

Original article

LIV-3D-QSAR models for PGI₂ receptor ligands using multiple conformations

Rita Cristina Azevedo Martins, Magaly Girão Albuquerque *, Ricardo Bicca de Alencastro

Laboratório de Modelagem Molecular (LabMMol), Departamento de Química Orgânica, Instituto de Química, Universidade Federal do Rio de Janeiro, Cidade Universitária, CT, Bloco A, Sala 609, Rio de Janeiro, RJ 21949-900, Brasil

Received 3 November 2003; received in revised form 11 February 2004; accepted 12 February 2004

Abstract

A new 3D descriptor, the local intersection volume (LIV), was developed by our group and applied to the construction of 3D-QSAR models for ligands of the PGI₂ receptor (IP). The target compounds are a set of 42 aromatic heterocyclic derivatives [Meanwell et al., J. Med. Chem. 36 (1993), 3884], which show agonist activities in the IP receptor and are inhibitors of platelet aggregation. The LIV-3D-QSAR models were obtained through the analysis of 30% of the generated conformations for each compound, using a combined Genetic Algorithm (GA) and Partial Least Square (PLS) approach [Rogers and Hopfinger, J. Inf. Comput. Sci. 34 (1994) 854]. Statistically, Model 3 is the best as well as the most comprehensive in a mechanistic sense. Furthermore, it can be applied to design new IP ligands.

© 2004 Elsevier SAS. All rights reserved.

Keywords: Prostacyclin I₂ receptor; 3D-QSAR; Local Intersection Volume (LIV)**1. Introduction**

A new 3D descriptor, the local intersection volume (LIV), was developed by our group and applied to the construction of 3D-QSAR models [1,2]. The LIV can be classified as a 3D local shape descriptor. The LIV is the intersection volume between the molecular volume of each compound and the volume of each hard sphere of defined size, which composes a three-dimensional virtual grid matrix in analogy to the Grid [3], CoMFA [4], and 4D-QSAR methods [5]. The descriptors in the Grid and CoMFA methods are the van der Waals steric (Lennard-Jones) and electrostatic (Coulomb) interaction energies calculated with a probe atom over a 3D molecular grid surface. The descriptors in the 4D-QSAR method, in which the fourth dimension (4D) corresponds to the conformational sample as a time function, are the frequency of atomic occupancy of a grid matrix on the 3D space. The 3D virtual grid matrix mimics the space in the receptor (or enzyme) active site where the molecules are inserted according a pre-defined alignment.

The LIV descriptor was first applied [1] in a set of rigid compounds, using a grid matrix composed by 206 hard

spheres where five spheres correspond to a quadrangular base pyramid shaped unitary cell with arrest lengths of 3.08 Å, and the volume of each sphere was calculated using a radii length of 1.54 Å. Therefore, we composed a grid matrix of hard spheres where the volumes of each sphere do not overlap each other.

In the subsequent application of the LIV methodology [2] in a set of compounds with higher conformational freedom, we proposed modifications on the grid matrix shape and sphere size in order to have a more refined analysis. In this application, we used a grid matrix composed by 2197 hard spheres where eight spheres corresponding to a cubic shaped unitary cell with arrest lengths of 1.50 Å, and the volume of each sphere was calculated using a radii length of 1.00 Å; therefore, still avoiding an overlap among the spheres and a minimal loss of volume among them. In this present work, we proposed one more modification; the use of multiple conformations (e.g. 30%) in order to have a better representation for each compound, in place of just one conformation, as in our previous work [2].

Prostacyclin I₂ (PGI₂) (1) (Figure 1) is an potent vasodilator and inhibitor of platelet aggregation [6,7], interacting with its specific receptor (IP), located on cellular membranes, and stimulating the adenylate cyclase [7,8]. PGI₂ is fundamental in vascular homeostasis, but it is not a useful therapeutic agent, since it has a half-life of approximately 3 min-

* Corresponding author.

E-mail address: magaly@iq.ufrj.br (M.G. Albuquerque).

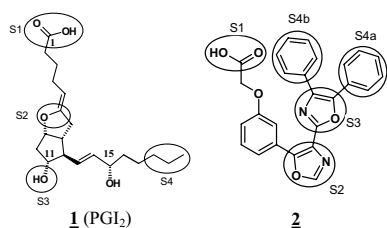


Fig. 1. Structures of prostacyclin I₂ (**1**) and compound **2** [9–13]. Pharmacophoric sites (S1, S2, S3, and S4) definition according SAR studies [9–14].

utes in physiological conditions, mainly due a labile enol-ether function, being, thus, of limited clinical usage [8]. In order to develop antithrombotic agents with a better pharmacokinetic profile, Meanwell and coworkers [9–13] synthesized and evaluated a series of aromatic heterocyclic derivatives (see compound **2**, Figure 1), which showed activity as IP receptor agonists and inhibitors of platelet aggregation.

Meanwell and coworkers [9–13] defined in structure-activity relationship (SAR) studies the correlation between PGI₂ [14] and nonprostanoid PGI₂ mimetics functional groups. They also identified the minimal structural features essential for effective PGI₂ mimicry [9]; that is, the pharmacophoric groups and their spatial disposition, proposing two-dimensional pharmacophore models [9–11]. Taking into account these studies, we labeled the PGI₂ (**1**) and the aromatic heterocyclic derivatives (e.g., compound **2**) pharmacophoric groups as S1–S4.

Site S1 corresponds, in both structures (compounds **1** and **2**), to the carbon atom of the carboxylic acid group, since the CO₂H group, in its ionized form, could interacts by electrostatic interaction with a cationic residue of the receptor or, in its neutral form, could interacts by hydrogen bonding. Site S2 corresponds to the oxygen atom of the enol-ether function of **1** or the heteroatoms of the A ring of **2**, representing a possible hydrogen bonding interaction. Site S2 also corresponds a spacer group, since an active analogue of PGI₂, Iloprost, has a methylene group in this position [15]. Site S3 corresponds to the oxygen atom of the hydroxyl group binding at C11 in PGI₂ and heteroatoms of the B ring of **2**, representing a hydrogen bonding interaction. Finally, site S4 corresponds to the hydrophobic chain of PGI₂ (ω chain) and the phenyl groups of C4 and C5 positions of the B ring, representing a hydrophobic interaction. The hydroxyl group at the C15 position of PGI₂ was not labeled as a pharmacophoric group because we could not correlate it with any functional group of the aromatic heterocyclic derivatives.

