

# Novel Potent and Selective Bile Acid Derivatives as TGR5 Agonists: Biological Screening, Structure–Activity Relationships, and Molecular Modeling Studies

Hiroyuki Sato,<sup>†,×</sup> Antonio Macchiarulo,<sup>‡,×</sup> Charles Thomas,<sup>†</sup> Antimo Gioiello,<sup>‡</sup> Mizuho Une,<sup>§</sup> Alan F. Hofmann,<sup>#</sup> Régis Saladin,<sup>△</sup> Kristina Schoonjans,<sup>†</sup> Roberto Pellicciari,<sup>\*,‡</sup> and Johan Auwerx<sup>\*,†,△,▽</sup>

*Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS/INSERM/ULP, 67404 Illkirch, France, Hôpitaux Universitaires de Strasbourg, Laboratoire de Biochimie Générale et Spécialisée, 67000 Strasbourg, France, Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, 06123 Perugia, Italy, Laboratory of Organic and Biomolecular Chemistry, Faculty of Pharmaceutical Sciences, Hiroshima International University, Kure, Hiroshima 737-0112, Japan, Department of Medicine, Division of Gastroenterology, University of California, San Diego, California 92093-0063, PhytoDia, 67405 Illkirch, France, and Institut Clinique de la Souris, 67404 Illkirch, France*

Received December 18, 2007

TGR5, a metabotropic receptor that is G-protein-coupled to the induction of adenylate cyclase, has been recognized as the molecular link connecting bile acids to the control of energy and glucose homeostasis. With the aim of disclosing novel selective modulators of this receptor and at the same time clarifying the molecular basis of TGR5 activation, we report herein the biological screening of a collection of natural occurring bile acids, bile acid derivatives, and some steroid hormones, which has resulted in the discovery of new potent and selective TGR5 ligands. Biological results of the tested collection of compounds were used to extend the structure–activity relationships of TGR5 agonists and to develop a binary classification model of TGR5 activity. This model in particular could unveil some hidden properties shared by the molecular shape of bile acids and steroid hormones that are relevant to TGR5 activation and may hence be used to address the design of novel selective and potent TGR5 agonists.

## Introduction

In light of their amphipathic properties, bile acids (BAs<sup>a</sup>) have for a long time been viewed as detergents able to solubilize cholesterol, fatty acids, and liposoluble vitamins, thus facilitating their digestion and transport and endowed with limited therapeutic applications.<sup>1,2</sup> A number of startling discoveries have shifted the above definition to the currently accepted notion of BAs as signaling hormones endowed with a wide array of endocrine functions.<sup>3</sup> Milestones of this more appealing redefinition of BA function are the discoveries in 1999 of the farnesoid X receptor (FXR) as the endogenously specific nuclear receptor of BAs<sup>4–6</sup> and in 2002 of TGR5 (GPBAR1 or M-BAR) as the endogenously specific metabotropic receptor of BAs.<sup>7,8</sup> Interestingly, the presence of both nuclear and membrane receptors for the same ligand is emerging as a common paradigm that helps to explain the pleiotropic activities of hormonal signaling pathways.<sup>9–11</sup> In our case, accordingly, the modulation of the intracellular FXR receptor by BAs constitutes a genomic pathway to maintain cholesterol, BA, lipid, and carbohydrate

homeostasis,<sup>12–14</sup> while the activation of the membrane TGR5 receptor may be the gateway to the nongenomic functions of BAs.

TGR5 is a G-protein-coupled receptor associated with the intracellular accumulation of cAMP, which is widely expressed in diverse cell types. While the activation of this membrane receptor in macrophages decreases proinflammatory cytokine production,<sup>8</sup> the stimulation of TGR5 by BAs in adipocytes and myocytes enhances energy expenditure.<sup>15</sup> This latter effect involves the cAMP-dependent induction of type 2 iodothyronine deiodinase (D2), which by local conversion of T4 into T3 gives rise to increased thyroid hormone activity. Consistent with the role of TGR5 in the control of energy metabolism, female TGR5 knockout mice show a significant fat accumulation with body weight gain when challenged with a high-fat diet, indicating that the lack of TGR5 decreases energy expenditure and elicits obesity.<sup>16</sup> In addition and in line with the involvement of TGR5 in energy homeostasis, bile acid activation of the membrane receptor has also been reported to promote the production of glucagon-like peptide 1 (GLP-1) in murine enteroendocrine cell lines.<sup>17</sup> On the basis of all the above observations, TGR5 is emerging as an attractive target for the treatment of obesity, diabetes, and the metabolic syndrome.

Few examples of TGR5 agonists have been so far described in literature. We have recently reported that 23-alkyl-substituted and 6,23-alkyl-disubstituted derivatives of chenodeoxycholic acid, such as the 6 $\alpha$ -ethyl-23(*S*)-methylchenodeoxycholic acid (**1**), are potent and selective agonists of TGR5 (Figure 1).<sup>18</sup> In particular, we have shown that methylation at the C<sub>23</sub>-(*S*) position of natural BAs confers a marked selectivity to TGR5 over FXR activation, whereas the 6 $\alpha$ -alkyl substitution increases the potency at both receptors. The screening of libraries of nonsteroidal compounds and natural products has led to the disclosure of 6-methyl-2-oxo-4-thiophen-2-yl-1,2,3,4-tetrahydropyrimidine-5-carboxylic acid benzyl ester (**2**) (WO2004067008,

\* To whom correspondence should be addressed. For R.P.: address, Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, Via del Liceo 1, 06123 Perugia, Italy; phone, +39 075 5855120; fax, +39 075 5855124; e-mail, rp@unipg.it. For J.A.: Johan Auwerx, M.D., PhD, IGBMC, 1 Rue Laurent Fries, BP 10142, 67404 Illkirch, France; phone, +33 388.653425; fax, +33 388.653299; e-mail, auwerx@igbmc.u-strasbg.fr.

<sup>†</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC).

<sup>▽</sup> Hôpitaux Universitaires de Strasbourg.

<sup>×</sup> These authors equally contributed to the work.

<sup>‡</sup> Università di Perugia.

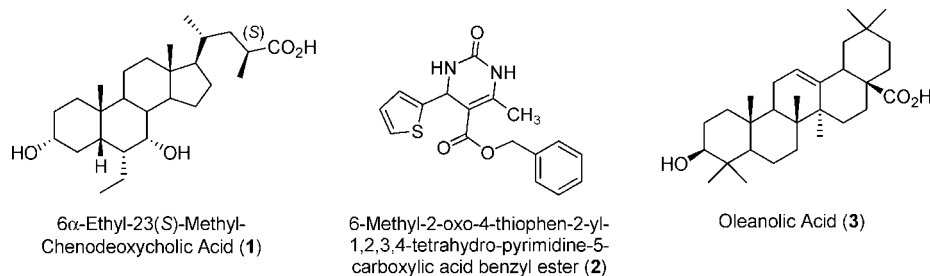
<sup>§</sup> Hiroshima International University.

<sup>#</sup> University of California, San Diego.

<sup>△</sup> PhytoDia.

<sup>△</sup> Institut Clinique de la Souris.

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

**Figure 1.** Natural and synthetic currently available TGR5 agonists.**Table 1.** TGR5 Agonist Potency of Natural Bile Acids and Their Tauro- or Glyco-Conjugated Forms<sup>a</sup>

compd	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	acid form		tauro form		glyco form	
				EC <sub>50</sub>	Efficacy	EC <sub>50</sub>	efficacy	EC <sub>50</sub>	efficacy
CDCA ( <b>4</b> )	—H	$\alpha$ -OH	—H	6.71	105	1.92	103	3.88	105
CA ( <b>5</b> )	—H	$\alpha$ -OH	$\alpha$ -OH	13.6	101	4.95	104	13.6	103
LCA ( <b>6</b> )	—H	—H	—H	0.58	101	0.29	106	0.54	92
DCA ( <b>7</b> )	—H	—H	$\alpha$ -OH	1.25	105	0.79	103	1.18	105
lago-DCA ( <b>8</b> )	—H	—H	$\beta$ -OH	4.39	114				
UDCA ( <b>9</b> )	—H	$\beta$ -OH	—H	36.4	74.9	30.0	97	33.9	91
HDCA ( <b>10</b> )	$\alpha$ -OH	—H	—H	31.6	79.7	24.2	107	36.7	101
muro-CA ( <b>11</b> )	$\beta$ -OH	—H	—H	4.89	94.4				

<sup>a</sup> Data represent average values of at least three independent experiments of CRE-driven luciferase reporter assays in TGR5-transfected CHO cells. Units are  $\mu$ M for EC<sub>50</sub> and % of 10  $\mu$ M LCA value for efficacy.

