010	2.899	0.007	0.008	0.048
H68	0.054	0.011	4.884	<0.0001	0.032	0.077
M31	0.040	0.013	3.159	0.003	0.014	0.065
M50	−0.025	0.007	−3.637	0.001	−0.040	−0.011
O37	−0.027	0.006	−4.280	0.000	−0.040	−0.014

^a Intercept and independent variables of the 3D-QSAR equation. ^b Coefficients of intercept and independent variables. ^c Standard error of intercept and the coefficients. ^d Student's t. ^e Probability of intercept and variables. ^f Confidence intervals.

Although model “d”, which was obtained using the combination of the proton and methyl interaction fields, yielded the better regression (R²) and predictive (R_{XV}²) coefficients, it suffered for chance correlation due to the elevated number of selected variables compared to the size of the training set (Obs/Vi = 4.8).³⁰ Consequently, we deemed it advisable to consider model “g” as a good starting point being less complex than model “d” (Obs/Vi = 5.9) and presenting relatively good regression (R²) and predictive (R_{XV}²) coefficients. Model “g”, in particular, presented two outliers: CA (**6**; residual = −0.95) and Tauro-HDCA (**11a**; residual = −1.26). The removal of these compounds from the training set led to the model “g_{opt}” (objects = 41; variables = 7; Table 4) which is endowed with good regression (R² = 0.75) and predictive (R_{XV}² = 0.69) coefficients and a root mean square of prediction errors

(RMSE) of 0.386. The two outliers were inserted in the test set and are discussed below. The QSAR equation associated with model “g_{opt}” is reported in eq 1:

$$\text{pEC}_{50} = -0.512(\pm 0.413) + 0.025(\pm 0.005) \times H30 - 0.036(\pm 0.010) \times H49 + 0.028(\pm 0.010) \times H60 + 0.054(\pm 0.011) \times H68 + 0.040(\pm 0.040) \times M31 - 0.025(\pm 0.007) \times M50 - 0.027(\pm 0.006) \times O37 \quad (1)$$

The plot of experimental pEC₅₀ versus calculated pEC₅₀ for the molecules of the training set is shown in Figure 2, while statistics of the model are reported in Table 3.

The variables composing the equation are independent, as the largest Pearson's correlation coefficient is 0.31 for H30 and H49 (Table 5).

Validation Set. The reliability and predictivity of model “g_{opt}”, was evaluated using an external set of compounds.

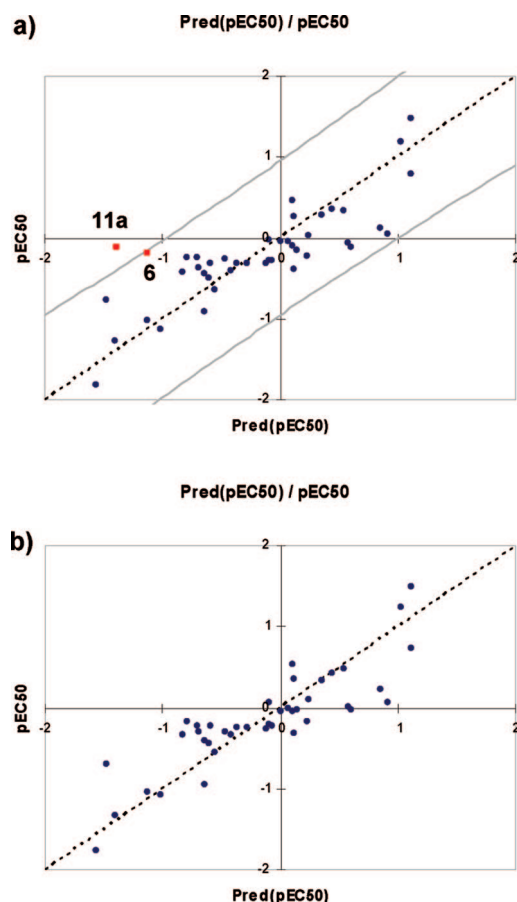


Figure 2. a) Plot of experimental pEC_{50} (x axis) versus predicted pEC_{50} (y axis) according to Model G. The outliers, CA (6: residual = -0.95) and Tauro-HDCA (11a: residual = -1.26), are labeled and highlighted with red dots. b) Plot of experimental pEC_{50} (x axis) versus predicted pEC_{50} (y axis) according to Model G_{opt} (objects = 41; variables = 7; Table 4).

The validation set, in particular, was composed of 20 compounds (6, 11a, outliers of the training set; 7b, 10–12, 10b, 13, 14a, 19, 21, 27, 35, 36, 43–48), including six bile acid derivatives (43–48) whose unknown TGR5 activities were assessed using an *in vitro* luciferase assay of cAMP formation. Compounds 44, 45, 47 and 48, in particular, were synthesized according to the Experimental Section.

Conversely to the training set, the validation set was intentionally designed to contain also compounds with a poor TGR5 efficacy in order to assess the influence of such parameter on the predictivity of model “ g_{opt} ”. The analysis of the residuals (Table 6) shows that six compounds out of 20 bile acid derivatives are not optimally predicted: UDCA (10: residual = -1.30) and Glyco-UDCA (10b: residual = -2.51); CA (6: residual = -1.05) and Tauro-HDCA (11a: residual = -1.34), being outliers also in the training set; muroCA (12: residual = -1.07) and Tauro-LCA-S (14a: residual = -1.82). Remarkably, no correlation is observed between the residual of prediction and the trend of efficacy in the validation set, hence indicating a negligible effect of this parameter on the predictivity of the 3D-QSAR model.