Based on data from Meanwell and coworkers research [9–13], we developed, in a previous work [2], LIV-3D-QSAR models for ligands of IP receptors using only one conformation of each compound, and the best model showed predictive capacity and only two outlier compounds. However, outliers analyses showed us that, the selected conformation for these compounds (outliers), though executed applying the same criteria, might not be the best ones to describe the observed activity.

Considering these results, we developed, in this work, new LIV-3D-QSAR models employing the same set of com-

pounds, but using thirty percent (30%) of the different conformations of each compound; including consequently, a representative number of conformations as mentioned before. The LIV-3D-QSAR models obtained in this work are more predictive than the previous ones and describe better the relationship between the selected LIVs and the pharmacophoric groups, which can be correlated with the drug-receptor interaction. The new LIV-3D-QSAR models can be used in the rational design of new IP receptor ligands with potential platelet anti-aggregation activity.

2. Methods

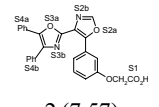
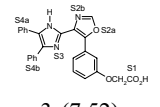
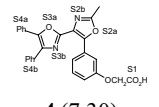
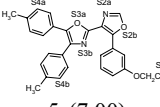
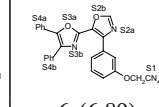
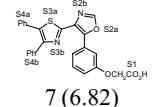
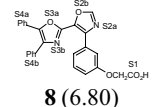
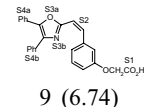
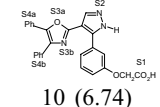
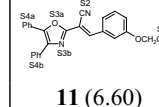
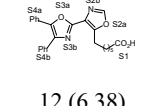
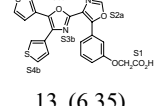
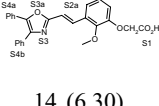
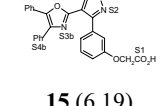
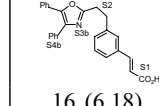
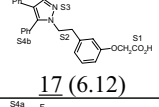
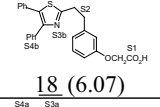
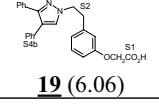
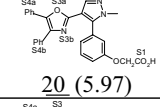
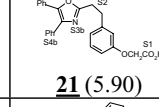
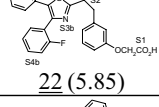
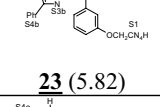
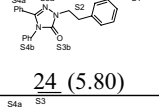
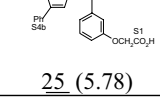
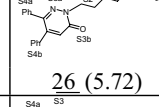
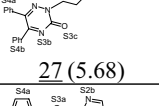
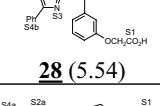
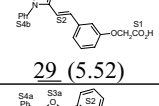
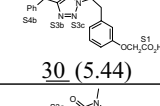
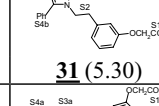
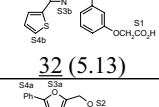
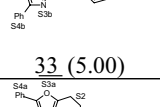
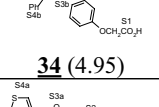
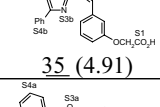
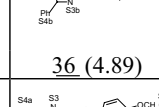
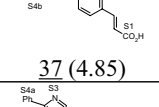
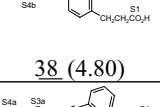
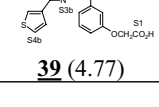
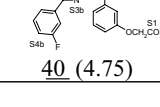
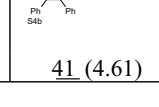
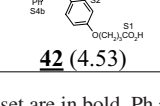
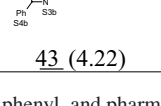
The eight steps involved in the current LIV-3D-QSAR model development are as follows: **Step 1, Biological Data Selection:** The target compounds are a set of 42 aromatic heterocyclic derivatives from Meanwell and coworkers [9–13] divided into a training data set (30 compounds) and a test data set (12 compounds) (Table 1): i.e., the same data sets from our previous work [2]. The biological activity was measured according to the pharmacological protocol established by Seiler and coworkers [16], where the inhibition of the platelet aggregation induced by ADP in human platelet rich plasma (PRP) is reported as IC₅₀ (μ M). The IC₅₀ values were then converted to pIC₅₀, and are show in Table 1. The compounds present a range of approximately four logarithms unites with regular distribution of pIC₅₀ values.

Step 2, Conformational Analysis and the PGI₂ Reference Conformation Selection: The conformational analysis was performed applying the systematic search tool available in the PC Spartan Pro v.1.0.5 [17] using the MMFF force field [18]. Default options were used, including the maximum number of 100 conformations and an energy cutoff of 10 kcal/mol from the minimum energy conformation found. The conformations generated were optimized using the AM1 [19] Hamiltonian, and we excluded the most similar conformations from each compound, according to the root mean square (RMS) deviation using all atoms superposition in the *Search_Compare* module available in the Insight II program [20]. One (Analysis 1) [2] and thirty percent (Analysis 2) of the different conformations for each compound were selected from this new set of conformations according to their minimal RMS value and to their lowest energy values, respectively.

In order to select a PGI₂ reference conformation, we performed a conformational analysis for PGI₂ as was described above, and we selected the conformation that has the lowest RMS value (RMS = 0.325) derived from the superposition of the conformations of PGI₂ and compound **5** generated using three atoms for the alignment. This means a common conformation, and probably a bioactive one. The three atoms used for the superposition were selected according to structure-activity relationship (SAR) studies [9–14] in which the pharmacophoric groups were labeled as S1, S2, and S3. In addition, we observed that the PGI₂ reference

Table I

Structures of the 42 aromatic heterocycle derivatives (**2–43**) used in this work and their pIC_{50} values [9–13].