Takeda Chemical Industries LTD, Japan, 2004)<sup>24</sup> and oleanolic acid (**3**) as structurally diverse TGR5 agonists (Figure 1).<sup>19,20</sup> More recently, the first synthesis of enantiomeric chenodeoxycholic acid (CDCA) and lithocholic acid (LCA) has allowed assessment of the specificity of the interaction of natural BAs to TGR5.<sup>21</sup>

While these chemical tools have provided for the first time a pharmacological differentiation of genomic versus nongenomic effects of BAs, some of them also allowed us to draw a first structure–activity relationship study where the presence of an accessory binding pocket in TGR5 plays a pivotal role in determining ligand selectivity.<sup>18</sup> In this context, the availability of more potent and selective TGR5 modulators may be instrumental to further identify additional features affecting receptor activation and to characterize the physiological and pharmacological actions of this receptor.

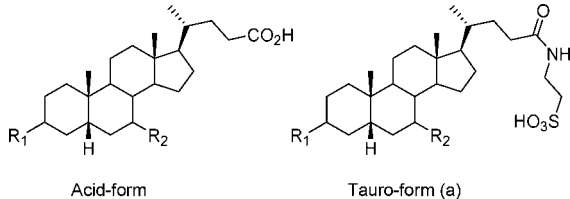
With this aim in mind we have evaluated for TGR5 activity a library of steroidal compounds consisting of natural occurring BAs (**4–11**), semisynthetic BA derivatives (**12–41**), and a number steroid hormones (**42–55**) and utilized the results thus obtained to define a SAR profile of the TGR5 agonists and to develop for them a binary classification model useful for unveiling some hidden properties of the molecular shape of bile acids and steroid hormones relevant to TGR5 activation. The results of these studies are reported herein.

## Results

Sixty-eight compounds comprising (i) natural occurring BAs (Table 1), (ii) semisynthetic BA derivatives (Tables 2 and 3), and (iii) some steroid hormones (Table 4) were evaluated for TGR5 activity. To this end, luciferase activity was determined

in CHO cells stably expressing hTGR5 or transiently cotransfected with a hTGR5 expression vector and a cAMP-responsive element (CRE)-driven luciferase reporter gene. In addition, cAMP formation was tested in two different cellular models. Finally, since BAs are well-recognized ligands for FXR, we also tested a selection of the most potent TGR5 agonists in FXR luciferase reporter assays.

**Luciferase Assays. (i) Natural Occurring Bile Acids.** Natural BAs differ from each other by distinct hydroxylation patterns of positions 6, 7, and 12 and by the presence (conjugated) or absence (unconjugated) of an amidic bond linking their acidic side chain with a taurine or glycine moiety. This particular set of compounds comprises eight endogenous BAs (**4–11**) and the respective tauro- (**4a–11a**) and glyco- (**4b–11b**) conjugated forms. The inspection of the biological results (Table 1) indicates that the unconjugated BAs are endowed with the following rank of potency: LCA (**6**) > deoxycholic acid (DCA, **7**) > murocholic acid (Muro-CA, **11**) = lagodeoxycholic acid (lago-DCA) (**8**) > CDCA (**4**) > cholic acid (CA, **5**) > hyodeoxycholic acid (HDCA, **10**) > ursodeoxycholic acid (UDCA, **9**). These results indicate that the absence of hydroxylation in positions 6, 7, and 12 of the steroid nucleus (LCA, **6**) favors the potency of the BA to TGR5. Conversely, either the presence of one hydroxyl group in any of these positions (DCA, **7**; CDCA, **4**; HDCA, **10**) or the presence of two hydroxyl groups (CA, **5**) in positions 7 and 12 decreases the EC<sub>50</sub> of the parent BA. As a general trend, while the conjugation of the BA with the taurine moiety slightly increases the agonist potency to TGR5, the conjugation with the glycine moiety has only a negligible effect on TGR5 activity (Table 1). Thus, on the basis of these biological results, tauro-

**Table 2.** TGR5 Agonist Potency of Body-Modified Bile Acids Derivatives and Their Tauro-Conjugated forms<sup>a</sup>


The image shows two chemical structures. The left structure is labeled 'Acid-form' and shows a steroid nucleus with substituents R<sub>1</sub> and R<sub>2</sub>, and a side chain ending in a carboxylic acid group (-CO<sub>2</sub>H). The right structure is labeled 'Tauro-form (a)' and shows the same steroid nucleus with a side chain ending in a tauro-conjugated group (-NH-C(=O)-CH<sub>2</sub>-CH<sub>2</sub>-SO<sub>3</sub>H).

compd	R <sub>1</sub>	R <sub>2</sub>	acid- form		tauro form	
			EC <sub>50</sub>	efficacy	EC <sub>50</sub>	efficacy
LCA ( <b>6</b> )	α-OH	-H	0.58	101	0.29	106
5β-cholanic acid ( <b>12</b> )	-H	-H	6.08	76		
LCA-S ( <b>13</b> )	α-OSO <sub>3</sub> H	-H	>100	0 (100μM)	47.8	92
LCA-Ac ( <b>14</b> )	α-OCOCH <sub>3</sub>	-H	>102	0 (50μM)		
dehydro-LCA ( <b>15</b> )	=O	-H	0.27	106		
iso-LCA ( <b>16</b> )	β-OH	-H	1.25	99		
7ξ-Me-LCA ( <b>17</b> )	α-OH	-CH <sub>3</sub>	0.076	106		
7α-F-LCA ( <b>18</b> )	α-OH	α-F	0.25	99		
7β-F-LCA ( <b>19</b> )	α-OH	β-F	2.29	92		
CDCA ( <b>4</b> )	α-OH	α-OH	6.71	105	1.92	103
3-dehydro-CDCA ( <b>20</b> )	=O	α-OH	3.98	107		
3-deoxy-CDCA ( <b>21</b> )	-H	α-OH	14.5	81		
7β-Me-CDCA ( <b>22</b> )	-H	β-CH <sub>3</sub>	6.18	105	2.36	112
7β-Et-CDCA ( <b>23</b> )	-H	β-CH <sub>2</sub> CH <sub>3</sub>	2.63	99	0.73	99
7β-Pr-CDCA ( <b>24</b> )	-H	β-CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	0.78	108	0.36	104

<sup>a</sup> 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 EC<sub>50</sub> and % of 10 μM LCA value for efficacy.

LCA (**6a**) is the most active compound in the set of endogenous BAs (EC<sub>50</sub> = 0.285 μM, efficacy = 106%).

**(ii) Semisynthetic Bile Acid Derivatives.** The semisynthetic BA derivatives can be classified into body-modified (subset A, **12–24**, Table 2) and side chain modified BA derivatives (subset B, **25–41**, Table 3).

The analysis of the biological results for the compounds in subset A indicates that the C<sub>3</sub> position of the BA nucleus is important for TGR5 activity. Indeed, while the oxidation of the 3α-hydroxyl group leads to the enhancement of the activity (dehydro-LCA, **15**; 3-dehydro-CDCA, **20**), its elimination (**12**, **21**) or isomerization (iso-LCA, **16**) decreases the potency with respect to the parent natural BA (Table 2). Moreover, the substitution of the 3α-hydroxyl group either with an acetoxy group (LCA-Ac, **14**) or with a sulfate moiety (LCA-S, **13**; tauro-LCA-S, **13a**) is detrimental for the EC<sub>50</sub> and the efficacy of the compound. As shown in Table 2, the introduction of an alkyl group in C<sub>7</sub>-β position of either CDCA or its tauro-conjugate improves TGR5 activity in direct proportion to the size of the substituent used, with the propyl group (**24**, **24a**) being more active than the ethyl (**23**, **23a**) and methyl (**22**, **22a**) moieties. Interestingly, the 7-methyl derivative of LCA (**17**) turned out to be a very potent TGR5 agonist. The insertion of a fluorine atom in the C<sub>7</sub>-α position of LCA has a similar effect. Thus, the 7α-fluoro-LCA (**18**) is twice more potent than LCA (**6**), while the 7β-epimer (**19**) is about 4 times less potent than LCA (**6**).

For the side chain modified BA derivatives (subset B, Table 3), the appraisal of their biological activity indicates that the progressive shortening of the acidic side chain (**26**, **28**, **29**, **38**, **39**) reduces TGR5 activity. Of particular interest is the analysis of the effect of the replacement of the C<sub>24</sub>-carboxylic moiety with diverse polar groups. While the methyl ester derivatives (**25**, **27**) keep both the potency and the efficacy of the corresponding BA, the C<sub>24</sub>-sulfate (**31**, **37**, **41**) and C<sub>24</sub>-hydroxyl derivatives (**30**, **36**, **40**) were about 1-fold of magnitude more potent. The screening of the four C<sub>22</sub>–C<sub>23</sub> cyclopropyl isomers of CDCA (**32–35**) has afforded very useful information on the

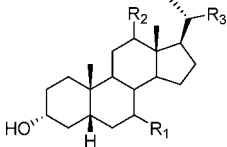
bioactive conformation adopted by the χ-torsional angle of BA side chain into the binding site of TGR5. In particular, it is the folded conformation of this dihedral angle in 22*S*,23*S*-CCDCA (**32**), which is slightly preferred over the extended conformation observed in **33**.

**(iii) Steroid Hormones.** Fourteen steroid hormones (**42–55**) have been examined with the aim to investigate whether other signaling molecules could modulate TGR5 activity.