DISCUSSION

TGR5 is a metabotropic receptor specifically modulated by BAs. Increasing evidence suggests links between the activation of this G protein-coupled receptor and the stimula-

tion of glucagon-like protein 1 production in enteroendocrine cells and activation of thyroid hormone through the stimulation of type 2 iodothyronine deiodinase (D2) in brown adipose tissue and muscle. Consequently, the development of potent and selective TGR5 modulators represents a major advance not only in the pharmacological characterization of bile acid nongenomic pathway but also in the validation of TGR5 as a target for the treatment of metabolic syndrome.

At this end, we previously reported a detailed structure–activity relationship (SAR) of TGR5 agonists by appraising the activities to the receptor of a collection of natural occurring BAs, semisynthetic BA derivatives, and steroid hormones. The results allowed us to depict the binding site of TGR5 hosting bile acid derivatives as composed of four regions (Figure 3a): a selectivity pocket lining the C-23(S) position of natural BAs (region A); a large polar site recognizing the acidic side chain of BAs (region B); a hydrophobic pocket hosting the C6 and C7 positions of the BA steroid nucleus (region C); and a narrow hydrogen bond donor site binding the 3-hydroxyl group of BAs (region D). The study was also instrumental in developing a binary classification model on the basis of the steric and lipophilic properties of the molecular shape of bile acids and steroid hormones.

As a continuation of our research efforts in the field of bile acid nongenomic pathways, herein we have developed a molecular interaction field analysis (MFA) and a 3D-quantitative structure–activity relationship study (3D-QSAR) of TGR5 agonists. The 3D-QSAR model shows good regression ($R^2 = 0.75$) and predictive ($R_{XV}^2 = 0.69$) coefficients and a low root-mean-square of prediction errors (RMSE = 0.386). The inspection of the 3D-QSAR equation reveals seven independent variables that encode as many interaction points with the receptor (Figure 3b). Four of these points result from the proton interaction fields (H30, H49, H60, H68). The point H68 has the highest coefficient in the equation and is placed near the C6- α position of bile acid derivatives that is hosted in region C of TGR5. In agreement with our previously SAR data, the positive value of H68 coefficient indicates that electrostatic interactions are not favored in this hydrophobic region. Similarly, points H30 and H60 reveal two additional unfavorable electrostatic interactions at positions C24 and C7- β of bile acid steroid nucleus. While the interaction point H30 at C24 nicely explains why conjugated, C24-hydroxy and methyl ester derivatives of bile acids do not require a free carboxylic charged group to potentially activate TGR5, the interaction point H60 further sustains the hydrophobic character of region C recognizing the C6 and C7 positions of the BA steroid nucleus. Interestingly, H49 is the only proton interaction point that shows a favorable electrostatic interaction with the receptor. This interaction point is located near to positions C20 and C22 of the bile acid scaffold and accounts for the activity of bile acid derivatives with a shorter acidic side chain such as Nor-LCA (26). The opposite effect on the activity of the steric interaction points M31 and M50 underlines the stereospecificity of the selectivity pocket of TGR5 (region A). Indeed, region A is accessible only from the position (S)-C23, where M50 is located, and not from the position (R)-C23 of the bile acid side chain, where M31 is placed. Finally, the favorable hydrogen bond interaction

Table 5. Correlation Matrix^a

variables	H30	H49	H60	H68	M31	M50	O37
H30	1.00	0.31	-0.13	-0.02	-0.10	-0.18	0.08
H49	0.31	1.00	0.03	-0.07	0.07	-0.03	0.10
H60	-0.13	0.03	1.00	-0.27	-0.01	0.11	-0.18
H68	-0.02	-0.07	-0.27	1.00	0.10	0.09	0.02
M31	-0.10	0.07	-0.01	0.10	1.00	0.04	-0.22
M50	-0.18	-0.03	0.11	0.09	0.04	1.00	-0.38
O37	0.08	0.10	-0.18	0.02	-0.22	-0.38	1.00

^a Pearson's correlation coefficients between selected variables are shown; negative values mean inverse correlations (range from -1 to 0), and positive values mean direct correlations (range from 0 to 1).