 2 (7.57)	 3 (7.52)	 4 (7.30)	 5 (7.00)	 6 (6.89)
 7 (6.82)	 8 (6.80)	 9 (6.74)	 10 (6.74)	 11 (6.60)
 12 (6.38)	 13 (6.35)	 14 (6.30)	 15 (6.19)	 16 (6.18)
 17 (6.12)	 18 (6.07)	 19 (6.06)	 20 (5.97)	 21 (5.90)
 22 (5.85)	 23 (5.82)	 24 (5.80)	 25 (5.78)	 26 (5.72)
 27 (5.68)	 28 (5.54)	 29 (5.52)	 30 (5.44)	 31 (5.30)
 32 (5.13)	 33 (5.00)	 34 (4.95)	 35 (4.91)	 36 (4.89)
 37 (4.85)	 38 (4.80)	 39 (4.77)	 40 (4.75)	 41 (4.61)
 42 (4.53)	 43 (4.22)			

a) Compounds from the test set are in bold, Ph = phenyl, and pharmacophoric groups (S1–S4) are show.

conformation selected in our study is in agreement with the conformation found by Tsai and coworkers [14]. In order to compare both conformations, we reproduced the PGI_2 conformation from Tsai and coworkers [14], and performed a superposition with the PGI_2 reference conformation selected for our work, using the same three atoms from the pharmacophoric groups described above. The RMS value observed was equal to 0.430 (see Figure A of Scheme 1 from reference 2). These results point out the high degree of similarity among the selected conformations.

Step 3, Grid Matrix Construction and Volume Calculation: The grid matrix is composed by cubic unitary cells, where the vertices of the cells correspond to the Cartesian coordinates of the eight carbon atoms. The vertices arrest lengths are 1.50\AA (that is, almost equal to the carbon van der Waals radii of 1.54\AA). The grid matrix is composed by a total of 2197 ($13 \times 13 \times 13$) carbon atoms, constructed on the Excel program, and imported into the Insight II program [20] (Fig-

ure 2, A). The volume for each hard sphere (Figure 2, B) of the grid matrix was calculated using a radii length of $1.54\text{\AA} \times 0.65$, where 0.65 is the scale factor used to avoid a large overlap among the spheres (although still allowing a small one), and, consequently, a minimal loss of volume among the hard spheres [2]. The volumes were calculated on the *Search-Compare* module of the Insight II program [20].

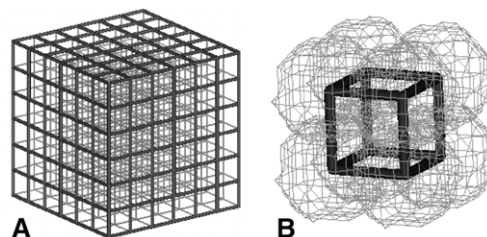


Fig. 2. (A) The grid matrix of hard spheres composed by carbon atoms subdivided in (B) cubic unitary cells of 1.5\AA of length and the associated carbon atom volumes used to calculate the LIVs.

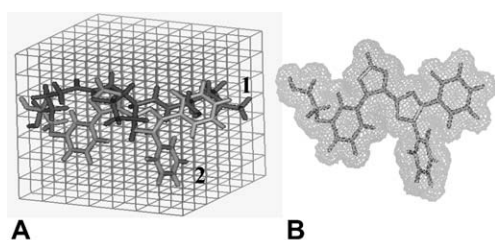


Fig. 3. (A) Superimposition of PGI₂ (1) and compound 2 into the grid matrix and (B) molecular volume of compound 2.

Step 4, Alignment: In this step, we have selected three atom pairs for the RMS alignment according SAR studies [9–13], where the atoms are from the pharmacophoric groups labeled S1–S3 (see Introduction and Figure 1). The S4 site was not selected for the alignment since it represents a group with high conformational freedom in PGI₂. A selected conformation of PGI₂ was used as a reference compound [2] (Figure 3), where this conformation was inserted into the grid matrix using the center of mass of the PGI₂ and the grid matrix structures. Subsequently, each conformation of each molecule was inserted into the grid matrix by superposition with the reference compound by RMS fitting, according to alignment rules, using the Search_Compare module in the Insight II program.

Step 5, Molecular Volume and Local Intersection Volume Calculations: The molecular volumes of 30% of the selected conformations for each compound were calculated using the van der Waals radii without scale factors. Then, the intersection volumes were calculated using the molecular volume of each compound and the volume of each hard sphere that composes the grid matrix. The intersection volume of each compound is an average value, considering all intersection volumes for that compound. These intersection volumes are labeled as local intersection volumes (LIV), and represent the independent variables (volume descriptors) of the database (DB). All volumes were calculated on the Search_Compare module of the Insight II program.

Step 6, Data Reduction: In order to exclude data noise on the database, we performed a data reduction where the original database containing 2197 variables (or LIVs descriptors) generated three databases A, B, and C as follows: DB-A, with 1233 LIVs, was constructed from the original DB, excluding the variables in which LIV equals zero in all molecules; DB-B, with 799 LIVs, was constructed from DB-A, excluding the variables in which LIV is different from zero in three or less than three molecules; and DB-C, with 694 LIVs, was constructed from DB-B, excluding the variables in which LIV is different from zero in five or less than five molecules. In DB-A, we excluded useless variables and, in DB-B and DB-C, we harmonized the data, removing variables that were not properly represented in the data set. That meant we did take into account the structural peculiarities of a few compounds.