When compared to BAs (**4–11**), almost the entire set of steroid hormones shows comparable activities to TGR5 but with a lower efficacy. The exceptions to this rule are dehydroepiandrosterone 3-sulfate (DHEA-S, **50**), testosterone (**54**), and 17β-estradiol (**55**), which are poorly active TGR5 compounds. Interestingly, pregnandiol (**42**), etiocholanolone (**45**), and etiocholandiol (**51**) can be envisaged as side chain truncated derivatives of LCA (**6**). In line with what we have reported for nor-LCA (**26**), these steroidal compounds display a lower agonist potency than LCA (**6**), with the EC<sub>50</sub> values being 0.85, 3.76, and 2.44 μM, respectively.

A careful comparison of the activities of etiocholanolone (**45**), epietiocholanolone (**46**), androsterone (**47**), epiandrosterone (**48**), etiocholandiol (**51**), and androstadiol (**52**) reveals that while the EC<sub>50</sub> of the steroid hormones is slightly affected by the configurations of the hydroxyl group at C<sub>3</sub> position and of the A/B ring junction, the TGR5 efficacy is favored with the hydroxyl group in the 3α-position and the A/B ring junction in the cis configuration. As noticed for LCA (**6**), oxidation of the 3α-hydroxyl group enhanced the activity on TGR5. Moreover, as a general trend, the presence of double bonds in ring A (**44**, **54**) or ring B (**49**, **50**) of the steroid nucleus leads to a reduction of TGR5 activity.

**cAMP Assays.** In order to further confirm the capacity of our library of compounds to modulate intracellular signaling, as a predictive parameter of biological activity, a selection of the most potent TGR5 agonists has been evaluated in two different in vitro cell models of cAMP formation (Figure 2). We first analyzed cAMP production in CHO cells overexpressing TGR5. All the tested BA derivatives are able to induce

**Table 3.** TGR5 Agonist Potency of Side Chain Modified Bile Acids Derivatives<sup>a</sup>


Trivial Name	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	EC <sub>50</sub>	Efficacy
LCA ( <b>6</b> )	-H	-H	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	0.58	101
LCA Me Ester ( <b>25</b> )	-H	-H	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> CH <sub>3</sub>	0.59	102
Nor-LCA ( <b>26</b> )	-H	-H	CH <sub>2</sub> CO <sub>2</sub> H	0.77	102
CDCA ( <b>4</b> )	α-OH	-H	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	6.71	105
CDCA Me Ester ( <b>27</b> )	α-OH	-H	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> CH <sub>3</sub>	4.39	114
Nor-CDCA ( <b>28</b> )	α-OH	-H	CH <sub>2</sub> CO <sub>2</sub> H	10.4	102
Dinor-CDCA ( <b>29</b> )	α-OH	-H	CHCO <sub>2</sub> H	>100	50 (316 μM)
CDC-OH ( <b>30</b> )	α-OH	-H	CH <sub>2</sub> CH <sub>2</sub> OH	0.12	103
CDC-Sul ( <b>31</b> )	α-OH	-H	CH <sub>2</sub> CH <sub>2</sub> SO <sub>3</sub> H	0.44	103
22 <i>S</i> ,23 <i>S</i> -CCDCA ( <b>32</b> )	α-OH	-H	CH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> H	1.33	110
22 <i>S</i> ,23 <i>R</i> -CCDCA ( <b>33</b> )	α-OH	-H	CH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> H	2.91	102
22 <i>R</i> ,23 <i>R</i> -CCDCA ( <b>34</b> )	α-OH	-H	CH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> H	75.7	5
22 <i>R</i> ,23 <i>S</i> -CCDCA ( <b>35</b> )	α-OH	-H	CH(CH <sub>3</sub> )CH <sub>2</sub> CO <sub>2</sub> H	>100	4
CA ( <b>5</b> )	α-OH	α-OH	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	13.6	101
C-OH ( <b>36</b> )	α-OH	α-OH	CH <sub>2</sub> CH <sub>2</sub> OH	0.87	103
C-Sul ( <b>37</b> )	α-OH	α-OH	CH <sub>2</sub> CH <sub>2</sub> SO <sub>3</sub> H	1.00	103
UDCA ( <b>9</b> )	α-OH	β-OH	CH <sub>2</sub> CH <sub>2</sub> CO <sub>2</sub> H	36.4	75
Nor-UDCA ( <b>38</b> )	α-OH	β-OH	CH <sub>2</sub> CO <sub>2</sub> H	47.2	79
Dinor-UDCA ( <b>39</b> )	α-OH	β-OH	CHCO <sub>2</sub> H	>316	0 (316 μM)
UDC-OH ( <b>40</b> )	α-OH	β-OH	CH <sub>2</sub> CH <sub>2</sub> OH	2.18	79
UDC-Sul ( <b>41</b> )	α-OH	β-OH	CH <sub>2</sub> CH <sub>2</sub> SO <sub>3</sub> H	5.02	109

<sup>a</sup> 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 EC<sub>50</sub> and % of 10 μM LCA value for efficacy.

cAMP formation showing a good correlation with their TGR5 potency in luciferase assays (Figure 2a). Then the ability of the above set of BAs to increase cAMP levels has been assessed using the NCI-H716 intestinal enteroendocrine cell line that endogenously expresses high levels of TGR5.<sup>7</sup> Similar to what we observed in CHO cells that overexpress TGR5, BA derivatives also promote cAMP production in NCI-H716 cells (Figure 2b), with 7ξ-methyl-LCA (**17**) being the most active compound. Again, the rank order of potency of BA derivatives for the induction of cAMP in NCI-H716 cells is in good correlation with the TGR5 activities reported for this set of compounds in the luciferase assays. These results further support the hypothesis that the effects of bile acids on cAMP production in intestinal enteroendocrine cells are mainly TGR5-dependent.<sup>17</sup>

**FXR Activation Assays.** A subset of BA derivatives comprising the most active TGR5 compounds was further submitted to a luciferase reporter assay to score for their capacity to activate the nuclear BA receptor FXR (Table 5). Albeit different levels of TGR5 and FXR gene expression may affect the observed relative activities in the luciferase assays, all the

tested BAs display a selective agonistic activity to TGR5 over FXR receptor.

In particular, LCA derivatives (**6**, **17**, and **18**) are potent and selective TGR5 agonists with selectivity ratios ranging from 0.029 for LCA (**6**) to 0.0025 in the case of 7α-fluorine-LCA (**18**). The reduction of the BA carboxylic moiety to the hydroxyl group (**30**, **40**) and its replacement with an ethanesulfonic acid moiety (**31**, **41**) improve the selectivity to TGR5 by at least 1 order of magnitude.

**Linear Discriminant Analysis (LDA).** LDA is a pattern recognition method that provides a classification model based on the selection of independent variables from an original pool of molecular descriptors that best predict the class or category to which a given compound belongs. Therefore, the task of LDA is to generate linear discriminant functions using linear combinations of the selected variables in order to divide a *n*-dimensional space into regions where compounds are grouped according to the calculated value of these functions.

In this part of the study, starting from a pool of eight noncorrelated molecular shape descriptors ( $r^2 < 0.8$ ; see

**Table 4.** TGR5 Agonist Potency of Steroid Hormones<sup>a</sup>

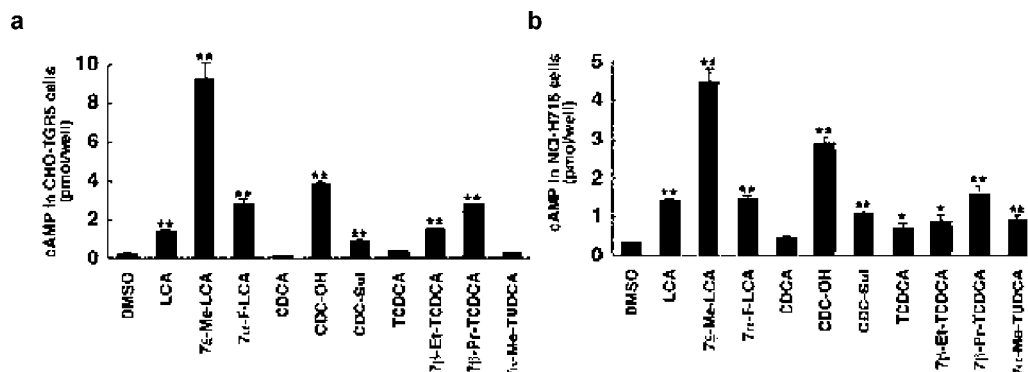
Trivial Name	Structure	EC <sub>50</sub>	Efficacy
Pregnandiol ( <b>42</b> )		0.85	84
5α-Pregnandione ( <b>43</b> )		0.29	72
Progesterone ( <b>44</b> )		2.77	98
Etiocholanolone ( <b>45</b> )		3.76	92
Epietiocholanolone ( <b>46</b> )		2.62	79
Androsterone ( <b>47</b> )		6.22	77
Epiandrosterone ( <b>48</b> )		3.20	87
Dehydroepiandrosterone ( <b>49</b> )		3.33	75
Dehydroepiandrosterone-Sul ( <b>50</b> )		127	83
Etiocholandiol ( <b>51</b> )		2.44	95
Androstandiol ( <b>52</b> )		7.01	72
Androstanolone ( <b>53</b> )		4.49	83
Testosterone ( <b>54</b> )		18.3	70
17β-Estradiol ( <b>55</b> )		38.4	11

<sup>a</sup> Data represent average values of at least three independent experiments of CRE-driven luciferase reporter assays in TGR5-transfected CHO cells. Units are  $\mu\text{M}$  for EC<sub>50</sub> and % of 10  $\mu\text{M}$  LCA value for efficacy.