Table 6. Validation Set

name	EXP.pEC ₅₀ ^a	efficacy ^b	PRED.pEC ₅₀ ^c	residual ^d
6 α -Me-23(S)-Me-CA (44)	0.097	76	-0.505	0.6
23(R)-Me-LCA (48)	0.256	102	-0.193	0.45
glyco-LCA (7b)	0.268	92	0.001	0.27
7 β -F-LCA (19)	-0.36	92	-0.469	0.11
7 α -F-muroCA (46)	-0.845	77	-0.666	-0.18
Nor-UDCA (35)	-1.674	79	-1.402	-0.27
6 α -Me-CDCA (43)	-0.279	165	0.035	-0.31
5 β -cholanic acid (13)	-0.784	76	-0.467	-0.32
6 α -Me-23(R)-Me-CA (45)	-1.093	81	-0.738	-0.35
CDCA Me ester (27)	-0.642	91	-0.283	-0.36
UDC-OH (26)	-0.338	79	0.03	-0.37
23(S)-Me-LCA (47)	-0.598	195	-0.099	-0.5
3-deoxy-CDCA (21)	-1.161	81	-0.58	-0.58
HDCA (11)	-1.5	80	-0.76	-0.74
CA (6)	-1.134	101	-0.086	-1.05
muroCA (12)	-0.689	94	0.38	-1.07
UDCA (10)	-1.561	75	-0.258	-1.3
tauro-HDCA (11a)	-1.384	107	-0.039	-1.34
tauro-LCA-S (14a)	-1.679	92	0.14	-1.82
glyco-UDCA (10b)	-1.53	91	0.977	-2.51

^a Experimental pEC₅₀ represents an average value of at least three independent experiments of CRE-driven luciferase reporter assays in TGR5-transfected CHO cells. ^b Efficacy units are % of 10 μ M LCA value. ^c Predicted pEC₅₀ values. ^d Residual of prediction obtained by the difference between EXP.pEC₅₀ and PRED.pEC₅₀.

point O37, being located close to position C12- α , accounts for the good activity shown by DCA (**8**) at TGR5.

Although we are aware of a general risk of occurring fallacy with QSAR models,³¹ the good agreement found between the interpretation of the model and the previously reported structure-activity relationships of TGR5 agonists (Figure 3) sustains the good reliance of our QSAR model.

In particular, the selected interaction points underline the roles of three of the four reported TGR5 binding regions, with region D hosting the 3-hydroxyl group of BAs being the only neglected. Thus, the steric interaction points sustain the role of region A in TGR5 selectivity, whereas the unfavorable high coefficient of the electrostatic interaction points H68 and H60 point out the hydrophobic character of region C as a key player in affecting the agonist potency. The low coefficient of the interaction point H30 that occupies part of region B indicates only a marginal contribution of this region to the activity. Interestingly, the QSAR model discloses two additional sites of bile acid recognition at TGR5: (i) a hydrogen bond interaction site hosting the position C12- α of the bile acid scaffold (interaction point O37) and (ii) a favorable electrostatic interaction point located at the beginning of the acidic side chain of the bile acid (interaction point H49).

The predictivity of the model has been assessed using an external set of twenty-two compounds, eight of which are

novel synthesized TGR5 agonists. While no correlation is observed between the residual of prediction and the efficacy of the compounds to the receptor, six out of twenty-two compounds are overpredicted with respect to their experimental EC₅₀. A close inspection of the EC₅₀ of such compounds reveals that four of them have TGR5 activity over 20 μ M (Tauro-HDCA **11a**, Glyco-UDCA **10b**, UDCA **10**, Tauro-LCA-S **14a**). Although we cannot rule out that for these compounds the poor prediction may arise from a limited compound sampling of the training set in that range of potency, it should be mentioned that the combination of this 3D-QSAR model with our previously reported binary active/inactive classification model²⁹ is able to improve these predictions with compound **11a** being the only outlier remaining (see below).

Cholic acid (**6**) was an outlier of the training set and is still predicted with higher activity in the test set; this might be expected if we compare the activities of the natural bile acids CA (**6**: EC₅₀ = 13.60 \pm 2.28 μ M) and CDCA (**5**: EC₅₀ = 6.71 \pm 1.35 μ M) with their relative 23(S)-methyl derivatives (23(S)-methyl-CA, **38** EC₅₀ = 4.39 \pm 1.12 μ M; 23(S)-methyl-CDCA, **39** EC₅₀ = 3.58 \pm 1.06 μ M). The increase of the potency observed in the case of CA after the introduction of the methyl group is unexpectedly higher than that observed in CDCA upon the same replacement. According to the idea of Maggiora about outliers in QSAR