Step 7, Models Calculation: The models were calculated using a combined GA-PLS [21,22] approach implemented in the Wolf 6.2 program [23]. We created 800, 300, and 200 ran-

domly generated models for DB-A, DB-B, and DB-C, respectively. Initially, each model contained four independent variables. The mutation operator was set to 400%, 100%, and 50% to DB-A, DB-B, and DB-C, respectively, for each 10-crossover operation. The smoothing factor (the variable that controls the number of independent variables in the models) was set to 0.1, and the maximal number of components for the PLS regression analysis was set to three. We performed 30,000 crossover operations for each database. All other options were left in their default values.

Step 8, Model Validation and Selection: a) **Cross-validation.** The ten best LIV-3D-QSAR models as scored by the “lack-of-fit” (LOF) measure [21,22] from the GA-PLS analysis were evaluated by “leave-one-out” (LOO) cross-validation procedure [17,18] using the Wolf 6.2 program [23]. b) **External Validation.** The best model from each database were individually applied to the test data set of 12 compounds that were not included in developing the LIV-3D-QSAR models. c) **Randomized Activity Data Set.** To explore the possibility of chance correlation in constructing the 3D-QSAR models, as well as to test the robustness of the models, the pIC₅₀ values for the training data set were randomized [24,25] with respect to compound structures, and we performed step 7. If low R² and Q² values are obtained for the 3D-QSAR models from the randomized data set, the probability of chance correlation is low.

3. Results and Discussion

3.1. Analysis of Models 1, 2, and 3

Models 1, 2, and 3, described below, correspond to the best models for DB-A, DB-B, and DB-C, respectively. These models are composed by six independent variables (LIV) with the square of the coefficient of linear correlation (R²) varying from 0.86 to 0.92 and R² after cross-validation (Q²), which means predictive capacity, varying from 0.78 to 0.81.

Model 1

$$\begin{aligned} \text{pIC}_{50} = & 6.77 + 1.50 (\text{LIV}_{280}) + 0.96 (\text{LIV}_{425}) - 0.21 (\text{LIV}_{1618}) - \\ & - 0.76 (\text{LIV}_{1684}) - 5.42 (\text{LIV}_{1636}) - 6.08 (\text{LIV}_{1953}) \\ & (N = 30; R^2 = 0.88; Q^2 = 0.78; S = 0.40; F = 40.91) \end{aligned}$$

Model 2

$$\begin{aligned} \text{pIC}_{50} = & 5.54 + 1.52 (\text{LIV}_{425}) - 1.30 (\text{LIV}_{925}) + 0.50 (\text{LIV}_{947}) + \\ & + 3.50 (\text{LIV}_{1142}) + 1.30 (\text{LIV}_{1673}) - 0.91 (\text{LIV}_{1684}) \\ & (N = 30; R^2 = 0.89; Q^2 = 0.79; S = 0.39; F = 32.40) \end{aligned}$$

Model 3

$$\begin{aligned} \text{pIC}_{50} = & 6.95 + 1.12 (\text{LIV}_{425}) - 1.03 (\text{LIV}_{743}) + 1.05 (\text{LIV}_{752}) - \\ & - 0.45 (\text{LIV}_{800}) - 0.82 (\text{LIV}_{1684}) - 0.57 (\text{LIV}_{1812}) \\ & (N = 30; R^2 = 0.90; Q^2 = 0.81; S = 0.37; F = 32.95) \end{aligned}$$

Table 2

Experimental and calculated pIC₅₀ values, residual values, and standard deviation of residues (SD) of models 1, 2, and 3

No. ^a	pIC ₅₀ exp.	pIC ₅₀ calculated			Residual ^c		
		Mod. 1	Mod. 2	Mod. 3	Mod. 1	Mod. 2	Mod. 3
2	7.57	7.38	7.75	7.72	-0.19	0.18	0.15
3	7.52	7.59	7.05	7.15	0.07	-0.47	-0.37
4	7.30	6.41	6.46	6.61	-0.89	-0.84	-0.69
5	7.00	7.09	6.70	7.15	0.09	-0.30	0.15
6	6.89	6.65	6.89	6.84	-0.24	0.00	-0.05
7	6.82	7.13	6.82	7.43	0.31	0.00	0.61
8	6.80	6.89	6.53	6.40	0.09	-0.27	-0.40
9	6.74	6.36	6.73	6.34	-0.38	-0.01	-0.40
10	6.74	6.74	6.47	6.57	0.00	-0.27	-0.17
11	6.60	5.38	6.18	5.29	-1.22	-0.42	-1.31
12	6.38	6.13	6.21	5.77	-0.25	-0.17	-0.61
13	6.35	6.20	6.27	6.33	-0.15	-0.08	-0.02
14	6.30	5.76	5.85	5.91	-0.54	-0.45	-0.39
15	6.19	6.77	6.16	6.11	0.58	-0.03	-0.08
16	6.18	6.22	6.28	5.76	0.04	0.10	-0.42
17	6.12	5.70	6.02	6.13	-0.42	-0.10	0.01
18	6.07	5.96	5.79	5.64	-0.11	-0.28	-0.43
19	6.06	6.00	5.70	5.48	-0.06	-0.36	-0.58
20	5.97	5.72	6.01	5.97	-0.25	0.04	0.00
21	5.90	5.51	5.45	5.57	-0.39	-0.45	-0.33
22	5.85	6.09	5.62	6.07	0.24	-0.23	0.22
23	5.82	6.67	5.68	5.74	0.85	-0.12	-0.06
24	5.80	5.91	6.62	6.46	0.11	0.80	0.64
25	5.78	6.04	6.33	5.78	0.26	0.55	0.00
26	5.72	5.83	6.10	5.57	0.11	0.38	-0.15
27	5.68	5.65	5.88	5.85	-0.03	0.20	0.17
28	5.54	4.93	5.62	4.83	-0.61	0.08	-0.71
29	5.52	5.64	5.51	5.71	0.12	-0.01	0.19
30	5.44	5.51	5.67	5.64	0.07	0.23	0.20
31	5.30	5.33	4.99	5.23	0.03	-0.31	-0.07
32	5.13	5.93	5.75	5.67	0.80	0.62	0.54
33	5.00	5.20	5.49	5.05	0.20	0.49	0.05
34	4.95	5.65	5.63	5.41	0.70	0.68	0.46
35	4.91	4.82	5.01	4.80	-0.09	0.10	-0.11
36	4.89	4.91	4.66	4.80	0.02	-0.23	-0.09
37	4.85	5.09	4.53	5.32	0.24	-0.32	0.47
38	4.80	4.99	4.42	4.76	0.19	-0.38	-0.04
39	4.77	5.90	5.60	5.91	1.13	0.83	1.14
40	4.75	4.70	4.99	4.95	-0.05	0.24	0.20
41	4.61	4.50	4.64	4.84	-0.11	0.03	0.23
42	4.53	4.88	4.94	4.77	0.35	0.41	0.24
43	4.22	4.16	4.50	4.36	-0.06	0.28	0.14
SD ^b					0.43	0.38	0.43

