

Elaine Fontes Ferreira da Cunha ·
Rita Cristina Azevedo Martins ·
Magaly Girão Albuquerque ·
Ricardo Bicca de Alencastro

LIV-3D-QSAR model for estrogen receptor ligands

Received: 12 January 2004 / Accepted: 27 May 2004 / Published online: 31 July 2004
© Springer-Verlag 2004

Abstract We have employed the *Local Intersection Volume* (LIV), a three-dimensional (3D) local shape descriptor, to study quantitative structure–activity relationships (QSAR) of 2-arylbenzothiophene analogs of raloxifene (**1**), a selective estrogen-receptor modulator (SERM), using the raloxifene bound conformation (PDB code: 1ERR) to build the structures of all ligands. The best LIV-3D-QSAR model obtained by a combined GA-PLS optimization, **Model 1**, was derived from RMS three-atom alignment using a training set of 44 compounds. The fit of the pIC_{50} values, expressed by the squared correlation coefficient, R^2 , was 0.78. After LOO-cv, a predictive squared correlation coefficient, Q^2 , of 0.63 was obtained. **Model 1** has three LIVs (1091, 1554, and 1654) with positive coefficients and three LIVs (597, 1463, and 1655) with negative coefficients. Four descriptors (LIVs) show excellent correspondence with pharmacophoric groups of the raloxifene series of compounds in accordance with SAR studies. Most interesting is the result of the prediction for 14 compounds (test set) used for external validation. The results provide the tools for predicting the pIC_{50} values of related compounds and for the design and synthesis of new ER ligands.

Keywords 3D-QSAR · Benzothiophene · Estrogen receptor · Local intersection volume · Raloxifene

Introduction

Raloxifene (**1**) (Fig. 1) acts as an estrogen-receptor (ER) antagonist in mammary tissue and the uterus, but also mimics the agonist effects of estrogen on bone and in the cardiovascular system. [1, 2] Because of its unique profile of tissue specificity, raloxifene (**1**) is considered to be a selective estrogen receptor modulator (SERM), [3] a class of compounds that can act as ER antagonists in some tissues while acting as ER agonists in others. SERMs are being evaluated and used to treat and prevent such diseases as breast cancer, osteoporosis, and cardiovascular disease. [4] The US Food and Drug Administration (FDA) recently approved raloxifene (**1**), under the trade name Evista, for the treatment and prevention of osteoporosis in postmenopausal women. [5]

The raloxifene (**1**) molecular structure has four important characteristics, labeled as **I-IV**, responsible for its activity, as shown in Fig. 1. The two phenolic hydroxy groups are labeled **Ia** and **Ib**; the basic aliphatic amine chain, **II**; the benzothiophene aromatic ring, **III**; and the carbonyl “hinge” between the side chain containing the basic aliphatic amine and the benzothiophene aromatic ring, **IV**. [2] These structural elements give raloxifene (**1**) a distinct molecular conformation, which may affect the

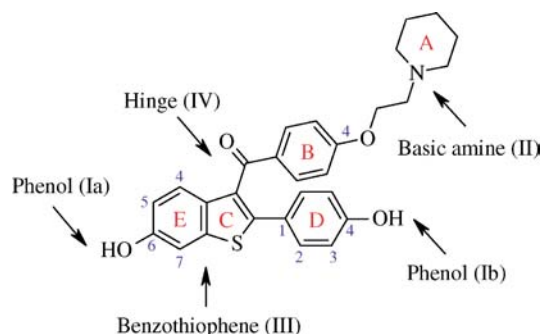


Fig. 1 Structure of raloxifene (**1**) and definition of the pharmacophoric groups (**Ia**, **Ib**, **II**, **III**, and **IV**) according SAR studies. [2, 6]

E. F. F. da Cunha · R. C. A. Martins · M. G. Albuquerque (✉) ·
R. B. de Alencastro
CCMN, Laboratório de Modelagem Molecular (LabMMol),
Departamento de Química Orgânica, Instituto de Química,
Universidade Federal do Rio de Janeiro,
Centro de Tecnologia, Bloco A, 6o andar, Sala 609, Ilha do Fundão,
Cidade Universitária, 21949-900 Rio de Janeiro, RJ, Brazil
e-mail: magaly@iq.ufrj.br
Tel.: +55-21-2562-7132
Fax: +55-21-2562-7132

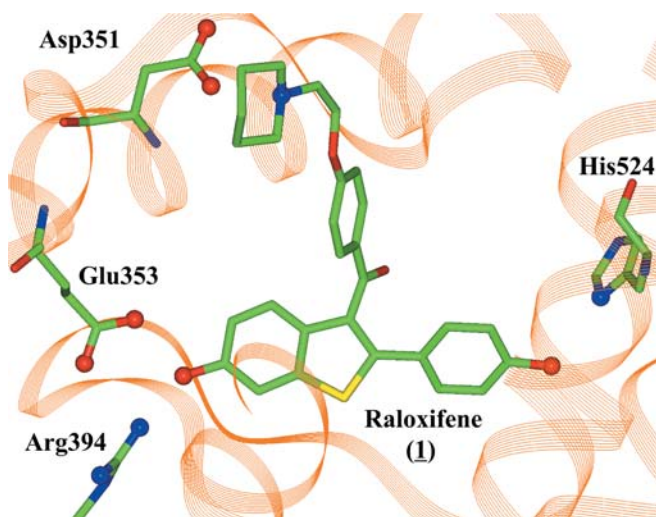


Fig. 2 Schematic representation of the interactions made by raloxifene (**1**) within the binding cavity from the crystallographic data. [7] Residues that make direct hydrogen bonds are depicted in ball-and-stick style. The observed hydrogen bonds are between nitrogen and oxygen atoms of raloxifene and the following amino acid residues: Asp351 (2.66 Å), Glu353 (2.41 Å), Arg394 (2.97 Å), and His524 (2.71 Å)

conformation of the raloxifene-ER complex, making it responsible for the unique tissue selectivity shown by this ER modulator. [6]

