Two- and three-dimensional quantitative structure-activity relationships studies on a series of liver X receptor ligands

http://www.producao.usp.br/handle/BDPI/49331

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo
Two- and Three-Dimensional Quantitative Structure-Activity Relationships Studies on a Series of Liver X Receptor Ligands

Káthia M. Honório\(^*\)\(^1\), Lívia B. Salum\(^2\), Richard C. Garratt\(^2\), Igor Polikarpov\(^2\) and Adriano D. Andricopulo\(^*\)\(^2\)

\(^1\)Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, Av. Arlindo Bettio, 1000, 03828-000, São Paulo, SP, Brazil

\(^2\)Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-Carlense 400, 13560-970, São Carlos, SP, Brazil

Abstract: Liver X receptor (LXR) is an attractive drug target for the development of novel therapeutic agents for the treatment of dyslipidaemia and cholestasis. In the present work, comparative molecular field analysis (CoMFA) and hologram quantitative structure-activity relationship (HQSAR) studies were conducted on a series of potent LXR ligands. Significant correlation coefficients (CoMFA, \(r^2 = 0.98\) and \(q^2 = 0.69\); HQSAR, \(r^2 = 0.99\) and \(q^2 = 0.85\)) were obtained, indicating the potential of the models for untested compounds. The models were then used to predict the potency of an external test set, and the predicted values obtained from the 2D and 3D models were in good agreement with the experimental results. The final QSAR models, along with the information obtained from 3D steric and electrostatic contour maps and 2D contribution maps should be useful for the design of novel LXR ligands having improved potency.

Keywords: LXR, bioactive ligands, hologram QSAR, CoMFA analyses.

INTRODUCTION

Nuclear receptors (NRs) are ligand-dependent intracellular transcription factors that have important roles in several biological processes, including cell proliferation, differentiation and cellular homeostasis [1-4]. NRs are implicated in important diseases such as cancer, diabetes, and osteoporosis, and thus have been considered attractive targets for drug discovery and development [5]. Typically, NRs act as ligand-inducible transcription factors, responding to endogenous and exogenous chemicals. However, there are some NRs whose endogenous ligands have not yet been identified and these are known as orphan receptors. The chemical diversity of the NR ligands is considerably high as a natural result of the structural diversity of the protein themselves. For example, endogenous ligands include a diversity of lipophilic compounds such as steroid hormones, thyroid hormone, vitamin D3, fatty acids, eicosanoids, bile acids and oxysterols [6,7]. Therefore, recognition of the importance of NRs as vital regulators of genes involved in metabolic control has resulted in intensive search for novel selective ligands to combat major human diseases and disorders [8].

Liver X receptors (LXRs) form a subgroup of NRs called metabolic receptors, which have attracted increasing interest in medicinal chemistry. The two LXR subtypes \(\alpha\) and \(\beta\) (LXR\(\alpha\) and LXR\(\beta\), respectively) share a high degree of amino acid sequence similarity (78% in their DNA- and ligand-binding domains), but differ substantially in their tissue distribution. LXR\(\alpha\) was first identified in the liver, but is also expressed in other metabolically active tissues such as kidney, intestine, adipose tissue, and macrophages. In contrast, LXR\(\beta\) is ubiquitously expressed [9]. The majority of the LXR target genes appear to have one of two biological functions; firstly, removal of excess cholesterol through efflux, catabolism or decreased absorption; and secondly, synthesis of fatty acids [10,11]. Alterations in cholesterol and fatty acid metabolism play an important role in the development of cardiovascular diseases. The established functions of many of these LXR targets have fueled speculation as to how their pharmacological regulation by LXR ligands might impact metabolic disease.

Integration of structure- and ligand-based approaches have become vital components of modern drug design [12,13] The main objective of our study was to investigate the quantitative structure-activity relationships (QSAR) of a series of LXR ligands by employing the comparative molecular field analysis (CoMFA) and hologram QSAR methods. The results obtained from the predictive 3D and 2D QSAR models generated in this work should be useful in the design of new LXR ligands. It is interesting to note that the number of QSAR studies concerning the chemical structure and corresponding biological information for NRs is significant, as a result of the SAR available information. However, to the best of our knowledge, the majority of QSAR models reported in the literature is associated with the estrogen receptor (ER) [14,15], while a few other models have been...
performed for distinct NRs [13,16-18]. This fact shows the
importance of further QSAR studies involving other NRs
such as LXR, the macromolecular target of our studies.