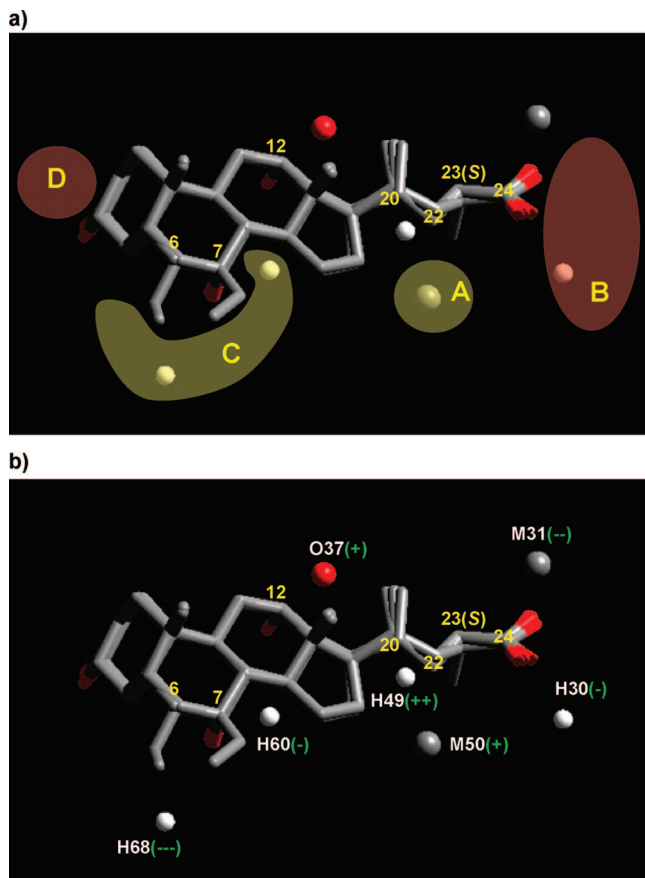


Figure 3. a) Map of the binding site of TGR5 as resulted from previously reported structure–activity relationships of bile acid derivatives (see ref 29). Red areas show polar binding pockets, and yellow regions indicate the presence of hydrophobic or sterically allowed sites. The selected interaction points of Model G_{opt} are depicted as color coded spheres in white (proton probe), red (hydroxyl probe), and gray (methyl probe). b) The selected interaction points are labeled according to the equation model G_{opt} . In bracket, the coefficients of the variables are encoded with plus signs (they positively affect the activity) or negative signs (they negatively affect the activity). The number of the signs encode a heavy or light effect depending on the value of the corresponding coefficient in the equation.

studies,³² the poor prediction of CA may be hence ascribed to the presence of activity cliffs around the insertion of a methyl group in the structure–activity optimization surface of this compound.

While a similar consideration can be ascribed also to muroCA (**12**), it is possible that in such a case the overprediction of the compound may arise from a poor estimation of the electrostatic interaction occurring at the C6- β position of the bile acid nucleus compared to the C6- α position.

Overall, the predictivity of the QSAR model can be considered satisfactory as far as the correct predictions cover the 70% of the validation set, and the incorrect predictions refer mainly to compounds with $EC_{50} > 10 \mu\text{M}$. In this regard, it is worth noting that if we combine the QSAR model with our previously reported discriminant model of TGR5 activity,²⁹ the predictivity further increases to 80% of correct predictions, with muroCA (**12**), 7 α -muroCA (**46**), and Glyco-LCA (**7b**) being false negatives (predicted $EC_{50} > 10 \mu\text{M}$) and Tauro-HDCA (**11a**) being the only false positive (residual of prediction = -1.34). This latter result is very

encouraging for future applications of the combination of both models, rather than one single model, to the activity prediction of unknown compounds.

CONCLUSIONS

BAs are endocrine molecules that trigger both genomic and nongenomic signaling pathways.¹ TGR5 is a metabotropic receptor G protein coupled to the formation of cAMP that mediates some of the nongenomic actions of BAs. The emerging roles of TGR5 in the control of energy homeostasis are paving the way for the development of modulators of this receptor as viable therapeutic strategies for the treatment of obesity, diabetes, and the metabolic syndrome.

In this work we have reported for the first time a quantitative structure–activity relationship of TGR5 agonists belonging to the class of bile acid derivatives. In particular, using molecular interaction fields and multiple linear regression analyses, we have provided a QSAR model that sheds new light on the molecular recognition of bile acids at TGR5. The interaction points of the QSAR equation, while in agreement with the previously reported structure–activity relationship of TGR5 agonists, disclose additional features of the TGR5 binding site. The results herein reported will be used in combination with our recently reported discriminant model of TGR5 activity to aid the design of novel selective agonists and to screen virtual libraries of compounds. Furthermore, the good predictivity shown by the QSAR model will be fruitful to prioritize the synthesis of a list of potential bile acid derivatives.

EXPERIMENTAL SECTION

Chemistry. General Method. Melting points were determined with a Buchi 535 electrothermal apparatus and are uncorrected. NMR spectra were obtained with a Bruker AC 200 MHz or 400 MHz spectrometer, and the chemical shifts are reported in parts per million (ppm). The abbreviations used are as follows: s, singlet; bs, broad singlet; d, doublet; dd, double doublet; m, multiplet. Flash column chromatography was performed using Merck silica gel 60 (0.040–0.063 mm). TLC were carried out on precoated TLC plates with silica gel 60 F-254 (Merck). Spots were visualized by staining and warming with a phosphomolybdate reagent (5% solution in EtOH). All reactions were carried out under a nitrogen atmosphere.

Synthesis and Characterization (Elemental Analysis, mp, NMR) of Target Compounds. 3 α ,7 α -Dihydroxy-6 α -ethyl-5 β -cholan-24-oic acid (**43**) and 3 α ,6 α -dihydroxy-7 α -fluoro-5 β -cholan-24-oic acid (**46**) were prepared as reported elsewhere.^{18,33} 23(S)-Methyl-3 α ,7 α ,12 α -trihydroxy-6 α -methyl-5 β -cholan-24-oic acid (**44**), 23(R)-methyl-3 α ,7 α ,12 α -trihydroxy-6 α -methyl-5 β -cholan-24-oic acid (**45**), 23(S)-methyl-3 α -hydroxy-5 β -cholan-24-oic acid (**47**), 23(R)-methyl-3 α -hydroxy-5 β -cholan-24-oic acid (**48**) were prepared according to the previously reported procedure.²⁵