Brzozowski and co-workers [7] described the crystal structures of the ligand-binding domain (LBD) of ER α in complex with 17 β -estradiol (the endogenous estrogen) and raloxifene (**1**). The crystal structures show that raloxifene binds at the same site as estradiol within the LBD, with the two phenolic hydroxy groups (**Ia** and **Ib**) of the benzothiophene moiety (Fig. 2) mimicking the hydroxyl groups of the endogenous estrogen. [7]

Raloxifene (**1**) is anchored to the protein by direct hydrogen bonds between: (i) the piperidine nitrogen atom and Asp351 (N...O distance of 2.71 Å); (ii) the phenolic hydroxy group (**Ia**) of the benzothiophene ring and Glu353 (O...O distance of 2.38 Å) and Arg394 (O...N distance of 2.93 Å); and (iii) the phenolic hydroxy group (**Ib**) and the imidazole ring of His524 (O...N distance of 2.62 Å). [7] Moreover, the side chain of raloxifene makes extensive hydrophobic interactions, but it is too long (at over 11 Å in length) to be contained within the confines of the binding cavity. Therefore, one of its ends displaces helix H12 and protrudes from the pocket between helices H3 and H11. [7] Brzozowski and co-workers argue that the antagonistic response of raloxifene is based on its ability to prevent the formation of a raloxifene-ER α complex active conformation, which is dependent on H12. [7]

Asp351 in the LBD of ER α plays a pivotal role in regulating the estrogen-like activity of SERM-ER α complexes. Raloxifene acts as a full antagonist in MDA-MB-231 human breast cancer cells stably transfected with the wild-type ER α (Asp351). [8] However, the agonist activity of the raloxifene-ER α complex is enhanced with Glu351 and Tyr351, but lost with Phe351. [8] In addition,

a raloxifene derivative in which the piperidine ring is replaced by a cyclohexane ring is a potent agonist with the wild-type ER α . [8] According to these results, it is postulated that the side chain of raloxifene (basic piperidine ring) shields and neutralizes Asp351 to produce an anti-estrogenic ER α complex (inactive conformation), and the distance between the piperidine nitrogen of raloxifene and the negative charge of amino-acid-351 is critical for estrogen-like action. [8]

Based on the raloxifene (**1**) structure, Grese and co-workers [9] synthesized and evaluated a series of 2-arylbenzothiophene raloxifene analogs as ligands of the estrogen receptor (ER) and inhibitors of the MCF-7 breast cancer cell proliferation in vitro. In order to develop three-dimensional quantitative structure-activity relationship (3D-QSAR) models for estrogen receptor ligands, we have selected this series of 2-arylbenzothiophene raloxifene analogs as a case study.

Quantitative structure-activity relationships (QSAR) are a mathematical methodology, statistically validated, and mostly used to correlate experimental or calculated properties derived from chemical structures with biological activities. QSAR also may be applied to predict the activity values of non-synthesized compounds structurally related to training sets. With the advent of molecular modeling, three-dimensional (3D) descriptors have replaced the didactical physicochemical (experimental or calculated) and bi-dimensional (calculated) descriptors. [10]

Recently, we developed a new descriptor for 3D-QSAR named the local intersection volume (LIV), which can be classified as a local shape descriptor. [11, 12] The LIV is the intersection volume between molecule atoms and a set of carbon-atom sized spheres, composing a virtual three-dimensional box. [11, 12] The molecules are, then, after a previous alignment derived from the best steric-electrostatic fit according Good and co-workers' method, [13, 14] inserted in this box.

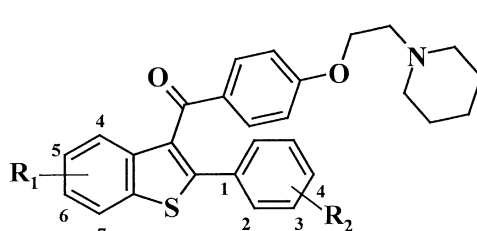
There is a strong analogy between the concept behind the LIV descriptor and the INVOL descriptor, the latter developed by Sulea and co-workers in 1997, [15] where the authors describe the van der Waals envelopes' intersection volumes as a steric potential field in a 3D-QSAR CoMFA study, and termed this intersection volumes field INVOL. [15]

In our work, we have developed a LIV-3D-QSAR model for estrogen-receptor ligands using a series of 2-arylbenzothiophene raloxifene analogs [9] as a case study. The compounds were modeled using the raloxifene (**1**) crystal structure complex with the LBD of ER α [7] as a template. The LIV-3D-QSAR model may be used to design new ER ligands.