METHODOLOGY

Data Set

The data set of 20 LXR ligands used for the HQSAR and
CoMFA analyses was selected from the literature and
consists of oxysterol and tertiary amine derivatives [19, 20]. The
generation of the ligand structures, as well as the QSAR
modeling analyses, calculations and visualizations were per-
duced using the SYBYL 8.0 package (Tripos Inc., St.
Louis, USA).

CoMFA Analyses

In 3D QSAR studies, it is crucial to consider the molecu-
lar alignment of the data set ligands, and a variety of useful
approaches has been described in the literature for this pur-
pose [12, 13, 18, 21]. In this work, the docking protocol as
implemented in GOLD 3.1 (Cambridge Crystallographic
Data Centre, Cambridge, UK) was employed to search for the
possible binding conformations of the ligands into the LXR
active site. Initially, the single optimized conformation of
each molecule in the data set was energetically minimized
employing the atom-centered partial charge AM1-ESP calcula-
tions implemented in MOPAC 6.0. The X-ray crystallo-
graphic data for LXR in complex with the small molecule
agonist (T0901317) used in docking simulations and QSAR
simulations were retrieved from the Protein Data Bank (PDB
ID 1UPV, with resolution of 2.1 Å) [22]. During the process
of structure preparation, the ligand was removed from the
complex structure and hydrogen atoms were added using the
program Pymol (DeLano Scientific, San Carlos, USA). The
active site for the docking simulations was defined consider-
ing a radius of 10 Å around the C28 atom of the ligand
bound in the X-ray structure. Default parameters and
GOLDscore function were employed in all docking runs, and
only the best ranked conformation was considered for 3D
QSAR studies.

CoMFA analyses also permit to explore the main contri-
butions of electrostatic and steric fields in the intermolecu-
lar interactions between the data set ligands and the binding site
of the LXR protein. In this context, CoMFA calculates the
steric and electrostatic properties according to Lennard-Jones
and Coulomb potentials, respectively [23]. Thus, the aligned
training set molecules were placed in a 3D grid box of 2.0 Å
in the x, y, and z directions, and a grid region was automati-
cally generated by the CoMFA routine to encompass all
molecules with an extension of 4.0 Å in each direction.
The CoMFA fields were generated at each grid point with Tripos
force field using an sp³ carbon atom probe carrying a +1 net
charge. CoMFA region focusing method was applied to in-
crease the resolution of CoMFA models. The default value
of 30 kcal mol⁻¹ was set as the maximum steric and elec-
tronic energy cutoff. Minimum-sigma (column filtering) was
set to be 2.0 kcal mol⁻¹ in order to improve the signal-to-
noise ratio by omitting those lattice points with energy varia-
tion below the threshold. All models were investigated using
noncross-validated correlation coefficient ($r^2$) and the full
cross-validated $r^2$ ($q^2$) partial least squares (PLS) leave-one-
out (LOO) method. Leave-many-out (LMO) crossvalidation
with 10 randomly selected groups were used as a more rig-
orous test to assess model stability and statistical signifi-
cance. Progressive scrambling method was applied to deter-
mine the sensitivity of the QSAR models to chance corre-
lations. The CoMFA contour maps were generated by interpo-
lation of the pairwise products between the PLS coefficients
and standard deviations of the corresponding descriptors
values. External model validation was performed with a test
set of compounds, which were not considered for QSAR
model generation. After generation of the PLS training set
models, the dependent variables (pEC₅₀) were predicted for the
test set compounds, allowing predictive-$r^2$ values to be
determined for the individual 3D QSAR models.