a) Compounds from the test set are in bold. b) Standard deviation of residues for the entire group (training set and test set). c) Standard deviation of residues for the training set.

On Table 2, we show the compound numbering, the corresponding experimental, calculated (training set), and predicted (test set) pIC₅₀ values, residual values, and the standard deviation of residues (SD) for models 1, 2, and 3. The outlier compounds were defined as the compounds whose residuals are more than twice the SD of the residual of fit considering the 42 compounds (training and test sets).

In order to evaluate the similarity among the three models, the correlation matrix of the residual values of each model was constructed (Table 3), and we observed that the three models are reasonably correlated, probably due by the pres-

ence of LIV₄₂₅ and LIV₁₆₈₄ in the three models, which means that, we can chose one model as representative of the others.

Table 3

Correlation matrix among the residual values of the models 1, 2, and 3

	Res. 1	Res. 2	Res. 3
Res. 1	1.00		
Res. 2	0.56	1.00	
Res. 3	0.75	0.65	1.00

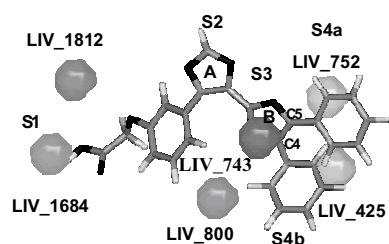


Fig. 4. Graphical representation of the LIV-3D-QSAR Model 3 using compound 2 as reference. LIV_425 and LIV_752 correspond to positive (gray) contributions for the activity. LIV_743, LIV_800, LIV_1684 and LIV_1812 correspond to negative (dark gray) contributions for the activity. The LIVs descriptors are represented in their maximum size.

Therefore, we have chosen Model 3 as the best model for three reasons: (i) it is statistically better, presenting the highest Q^2 value ($Q^2 = 0.81$) and only two outlier compounds (Table 2), while models 1 and 2 have four and three outlier compounds, respectively; (ii) it is more comprehensive in a mechanistic sense, due the pharmacophoric groups that could be correlated to the selected LIVs (see subsequent discussion of Model 3); and (iii) it is the most representative of both models 1 and 2.

3.2. Graphic Representation of Model 3

Figure 4 shows the graphic representation of LIVs from Model 3, using compound 2, the most active compound, as a template, where the LIVs are represented in their maximum size. Therefore, this is not the representation of the LIV values for compound 2. From the six LIV descriptors selected in Model 3 (Figure 4), five are straightly correlated to three (S1, S3, and S4) of the four-pharmacophoric sites (Figure 1).

Site S1 is related to LIV_1684, which represents a negative contribution to the biological activity. However, it is interesting to observe that in Model 2 of our previous work [2] the LIV related to site S1, LIV_1504, represents a positive contribution, but in another region of space. This difference in contribution could be explained by an opposite carboxylic acid chain orientation of the compound 2 in the models. When that function is oriented to the same side of the heterocyclic A ring, there is a positive contribution [2]; when it is oriented to the opposite side, a negative contribution is observed (Figure 4). These results indicate that there is a preferential orientation of the carboxylic acid group in the receptor site, which is important for the biological activity. This is reasonable, since the carboxylic acid group, in its ionized form, could interact by electrostatic interaction with a cationic residue of the receptor, which is an interaction strong enough to orientate this group to the “correct” position.

The S3 site is related to LIV_743, which represents a negative contribution to the biological activity, this LIV is located over the heterocycle B ring, but it occurs only in compounds that present a single bond between the carboxylic acid group and the heterocycle B ring, the less potent compounds.

Site S4 is split into two sub-sites: S4a and S4b. S4a is related to LIV_752, which represents a positive contribution to the biological activity located around the phenyl group at the C5 position of the heterocycle B ring. S4b is related to LIV_425 and LIV_800, which represent a positive and negative contribution to the biological activity, respectively. Both LIVs are located around the phenyl ring at the C4 position of the heterocycle B ring. The contribution degree of these LIVs depends on the relative orientation between these rings (as will be discussed for compound 2 in the next section). It is interesting to reinforce that the S4 site was not selected for the alignment (see Step 4 in Methods) since it represents a group with high conformational freedom in PGI₂, but, in Model 3, it is correlated to the above-discussed LIVs, pointing the importance of a hydrophobic interaction in this site.

Finally, in Model 3 (Figure 4), we could not find any correlation between site S2 and any LIV descriptor, nor between LIV_1812 and any pharmacophoric site. However, LIV_1812, which represents a negative contribution to the biological activity, is located around phenoxy group, but it occurs only in compounds that present a double (*trans* isomer) or a single (antiperiplanar conformation) bond between the carboxylic acid group and the heterocycle B ring, which are the less potent compounds, as compounds 36 and 39, respectively.