Methods

Biological data

A data set of 55 raloxifene derivatives was taken from published results [9] and the 3D-QSAR model was developed using a training

Table 1 Structures and experimental IC₅₀ (nM)^a values for estrogen receptor antagonists in the data set (55 compounds) [9]


N	IC ₅₀ (nM)	R ₁				R ₂				
		4-	5-	6-	7-	2-	3-	4-	5-	
1	0.2			OH				OH		
2	0.3			OH			F	OH		
3	0.7			OH		Me				
4	0.8			OH				C≡CH		
5	1.0			OH			Me	OH		
6	1.0			OH				Cl		
7	2.0			OH		Me		OH		
8	2.0			OH		OMe		OH		
9	2.3			OH			Cl	OH		
10	2.3			OH				F		
11	2.5			OH						
12	2.5			OH			F			
13	3.0		F	OH				OH		
14	5.0			OH				Et		
15	7.0			OH				CH=CH ₂		
16	10.0			OH		OH				
17	10.0			OH				n-Bu		
18	20.0			OH				CONMe ₂		
19	20.0			C≡CH				OH		
20	30.0			OH				i-Pr		
21	30.0			CO ₂ Me				OH		
22	32.0			OH				COMe		
23	35.0							OH		
24	40.0			OH				CONHMe		
25	50.0			OH				Me		
26	50.0			OH				CO ₂ Me		
27	50.0			OH				CO ₂ Et		
28	60.0			COMe				OH		
29	100.0		Me	OH				OH		
30	100.0	Me		OH	Me			OH		
31	100.0							OMe		
32	100.0		OH					OH		
33	100.0			OH				Ph		
34	100.0			OH				CH ₂ SEt		
35	100.0			OH			Me	OH	Me	
36	190.0	OH						OH		
37	200.0			OH				CONH ₂		
38	250.0			OMe				OH		
39	300.0									
40	300.0			OMe				OMe		
41	300.0			Me				OH		
42	300.0				OH			OH		
43	325.0			OH				CO ₂ Et		
44	325.0		OMe	OMe	OMe			OMe		
45	350.0	OH		OH				OH		
46	400.0		OH	OH				OH		
47	500.0			OH				NO ₂		
48	500.0	-C H ₄ -		OH				OH		
49	500.0			OMe				-OCH ₂ O-		
50	500.0		Me	OH	Me			OH		
51	600.0			OMe				CH ₂ OH		
52	1000.0			OH				OMe		
53	1000.0			CONH ₂				OH		
54	1000.0			Cl				OH		
55	1000.0			OH				CF ₃		

^a The potency was defined as $\log(1/C)(pIC_{50})$, where C is the effective inhibitory concentration of the compound required to achieve 50% (IC₅₀) inhibition of MCF-7 cell proliferation. [9] Underlined numbers correspond to the 14 compounds of the test set

data set of 41 compounds (1–7, 9–13, 15, 16, 19, 21–25, 27, 28, 30, 31, 33, 35–40, 43–46, 48, 50, 51, and 53–55), randomly selected from the original 55 compounds. The model also was externally validated using a test data set of 14 compounds (8, 14, 17, 18, 20, 26, 29, 32, 34, 41, 42, 47, 49, and 52), randomly selected from the original 55 compounds. These compounds are structurally similar to the compounds in the training set. Table 1 reports the 55 compounds' structures (training set and test set) and the related potency, which is defined as IC₅₀ (nM), where C is the effective inhibitory concentration of compound required to achieve 50% (IC₅₀) inhibition of MCF-7 cell proliferation. [9] In addition, all pharmacological data were obtained from the same laboratory, eliminating all potential noise by the pooling of data sets from different sources.

Compounds model building

Three-dimensional (3D) structures of each of the 54 compounds (Table 1) were based on the structure of the compound 1 (raloxifene) co-crystallized with ER retrieved from the Protein Data Bank [16] (PDB code: 1ERR). [7] This structure is the conformation bound to the receptor-binding site or the active conformation of the compound 1. The 3D models for each compound in their neutral forms were constructed using the PC Spartan Pro v.1.0.5 software [17] based on the conformation of the bound structure of compound 1. Each structure, including raloxifene (1), was geometry-optimized using the PC Spartan Pro PM3 semiempirical method without any restriction in vacuum, and the partial atomic charges were assigned by the same method available on the AMPAC/MOPAC module from the *InsightII* software. [18]

Grid matrix assembling and volumes calculation

We constructed a grid matrix composed by 2,197 cubic unitary cells, where the vertices correspond to the Cartesian coordinates of the eight carbon atoms or hard spheres, and the vertices arrest lengths are 1.50 Å. [12] The volume for each sphere was calculated using the carbon van der Waals radii length of 1.54 Å×0.65, where 0.65 is the scale factor used to avoid a large overlap among the spheres, and, consequently, a minimal loss of volume among the hard spheres. [12] The volumes were calculated using the Search-Compare module of the *InsightII* program. [18]

Alignment rules

The overall alignment step was performed using the standard tools available in the Search-Compare module of the *InsightII* program. [18] The most active compound (1), the reference compound, was inserted in the grid matrix considering both centers of mass. All the other 40 compounds were superimposed to the reference compound by two different alignment procedures, using three atoms according to the raloxifene numbering as shown in Fig. 1: C6 (E-ring), C1 (D-ring) and C4 (B-ring). The three selected atoms are from the rigid framework (aromatic rings) and they are common for all compounds. The first alignment procedure was done by root mean squares (RMS) deviation using the three selected atoms. The second alignment procedure was done in two steps: a pre-alignment by RMS deviation (using the same three selected atoms) was used, followed by a subsequent alignment by steric (50%) and electrostatic (50%) fit. The similarity index of the steric and electrostatic fit is in accordance with the method proposed by Good and co-workers. [13, 14]

Molecular volume and LIV calculations

The molecular volumes for each compound were calculated using the van der Waals radii without scale factors. After the alignment step, the intersection volumes were calculated using the molecular volume of each compound and the volume of each hard sphere that composes the grid matrix. These intersection volumes were labeled local intersection volumes (LIV), and they represent the indepen-

dent variables (volume descriptors) of the database. The molecular volume and the LIV calculations were performed using the tools available in the Search-Compare module of the *InsightII* program. [18]