HQSAR Analyses

Since the integrated information obtained from 2D and
3D methods is recognized as a valuable strategy in drug de-
design [12, 13], we have explored the 2D molecular features
related to the potency of this series of LXR agonists. Predic-
tive 2D QSAR models were generated using the HQSAR
method, which only requires 2D structures and the corre-
sponding biological activity as input, allowing the investiga-
tion of a wide variety of bioactive compounds, even when the
3D biological target information is not available. Be-
sides, HQSAR typically produces fast statistical correlations
that are comparable in quality to 3D QSAR techniques, such
as CoMFA, but avoids the time consuming step of 3D model
generation and mutual alignment in 3D space [15-17, 24]. The
methodology employed in HQSAR consists of some
basic steps: (i) data set preparation; (ii) substructural frag-
mentation of the training set molecules; (iii) molecular holo-
gram generation; (iv) statistical analysis and model genera-
tion; and, (v) test set selection and external validation [13,
15-17]. In HQSAR, each molecule is hashed to a molecular
fingerprint which encodes the frequency of occurrence of
various molecular fragment types using a pre-defined set of
rules. To construct a molecular hologram, the fingerprint is
cut into strings at a fixed interval as specified by a hologram
length (HL) parameter. The strings are then aligned and the
sum of each column constitutes the individual component of
the molecular hologram of a particular length. The progress
of incorporating information about each fragment, and each of
its constituent sub-fragments, implicitly encodes 3D struc-
tural information (e.g., hybridization, chirality) [12, 25]. The
final HQSAR models can be affected by a number of pa-
rameters concerning hologram generation: hologram length,
fragment size and fragment distinction [12, 13, 15-17]. In
our studies, holograms were generated using the standard
parameters implemented in SYBYL 8.0. The models were
investigated in a similar way to that described for the
CoMFA models.

RESULTS AND DISCUSSION

The chemical structures and biological properties for the
complete set of compounds are listed in Table I. The values
of EC₅₀ employed in this study were measured under the
Table 1. Chemical Structures and EC$_{50}$ Values for the Series of LXR Ligands Studied

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Structure</th>
<th>EC$_{50}$ (μM)</th>
<th>Ligand</th>
<th>Structure</th>
<th>EC$_{50}$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training Set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td><img src="image" alt="Structure" /></td>
<td>0.46</td>
<td>2</td>
<td><img src="image" alt="Structure" /></td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td><img src="image" alt="Structure" /></td>
<td>0.32</td>
<td>4</td>
<td><img src="image" alt="Structure" /></td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td><img src="image" alt="Structure" /></td>
<td>0.22</td>
<td>6</td>
<td><img src="image" alt="Structure" /></td>
<td>0.18</td>
</tr>
<tr>
<td>7</td>
<td><img src="image" alt="Structure" /></td>
<td>0.10</td>
<td>8</td>
<td><img src="image" alt="Structure" /></td>
<td>0.26</td>
</tr>
<tr>
<td>9</td>
<td><img src="image" alt="Structure" /></td>
<td>0.26</td>
<td>10</td>
<td><img src="image" alt="Structure" /></td>
<td>0.86</td>
</tr>
<tr>
<td>11</td>
<td><img src="image" alt="Structure" /></td>
<td>0.66</td>
<td>12</td>
<td><img src="image" alt="Structure" /></td>
<td>0.25</td>
</tr>
<tr>
<td>13</td>
<td><img src="image" alt="Structure" /></td>
<td>0.08</td>
<td>14</td>
<td><img src="image" alt="Structure" /></td>
<td>0.04</td>
</tr>
</tbody>
</table>
same experimental conditions [19, 20], a fundamental requirement for successful QSAR studies [12, 13, 15]. The generation of reliable statistical models is dependent on the creation of appropriate training and test sets. From the original data set of 20 LXR modulators, 16 ligands were selected as members of the training set for model construction (1-16, Table 1) and the remaining 4 compounds (17-20, Table 1) were defined as members of the test set for external model validation. Hierarchical cluster analyses performed with Tsar 3D (Accelrys, San Diego, USA) was used as previously described to guide an appropriate compound selection [12,15,17]. Training and test sets were selected in such a way that structurally diverse molecules possessing activities of a wide range were included in both sets. Thus, the data set is appropriate for the purpose of QSAR model development. The same training and test sets were employed for all 2D and 3D QSAR analyses. The EC50 values were converted to the corresponding pEC 50 (-logEC50) and used as dependent variables in the QSAR investigations.

CoMFA Analyses

Fig. (1) shows the alignment obtained from the docking simulations performed for the complete set of LXR ligands. As can be seen, the data set molecules presented a unique molecular alignment, indicating the robustness of the method employed using the program GOLD. From the 3D molecular alignment obtained for the training set molecules, several PLS analyses were carried out and the statistical results obtained are presented in Table 2.