23(S)-Methyl-3 α ,7 α ,12 α -trihydroxy-6 α -methyl-5 β -cholan-24-oic Acid (**44**). mp: 165–166 °C. Elemental Analysis: Calcd for $C_{26}H_{44}O_5$: C, 71.52; H, 10.16. Found: C, 71.85, H, 9.93. ¹H NMR ($CDCl_3 + CD_3OD$) δ : 0.66 (s, 3H, CH_3 -18), 0.87 (s, 3H, CH_3 -19), 0.97–1.00 (d, 3H, CH_3 -6), 1.14–1.18 (d, 3H, CH_3 -23), 3.39 (m, 1H, CH-3), 3.57 (brs, 1H, CH-7), 3.97 (brs, 1H, CH-12). ¹³C NMR ($CDCl_3 +$

CD₃OD) δ : 12.40, 15.59, 17.53, 18.94, 22.71, 23.12, 26.21, 27.43, 28.02, 29.99, 33.44, 34.11, 34.44, 35.27, 35.33, 36.89, 39.97, 40.60, 41.75, 46.44, 47.24, 47.53, 71.97, 72.69, 72.94, 180.03.

23(R)-Methyl-3 α ,7 α ,12 α -trihydroxy-6 α -methyl-5 β -cholan-24-oic Acid (45). mp: 97–98 °C. Elemental Analysis: Calcd for C₂₆H₄₄O₅: C, 71.52; H, 10.16. Found: C, 71.79, H, 9.90. ¹H NMR (CD₃OD) δ : 0.73 (s, 3H, CH₃-18), 0.91 (s, 3H, CH₃-19), 0.97–1.00 (m, 6H, CH₃-21, CH₃-6), 1.06–1.10 (d, 3H, CH₃-23), 2.45 (m, 1H, CH-23), 3.30 (m, 1H, CH-3), 3.53 (brs, 1H, CH-7), 3.97 (brs, 1H, CH-12). ¹³C NMR (CD₃OD) δ : 13.04, 16.48, 16.74, 17.57, 23.48, 24.20, 27.72, 29.07, 29.62, 30.77, 31.06, 34.36, 35.19, 35.76, 36.50 (\times 2), 36.63, 37.99, 41.15, 41.70, 43.12, 47.73, solvent (C \times 2), 73.19 (\times 2), 74.06, 181.52.

23(S)-Methyl-3 α -hydroxy-5 β -cholan-24-oic Acid (47). mp: 177–178 °C. Elemental Analysis: Calcd for C₂₅H₄₂O₃: C, 76.87; H, 10.84. Found: C, 76.49, H, 10.11. ¹H NMR (CD₃Cl₃) δ : 0.65 (3H, s, CH₃-18), 0.85–0.90 (6H, m, CH₃-21, CH₃-6), 2.55 (1H, m, CH-23), 3.60 (1H, m, CH-3), 4.2–4.7 (2H, br, COOH, 3-OH). ¹³C NMR (CD₃Cl₃) δ : 11.97, 18.52, 18.87, 20.73, 23.29, 24.13, 26.33, 21.10, 28.15, 30.18, 34.48 (\times 2), 35.23, 35.73, 36.05, 37.01, 40.12, 40.33, 40.73, 41.98, 42.68, 56.43, 56.74, 71.70, 181.41.

23(R)-Methyl-3 α -hydroxy-5 β -cholan-24-oic Acid (48). mp: 143–144 °C. Elemental Analysis: Calcd for C₂₅H₄₂O₃: C, 76.87; H, 10.84. Found: C, 76.52, H, 10.33. ¹H NMR (CD₃Cl₃) δ : 0.65 (3H, s, CH₃-18), 0.80–0.90 (6H, m, CH₃-21, CH₃-6), 2.45 (1H, m, CH-23), 2.7–3.0 (2H, br, COOH, 3-OH), 3.65 (1H, m, CH-3). ¹³C NMR (CD₃Cl₃) δ : 11.98, 15.96, 18.00, 20.75, 23.31, 24.14, 26.34, 27.11, 28.48, 30.25, 33.67, 34.50, 35.25, 35.77, 36.14, 36.45, 39.59, 40.13, 40.36, 42.01, 42.79, 56.44, 56.76, 71.71, 181.01.

Biology. Plasmids. The NIH Mammalian Gene Collection clone MGC:40597 (also named pCMVSPORT6/hTGR5 or pTGR5) and pcDNA3.1(+) were obtained from Invitrogen (Carlsbad, CA). pCRE-Luc and pCMV β were obtained from Clontech (Palo Alto, CA). pEcRE \times 7-Luc was a kind gift from Dr. Richard A. Heyman (X-ceptor Therapeutics, CA).

Cell Culture. Chinese hamster ovary (CHO) cells were obtained from American type Culture Collection (Manassas, VA). Cell culture medium, serum, and supplements were from Invitrogen or Sigma-Aldrich. All CHO cells were maintained in minimum essential medium α (α -MEM) supplemented with 10%(v/v) fetal bovine serum (FBS) and 100 μ M nonessential amino acids (NEAA). All cell culture medium was supplemented with 100 units/mL penicillin and 100 μ g/mL streptomycin sulfate. Cells were grown at 37 °C in an atmosphere of 5% CO₂, passed every 2–6 days, and freshly plated for each experiment.