Data reduction

In order to exclude data noise from both databases, generated by the RMS alignment and the combined RMS/steric-electrostatic fit alignment, named alignment-1 and alignment-2, respectively, we have done two types of data reduction (A and B) generating four databases, named DB-1A, DB-1B, DB-2A, and DB-2B. Databases A were constructed excluding the variables where LIV is equal to zero for all molecules, and databases B were constructed excluding the variables where LIV is different from zero in three or less than three molecules. In DBs-A, we excluded useless variables and, in DBs-B, we harmonized the data, removing variables that were not properly represented in the data set. That meant we took into account the structural peculiarities of a few compounds. [11, 12]

LIV-3D-QSAR models calculation

The QSAR analysis relates chemical descriptors (x) to a response variable (y) by a mathematical equation, most often in the form of Eq. (1):

$$y = f(x) = c + \sum a_{ij}B_j(x_i) \quad (1)$$

where c is a constant, a is the (regression) coefficient, i ranges over the set of descriptors, and the j th basis function type varies depending on the i th descriptor. The function (operator) B may be, for example, an algorithmic or quadratic transform, a half-space spline, or unity. [19]

The models were calculated using a combined Genetic Algorithm (GA) and Partial Least Square (PLS) [20, 21] approach, implemented in the Wolf 6.2 program. [22] The GA-PLS optimizations were applied for the four databases (DB-1A, DB-1B, DB-2A, and DB-2B) using 300 randomly generated models (functions or equations), each model depending initially on four basis functions, where each is composed by one independent variable. Only linear basis functions were used. Mutation probability over the crossover optimization was set to a rate of 100% at each ten-crossover operation. Since the training set is composed of 41 compounds, in order to have no more than eight descriptors in each equation to avoid data over-fitting, the smoothing factor (the variable that controls the number of independent variables in the models) was set to 0.1. The maximal number of components for the PLS regression analysis was four, and 40,000-crossover operations were used. All other options were left in their default values.

The models were scored using Friedman's lack-of-fit (LOF) measure, [23] which is given by Eq. (2):

$$\text{LOF} = \text{LSE} / \{1 - (c + d \cdot p) / m\}^2 \quad (2)$$

where LSE is the least-squares error (calculated from the difference between actual and calculated values for the activity index over the data set), c is the number of basis functions in the model, d is the smoothing factor, p is the total number of variables contained in all basis functions, and m is the number of samples (compounds) in the training set. The LOF measure penalizes appropriately for the addition of terms to the equation (and consequent loss of degrees of freedom) in such a way to resist over-fitting.

Model validation

(a) Cross-validation (internal validation): the ten best 3D-QSAR models, as scored by the LOF measure from the GA-PLS analysis, were evaluated by the leave-one-out cross-validation (LOO-cv) procedure using the *Wolf* 6.2 program. [21]

(b) External validation: the best model was applied to the test data set of 14 compounds that were not included in developing the LIV-3D-QSAR models. This step was performed only for the best model.

Results

The GA-PLS analysis for DB-1A, DB-1B, DB-2A, and DB-2B generated several models or equations. However, DB-1B was the only one that gave a model with a squared coefficient of linear correlation after cross-validation (Q^2) higher than 0.5. According to these results, we will only present the analysis of the best model derived from DB-1B, Model 1 (Eq. 3). It is interesting to note that the best model was derived from alignment-1 (the RMS alignment without the steric-electrostatic fit) using data reduction procedure B. This suggests that the RMS alignment, in this particular case, is superior to alignment-2 (the RMS/steric-electrostatic fit alignment), probably because there are specific interactions governing the ligand-receptor binding that might change after the steric-electrostatic fit alignment.

In fact, the crystallographic structure of the raloxifene-ER α complex [7] shows at least four hydrogen bonds (Fig. 2). However, we need to be careful with this hypothesis because the visual observation of the superimposition after both alignment procedures gives similar results. In addition, we may conclude, as was to be expected, that data reduction B is superior to data reduction A because data reduction A excludes useless variables and, data reduction B harmonizes the data, removing variables that were not properly represented in the data set.

According Eq. (3), the best model, **Model 1**, is composed by six LIV descriptors (independent variables), with both positive and negative contributions to the activity.

Model 1

$$\begin{aligned} \text{pIC}_{50} = & 16.242 - 4.130(\text{LIV}_{597}) + 10.504(\text{LIV}_{1091}) \\ & - 2.234(\text{LIV}_{1463}) + 0.347(\text{LIV}_{1554}) \\ & + 9.277(\text{LIV}_{1654}) - 15.646(\text{LIV}_{1655}) \end{aligned}$$

$$(N = 41 \quad R^2 = 0.76 \quad Q^2 = 0.68 \quad \text{SD} = 0.61) \quad (3)$$

Model 1 represents the best LIV-3D-QSAR equation derived from 41 compounds of the training set. The conventional squared linear correlation coefficient, R^2 , of **Model 1** is equal to 0.76. This means that the analyzed results have a small fitness compared to the biological in vitro test results. Table 2 shows the calculated pIC_{50} values for the training set. The statistical significance of the relationship between the biological response and the chemical structure descriptors was further demonstrated by a cross-validation analysis. Leave-one-out cross-validation (LOO-cv) analysis of **Model 1** had a Q^2 value of 0.68 with standard error of 0.61. This means that **Model 1** has a predictive capacity of 68%. LOO-cv correlation coefficient values over 0.5 reveals that the model is a

Table 2 Experimental [9] and calculated pIC₅₀ values for the estrogen receptor antagonists in the data set (55 compounds)