From Table 2, it can be seen that an initial analysis without applying the option “region focusing” (an advanced method of noise reduction) produced a non-significant q2 value of 0.39. Thus, we have applied the region focusing that was weighted by S.D. x coefficient values ranging from 0.3 to 1.2, with a grid spacing ranging from 0.5 to 1.5. The best statistical results (r2 = 0.98, q2 = 0.69) among all models were obtained with the region focusing weighted by a S.D. x coefficient values of 0.8 and a grid spacing of 1.0. The region focusing not only increased q2 values during the process of model generation, but also resulted in the refinement of contour maps. In order to check the reliability of the PLS models to small systematic perturbations of the response variable, we have used the progressive scrambling method [26], which is a nonparametric approach that does not disturb the underlying covariance structure of the data, being used to determine the sensitivity of the QSAR models to chance correlations. The results further confirmed consistency of the models as defined by the critical slope, and optimum statistics for cSDEP and Q**2 obtained at the end of different runs.

The predictive ability of the most significant CoMFA model derived for the training set molecules was assessed by using a test set of molecules (compounds 17-20, Table 1). The results of the external validation procedure employing the test set compounds are listed in Table 3, and the graphic results for the experimental versus predicted activities of both training set and test set are displayed in Fig. (2).

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Structure</th>
<th>EC50 (μM)</th>
<th>Ligand</th>
<th>Structure</th>
<th>EC50 (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td><img src="image1" alt="Structure" /></td>
<td>0.12</td>
<td>16</td>
<td><img src="image2" alt="Structure" /></td>
<td>0.19</td>
</tr>
<tr>
<td>Test Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td><img src="image3" alt="Structure" /></td>
<td>0.25</td>
<td>18</td>
<td><img src="image4" alt="Structure" /></td>
<td>0.17</td>
</tr>
<tr>
<td>19</td>
<td><img src="image5" alt="Structure" /></td>
<td>0.47</td>
<td>20</td>
<td><img src="image6" alt="Structure" /></td>
<td>0.08</td>
</tr>
</tbody>
</table>
log units. No outliers were detected in this series of compounds. From the low residual values, it can be seen that the CoMFA model obtained is highly reliable and can thus be used to predict the biological activity of novel compounds within this structural diversity.

Table 2. Statistical Results for the CoMFA Analyses

<table>
<thead>
<tr>
<th>Model</th>
<th>Statistical Parameters</th>
<th>r²</th>
<th>q²</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>No focusing</td>
<td></td>
<td>0.97</td>
<td>0.39</td>
<td>3</td>
</tr>
<tr>
<td>W = 0.8, d = 1.0*</td>
<td></td>
<td>0.98</td>
<td>0.69</td>
<td>3</td>
</tr>
</tbody>
</table>

*Region focusing was weighted by standard deviation coefficient (w) and grid spacing (d). Noncross-validated correlation coefficient (r²); cross-validated correlation coefficient (q²); optimal number of components (N).

In CoMFA methodology, visualization of the steric and electrostatic interaction fields is an essential tool to guide further SAR studies aimed at the design of new molecules with improved properties, using an association of medicinal chemistry knowledge with expertise in synthetic chemistry. Fig. (3) displays the CoMFA PLS S.D. x coefficient contour maps for the steric and electrostatic fields employing the most potent LXR ligand of the series (compound 14, pEC₅₀ = 7.35). CoMFA contour maps can be analyzed considering two main aspects: (i) steric, where favorable and unfavorable regions are represented in green and yellow; (ii) electrostatic, where red contours represent regions in which electronegative substituents may increase the biological activity, and blue contours indicate regions in which electropositive groups would contribute to enhance the activity.