Experimental Section for details), we have investigated which set of these variables is able to differentiate between active (EC<sub>50</sub>

< 10  $\mu\text{M}$ , class 1) and moderately active, inactive (EC<sub>50</sub> ≥ 10  $\mu\text{M}$ , class 2) TGR5 agonists. Since we used two categories of





**Figure 2.** CHO-TGR5 cells (a) and human intestinal NCI-H716 cells (b) were cultured and stimulated with vehicle (DMSO) only or various bile acids and bile acid derivatives at a concentration of 1.6  $\mu$ M for 30 min. cAMP production was then determined. The values represent the mean  $\pm$  SEM: (\*)  $p < 0.05$ ; (\*\*)  $p < 0.01$  versus vehicle (DMSO)-treated cells.

**Table 5.** TGR5 Selectivity of Bile Acids and Bile Acid Derivatives<sup>a</sup>

compd	Acid-form			TGR5		FXR		TGR5 selectivity index, EC <sub>50</sub> ratio (TGR5/FXR)
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	EC <sub>50</sub>	efficacy	EC <sub>50</sub>	efficacy	
LCA ( <b>6</b> )	—H	—H	—CO <sub>2</sub> H	0.58	101	20.0	30	0.029
7ξ-Me-LCA ( <b>17</b> )		—H, —CH <sub>3</sub>	—CO <sub>2</sub> H	0.076	106	16.2	32	0.0047
7α-F-LCA ( <b>18</b> )	—F	—H	—CO <sub>2</sub> H	0.25	99	>100	0 (100 $\mu$ M)	<0.0025
CDCA ( <b>4</b> )	—OH	—H	—CO <sub>2</sub> H	6.71	105	13.0	62	0.52
7β-Me-CDCA ( <b>22</b> )	—OH	—CH <sub>3</sub>	—CO <sub>2</sub> H	6.18	105	10.6	20	0.58
7β-Et-CDCA ( <b>23</b> )	—OH	—CH <sub>2</sub> CH <sub>3</sub>	—CO <sub>2</sub> H	2.63	99	>50	0 (50 $\mu$ M)	<0.053
7β-Pr-CDCA ( <b>24</b> )	—OH	—CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	—CO <sub>2</sub> H	0.78	108	>31.6	0 (31.6 $\mu$ M)	<0.025
CDC-OH ( <b>30</b> )	—OH	—H	—CH <sub>2</sub> OH	0.12	103	6.0	18	0.020
CDC-Sul ( <b>31</b> )	—OH	—H	—CH <sub>2</sub> SO <sub>3</sub> H	0.44	103	>100	0 (100 $\mu$ M)	<0.0044
UDCA ( <b>9</b> )	—H	—OH	—CO <sub>2</sub> H	36.4	75	>50	4.7 (50 $\mu$ M)	<0.36
UDC-OH ( <b>40</b> )	—H	—OH	—CH <sub>2</sub> OH	2.18	79	>31.6	0 (31.6 $\mu$ M)	<0.069
UDC-Sul ( <b>41</b> )	—H	—OH	—CH <sub>2</sub> SO <sub>3</sub> H	5.02	109	>100	0 (100 $\mu$ M)	<0.050

<sup>a</sup> Data represent average values of at least three independent experiments of CRE-driven luciferase reporter assays in TGR5-transfected CHO cells. Units are  $\mu$ M for EC<sub>50</sub> and % of 10  $\mu$ M LCA value for efficacy. FXR agonist activities were determined as describe in the Experimental Section; units are  $\mu$ M for EC<sub>50</sub> and % of 10  $\mu$ M 6-ECDCa value for efficacy. Regarding compounds that did not reach plateau activation level, maximum concentrations tested are in parentheses.

**Table 6.** Classification Model of TGR5 Agonists

variable	<i>F</i>	eigenvector	discriminant function coefficients	
			class 1	class 2
constant			−0.866	−1.595
(i) TASA	5.793	−0.314	−1.509	3.450
(ii) RASA	13.242	0.389	1.872	−4.279
(iii) RNCS	5.480	0.070	0.337	−0.771
(iv) COSV	17.646	0.527	2.535	−5.795
(v) NCOSV	5.791	0.118	0.569	−1.302

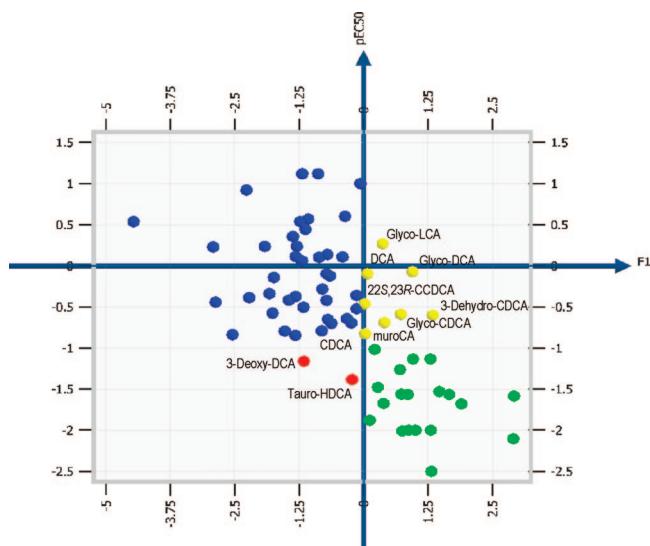
compounds, two discriminant functions have been generated by LDA using five selected variables (Table 6; see Experimental Section for details). Therefore, depending on the greatest value resulting from these functions, a compound is inserted into the group of active (class 1) or moderately active, inactive (class 2) molecules. The independent variables selected by LDA to generate the discriminant functions are the following molecular descriptors: (i) the total hydrophobic surface area (TASA); (ii) the relative hydrophobic surface area of the compound (RASA, total hydrophobic surface area divided by the total molecular solvent-accessible surface area); (iii) the relative negative charged surface area (RNCS, solvent-accessible surface area of

the most negative atom divided by the relative charge of most negative atom); (iv) the common overlapping volume (COSV) calculated with the tauro-conjugated form of LCA (**6a**); (v) the noncommon overlapping volume (NCOSV) calculated with the tauro-conjugated form of LCA (**6a**). The *F* values, eigenvectors, and classification function coefficients relative to the five descriptors are reported in Table 6 and discussed in the following section.

The classification model is endowed with the following statistical parameters:  $n_{\text{class-1}} = 48$ ;  $n_{\text{class-2}} = 21$ ;  $F = 12.241$ ; eigenvalue = 1;  $p < 0.0001$ . It is mentioned that the total number of observations is 69, since both the 7α- and 7β-Me-LCA (**17**) are considered.

The above model is able to correctly classify 83.33% of active compounds (class 1) and 90.48% of inactive compounds (class 2). Figure 3 shows the plot of the observed activity (pEC<sub>50</sub>) versus the scores (F1) of the compounds resulting from the classification model.

After the leave-one-out cross-validation, the model is still able to classify well 79.17% of active TGR5 agonists and 85.71% of inactive TGR5 agonists. The overall performance of the classification model, as evaluated using the value of the



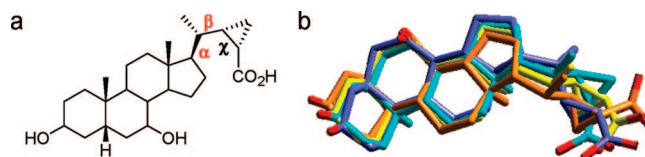
**Figure 3.** Plot of TGR5 activity ( $pEC_{50}$ ) versus the scores of the compounds (F1) obtained from the LDA. Correctly predicted class 1 and class 2 compounds are shown as blue and green spheres, respectively. False positive and false negative predicted compounds are labeled and shown as red and yellow spheres, respectively.

area under the ROC curve ( $AUC = 0.928$ ; see Experimental Section for details), indicates the presence of good discriminant functions.