3.3. Graphic Representation of Compound 2 on Model 3

The graphic representation of the LIV values for compound 2 according to Model 3 is shown on Figure 5. In order to represent this compound, we have selected the conformation that presents the higher number of LIVs (this conformation is the same used on Figure 4), and this criteria was used for all compounds. This compound was chosen because it is the most potent compound, and presents a small residual value (residual = 0.15, Table 2). The LIVs for compound 2 in Model 3 are: LIV_1684 (S1 site), which has a large negative contribution for the activity; LIV_425 (S4b site), which has also a large but positive contribution; and LIV_800 (S4b site), which has a small negative contribution. Although the negative contribution of LIV_1684 is large, its coefficient is smaller (0.82) than the coefficient of LIV_425 (1.12).

It is interesting to observe that for the hydrophobic region, we have both contributions in S4b site, LIV_425 and

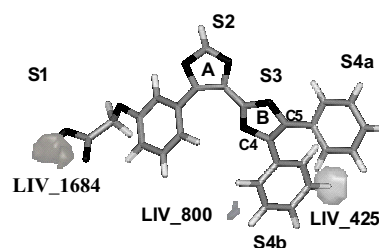


Fig. 5. Graphical representation of the LIV-3D-QSAR Model 3 for compound 2. LIV_425 corresponds to positive (gray) contribution for the activity. LIV_800 and LIV_1684 correspond to negative (dark gray) contributions for the activity. The LIVs descriptors are represented in their maximum size.

LIV_1684, which may be explained by the relative orientation of the phenyl rings at C4 and C5 position with the heterocycle B ring, as we could observe for compound 2, measuring the torsion angle between them. Therefore, the torsion angle between the phenyl group in C5 position (S4a) and the heterocyclic B ring is -22.14° (that is almost a coplanar arrangement between both rings) and the torsion angle between the phenyl group in C4 position (S4b) and the heterocyclic B ring is 44.21° (being less coplanar than the first arrangement). Measuring these angles for the other compounds (data not show), we observed that the more perpendicular this orientation is, the greater the positive contribution (LIV_425) will be; on the other hand, the more coplanar it is, the greater the negative contribution (LIV_800) will be. However, in an intermediate orientation both contributions will be observed. These results are supported by X-ray crystallographic studies [8], where the activity increases when the coplanarity between the heterocycle B and the phenyl ring at C5 position is enhanced (as in compound 2), and the greater this coplanarity, the more perpendicular is the orientation between the phenyl group at C4 position and the heterocycle B ring. This result was also observed in our previous work [2], but for both S4 sub-sites.

3.4. Outlier Compounds of Model 3

As we can see on Table 2, there are two outlier compounds for Model 3: compounds 11 and 39, both from the test data set. In order to understand better the explanation about the outliers, the entire data set was divided in four groups, according to the spatial disposition between the heterocycle B ring and the carboxylic acid group (Table 1). i) pseudo-*cis* arrangement (2-8, 10, 12, 13, 15, 20, 32, 34); ii) *cis* (9, 35) and *trans* (11, 14, 29, 33 and 36) isomers; and iii) linear arrangement (42 and 43); iv) no fixed arrangement (there is a free rotation allowing all conformations, e.g. *antiperiplanar* and *synclinal*) (16, 17, 18, 19, 21-28, 30, 31, 37-41). According to Meanwell and co-workers results [12], the pseudo-*cis* compounds and *cis* isomers are more potent than the compounds with free rotation and *trans* isomers, e.g. compound 9 (*cis* isomer) is more potent than compound 36 (*trans* isomer) (Table 1).

Figure 6 shows the graphic representation of the LIVs (LIV_1684 e LIV_1812) of compound 11 according to

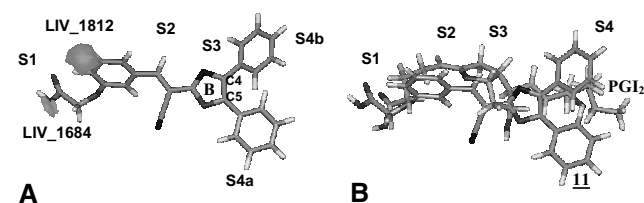


Fig. 6. (A) Graphical representation of the LIV-3D-QSAR Model 3 for compound 11 (outlier). LIV_1684 and LIV_1812 correspond to negative (dark gray) contributions for the activity. (B) Superposition between compound 11 (gray) and PGI₂ (gray). The LIVs descriptors are represented in their maximum size.

Model 3. The residual value (calculated minus experimental pIC₅₀) for compound 11 is -1.31 : i.e., the predicted activity is lower than the experimental one (Table 2). The predicted activity of compound 11 is consistent with the experimental activity for the other *trans* compounds (14, 29, 33, and 36). The discrepancy is the experimental activity observed for compound 11 and, in a lesser degree, for compound 14. In order to explain the unexpected experimental activity for compound 11, Meanwell e co-workers [12] suggested that the nitrogen atom of the nitrile group could interact by a hydrogen bond in S2 site.

However, we observed that the distance between S2 site of the PGI₂ (oxygen atom of the endoperoxide ring) and the compound 11 nitrile group (nitrogen atom) is relatively big (3.87\AA) (Figure 6). Hence, the nitrile group could not correspond to the S2 site. We need to emphasize that nitrogen atom of the nitrile group was not utilized as S2 site on the alignment, because when it was used, the RMS value from the superposition with the PGI₂ was 0.786 : i.e., higher than when the carbon atom from the double bond adjacent of the nitrile group (RMS = 0.361) was used.

Therefore, we proposed in this work an additional site, without correspondence on the PGI₂ structure, in order to explain the experimental activity observed for compound 11, since that the nitrile group is not a bulky group and has extension enough to interact by an hydrogen bond with aminoacid residues a little more distant on the receptor site. For compound 17 (Table 1), we observed a lower distance (2.77\AA) between site S2 of the PGI₂ and the oxygen atom of the methoxy group of compound 17. That could explain the experimental activity observed, where this group could be correlated with site S2.