Transient Transfections. CHO cells were plated in 96-well plates at a density of 3.5×10^4 cells/well, cultured for 24 h, and then transfected with 10 ng of human (h) TGR5 expression plasmid (pCMVSPORT6/hTGR5), 50 ng of cAMP-responsive element (CRE)-driven luciferase reporter plasmid (pCRE-Luc), and pCMV β -galactosidase as internal control in each well using JetPEI (Polyplus) according to the manufacturer's instructions. 24 h after the transfection, medium was replaced with DMEM containing 0.1% (w/v) BSA for 3 h and then the cells were treated for 5 h with different concentrations of each compound in fresh DMEM containing 0.1% (w/v) BSA. After treatment, the cells were

lysed with 100 μ L of lysis buffer (25 mM Tris-Cl (pH7.6), 2 mM EDTA, 1 mM dithiothreitol (DTT), 10% (v/v) glycerol, and 1% (v/v) triton X-100) by a freeze–thaw cycle and subjected to luciferase assays as described below. Normalized luciferase values were determined by dividing the luciferase activity by the β -galactosidase activity.

Luciferase and β -Galactosidase Assays. For luciferase assays, 20 μ L of cell lysate was mixed with 100 μ L of luciferase reaction buffer [235 μ M luciferine, 265 μ M ATP, and 135 μ M coenzyme A (CoA)], and luminescence was determined with CentroXS3 LB960 (Berthold Technologies, Bad Wildbad, Germany). For β -galactosidase assays, 10 μ L of cell lysate was mixed with 100 μ L of Buffer Z [60 mM Na₂HPO₄, 10 mM KCl, 1 mM MgSO₄, 50 mM β -mercaptoethanol, and 0.75 mg/mL o-nitrophenyl- β -D-galactopyranoside (ONPG)] and incubated at 37 °C for 0.5–3 h. Reactions were stopped by adding 50 μ L of Stop buffer (1 M Na₂CO₃), and the optical density at 420 nm was determined.

50% Effective Concentrations (EC₅₀) and Efficacy Determination. Assays were performed in triplicate or quadruplicate for each condition. EC₅₀ values were determined by probit analysis. Efficacy was determined by calculating percentages of 10 μ M LCA value for TGR5 agonism. After subtracting the average value of the basal (vehicle-treated) condition, values were applied to EC₅₀ and/or efficacy determinations. Calculation of average EC₅₀ and comparison of the EC₅₀ between different compounds were done after transformation to logarithms.

Statistical Analysis. Statistical analysis was performed by Student's *t* test, and *p* < 0.05 was considered statistically significant.

Molecular Modeling. All molecules were drawn with the sketch module of Cerius-2³⁴ and optimized using Universal force-field v.1.2³⁵ and the Smart Minimizer protocol of the Open Force Field module (OFF). The acidic groups were considered in their protonated form, and, as consequence, all molecules were endowed with a formal charge of 0. We deemed it appropriate to adopt this approximation in the MFA study, in order to avoid any overestimation of the electrostatic interaction points with respect to the steric and hydrogen bond interaction points. Both 7 α - and 7 β -isomers of compound **17** (7 ξ -Me-LCA) were considered in the study. Atomic charges were calculated using the Gasteiger method,³⁶ as implemented in the polygraph version of Cerius-2.

The conformational analysis of the active C₂₂–C₂₃ cyclopropyl isomers of CDCA (**31**, **32**) was carried out using a systematic search procedure on the α and β torsional angles of the BA side chain. Briefly a window of 360° was explored for each dihedral angle using an incremental torsion of 10°. The resulting conformations were energetically minimized using the Universal force-field v.1.2 and the Smart Minimizer protocol of Cerius-2. After the removal of duplicate conformations, the global and local minima conformers of compounds **31** and **32** were mutually aligned using the atomic coordinates. The resulting rmsd values were used to identify the best fitting pair of conformations thus representing the proposed bioactive conformations of 22S,23S-CCDCA (**31**) and 22S,23R-CCDCA (**32**). Then, all the compounds of the collection were mutually aligned on the bioactive conformation of **31** using the subgraph matching routine of the Align

module of Cerius-2 which fitted all the conserved heavy atoms between the template (31) and the query compound.

The resulting alignment was manually refined to maximize the overlapping volume of the compounds. The aligned molecules were inserted in a cubic box endowed with the following dimensions: $X_{\min} = -12$, $X_{\max} = 36$; $Y_{\min} = 8$, $Y_{\max} = 30$; $Z_{\min} = 42$, $Z_{\max} = 72$. The step size of the interaction points was set to 2 Å. The total number of the interaction points was 4800. For each grid point, the interaction between three probe atoms and each compound was calculated yielding three molecular interaction fields. The probes were the electrostatic probe (hydrogen atom, net charge = +1), the steric probe (methyl group, net charge = 0), and the hydrogen bond probe (hydroxyl group, net charge = 0) as defined in the fragment library of Cerius-2.