N ^a	pIC ₅₀ exp	pIC ₅₀ calc	Residue	N ^a	pIC ₅₀ exp	pIC ₅₀ calc	Residue
<u>1</u>	9.69	8.30	-1.39	<u>29</u>	7.00	7.60	0.60
<u>2</u>	9.52	9.36	-0.16	<u>30</u>	7.00	6.92	-0.08
<u>3</u>	9.15	9.55	0.40	<u>31</u>	7.00	6.93	-0.07
<u>4</u>	9.09	8.05	-1.04	<u>32</u>	7.00	7.53	0.53
<u>5</u>	9.00	8.65	-0.35	<u>33</u>	7.00	7.04	0.04
<u>6</u>	9.00	9.42	0.42	<u>34</u>	7.00	7.80	0.80
<u>7</u>	8.69	8.57	-0.12	<u>35</u>	7.00	7.56	0.56
<u>8</u>	8.69	8.80	0.11	<u>36</u>	6.72	6.75	0.03
<u>9</u>	8.63	8.68	0.05	<u>37</u>	6.69	7.27	0.58
<u>10</u>	8.63	8.85	0.22	<u>38</u>	6.60	6.41	-0.19
<u>11</u>	8.60	7.55	-1.05	<u>39</u>	6.52	6.22	-0.30
<u>12</u>	8.60	8.00	-0.60	<u>40</u>	6.52	6.73	0.21
<u>13</u>	8.52	7.60	-0.92	<u>41</u>	6.52	7.42	0.90
<u>14</u>	8.30	8.03	-0.27	<u>42</u>	6.52	6.79	0.27
<u>15</u>	8.15	7.26	-0.89	<u>43</u>	6.48	7.44	0.96
<u>16</u>	8.00	7.96	-0.04	<u>44</u>	6.45	6.50	0.05
<u>17</u>	8.00	4.68	-3.32	<u>45</u>	6.45	6.52	0.07
<u>18</u>	7.69	8.40	0.71	<u>46</u>	6.39	6.94	0.55
<u>19</u>	7.69	7.73	0.04	<u>47</u>	6.30	7.25	0.95
<u>20</u>	7.52	7.78	0.26	<u>48</u>	6.30	6.29	-0.01
<u>21</u>	7.52	7.11	-0.41	<u>49</u>	6.30	5.41	-0.88
<u>22</u>	7.49	8.22	0.73	<u>50</u>	6.30	6.80	0.50
<u>23</u>	7.45	8.24	0.79	<u>51</u>	6.22	5.91	-0.31
<u>24</u>	7.39	7.97	0.58	<u>52</u>	6.00	7.42	1.42
<u>25</u>	7.30	7.10	-0.20	<u>53</u>	6.00	6.04	0.04
<u>26</u>	7.30	6.39	-0.91	<u>54</u>	6.00	6.38	0.38
<u>27</u>	7.30	7.43	0.13	<u>55</u>	6.00	6.40	0.40
<u>28</u>	7.22	7.58	0.36				

^a Underlined compounds' numbers correspond to the test set and residue values in bold correspond to outlier compound

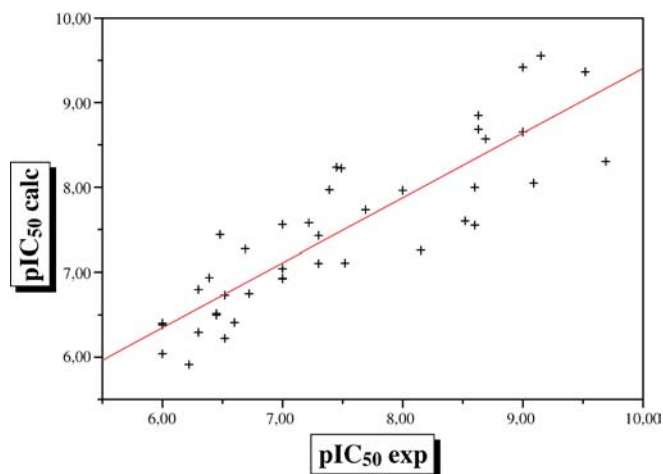


Fig. 3 Plot of the experimental [9] versus calculated pIC₅₀ values of the training set (41 compounds). The line represents the perfect correlation

useful tool for predicting affinities for new compounds in this set. Therefore, an external validation was performed with the test set compounds. Table 2 shows the predicted pIC₅₀ values for the test set compounds.

Figure 3 shows the experimental versus the calculated pIC₅₀ values for the training data set (41 compounds). The standard deviation (SD) of the residual values is 0.53, and there are not outlier compounds.

Discussion

Graphical representation of Model 1

The graphic representation of LIVs from **Model 1** is shown in Fig. 4, using the most active compound (**1**, raloxifene) as a template. LIVs are represented in their maximum

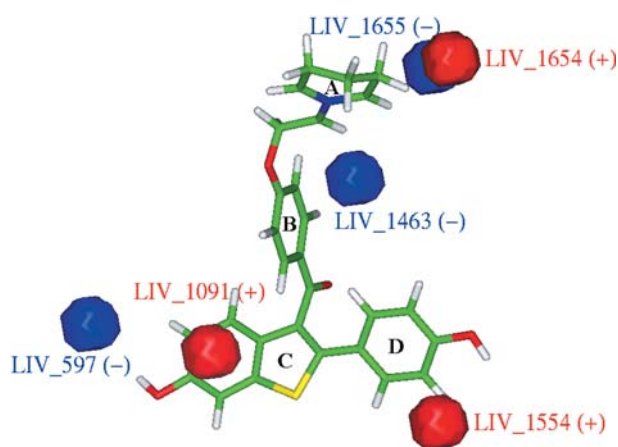


Fig. 4 Graphic representation of **Model 1**, the best LIV-3D-QSAR pharmacophoric model for ligands of the estrogen receptor proposed in this work, using compound **1** (raloxifene) as reference. LIVs with positive contribution (red): 1091, 1554, and 1654. LIVs with negative contribution (blue): 597, 1463, and 1655. LIVs descriptors are represented in their maximum size

size; therefore, Fig. 4 is not the representation of the LIV values for compound **1**.