According to the CoMFA/PLS analysis, the steric and electrostatic field properties contribute in a 47/53 ratio to the total variance, respectively. As it can be seen in Fig. (3),

Table 3. Experimental and Predicted Activities (pEC₅₀) with Residual Values for a Series of LXR Ligands by Using the HQSAR and CoMFA Methods

<table>
<thead>
<tr>
<th>Ligands</th>
<th>Experimental</th>
<th>Predicted CoMFA</th>
<th>Residual</th>
<th>Predicted HQSAR</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.34</td>
<td>6.30</td>
<td>-0.04</td>
<td>6.36</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>6.17</td>
<td>6.16</td>
<td>-0.01</td>
<td>6.14</td>
<td>-0.03</td>
</tr>
<tr>
<td>3</td>
<td>6.49</td>
<td>6.45</td>
<td>-0.04</td>
<td>6.53</td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>6.89</td>
<td>6.94</td>
<td>0.05</td>
<td>6.80</td>
<td>-0.09</td>
</tr>
<tr>
<td>5</td>
<td>6.66</td>
<td>6.67</td>
<td>0.01</td>
<td>6.71</td>
<td>0.05</td>
</tr>
<tr>
<td>6</td>
<td>6.74</td>
<td>6.78</td>
<td>0.04</td>
<td>6.77</td>
<td>0.03</td>
</tr>
<tr>
<td>7</td>
<td>6.98</td>
<td>6.98</td>
<td>0.00</td>
<td>6.96</td>
<td>-0.02</td>
</tr>
<tr>
<td>8</td>
<td>6.59</td>
<td>6.63</td>
<td>0.04</td>
<td>6.59</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>6.59</td>
<td>6.61</td>
<td>0.02</td>
<td>6.57</td>
<td>-0.02</td>
</tr>
<tr>
<td>10</td>
<td>6.07</td>
<td>6.17</td>
<td>0.10</td>
<td>6.09</td>
<td>0.02</td>
</tr>
<tr>
<td>11</td>
<td>6.18</td>
<td>6.20</td>
<td>0.02</td>
<td>6.14</td>
<td>-0.04</td>
</tr>
<tr>
<td>12</td>
<td>6.60</td>
<td>6.51</td>
<td>-0.09</td>
<td>6.63</td>
<td>0.03</td>
</tr>
<tr>
<td>13</td>
<td>7.07</td>
<td>7.10</td>
<td>0.03</td>
<td>7.05</td>
<td>-0.02</td>
</tr>
<tr>
<td>14</td>
<td>7.35</td>
<td>7.35</td>
<td>0.00</td>
<td>7.37</td>
<td>0.02</td>
</tr>
<tr>
<td>15</td>
<td>6.90</td>
<td>6.82</td>
<td>-0.08</td>
<td>6.89</td>
<td>-0.01</td>
</tr>
<tr>
<td>16</td>
<td>6.72</td>
<td>6.68</td>
<td>-0.04</td>
<td>6.74</td>
<td>0.02</td>
</tr>
<tr>
<td>17</td>
<td>6.60</td>
<td>6.72</td>
<td>0.12</td>
<td>6.67</td>
<td>0.07</td>
</tr>
<tr>
<td>18</td>
<td>6.77</td>
<td>6.85</td>
<td>0.08</td>
<td>6.70</td>
<td>-0.07</td>
</tr>
<tr>
<td>19</td>
<td>6.33</td>
<td>6.43</td>
<td>0.10</td>
<td>6.48</td>
<td>0.15</td>
</tr>
<tr>
<td>20</td>
<td>7.07</td>
<td>7.25</td>
<td>0.18</td>
<td>7.15</td>
<td>0.08</td>
</tr>
</tbody>
</table>
The electronegative groups surrounding the phenyl rings of the phenoxypropanamine and phenylmethanamine side chains are related to improved potency. In addition, the favorable steric contours suggest that aromatic bulky groups at the diphenylethanamine side chain may increase ligand potency. The understanding of the molecular determinants for ligand binding and affinity is critical in drug design. In cases where the crystal structure of a target protein in complex with a ligand is available, the strategy of combining the information of the 3D contour maps with the protein structure environment can be useful to investigate the role of the intermolecular interactions related to biological activity. Fig. (2) displays an example of this integrated approach, where it can be visualized the predicted binding interactions in the formation of the complex between the most potent LXR ligand of the series (14) and the LXR ligand-binding site. The 3D QSAR models generated in this study are compatible with the 3D protein environment in the LXR binding site, as also depicted in Fig. (4). The importance of the electronegative substituents in the phenylmethanamine moiety corroborates with the interaction between this group and a positively charged residue (His435) [22]. Furthermore, the role of the hydrophobic interactions with Ile309 and Met312 side chains is also in good agreement with previous results [22, 27]. This suggests that these models can be useful in the structure-based drug design of novel structurally related LXR ligands.