Figure 7 shows the graphic representation of the LIVs (LIV_425, LIV_1684, and LIV_1812) of compound 39 according Model 3, and it is represented by an *antiperiplanar* conformation torsion angle ($t_6 = 179.25^\circ$). The residual value for compound 39 is 1.14 (Table 2), and it also occurred as an outlier in Model 2 of our previous work [2], but its residual value was higher (4.31). According to the arrangement between the heterocycle B ring and the carboxylic acid group, this compound is classified as not having a fixed arrangement, and so it can present both types of conformations: *antiperiplanar* and *synclinal*, where the *synclinal* conformation corresponds to the pseudo-*cis* arrangement and *cis* iso-

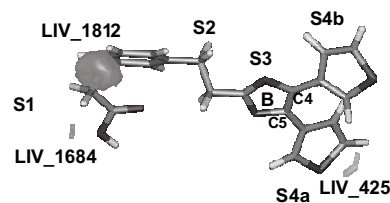


Fig. 7. Graphical representation of the LIV-3D-QSAR Model 3 for compound 39 (*antiperiplanar* conformation, $\theta = 179.25^\circ$). LIV_425 corresponds to positive (gray) contribution for the activity. LIV_1684 and LIV_1812 correspond to negative (dark gray) contributions for the activity. The LIVs descriptors are represented in their maximum size.

Table 4
Summary of results from Analysis 1 (one conformation) and Analysis 2 (30% conformations)

Model ^c	DB ^d	Analysis 1 ^a (one conformation)				Analysis 2 ^b (30% conformations)			
		R ²	Q ²	SD ^e	NO ^f	R ²	Q ²	SD ^e	NO ^f
1	A	0.88	0.78	0.75	4	0.88	0.78	0.43	4
2	B	0.92	0.84	0.95	2	0.89	0.79	0.38	3
3	C	0.86	0.73	0.74	2	0.90	0.81	0.43	2

a) Results from reference 2. b) This work. c) All models have the same number of selected LIVs (six), 30 compounds as the training set, and 12 compounds as the test set. d) Databases A, B, and C. e) Standard deviation of residues for the entire group (42 compounds). f) Number of outliers.

mers. In Model 2 of our previous work [2], compound **39** was represented by a unique conformation, *synclinal*, which is probably not the predominant one in the biological medium. In Model 3 of this work, where we have utilized 30% of the different conformations for the each compound, the LIVs values for compound **39** are an average from six conformations, where we have four *antiperiplanar* and two *synclinal* conformations. Therefore, the lower residual value for compound **39** (Model 3) can be explained by a higher percentage of *antiperiplanar* conformation, which is probably the predominant one on the biological medium. This also might explain why other analogs of compound **39** with predominant *antiperiplanar* conformations are not outliers.

3.5. Comparing Analysis 1 (one conformation) and Analysis 2 (30% conformations)

Comparing the results from the LIV-3D-QSAR models obtained through the analysis of a single conformation (Analysis 1, our previous work [2]) and a set of 30% of the generated conformations (Analysis 2) for each compound (Table 4), we observed that, in terms of R² and Q² values, the models are equivalent. When comparing the standard deviation (SD) of residues for the entire group (Table 4), we can see that the models from Analysis 2 are superior to the models from Analysis 1 since they have lower SD values.

Comparing the LIVs from the best models of Analysis 1 (Model 2) to Analysis 2 (Model 3), we observed that sites S1 and S4 correlate with at least one LIV in both models. Additionally, site S3 of Analysis 2 correlates with one LIV. In both models, site S2 was not correlated to any LIV, which is in agreement with the proposal that this site could be acting just as a spacer group, since an active analogue of PGI₂, Iloprost, has a methylene group in this position.

Comparing the outliers, we observed that compound **39** is an outlier in both analysis, but it has a lower residual value in Analysis 2. In Analysis 1, the selected conformation of compound **39** is a *synclinal* one (that is correlated to the *pseudo-cis* arrangement), which is better for the activity than the *antiperiplanar*. Compound 4, an unexpected outlier in Analysis 1, stopped behaving as an outlier in Analysis 2, showing the improvement of the Analysis 2.

Overall, Analysis 2 is superior to Analysis 1, showing that the use of 30% of the conformations, instead of only one, leads to better models not only in terms of statistical indices but also in terms of interpretability and mechanistic sense.

4. Conclusion

In this work, we developed 3D-QSAR models for ligands of the IP receptor using the local intersection volume (LIV) descriptor and 30% of the conformations generated for each compound, in order to compare with our previous work [2], where we used only one conformation. We obtained three LIV-3D-QSAR models by genetic algorithms (GA) [21] and partial least squares (PLS) methods [22], namely Model 1, Model 2, and Model 3, corresponding to the best models from databases A, B, and C, respectively.

Model 3 was chosen as the best model, since it has the highest Q² value (Q² = 0.81) and only two outlier (compounds 11 and **39**), and also because, in a mechanistic sense, it is more comprehensive, due the pharmacophoric groups that could be correlated to the selected LIVs. Observing the selected LIVs on Model 3, we can distinguish two with positive and four with negative contributions to the biological activity. LIV_1684 (negative coefficient) is correlated to the pharmacophoric S1 site (carboxylic acid group); LIV_743 (negative coefficient) is correlated to the pharmacophoric S3 (heterocyclic B ring); and LIV_425 and LIV_752 (positive coefficients) and LIV_800 (negative coefficient) are correlated to the pharmacophoric S4 site (hydrophobic group).

There is no correlation between the pharmacophoric S2 site and any LIV descriptor of Model 3, however, we proposed an additional site around site S2 without correspondence on the PGI₂ structure, in order to explain the unexpected experimental activity observed for compound **11** (outlier), which has a nitrile group that could interact by an hydrogen bond with aminoacid residues a little more distant on the receptor site.

In Model 3, there is no any correlation between LIV_1812 (negative coefficient) and any pharmacophoric site, though LIV_1812 occurs in compounds that present a double (*trans* isomer) or single (*antiperiplanar* conformation) bond between the carboxylic acid group and heterocycle B ring, which are the less potent compounds, as compounds **36** and **39**, respectively.