The multiple linear regression analysis was performed using the XLSTAT software. The selection of independent variables that best correlated with the activity was carried out using a stepwise procedure. This process starts by incrementally adding the independent variables with the largest contribution to the model on the basis of the "Student's t" statistic and a cutoff of probability for entry criterion ($p_{\text{in}} = 0.05$). After the addition of the first three variables to the model, the impact of removing each new variable added to the model was evaluated using the "Student's t" statistic and a cutoff of probability of removal criterion ($p_{\text{out}} = 0.1$). The cross-validation protocol was used to determine the statistical significance of the equations. The predictive index of the generated models was given by the cross-validated R_{XV}^2 (q^2).

Abbreviations: BA, bile acid; CA, cholic acid; cAMP, cyclic adenosine monophosphate; CDCA, chenodeoxycholic acid; FXR, farnesoid X receptor; GPCR, G protein-coupled receptor.

ACKNOWLEDGMENT

This work was supported by Intercept Pharmaceuticals (New York, U.S.A.), CNRS, INSERM, Hôpitaux Universitaires de Strasbourg, and the European Union.

REFERENCES AND NOTES

- Houten, S. M.; Watanabe, M.; Auwerx, J. Endocrine functions of bile acids. *EMBO J.* **2006**, *25*, 1419–1425.
- Wang, H. B.; Chen, J.; Hollister, K.; Sowers, L. C.; Forman, B. M. Endogenous bile acids are ligands for the nuclear receptor FXR BAR. *Mol. Cell* **1999**, *3*, 543–553.
- Parks, 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. Bile acids: Natural ligands for an orphan nuclear receptor. *Science* **1999**, *284*, 1365–1368.
- 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. Identification of a nuclear receptor for bile acids. *Science* **1999**, *284*, 1362–1365.
- Maruyama, T.; Miyamoto, Y.; Nakamura, T.; Tamai, Y.; Okada, H.; Sugiyama, E.; Nakamura, T.; Itadani, H.; Tanaka, K. Identification of membrane-type receptor for bile acids (M-BAR). *Biochem. Biophys. Res. Commun.* **2002**, *298*, 714–719.
- Kawamata, Y.; Fujii, R.; Hosoya, M.; Harada, M.; Yoshida, H.; Miwa, M.; Fukusumi, S.; Habata, Y.; Itoh, T.; Shintani, Y.; Hinuma, S.; Fujisawa, Y.; Fujino, M. A G protein-coupled receptor responsive to bile acids. *J. Biol. Chem.* **2003**, *278*, 9435–9440.
- Acconcia, F.; Kumar, R. Signaling regulation of genomic and nongenomic functions of estrogen receptors. *Cancer Lett.* **2006**, *238*, 1–14.
- Losel, R.; Wehling, M. Nongenomic actions of steroid hormones. *Nat. Rev. Mol. Cell. Biol.* **2003**, *4*, 46–56.
- Vasudevan, N.; Pfaff, D. W. Membrane-initiated actions of estrogens in neuroendocrinology: Emerging principles. *Endocr. Rev.* **2007**, *28*, 1–19.
- Watanabe, M.; Houten, S. M.; Matak, C.; Christoffoleto, M. A.; Kim, B. W.; Sato, H.; Messaddeq, N.; Harney, J. W.; Ezaki, O.; Kodama, T.; Schoonjans, K.; Bianco, A. C.; Auwerx, J. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **2006**, *439*, 484–489.
- Maruyama, T.; Tanaka, K.; Suzuki, J.; Miyoshi, H.; Harada, N.; Nakamura, T.; Miyamoto, Y.; Kanatani, A.; Tamai, Y. Targeted disruption of G protein-coupled bile acid receptor 1 (Gpbar1/M-Bar) in mice. *J. Endocrinol.* **2006**, *191*, 197–205.
- Katsuma, S.; Hirasawa, A.; Tsujimoto, G. Bile acids promote glucagon-like peptide-1 secretion through TGR5 in a murine enteroendocrine cell line STC-1. *Biochem. Biophys. Res. Commun.* **2005**, *329*, 386–390.
- Francis, G. A.; Fayard, E.; Picard, F.; Auwerx, J. Nuclear receptors and the control of metabolism. *Annu. Rev. Physiol.* **2003**, *65*, 261–311.
- Modica, S.; Moschetta, A. Nuclear bile acid receptor FXR as pharmacological target: Are we there yet. *FEBS Lett.* **2006**, *580*, 5492–5499.
- Fiorucci, S.; Rizzo, G.; Donini, A.; Distrutti, E.; Santucci, L. Targeting farnesoid X receptor for liver and metabolic disorders. *Trends Mol. Med.* **2007**, *13*, 298–309.
- Thomas, C.; Auwerx, J.; Schoonjans, K. Bile Acids and the Membrane Bile Acid Receptor TGR5-Connecting Nutrition and Metabolism. *Thyroid* **2008**, *18*, 167–174.
- Pellicciari, R.; Costantino, G.; Fiorucci, S. Farnesoid X receptor: from structure to potential clinical applications. *J. Med. Chem.* **2005**, *48*, 5383–5403.
- Pellicciari, R.; Fiorucci, S.; Camaioni, E.; Clerici, C.; Costantino, G.; Maloney, P. R.; Morelli, A.; Parks, D. J.; Willson, T. M. 6 α -ethyl-chenodeoxycholic acid (6-ECDCA), a potent and selective FXR agonist endowed with anticholestatic activity. *J. Med. Chem.* **2002**, *45*, 3569–3572.
- Costantino, G.; Entrena-Guadix, A.; Macchiarulo, A.; Gioiello, A.; Pellicciari, R. Molecular dynamics simulation of the ligand binding domain of farnesoid X receptor. Insights into helix-12 stability and coactivator peptide stabilization in response to agonist binding. *J. Med. Chem.* **2005**, *48*, 3251–3259.
- Downes, M.; Verdecia, M. A.; Roecker, A. J.; Hughes, R.; Hogenesch, 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. A chemical, genetic, and structural analysis of the nuclear bile acid receptor FXR. *Mol. Cell* **2003**, *11*, 1079–1092.
- Mi, L. Z.; Devarakonda, S.; Harp, J. M.; Han, Q.; Pellicciari, R.; Willson, T. M.; Khorasanizadeh, S.; Rastinejad, F. Structural basis for bile acid binding and activation of the nuclear receptor FXR. *Mol. Cell* **2003**, *11*, 1093–1100.
- Costantino, G.; Macchiarulo, A.; Entrena-Guadix, A.; Camaioni, E.; Pellicciari, R. Binding mode of 6ECDCA, a potent bile acid agonist of the farnesoid X receptor (FXR). *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1865–1868.
- Pellicciari, R.; Costantino, G.; Camaioni, E.; Sadeghpour, B. M.; Entrena, A.; Willson, T. M.; Fiorucci, S.; Clerici, C.; Gioiello, A. Bile acid derivatives as ligands of the farnesoid X receptor. Synthesis, evaluation, and structure-activity relationship of a series of body and side chain modified analogues of chenodeoxycholic acid. *J. Med. Chem.* **2004**, *47*, 4559–4569.
- Zhang, T.; Zhou, J. H.; Shi, L. W.; Zhu, R. X.; Chen, M. B. 3D-QSAR studies with the aid of molecular docking for a series of non-steroidal FXR agonists. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2156–2160.
- Pellicciari, R.; Sato, H.; Gioiello, A.; Costantino, G.; Macchiarulo, A.; Sadeghpour, B. M.; Giorgi, G.; Schoonjans, K.; Auwerx, J. Nongenomic actions of bile acids. Synthesis and preliminary characterization of 23- and 6,23-alkyl-substituted bile acid derivatives as selective modulators for the g-protein coupled receptor TGR5. *J. Med. Chem.* **2007**, *50*, 4265–4268.
- Sato, H.; Genet, C.; Strehle, A.; Thomas, C.; Lobstein, A.; Wagner, A.; Mioskowski, C.; Auwerx, J.; Saladin, R. Anti-hyperglycemic activity of a TGR5 agonist isolated from *Olea europaea*. *Biochem. Biophys. Res. Commun.* **2007**, *362*, 793–798.
- Ito, F.; Hinuma, K.; Kanzaki, N.; Miki, T.; Kawamata, Y.; Oi, S.; Tawaeaiishi, T.; Ishichi, Y.; Hirohashi, M. Preparation of aromatic ring-fused cyclic compounds as TGR5 receptor agonists. PN: WO2004067008, 2004.
- Katona, B. W.; Cummins, C. L.; Ferguson, A. D.; Li, T.; Schmidt, D. R.; Mangelsdorf, D. J.; Covey, D. F. Synthesis, Characterization, and Receptor Interaction Profiles of Enantiomeric Bile Acids. *J. Med. Chem.* **2007**, *50*, 6048–6058.