HQSAR Analyses

The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). In order to assess the process of hologram generation, several combinations of these parameters were considered during the QSAR modeling runs. The HQSAR analyses were performed by screening the 12 default series of hologram length values ranging from 53 to 401 bins. The patterns of fragment counts from the data set compounds were related to the experimental biological activity using the PLS analysis. The statistical results from the PLS analyses using several fragment distinction combinations and the fragment size default (4-7) are summarized in Table 4.

According to Table 4, the best statistical results among all models were obtained for model 5 ($r^2 = 0.99$ and $q^2 = 0.74$). This model was derived using a combination of A, B, C and Ch, with 6 being the optimum number of PLS components. The influence of different fragment sizes in the statistical parameters was further investigated for this HQSAR model. Fragment size parameters control the minimum and maximum length of fragments to be included in the hologram fingerprint, and can be varied to incorporate larger or smaller fragments [28, 29]. The HQSAR results for a variety of different fragment sizes are shown in Table 5. The variation of fragment size provided substantial improvements in the statistical parameters, as can be observed for the model with fragment size 3-6, which presents high correlation coefficients ($r^2 = 0.99$ and $q^2 = 0.85$) associated to low standard errors.

In terms of validation of a QSAR model, a measure of internal consistency is available in the form of $q^2$. However, the most important test of a QSAR model is its ability to predict the property value for new compounds. As the structure encoded in a 2D fingerprint is directly related to the biological activity of series of molecules, HQSAR models should be able to predict the activity of new structurally re-
lated LXR ligands from their fingerprints. Thus, the predictive power of the best HQSAR model derived using the training set molecules (fragment distinction A/B/C/Ch; fragment size 3-6, Table 5) was assessed by predicting pEC$_{50}$ values for the test set compounds (17-20, Table 1). The external validation process can be considered the most valuable validation method, as these compounds are completely excluded during the training of the model. For simplification, the results are also listed in Table 3 and the graphic results for the experimental versus predicted activities of both training and test sets are displayed in Fig. (5).

**Fig. (3).** CoMFA steric (green and yellow) and electrostatic (red and blue) contour maps. The most potent LXR modulator (14) of the series is shown in the background.

**Fig. (4).** Binding mode predicted for the most potent modulator (14, colored in pink) of the data set into the crystal structure of LXR. Important amino acid residues of the ligand-binding site are represented in green.
Table 4. HQSAR Analyses for Various Fragment Distinction on the Key Statistical Parameters Using Fragment Size Default (4-7)

<table>
<thead>
<tr>
<th>Model</th>
<th>Fragment Distinction</th>
<th>$r^2$</th>
<th>SEE</th>
<th>$q^2_{LOO}$</th>
<th>SEP</th>
<th>HL</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A/B/C</td>
<td>0.97</td>
<td>0.08</td>
<td>0.70</td>
<td>0.25</td>
<td>151</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>A/B/C/H</td>
<td>0.87</td>
<td>0.15</td>
<td>0.55</td>
<td>0.27</td>
<td>257</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>A/B/C/H/Ch</td>
<td>0.96</td>
<td>0.09</td>
<td>0.65</td>
<td>0.27</td>
<td>71</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>A/B/C/H/Ch/DA</td>
<td>0.97</td>
<td>0.07</td>
<td>0.56</td>
<td>0.30</td>
<td>53</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>A/B/C/Ch</td>
<td>0.99</td>
<td>0.04</td>
<td>0.74</td>
<td>0.23</td>
<td>59</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>A/B/C/Ch/DA</td>
<td>0.98</td>
<td>0.06</td>
<td>0.71</td>
<td>0.24</td>
<td>61</td>
<td>6</td>
</tr>
</tbody>
</table>

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

Table 5. HQSAR Analyses for the Influence of Various Fragment Sizes on the Key Statistical Parameters Using the Best Fragment Distinction (A, B, C and Ch)