Also in Model 3, compound **39** can be explained as an outlier because although it has a high percentage of *antiperiplanar* conformation, which is correlated to the *trans* isomers (less potent compounds), LIVs values are an average from all selected conformations, including the *synclinal* conformations, which are correlated to the *pseudo-cis* compounds and *cis* isomers (more potent compounds).

Finally, the use of 30% of the conformations for each compound (Analysis 2) instead of only one as in our previous

work (Analysis 1) [5] leads to better models, not only in terms of statistical indices but also in terms of interpretability and mechanistic sense. We may, thus, conclude that Model 3 can be applied to design new PGI₂ receptor ligands that could be potential inhibitors of platelet aggregation on the therapy of thrombotic diseases.

Acknowledgements

We thank the Brazilian agencies CAPES, CNPq, and FAPERJ for their support. We also thank Fundação Universitária José Bonifácio (FUJB) and Conselho de Ensino para Graduados e Pesquisa (CEPG) da Universidade Federal do Rio de Janeiro (UFRJ) for a grant under “Programa ALV/98”, Proc. No. 8436-1 (MGA).

References

- [1] H. Verli, M.G. Albuquerque, R.B. Alencastro, E.J. Barreiro, *Eur. J. Med. Chem.* 37 (2002) 219–229.
- [2] R.C.A. Martins, M.G. Albuquerque, R.B. Alencastro, *J. Braz. Chem. Soc.* 13 (2002) 816–821.
- [3] P.J. Goodford, *J. Med. Chem.* 28 (1985) 849–857.
- [4] R.D. Cramer III, D.E. Patterson, J.D. Brunce, *J. Am. Chem. Soc.* 110 (1988) 5959–5967.
- [5] A.J. Hopfinger, S. Wang, J.S. Tokarski, B. Jin, M. Albuquerque, P.J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.* 119 (1997) 10509–10524.
- [6] P.W. Collins, S.W. Djuric, *Chem. Rev.* 93 (1993) 1533–1564.
- [7] J.R. Vane, R.M. Botting, *Am. J. Cardiol.* 75 (1995) 3A–10A.
- [8] W.B. Campbell, 8th ed, in: A.G. Gilman, T.W. Rall, A.S. Nies, P. Taylor (Eds.), *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, Pergamon Press, New York, 1990, pp. 600–617.
- [9] N.A. Meanwell, M.J. Rosenfeld, J.J. Kim Wright, C.L. Brassard, J.O. Buchanan, M.E. Federici, J.S. Fleming, S.M. Seiler, *J. Med. Chem.* 35 (1992) 389–397.
- [10] N.A. Meanwell, M.J. Rosenfeld, A.K. Trehan, J.J. Kim Wright, C.L. Brassard, J.O. Buchanan, M.E. Federici, J.S. Fleming, M. Gamberdella, G.B. Zavoico, S.M. Seiler, *J. Med. Chem.* 35 (1992) 3483–3497.
- [11] N.A. Meanwell, M.J. Rosenfeld, A.K. Trehan, J.L. Romine, J.J. Kim Wright, C.L. Brassard, J.O. Buchanan, M.E. Federici, J.S. Fleming, M. Gamberdella, G.B. Zavoico, S.M. Seiler, *J. Med. Chem.* 35 (1992) 3498–3512.
- [12] N.A. Meanwell, M.J. Rosenfeld, J.L. Romine, J.J. Kim Wright, C.L. Brassard, J.O. Buchanan, M.E. Federici, J.S. Fleming, M. Gamberdella, K.S. Hartl, G.B. Zavoico, S.M. Seiler, *J. Med. Chem.* 36 (1993) 3871–3883.
- [13] N.A. Meanwell, J.L. Romine, M.J. Rosenfeld, S.W. Martin, A.K. Trehan, J.J. Kim Wright, M.F. Malley, J.Z. Gougoutas, C.L. Brassard, J.O. Buchanan, M.E. Federici, J.S. Fleming, M. Gamberdella, K.S. Hartl, G.B. Zavoico, S.M. Seiler, *J. Med. Chem.* 36 (1993) 3884–3903.
- [14] A. Tsai, E. Strobeljager, K.K. Wu, *J. Comp-Aided Mol. Design* 5 (1991) 135–148.
- [15] J.E. Merritt, T.J. Hallam, A.M. Brown, I. Boyfield, D.G. Cooper, D.M. Hickey, A.A. Jaxa-Chamiec, A.J. Kaumann, M. Keen, E. Kelly, *Br. J. Pharmacol.* 102 (1991) 251–259.
- [16] S. Seiler, C.L. Brassard, A.J. Arnold, N.A. Meanwell, J.S. Fleming, S.L. Keely Jr, *J. Pharmacol. Exp. Ther.* 255 (1990) 1021–1026.
- [17] P.C. Spartan, Pro, v.1.0.5, Wavefunction Inc., Irvine, CA 92612, 1999.
- [18] T.A. Halgren, *J. Comp. Chem.* 17 (1996) 490–519.
- [19] M.J.S. Dewar, E.G. Zebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.* 107 (1985) 3902–3909.
- [20] Insight II User Guide, Molecular Simulations Inc., San Diego, USA, 1995.
- [21] D. Rogers, A.J. Hopfinger, *J. Chem. Inf. Comp. Sci.* 34 (1994) 854–866.
- [22] W.J. Dunn III, D. Rogers, in: J. Devillers (Ed.), *Genetic Algorithms in Molecular Modeling*, Academic Press, London, 1996, pp. 109–130.
- [23] D. Rogers, *Wolf Genetic Function Approximation (GFA)*, v.6.2, Molecular Simulations Inc, 1994.
- [24] G. Klopman, A.N. Kalos, *J. Comp. Chem.* 6 (1985) 492–506.
- [25] M.G. Albuquerque, A.J. Hopfinger, E.J. Barreiro, R.B. Alencastro, *J. Chem. Inf. Comput. Sci.* 38 (1998) 925–938.