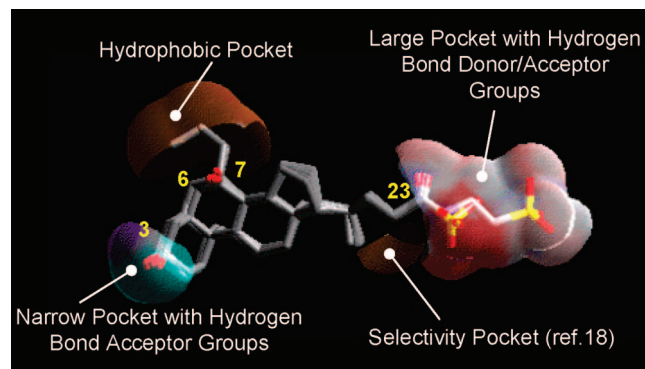
## Discussion

Unlike FXR, which mediates the activation of the genomic pathway of BAs, TGR5 is a newly discovered metabotropic receptor whose activation constitutes the gateway to the nongenomic functions of BAs. The identification of selective and potent modulators for FXR and TGR5 is crucial not only to unravel the physiological actions and pathological implications of genomic and nongenomic signaling pathways mediated by BAs but also to advance in the relative therapeutic arenas where these receptors play a pivotal role. With this aim, we recently reported that alkyl-substituted derivatives of CDCA were able for the first time to pharmacologically differentiate genomic and nongenomic activities of BAs.<sup>18</sup> We disclosed, in particular, that the introduction of a methyl group at the  $C_{23}$ -(*S*) position of natural BAs conferred selectivity for TGR5 over FXR activation and that 6,23-dialkyl substituted BA derivatives, such as 6-Et,23(*S*)-Me-CDCA (**1**), were very potent and selective agonists of TGR5.

In this paper, we have extended the structure–activity relationship of TGR5 agonists by appraising the activities to the receptor of a collection of natural occurring BAs, semisynthetic BA derivatives, and steroid hormones. As a general consideration, the inspection of both TGR5 activities and efficacy values of natural occurring BAs (**4**–**11**) and steroid hormones (**42**–**55**) indicates that while the former ones are the truly endogenous modulators of the receptor, some of the latter ones, such as pregnandiol (**40**) and  $5\alpha$ -pregnandione (**41**), may still markedly modulate the receptor. On this point, it should be mentioned that the use of supraphysiological concentrations of some steroid hormones, including DHEA (**49**), etiocholanolone (**45**), and epietiocholanolone (**46**), has been previously reported to show antiobesity and antidiabetic properties in rodent models, perhaps accounting for their activity on TGR5.<sup>22–25</sup> Further *in vivo* studies in TGR5-deficient mouse models are, however, required to identify the relative contributions of TGR5



**Figure 4.** (a) Dihedral angles used to carry out the conformational search are labeled with red symbols onto compound **32**. (b) Superposition of the best matching pair of minimum conformations between isomers **32** (yellow) and **33** (orange). The best fitting conformations of compounds **34** (cyan) and **35** (blue) are also shown.



**Figure 5.** Structure–activity relationships of bile acids as TGR5 agonists. The alignment of compounds **6a**, **18**, **23**, **24**, **30**, **31** is shown along with the features of TGR5 binding site. The surfaces of the pockets were generated using the Receptor module of Cerius-2.

modulation to the pharmacological properties of the above-mentioned steroid hormones.

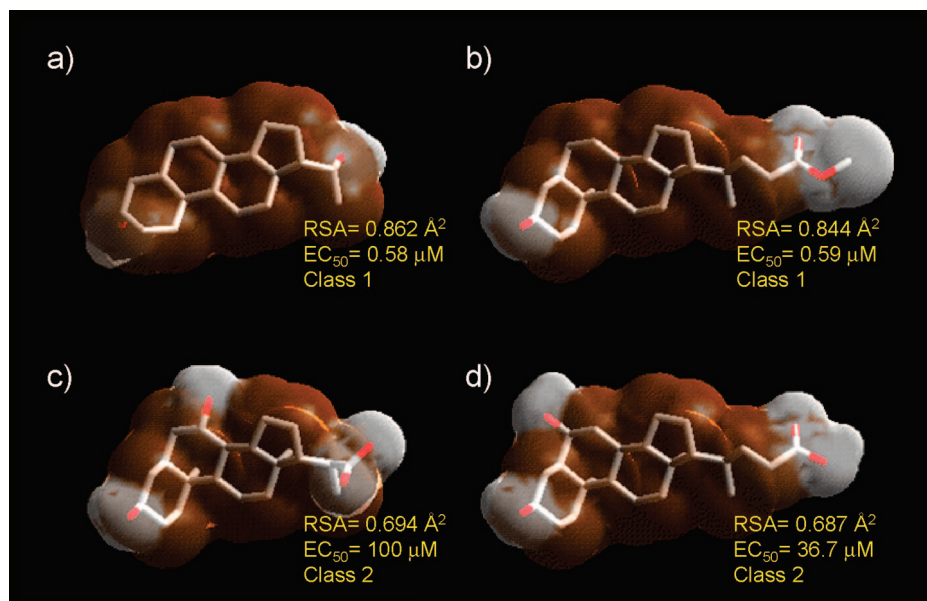
Among naturally occurring BA derivatives, the tauro-conjugated derivative of LCA (**6a**) is the most active compound, halving the  $EC_{50}$  of the glyco-conjugate (**6b**) and the free form of LCA (**6**), respectively. While the elongation of BA side chain with the conjugated derivatives is well tolerated by the receptor, its shortening slightly reduces TGR5 activity.

The biological results of the four  $C_{22}$ – $C_{23}$  cyclopropyl isomers of CDCA (**32**–**35**) are instrumental to suggest the bioactive conformation of the torsional  $\chi$ -angle of BA side chain into the binding site of TGR5 (Figure 4; see Experimental Section for details). While the folded conformation of this angle in 22*S*,23*S*-CCDCA (**32**) is preferred for TGR5 binding, the low micromolar activity shown by its extended conformation in 22*S*,23*R*-CCDCA (**33**) raises the hypothesis of a large binding pocket in TGR5 that can host this region of the molecule.

This observation further sustains the previous structure–activity relationship study that reported, converse to FXR, the presence of an accessory binding pocket in the receptor that played a key role in determining TGR5 selectivity of  $C_{23}$ -(*S*) methyl derivatives of CDCA.<sup>18</sup>

Interestingly, the molecular basis of the enhancement of the TGR5 potency of the tauro-conjugated derivatives of BAs as well as the sulfonic substituted BAs may be ascribed to the presence of the sulfonic moiety. Very recently, by statistically comparing a large amount of structural information about binding sites that recognize carboxylic, sulfonic, and phosphonic groups, we have found evidence that binding pockets specifically recognizing sulfonic moieties in general do not contain positively charged residues.<sup>26</sup> On this basis and from the above observations, we may speculate that TGR5, converse to FXR, is endowed with a neutral charged binding site (Figure 5).

A consolidation of this hypothesis comes both from the selectivity shown by compound **31** to TGR5 and from the



**Figure 6.** Graphical representation of the relative hydrophobic surface area (RSA, brown area) of compounds **40** (a), **25** (b), **39** (c), and **10** (d). The high value of this molecular shape property in compounds **42** and **25** favors TGR5 modulation over compounds **39** and **10**. The hydrophobic areas are plotted on the van der Waals surfaces as generated by the Receptor module of Cerius-2.

evidence that the reduction of the negatively formal charged carboxylic group to the neutrally formal charged alcohol group (**30**, **40**) further enhances the TGR5 potency of the parent BA while showing very poor efficacy at FXR.

As far as the body modified BA derivatives are considered, we previously reported that the introduction of an ethyl group in the 6 $\alpha$ -position considerably impacts both on TGR5 and FXR potency. In this paper we have demonstrated that the introduction of alkyl groups in position 7 $\beta$  of the BA steroidal structure elicits an improvement of both TGR5 potency and selectivity, with the 7 $\xi$ -methyl derivative of LCA (**17**) being one of the most active and selective TGR5 agonists currently available. It is worth noting that when compared to the parent BA, the insertion of a fluorine atom in the 7 $\alpha$  position of LCA (**18**) shows only a slight improvement of TGR5 activity but the selectivity to this receptor is significantly enhanced. On this basis, we can advance the hypothesis that the region of the TGR5 binding site recognizing this part of the molecule is a hydrophobic pocket that is roomy enough to host the propyl moiety of compound **24** (Figure 5).

Among the hydroxylation patterns of the steroid nucleus of BAs (positions C<sub>3</sub>, C<sub>7</sub>, C<sub>12</sub>), only the hydroxyl group in the 3 $\alpha$ -position favors the activity. Moreover, while the oxidation of such group (**15**, **20**) halves the EC<sub>50</sub> of the parent BA, its substitution with acetoxy or sulfate moieties abolishes TGR5 activity (**13**, **14**). These data allow us to envisage the presence of a narrow binding pocket that is enriched with hydrogen bonding donor groups that anchor the 3-keto or 3-hydroxyl groups (Figure 5).

In order to validate the TGR5 activities of the compounds observed in luciferase assays, we have additionally tested a selection of the most potent agonists in two different cell models of cAMP formation: CHO cells overexpressing TGR5; NCI-H716 cells that endogenously express high levels of the receptor.<sup>7</sup> The tested subset of TGR5 agonists was able to elicit cAMP formation in both cell lines with the same order of potency as that observed in our transfection assays. Furthermore, the fact that these compounds induced cAMP levels in the intestinal endocrine cell line (NCI-H716) suggests that these

compounds could induce GLP-1 secretion in vivo, supporting the concept that TGR5 agonists may be useful in the treatment of metabolic diseases such as diabetes.<sup>17,27</sup>

While the above observations have enabled us to extend the structure–activity relationships of bile acids as TGR5 agonists (Figure 5), they do not provide satisfactory insights into what properties relevant to TGR5 modulation are shared by bile acids and steroid hormones. At this point, we deemed it of interest to develop a binary classification model of TGR5 agonists using a linear discriminant analysis and a pool of molecular shape descriptors.