LIVs 1091, 1554, and 1654, with positive coefficients, correspond to favorable interactions between the molecule structure and an amino-acid residue in the estrogen receptor. Therefore, substituents in these positions lead to increase potency. LIVs 597, 1463 and 1655, with negative coefficients, correspond to unfavorable interactions between the molecule structure and an amino-acid residue in the estrogen receptor. Therefore, substituents in these positions lead to a decrease in potency. In Fig. 4, red solid volumes represent LIVs with a positive contribution and blue solid volumes, LIVs with a negative contribution.

Taking into account the pharmacophoric groups responsible for raloxifene (**1**) and related compounds activities from structure-activity relationship (SAR) studies, [2, 6] we found correlations between four LIVs (Fig. 4) of **Model 1** and the following pharmacophoric groups (Fig. 1): the phenol group (**Ia**) is related to LIV-597; the hinge (carbonyl group) (**IV**) is related to LIV-1463; and the amine group (**II**) is related to LIV-1654 and LIV-1655.

LIV-597 is located near the C6 position of the benzothiophene (E-ring) (Figs. 1 and 4), indicating that the hydroxyl substituent is preferred in this position, since bulky substituents reduce the potency. This is consistent with our LIV calculation in which we have greater LIV-597 values for compounds **38**, **40**, **48**, **50**, and **51** corresponding to methoxyl, methyl or phenyl group in the 5, 6 or 7 positions of the benzothiophene (E-ring). These compounds were well calculated and all of them have lower potency, in accordance with **Model 1**. This indicates that this area represents a specific hydrogen bond with the receptor, as revealed by the co-crystallized raloxifene-ER α structure, [7] corresponding specifically to the interaction with Glu353 residue [7] and the occupation will be detrimental to the activity.

LIV-1091 is located close to the C5 position of the benzothiophene (E-ring) (Figs. 1 and 4), but in front of the C5-H bond. Small substituents as hydrogen and fluorine at the C5 position are preferred. However larger substituents such as methoxyl and methyl reduce the potency, but they are not directly related to LIV-1091, which has a positive coefficient. This means that LIV-1091 is not directly correlated to any substituent, but it represents an intersection volume with benzothiophene (E-ring) (Fig. 1). The crystallographic structure of the raloxifene-ER α complex [7] shows that the C6-hydroxy group (**Ia**, Fig. 1) of the benzothiophene ring makes two hydrogen bonds, one with Glu353 and another with Arg394 (Fig. 2). There are also two amino-acid residues 4.5 Å around raloxifene C5-atom making hydrophobic interactions: Leu387, 3.94 Å distant from C5 and located above the E-ring, and Phe404, 4.32 Å distant from C5 and located under E-ring. Additionally, the carbonyl group of Phe404 backbone also makes a hydrogen bond with Arg394 (2.62 Å).

Therefore, the relative orientation of the benzothiophene, which is correlated to LIV-1091, is crucial for the

hydrogen-bonding network around the C6-hydroxy group. An incorrect alignment can promote different orientations of benzothiophene inside the binding pocket, disrupting hydrogen bonds with C6-hydroxy group. This effect may be clearly observed by comparing the activity among the compounds **44**, **46**, **48** and **50** (large substituents and low potency) and **13** (small substituent and high potency). This decrease in the potency caused by bulky substituents is probably due the steric hindrance of the hydroxyl group in the C6 position (E-ring). In fact, making a visual inspection of the superposition of the compounds that have higher LIV-1091 values, such as compounds **5** and **6**, we concluded that this effect is caused by the overall RMS alignment.

LIV-1463, located near the B-ring, is influenced indirectly by the substituents in position C4 of the benzothiophene (E-ring) and in position C2 of the D-ring (Figs. 1 and 4). Substituents in position C4 (E-ring) are detrimental to the activity because all compounds that have a substituent in this position have a greater LIV value. LIV-1463 was found for compounds **30**, **36**, **45** and **48**, which have lower potency. On the other hand, substituents at position C2 (D-ring) are favorable to the activity, because all compounds that have a substituent in this position do not have an intersection volume for LIV-1463, found for compounds **3**, **7** and **16**, which have higher potency.

LIV-1554, located near the C3 position of the D-ring (Figs. 1 and 4), indicates that substituents in this position are favorable to the activity, because compounds **2**, **5**, **9**, **12** and **35** have a greater LIV value in this position. In fact, compounds **2**, **5**, **9**, and **12** have a different substituent just in the C3 position (D-ring), as do the hydrophobic substituents fluorine, chlorine, and methyl. On the other hand, compound **35**, which has methyl groups at both C3 and C5 positions of the same ring, is 100 times less potent than these compounds, probably due to another effect.

LIV-1654 and LIV-1655 are located around the piperidine group (A-ring) (Figs. 1 and 4), indicating that the relative orientation of the amine group is important for the increase and decrease of potency, respectively, since LIV-1654 has a positive coefficient and LIV-1655 has a negative one. This region of the compounds is constant but has great conformational freedom, being modified mostly according to the alignment step. On the co-crystallized raloxifene-ER α structure, [7] the piperidine group interacts with Asp351 by hydrogen bond or ionic interaction (Fig. 2). However, the raloxifene (**1**) optimized conformation is different from the co-crystallized structure, mostly in this region, where the piperidine group is orientated to the receptor outside.