- (29) Sato, H.; Macchiarulo, A.; Thomas, C.; Gioiello, A.; Une, M.; Hofmann, A. F.; Saladin, R.; Schoonjans, K.; Pellicciari, R.; Auwerx, J. Novel Potent and Selective Bile Acid Derivatives as TGR5 Agonists: Biological Screening, Structure-Activity Relationships, and Molecular Modeling Studies. *J. Med. Chem.* **2008**, *51*, 1831–1841.
- (30) Topliss, J. G.; Edwards, R. P. Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.* **1979**, *22*, 1238–1244.
- (31) Johnson, S. R. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* **2008**, *48*, 25–26.
- (32) Maggiora, G. M. On outliers and activity cliffs-why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535.
- (33) Clerici, C.; Castellani, D.; Asciutti, S.; Pellicciari, R.; Setchell, K. D.; O'Connell, N. C.; Sadeghpour, B.; Camaioni, E.; Fiorucci, S.; Renga, B.; Nardi, E.; Sabatino, G.; Clementi, M.; Giuliano, V.; Baldoni, M.; Orlandi, S.; Mazzocchi, A.; Morelli, A.; Morelli, O. 3alpha-6alpha-Dihydroxy-7alpha-fluoro-5beta-cholanoate (UPF-680), physicochemical and physiological properties of a new fluorinated bile acid that prevents 17alpha-ethynyl-estradiol-induced cholestasis in rats. *Toxicol. Appl. Pharmacol.* **2006**, *214*, 199–208.
- (34) *Accelrys, Cerius-2, version 4.6*; Accelrys: San Diego, CA, 2001.
- (35) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF. a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (36) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.

CI800196H