The proposed model is able to correctly classify 83.33% of active compounds and 90.48% of inactive compounds (Figure 3). The influence of the selected descriptors on TGR5 activity is reflected by their relative eigenvectors (Table 6). The inspection of these parameters uncovers some hidden properties shared by the molecular shape of bile acids and steroid hormones that are relevant to TGR5 activation. In particular, the high positive values of the common overlapping volume (COSV, eigenvector = 0.527) with tauro-LCA (**6a**) and the relative hydrophobic surface area of the compounds (RSA, eigenvector = 0.389) support the following scenario: TGR5 agonists have highly conserved overlapping volumes despite the chemical diversity of bile acids and steroid hormones, which allow plenty of geometrical complementarities with the binding site of the receptor. The relative hydrophobic surface area of the compounds is a required property to tightly bind TGR5 into a pocket endowed with hydrophobic properties (Figure 6).

Furthermore, because the total hydrophobic surface area negatively affects the activity of the compounds (TASA, eigenvector = -0.314) whereas the relative negative charged surface slightly does the opposite (RNCS, eigenvector = 0.070), the above binding site of bile acids and steroid hormones in TGR5, unlike nuclear receptors, should be only partly buried into the core of the active conformation of the GPCR. From these observations, the influence of the molecular shape of bile acids and steroid hormones on TGR5 activity can mainly be ascribed to two aspects: steric complementarities (COSV and NCOSV) and lipophilic profile (RSA, TASA, RNCS). These



aspects combine with the stereospecificity of molecular recognition of natural bile acids to TGR5.<sup>21</sup>

Concerning the steric aspect, it has been reported that the nuclear receptor superfamily has evolved from a common ancestor through divergent evolution conserving the overall fold of the ligand binding domain and selecting new ligands with conserved van der Waals volumes in order to complementary fill the sterical properties of the binding cavity.<sup>28</sup> In line with this model, our results indicate that TGR5 could adopt a similar evolutionary strategy directed to the selection of endogenous ligands with conserved binding site-specific sterical properties of the molecular shape.

Thus, it is conceivable to envisage the presence of an evolutionary relationship between nongenomic and genomic pathways of BAs, despite their functional diversity.

## Conclusions

BAs are not only amphipathic molecules that act as solubilizing agents but are signaling molecules with both genomic and nongenomic functions.<sup>3</sup> TGR5, a recently discovered GPCR, mediates some of the nongenomic actions of BAs by promoting cAMP formation. In view of the reported role of TGR5 in metabolic homeostasis, its pharmacological modulation may furnish alternative therapeutic strategies to treat metabolic diseases including obesity, diabetes, and the metabolic syndrome.

In this paper we have extended the current knowledge of the structure–activity relationships of TGR5 agonists by devising the screening of a collection of natural BAs, semisynthetic BA derivatives, and steroid hormones. The screening has afforded the disclosure of 7 $\xi$ -Me-LCA (**17**), 7 $\alpha$ -F-LCA (**18**), and CDC-Sul (**31**) as some of the most potent and selective TGR5 agonists currently available. These compounds, while providing novel chemical tools to investigate the physiological and pathological roles of the nongenomic actions of BAs, will enable a thorough pharmacological characterization and validation of TGR5 in the above therapeutic areas.

Structure–activity relationship studies have indicated that the binding site of TGR5, additionally to the selectivity binding pocket,<sup>18</sup> is endowed with a narrow hydrogen bonding donor site recognizing the 3-hydroxyl group of BAs, a hydrophobic pocket lining C<sub>6</sub> and C<sub>7</sub> positions of BAs, and a large and neutrally formal charged pocket that anchors the acidic side chain. Furthermore, LDA study has provided some insights into what common features of the molecular shape of bile acids and steroid hormones affect TGR5 activation. These consist mainly in two aspects: steric complementarities and lipophilic profile, with the former being the most relevant to the activity of the compound. This novel information will be instrumental to design novel selective and potent TGR5 agonists and to further explore the role of TGR5 in health and disease.

## Experimental Section

**Chemistry.** We obtained general chemicals from Sigma-Aldrich (St. Louis, MO). Natural occurring BAs, BA derivatives, and steroid hormones were purchased from Sigma-Aldrich or Steraloids (Newport, RI) except the following compounds. Lagodeoxycholic acid (3 $\alpha$ ,12 $\beta$ -dihydroxy-5 $\beta$ -cholan-24-oic acid, lagoon-DCA, **8**), 3-dehydrochenodeoxycholic acid (3-oxo-7 $\alpha$ -hydroxy-5 $\beta$ -cholan-24-oic acid, 3-dehydro-CDCA, **20**), 3-deoxychenodeoxycholic acid (7 $\alpha$ -hydroxy-5 $\beta$ -cholan-24-oic acid, 3-deoxy-CDCA, **21**) were synthesized as described previously.<sup>29–31</sup> Nor-BAs, 3 $\alpha$ -hydroxy-24-nor-5 $\beta$ -cholan-23-oic acid (nor-LCA, **26**), 3 $\alpha$ ,7 $\alpha$ -dihydroxy-24-nor-5 $\beta$ -cholan-23-oic acid (nor-CDCA, **28**), and 3 $\alpha$ ,7 $\beta$ -dihydroxy-24-nor-5 $\beta$ -cholan-23-oic acid (nor-UDCA, **38**), and dinor-BAs, 3 $\alpha$ ,7 $\alpha$ -dihydroxy-23,24-dinor-5 $\beta$ -cholan-22-oic acid (dinor-CDCA,

**29**) and 3 $\alpha$ ,7 $\beta$ -dihydroxy-23,24-dinor-5 $\beta$ -cholan-22-oic acid (dinor-UDCA, **39**), were synthesized by side chain degradation of their corresponding BAs as described previously.<sup>32,33</sup> Bile alcohols, 5 $\beta$ -cholan-3 $\alpha$ ,7 $\alpha$ ,24-triol (CDC-OH, **30**), 5 $\beta$ -cholan-3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ ,24-tetrol (C-OH, **36**), and 5 $\beta$ -cholan-3 $\alpha$ ,7 $\beta$ ,24-triol (UDC-OH, **40**), and bile sulfonates, 3 $\alpha$ ,7 $\alpha$ -dihydroxy-5 $\beta$ -cholan-24-sulfonate (CDC-Sul, **31**), 3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ -trihydroxy-5 $\beta$ -cholan-24-sulfonate (C-Sul, **37**), and 3 $\alpha$ ,7 $\beta$ -dihydroxy-5 $\beta$ -cholan-24-sulfonate (UDC-Sul, **41**), were synthesized from the corresponding BAs as described previously.<sup>34–7</sup> Alkylated BA derivatives, 7 $\xi$ -methylthiocholic acid (3 $\alpha$ -hydroxy-7 $\xi$ -methyl-5 $\beta$ -cholan-24-oic acid, 7 $\xi$ -Me-LCA, **17**), 7 $\beta$ -methylchenodeoxycholic acid (3 $\alpha$ ,7 $\alpha$ -dihydroxy-7 $\beta$ -methyl-5 $\beta$ -cholan-24-oic acid, 7 $\beta$ -Me-CDCA, **22**), 7 $\beta$ -methyltaurochenodeoxycholic acid (3 $\alpha$ ,7 $\alpha$ -dihydroxy-7 $\beta$ -methyl-5 $\beta$ -cholan-24-oic acid *N*-(2-sulfoethyl)amide, 7 $\beta$ -Me-TCDCa, **22a**), 7 $\beta$ -ethyl-CDCA (3 $\alpha$ ,7 $\alpha$ -dihydroxy-7 $\beta$ -ethyl-5 $\beta$ -cholan-24-oic acid, 7 $\beta$ -Et-CDCA, **23**), 7 $\beta$ -ethyl-TCDCa (3 $\alpha$ ,7 $\alpha$ -dihydroxy-7 $\beta$ -ethyl-5 $\beta$ -cholan-24-oic acid *N*-(2-sulfoethyl)amide, 7 $\beta$ -Et-TCDCa, **23a**), 7 $\beta$ -propyl-CDCA (3 $\alpha$ ,7 $\alpha$ -dihydroxy-7 $\beta$ -propyl-5 $\beta$ -cholan-24-oic acid, 7 $\beta$ -Pr-CDCA, **24**), and 7 $\beta$ -propyl-TCDCa (3 $\alpha$ ,7 $\alpha$ -dihydroxy-7 $\beta$ -propyl-5 $\beta$ -cholan-24-oic acid *N*-(2-sulfoethyl)amide, 7 $\beta$ -Pr-TCDCa, **24a**) were synthesized as described previously.<sup>35–37</sup> 7-Fluorinated BA derivatives, 7 $\alpha$ -fluoro-LCA (3 $\alpha$ -hydroxy-7 $\alpha$ -fluoro-5 $\beta$ -cholan-24-oic acid, 7 $\alpha$ -F-LCA, **18**) and 7 $\beta$ -fluoro-LCA (3 $\alpha$ -hydroxy-7 $\beta$ -fluoro-5 $\beta$ -cholan-24-oic acid, 7 $\beta$ -F-LCA, **19**), were synthesized as described previously.<sup>38</sup> 22,23-Methylene-CDCA derivatives (3 $\alpha$ ,7 $\alpha$ -dihydroxy-22,23-methylene-5 $\beta$ -cholan-24-oic acid (CCDCA) derivatives (**32–35**) were prepared as described previously.<sup>39</sup>

**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 $\beta$  were obtained from Clontech (Palo Alto, CA). pCMX-hFXR and pCMX-mRXR $\alpha$  were kind gifts from Dr. David J. Mangelsdorf (Howard Hughes Medical Institute, University of Texas Southwestern Medical Center). pEcRE $\times$ 7-Luc was a kind gift from Dr. Richard A. Heyman (X-ceptor Therapeutics, CA).