Actually, in the raloxifene-ER α complex, [7] the piperidine group is close to the borderline of the receptor-site entrance, composed of the Asp351, Leu539, Pro535, and Tyr526 residues (Fig. 5a and b). The N...O distances of piperidine/Asp351 and piperidine/Tyr526 are 2.66 Å and 8.05 Å, respectively, and there is a "vacant" space between the piperidine group and Tyr526 residue. There-

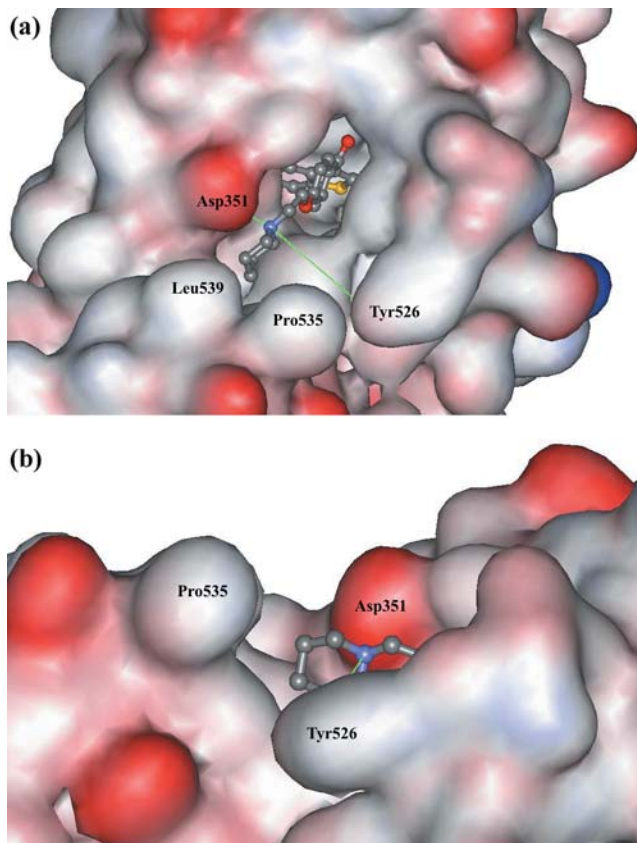
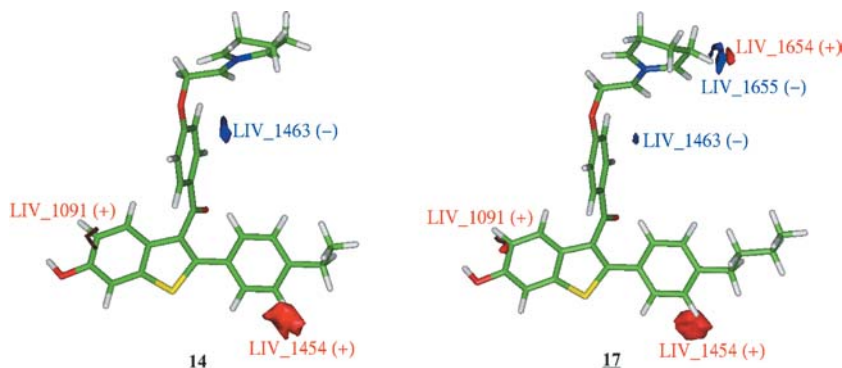


Fig. 5a–b Molecular electrostatic surface of ligand binding domain of ER in complex with raloxifene (**1**) showing the receptor site entrance, delimited by Asp351, Leu539, Pro535, and Tyr526 (PDB code 1ERR). [7] **a** Distances between piperidine nitrogen atom of raloxifene and Asp351 (2.66 Å) and Tyr526 (8.05 Å) are shown as green lines. **b** Close up view of the receptor entrance looking by the axis of line connecting Tyr526-N-Asp351. In this view, Leu539 is located behind Pro535

fore, the piperidine group can make an ionic interaction with Asp351, as in the 1ERR complex, [7] or it can have an alternative interaction with the phenol group of Tyr526 (cation- π or hydrogen bonding) if the amine-chain of raloxifene moves from the original position to closer Tyr526 (as in the minimized conformation). Of course, since the ionic interaction is stronger, the original conformation will be the preferential one.

Fig. 6 Graphical representation of LIV 3D-QSAR **Model 1** for compounds **14** and **17**. LIVs with positive contribution (red): 1091, 1454, and 1654. LIVs with negative contribution (blue): 1463 and 1655



External validation

The test set was used in order to test the predictive ability of our best model, **Model 1**, since the 14 compounds were not included in developing the LIV-3D-QSAR models. It was interesting to see how well the experimentally observed activity agrees with the predicted one.

Table 2 shows the actual activity values for the entire data set (training and test set), the calculated and predicted activity values for the training set (41 compounds) and the test set (14 compounds), respectively, according to **Model 1**, where the numbers of test-set compounds are underlined. Table 2 also gives the residual value, i.e., the experimental activity minus the calculated activity.

Model 1 produces accurate predictions for 13 of the 14 test-set compounds. The standard deviation (SD) of the residual values is 1.19, and the outlier compounds were defined as those compounds in which the residual value is higher than twice the SD value.

As we can see on Table 2, there is just one outlier compound for **Model 1**, namely compound **17** from the test set, with a residual value of -3.32 , meaning that the predicted activity ($\text{pIC}_{50}=4.68$) is lower than the experimental one ($\text{pIC}_{50}=8.00$). Compound **17** is a raloxifene (**1**) analogue where the hydroxyl group at C4 (D-ring) is replaced by an *n*-butyl group, decreasing the potency by 1.69 log units. The experimental activity of **17** is similar to **14** ($\text{pIC}_{50}=8.30$) where the hydroxyl group at C4 (D-ring) is replaced by an ethyl group, but different to **17**, the potency of compound **14** is well predicted, with a residual value of -0.27 . Therefore, compound **17** is an unexpected outlier.