<table>
<thead>
<tr>
<th>Fragment Size</th>
<th>$r^2$</th>
<th>SEE</th>
<th>$q^2_{LOO}$</th>
<th>SEP</th>
<th>HL</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-5</td>
<td>0.98</td>
<td>0.06</td>
<td>0.70</td>
<td>0.25</td>
<td>61</td>
<td>6</td>
</tr>
<tr>
<td>3-6</td>
<td>0.99</td>
<td>0.05</td>
<td>0.85</td>
<td>0.18</td>
<td>53</td>
<td>6</td>
</tr>
<tr>
<td>4-7</td>
<td>0.99</td>
<td>0.04</td>
<td>0.74</td>
<td>0.23</td>
<td>59</td>
<td>6</td>
</tr>
<tr>
<td>5-8</td>
<td>0.99</td>
<td>0.04</td>
<td>0.64</td>
<td>0.27</td>
<td>59</td>
<td>6</td>
</tr>
<tr>
<td>6-9</td>
<td>0.99</td>
<td>0.03</td>
<td>0.60</td>
<td>0.29</td>
<td>401</td>
<td>6</td>
</tr>
</tbody>
</table>

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

The results show that the test set compounds, which represent the different structural properties incorporated within the training set, are well predicted without any outliers. From the low residual values, it can be seen that the HQSAR model obtained is highly reliable and can be used in further medicinal chemistry studies. The predicted values fall very close to the experimental pEC$_{50}$ values, deviating by less than 0.15 log units.

Besides predicting the potency of LXR ligands, an HQSAR analysis provides important hints about what molecular fragments are directly related to biological activity. In the HQSAR method, it is possible to visualize the individual contributions to activity of each atom in a given molecule of the data set through the generation of contribution maps. The HQSAR module implemented in SYBYL 8.0 uses a color code in order to discriminate the main atomic contributions to activity. The colors at the red end of the spectrum (red, red orange and orange) reflect poor contributions, whereas colors at the green end (yellow, green blue and green) reflect favorable contributions. Atoms with intermediate contributions are colored white. The individual atomic contributions for the most potent compound of the data set (14) are presented in Fig. (6). In particular, the molecular fragments represented by the phenyl ring of the phenoxypyramamine side chain and the chloro group attached to the aromatic ring at the ortho-position, are indicated to be strongly related to biological potency. Moreover, in any molecule, regions with intermediate or poor contributions can be identified as potential targets for molecular modification and further SAR studies. In this case, the main region that negatively contributes for biological activity is the methyl group linked to the tertiary amine, which could be replaced by another moiety with different spatial and physicochemical features in order to increase affinity and potency.

![Fig. (5)](image-url) Plot of predicted versus experimental values of pEC$_{50}$ for all LXR ligands (training and test sets) by using the HQSAR method.
CONCLUSIONS

The 3D and 2D QSAR statistical models described in this work show both good internal and external consistency, and represent important contribution to the QSAR field in the area of NRs. The good correlation between experimental and predicted pEC \textsubscript{50} values for the test set compounds further proved the reliability of the constructed QSAR models. It is worth noting that we have employed the same training and test sets for all QSAR analyses, and the results showed that investigations can be carried out concomitantly to search for synergies between 3D and 2D QSAR technologies. Moreover, the 3D QSAR models generated are compatible with the 3D protein environment in the LXR binding-site. From the CoMFA steric and electrostatic contour maps we can conclude that electronegative groups surrounding the phenyl rings of the phenoxypropanamine and phenylmethanamine side chains are related to improved potency. In addition, the favorable steric contours suggest that aromatic bulky groups at the diphenylethanimine side chain may increase ligand potency. Regarding the HQSAR results, it was possible to obtain important hints about what molecular fragments are directly related to biological activity, i.e. the molecular fragments represented by the phenyl ring of the phenoxypropanamine side chain and the chloro group attached to the aromatic ring at the orto-position, are strongly related to biological potency. Besides, the main region that negatively contributes for biological activity is the methyl group linked to a tertiary amine, which could be replaced by another moiety with different spatial and physicochemical features in order to increase potency. The CoMFA and HQSAR models and the information obtained from the 3D contour and 2D contribution maps should be useful for the design of new structurally related LXR activators having improved affinity and potency. The integration of 3D and 2D QSAR methods is a powerful tool in medicinal chemistry and drug design studies.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from FAPESP (The State of São Paulo Research Foundation) and CNPq (The National Council for Scientific and Technological Development), Brazil.

REFERENCES

[19] Spencer, T.A.; Li, D.; Russel, J.S.; Collins, J.L.; Bledsoe, R.K.; Consler, T.G.; Moore, L.B.; Galardi, C.M.; McKee, D.D.; Moore,