**Cell Culture.** Chinese hamster ovary (CHO) cells, NCI-H716 cells, Hep3B cells, and COS1 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  $\alpha$  ( $\alpha$ -MEM) supplemented with 10% (v/v) fetal bovine serum (FBS) and 100  $\mu$ M nonessential amino acids (NEAA). NCI-H716 cells were maintained in suspension in RPMI-1640 supplemented with 10% (v/v) FBS, 10 mM HEPES, and 1 mM sodium pyruvate. Hep3B cells were maintained in Eagle's medium supplemented with 10% (v/v) FBS and 100  $\mu$ M NEAA. COS1 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) FBS. All cell culture medium was supplemented with 100 units/mL penicillin and 100  $\mu$ g/mL streptomycin sulfate. Cells were grown at 37 °C in an atmosphere of 5% CO<sub>2</sub>, 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  $3.5 \times 10^4$  cells/well, cultured for 24 h, and then transfected with 150 ng of human (h) TGR5 expression plasmid (pCMVSPORT6/hTGR5) and 100 ng of cAMP-responsive element (CRE)-driven luciferase reporter plasmid (pCRE-Luc) in each well using lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. After 6 h of incubation, cells were washed once with phosphate-buffered saline (PBS) and medium was exchanged for DMEM containing 0.1% (w/v) bovine serum albumin (BSA). After incubation for another 18 h, 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 50  $\mu$ L of lysis buffer (25 mM Tris-Cl (pH 7.6), 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.

COS1 cells were plated in 96-well plates at a density of  $2.5 \times 10^4$  cells/well in DMEM supplemented with 10% (v/v) charcoal-

stripped FBS, cultured for 24 h, and then transfected with 25 ng of hFXR expression plasmid (pCMX-hFXR), 25 ng of mouse (m) retinoid X receptor  $\alpha$  (RXR $\alpha$ ) expression plasmid (pCMX-mRXR $\alpha$ ), 50 ng of reporter plasmid (pEcRE $\times$ 7-Luc), and 50 ng of pCMV $\beta$  as internal control in each well, using the lipofectamine 2000 reagent. After 24 h, cells were washed twice with PBS and treated with different concentrations of each compound in fresh DMEM supplemented with 10% (v/v) charcoal-stripped FBS for 24 h. After treatment, the cells were lysed with 50  $\mu$ L of lysis buffer by a freeze-thaw cycle and subjected to both luciferase and  $\beta$ -galactosidase assays as described below. Normalized luciferase values were determined by dividing the luciferase activity by the  $\beta$ -galactosidase activity.

**Luciferase and  $\beta$ -Galactosidase Assays.** For luciferase assays, 20  $\mu$ L of cell lysate was mixed with 100  $\mu$ L of luciferase reaction buffer [235  $\mu$ M luciferine, 265  $\mu$ M ATP, and 135  $\mu$ M coenzyme A (CoA)] and luminescence was determined with CentroXS3 LB960 (Berthold Technologies, Bad Wildbad, Germany). For  $\beta$ -galactosidase assays, 10  $\mu$ L of cell lysate was mixed with 100  $\mu$ L of buffer Z [60 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM KCl, 1 mM MgSO<sub>4</sub>, 50 mM  $\beta$ -mercaptoethanol, and 0.75 mg/mL *o*-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG)] and incubated at 37 °C for 0.5–3 h. Reactions were stopped by adding 50  $\mu$ L of stop buffer (1 M Na<sub>2</sub>CO<sub>3</sub>), and the optical density at 420 nm was determined.

**Establishing CHO Cells Stably Expressing Human TGR5 (CHO-TGR5 Cells).** CHO cells were transfected with 3.8  $\mu$ g of hTGR5 expression plasmid (pCMVSPORT6/hTGR5), 3.8  $\mu$ g of CRE-driven luciferase reporter plasmid (pCRE-Luc), and 0.4  $\mu$ g of neomycin-resistant gene expression plasmid [pcDNA3.1(+)] using lipofectamine 2000. The transfectants were selected with 400  $\mu$ g/mL G418 sulfate, and single clones were grown in 96-well plate independently. TGR5-expressing CHO cell lines were screened by LCA treatments, followed by luciferase assays.

**cAMP Production Analysis.** NCI-H716 cells were plated in 96-well plates coated with 0.75 mg/mL Matrigel (BD Biosciences) according to manufacturer's instructions just prior to use, at a density of  $6 \times 10^4$  cells/well in DMEM supplemented with 10% (v/v) FBS, 100 units/mL penicillin, and 100  $\mu$ g/mL streptomycin sulfate and cultured for 24 h, which allowed cell adhesion to the bottom of the plate. CHO-TGR5 cells were plated in 96-well plates at a density of  $3.5 \times 10^4$  cells/well in  $\alpha$ -MEM supplemented with 10% (v/v) FBS, 100  $\mu$ M NEAA, 100 units/mL penicillin, and 100  $\mu$ g of streptomycin sulfate and cultured for 24 h. The cells were washed twice with PBS, and medium was exchanged for cAMP assay medium [DMEM containing 0.1% (w/v) BSA and 0.5 mM 3-isobutyl-1-methylxanthine (IBMX)]. After incubation for 30 min at 37 °C, the cells were treated with each compound in fresh cAMP assay medium for 30 min. After treatment, medium was discarded and cAMP amounts were determined using cAMP screen kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions.

**50% Effective Concentrations (EC<sub>50</sub>) and Efficacy Determination.** Assays were performed in triplicate or quadruplicate for each condition. EC<sub>50</sub> values were determined by probit analysis. Efficacy was determined by calculating percentages of 10  $\mu$ M LCA value for TGR5 agonist study and 10  $\mu$ M 6 $\alpha$ -Et-CDCA value for FXR agonist study, respectively. After the average value of the basal (vehicle-treated) condition was subtracted, values were applied to EC<sub>50</sub> and/or efficacy determinations. Calculation of average EC<sub>50</sub> and comparison of the EC<sub>50</sub> 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-2<sup>40</sup> and optimized using Universal Force Field, version 1.2,<sup>41</sup> and the Smart Minimizer protocol of the Open Force Field module (OFF). Atomic charges were calculated using the Gasteiger method,<sup>42</sup> implemented in the polygraph version of Cerius-2.

The conformational analysis of the four C<sub>22</sub>–C<sub>23</sub> cyclopropyl isomers of CDCA (**32**–**35**) was carried out using a systematic

search procedure on the  $\alpha$  and  $\beta$  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 Universal Force Field, version 1.2, and the Smart Minimizer protocol of Cerius-2. After the removal of duplicate conformations, the global and local minima conformers of compounds **32** and **33** 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 22*S*,23*S*-CCDCA (**32**) and 22*S*,23*R*-CCDCA (**33**). Then, all the compounds of the collection were mutually aligned on the bioactive conformation of **32** using the subgraph matching routine as implemented in the Align module of Cerius-2. The resulting alignment was manually refined to maximize the overlapping volume of the compounds. For each compound, a total of 32 molecular shape descriptors (30 Jurs descriptors and 2 excluded volume descriptors) were calculated using the Descriptor+ module of Cerius-2. While Jurs descriptors capture the shape and electronic information of the molecules by mapping atomic partial charges on solvent-accessible surface areas of individual atoms,<sup>43</sup> excluded volume descriptors encode the spatial arrangement of the molecules with respect to a compound chosen as reference. In particular, the latter were calculated assessing the common (COSV) and noncommon (NCOSV) overlapping volumes of each compound with the most active compound endowed with the larger shape (TLCA, EC<sub>50</sub> = 0.29  $\mu$ M). All the descriptors were standardized using the standard deviation. Then, the pool of descriptors was checked to avoid the presence of highly correlated variables ( $r^2 > 0.8$ ). The analysis led to a reduction of the descriptors to eight poorly correlated independent variables (six Jurs descriptors and two excluded volume descriptors).