In order to understand this behavior, we superimposed compounds **14** and **17**, using the same three-atom alignment procedure, resulting in an almost perfect superimposition, differing slightly on the piperidine ring. Consequently, we may assume that the overall alignment was not responsible for this divergence, which is better explained comparing the LIV values of compound **14**, which does not have an intersection volume for LIV-1654 and LIV-1655, to **17** with intersection volumes for both LIVs. The intersection volume of LIV-1655, with the largest and negative coefficient (**Model 1**), is twice the intersection volume of LIV-1654 with a positive coefficient (Fig. 4). Fig. 6 shows the graphic representation of the selected LIVs for compounds **14** and **17**.

Proposal of novel ER ligands

In the present study, a 3D-QSAR method using the local intersection volume (LIV) as descriptor was used to construct LIV-3D-QSAR models using 41 estrogen receptor ligands. The best model was employed to predict the potency of 14 ligands. The raloxifene-based alignment generated reflects a mutual superposition of the ligands along with their relative orientation towards the binding pocket. In this way, information about the geometry of the binding pocket is included in the derived LIV-3D-QSAR model.

Overall, the best model, **Model 1**, comprises important features that may be applied in the development of new estrogen receptor modulators. Therefore, we propose analogs of raloxifene (**1**) with simultaneous substituents at positions C2 (C6) and C3 (C5) of the D-ring as novel ER ligands, in order to explore the region highlighted by LIV-1554 (Fig. 4). These substituents could be methyl, methoxyl, fluorine, and chlorine groups, which showed good results in the raloxifene series of compounds studied. Additionally, we propose to reduce the length of the alkyl chain or change the whole 4-(1-piperidinoethoxy)phenyl group by less flexible groups like 4-(4-piperidine)-phenyl or 4-(1-piperazine)-phenyl groups. This proposal is because the alkyl group bonded to the piperidine group has higher conformational freedom and the relative orientation of the amine group may change the potency of these compounds.

Acknowledgments We thank the Brazilian agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) for their support.

References

- Jones CD, Jevnikar MG, Pike AJ, Peters MK, Black LJ, Thompson AR, Falcone JF, Clemens JA (1984) *J Med Chem* 27:1057–1066
- Grese TA, Sluka JP, Bryant HU, Cullinan GJ, Glasebrook AL, Jones CD, Matsumoto K, Palkowitz AD, Sato M, Termine JD, Winter MA, Yang NN, Dodge JA (1997) *Proc Natl Acad Sci USA* 94:14105–14110
- Heringa M (2003) *Int J Clin Pharmacol Ther* 41:331–345
- Morello KC, Wurz GT, DeGregorio MW (2002) *Crit Rev Oncol Hematol* 43:63–76
- Sarrel PM, Nawaz H, Chan W, Fuchs M, Katz DL (2003) *Am J Obstet Gynecol* 188:304–309
- Grese TA, Sluka JP, Bryant HU, Cole HW, Kim JR, Magee DE, Rowley ER, Sato M (1996) *Bioorg Med Chem Lett* 6:903–908
- Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engström O, Ohman L, Greene GL, Gustafsson JA, Carlquist M (1997) *Nature* 389:753–758
- Liu H, Park WC, Bentrem DJ, McKian KP, Reyes Ade L, Loweth JA, Schafer JM, Zapf JW, Jordan VC (2002) *J Biol Chem* 277:9189–9198
- Grese TA, Cho S, Finley DR, Godfrey AG, Jones CD, Lugar CW, Martin MJ, Matsumoto K, Pennington LD, Winter MA, Adrian MD, Cole HW, Magee DE, Phillips DL, Rowley ER, Short LL, Glasebrook AL, Bryant HU (1997) *J Med Chem* 40:146–167
- Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C (1997) *J Am Chem Soc* 119:10509–10524
- Verli H, Albuquerque MG, Alencastro RB, Barreiro EJ (2002) *Eur J Med Chem* 37:219–229
- Martins RCA, Albuquerque MG, Alencastro RB (2002) *J Braz Chem Soc* 13:816–821
- Good AC, Hodgking EE, Richards WG (1992) *J Chem Inf Comput Sci* 32:188–191
- Good AC (1992) *J Mol Graph* 10:144–151
- Sulea T, Oprea TI, Muresan S, Chan SL (1997) *J Chem Inf Comput Sci* 37:1162–1170
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) *Nucleic Acids Res* 28:235–242
- Hehre WJ, Deppmeier BJ, Klunzinger PE (1999) PC Spartan Pro. Wavefunction Inc, Irvine, USA
- Insight II (1995) Biosym, Molecular Simulations Inc, Accelrys, San Diego, USA
- Kubinyi H (1995) Burger's medicinal chemistry and drug discovery, 5th edn. In: Wolff ME (ed) Principles and practice, vol 1. New York
- Dunn III WJ, Rogers, D (1996) In: Devillers J (ed) Genetic algorithms in molecular modeling. Academic Press, London, pp 109–130
- Rogers D, Hopfinger AJ (1994) *J Chem Inf Comput Sci* 34:854–866
- Rogers D, Molecular Simulation Inc (1994) Wolf v.6.2, Genetic Function Approximation, USA
- Karki RG, Kulkarni VM (2001) *Bioorg Med Chem* 9:3153–3160