The linear discriminant analysis was performed using the XLSTAT software. In particular, the selection of molecular descriptors that best predicts the active or inactive class to which a given compound belongs was carried out with the forward stepwise method. Briefly, this method is based on the *F* value of the variable and the Wilk's  $\lambda$  test. According to this test, at each step the variable that minimizes the overall Wilk's  $\lambda$  is entered. The *F* value allows the appraisal of the relative importance of the variable among the original pool of descriptors to the discriminant function. In particular, a variable is entered into the function if its *F* value is greater than a defined entry value ( $F_{\min} = 0.05$ ) and minimizes the Wilk's  $\lambda$ ; thereafter, *F* values of the rest of the variables in the model are recalculated and those with *F* values less than the Removal value ( $F_{\max} = 0.10$ ) are removed. This procedure is continued until the *F* values of the remaining variables out of the discriminant function are all less than the defined  $F_{\min}$ .

The predictive ability of the discriminant function was assessed using a leave-one-out (LOO) cross-validation procedure, whereas the performance of the classification model was evaluated using the ROC curve (receiver operating characteristics). In a ROC curve, the proportion of well-classified positive events (correctly classified active compounds) is called the sensitivity and the proportion of well-classified negative events (correctly classified inactive compounds) is termed specificity. The ROC curve is obtained by plotting "1 – specificity" versus the "sensitivity". The value of the area under the curve (AUC) indicates the quality of the model. While for an ideal model AUC = 1, for a random model AUC = 0.5. A model is usually considered good when the AUC value is greater than 0.7. The data set of compounds and their descriptors values, observed class membership, and predicted class membership are provided in Tables s1–s2 in Supporting Information.

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

**Supporting Information Available:** Table s1 listing the data set used for LDA analysis with the calculated and normalized molecular shape descriptors; Table s2 listing prior and posterior classification, membership probabilities (Pr1 and Pr2), and scores



(F1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Green, G. M.; Nasset, E. S. Importance of bile in regulation of intraluminal proteolytic enzyme activities in the rat. *Gastroenterology* **1980**, *79*, 695–702.
- (2) Gass, J.; Vora, H.; Hofmann, A. F.; Gray, G. M.; Khosla, C. Enhancement of dietary protein digestion by conjugated bile acids. *Gastroenterology* **2007**, *133*, 16–23.
- (3) Houten, S. M.; Watanabe, M.; Auwerx, J. Endocrine functions of bile acids. *EMBO J.* **2006**, *25*, 1419–1425.
- (4) 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.
- (5) 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.
- (6) Makishima, M.; Okamoto, A. Y.; Repa, J. J.; Tu, H.; Learned, R. M.; Luk, A.; Hull, M. V.; Lustig, K. D.; Mangelsdorf, D. J.; Shan, B. Identification of a nuclear receptor for bile acids. *Science* **1999**, *284*, 1362–1365.
- (7) 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.
- (8) 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.
- (9) Acconcia, F.; Kumar, R. Signaling regulation of genomic and nongenomic functions of estrogen receptors. *Cancer Lett.* **2006**, *238*, 1–14.
- (10) Losel, R.; Wehling, M. Nongenomic actions of steroid hormones. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 46–56.
- (11) Vasudevan, N.; Pfaff, D. W. Membrane-initiated actions of estrogens in neuroendocrinology: merging principles. *Endocr. Rev.* **2007**, *28*, 1–19.
- (12) Francis, G. A.; Fayard, E.; Picard, F.; Auwerx, J. Nuclear receptors and the control of metabolism. *Annu. Rev. Physiol.* **2003**, *65*, 261–311.
- (13) Modica, S.; Moschetta, A. Nuclear bile acid receptor FXR as pharmacological target: Are we there yet? *FEBS Lett.* **2006**, *580*, 5492–5499.
- (14) 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.
- (15) 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.
- (16) 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.
- (17) 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.
- (18) 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.
- (19) 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.
- (20) 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. Patent WO2004067008, 2004.
- (21) 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.
- (22) Cleary, M. P.; Shepherd, A.; Jenks, B. Effect of dehydroepiandrosterone on growth in lean and obese Zucker rats. *J. Nutr.* **1984**, *114*, 1242–1251.
- (23) Coleman, D. L. Antiobesity effects of etiocholanolones in diabetes (db), viable yellow (Avy), and normal mice. *Endocrinology* **1985**, *117*, 2279–2283.
- (24) Coleman, D. L. Therapeutic effects of dehydroepiandrosterone (DHEA) and its metabolites in obese-hyperglycemic mutant mice. *Prog. Clin. Biol. Res.* **1988**, *265*, 161–175.
- (25) Yen, T. T.; Allan, J. A.; Pearson, D. V.; Acton, J. M.; Greenberg, M. M. Prevention of obesity in Avy/a mice by dehydroepiandrosterone. *Lipids* **1977**, *12*, 409–413.
- (26) Macchiarulo, A.; Pellicciari, R. Exploring the other side of biologically relevant chemical space: insights into carboxylic, sulfonic and phosphonic acid bioisosteric relationships. *J. Mol. Graphics Modell.* **2007**, *26*, 728–739.
- (27) Reimer, R. A.; Darimont, C.; Gremlich, S.; Nicolas-Metral, V.; Ruegg, U. T.; Mace, K. A human cellular model for studying the regulation of glucagon-like peptide-1 secretion. *Endocrinology* **2001**, *142*, 4522–4528.
- (28) Bogan, A. A.; Cohen, F. E.; Scanlan, T. S. Natural ligands of nuclear receptors have conserved volumes. *Nat. Struct. Biol.* **1998**, *5*, 679–681.
- (29) Schmassmann, A.; Angellotti, M. A.; Clerici, C.; Hofmann, A. F.; Ton-Nu, H. T.; Schteingart, C. D.; Marcus, S. N.; Hagey, L. R.; Rossi, S. S.; Aigner, A. Transport, metabolism, and effect of chronic feeding of lagodeoxycholic acid. A new, natural bile acid. *Gastroenterology* **1990**, *99*, 1092–1104.
- (30) Nakagaki, M.; Danzinger, R. G.; Hofmann, A. F.; DiPietro, R. A. Biliary secretion and hepatic metabolism of taurine-conjugated 7 alpha-hydroxy and 7 beta-hydroxy bile acids in the dog. Defective hepatic transport and bile hyposecretion. *Gastroenterology* **1984**, *87*, 647–659.
- (31) Tserng, K. Y. A convenient synthesis of 3-keto bile acids by selective oxidation of bile acids with silver carbonate-Celite. *J. Lipid Res.* **1978**, *19*, 501–504.
- (32) Yeh, H. Z.; Schteingart, C. D.; Hagey, L. R.; Ton-Nu, H. T.; Bolder, U.; Gavrilkina, M. A.; Steinbach, J. H.; Hofmann, A. F. Effect of side chain length on biotransformation, hepatic transport, and choleretic properties of chenodeoxychyl homologues in the rodent: studies with dinorchenodeoxycholic acid, norchenodeoxycholic acid, and chenodeoxycholic acid. *Hepatology* **1997**, *26*, 374–385.
- (33) Schteingart, C. D.; Hofmann, A. F. Synthesis of 24-nor-5 beta-cholan-23-oic acid derivatives: a convenient and efficient one-carbon degradation of the side chain of natural bile acids. *J. Lipid Res.* **1988**, *29*, 1387–1395.
- (34) Kihira, K.; Yoshii, M.; Okamoto, A.; Ikawa, S.; Ishii, H.; Hoshita, T. Synthesis of new bile salt analogues, sodium 3 alpha, 7 alpha-dihydroxy-5 beta-cholane-24-sulfonate and sodium 3 alpha, 7 beta-dihydroxy-5 beta-cholane-24-sulfonate. *J. Lipid Res.* **1990**, *31*, 1323–1326.
- (35) Une, M.; Yamanaga, K.; Mosbach, E. H.; Kuroki, S.; Hoshita, T. Synthesis of bile acid analogs: 7-alkylated chenodeoxycholic acids. *Steroids* **1989**, *53*, 97–105.
- (36) Une, M.; Cohen, B. I.; Mosbach, E. H. New bile acid analogs: 3 alpha, 7 alpha-dihydroxy-7 beta-methyl-5 beta-cholanoic acid, 3 alpha, 7 beta-dihydroxy-7 alpha-methyl-5 beta-cholanoic acid, and 3 alpha-hydroxy-7 xi-methyl-5 beta-cholanoic acid. *J. Lipid Res.* **1984**, *25*, 407–410.
- (37) Tserng, K. Y.; Hachey, D. L.; Klein, P. D. An improved procedure for the synthesis of glycine and taurine conjugates of bile acids. *J. Lipid Res.* **1977**, *18*, 404–407.
- (38) Kauffman, J. M.; Pellicciari, R.; Carey, M. C. Interfacial properties of most monofluorinated bile acids deviate markedly from the natural congeners: studies with the Langmuir–Pockels surface balance. *J. Lipid Res.* **2005**, *46*, 571–581.
- (39) 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.
- (40) Cerius-2; Accelrys: San Diego, CA, 2001.
- (41) 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.
- (42) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- (43) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies. *Anal. Chem.* **1990**, *62*, 2323–2329.