Eur. J. Med. Chem. 36 (2001) 1–19

PII: S0223-5234(00)01183-1/FLA

Original article

A molecular modeling and 3D QSAR study of a large series of indole inhibitors of human non-pancreatic secretory phospholipase A_2

Philippe Bernard[a,b], Marco Pintore[a], Jean-Yves Berthon[b], Jacques R. Chrétien[a]*

[a]Laboratory of Chemometrics and BioInformatics, University of Orléans, BP 6759, 45067 Orléans Cedex 2, France
[b]Greentech S. A., Biopôle Clermont Limagne, 63360 Saint-Beauzire Cedex, France

Received 24 November 1999; revised 6 July 2000; accepted 13 July 2000

Abstract – Automated docking allowing protein-based alignment was performed for a series of 188 indole inhibitors of the human non-pancreatic secretory phospholipase A_2 (hnps-PLA_2). All the substituted indoles were docked to the crystal structure of hnps-PLA_2 and a three-dimensional QSAR model was then established using the CoMFA method. The set of 188 compounds was divided into two subsets, the first one constituting the training set (126 compounds), while the second constituted the test set (62 compounds). The established CoMFA model derived from the training set was then applied to the test set. A good correlation between predicted and experimental activity data allows to validate the 3D QSAR model. A second and global 3D QSAR including all the compounds was established, allowing the creation of the hnps-PLA_2 pharmacophore. © 2001 Éditions scientifiques et médicales Elsevier SAS

3D QSAR / CoMFA / docking / protein-based alignment / human non-pancreatic phospholipase

1. Introduction

A research program of 'new leads' [1], combining ethnopharmacology studies with bioinformatic ones, was initiated to find new molecules implicated in the modulation of inflammation processes. The first step of this program was to define the biomolecule target (enzyme, receptors, ...). The criteria for the selection of this target were (i) the experimental conditions (availability of enzyme sources, completion of the enzymatic tests, ...) and (ii) the quantity of available experimental data for the bioinformatic treatment. Phospholipase A_2 (PLA_2) was found to be the best target according to recently performed enzymatic tests carried out on this enzyme and available experimental data. Furthermore, the PLA_2 plays a crucial role by modulating both the cyclooxygenase and the 5-lipoxygenase pathways which are simultaneous pathways of the inflammation process.

* Correspondence and reprints.
E-mail address: jacques.chretien@univ-orleans.fr (J.R. Chrétien).

PLA_2 corresponds to a class of enzymes which catalyze the hydrolysis of membrane glycerophospholipids at the *sn*-2 position to release fatty acids and lysophospholipids. When the fatty acid is the arachidonic acid, a complementary metabolism leads to pro-inflammatory mediators such as prostaglandins, leukotrienes, thromboxanes and platelet-activating factors [2]. Thus, modulating pro-inflammatory lipid mediator production by inhibiting PLA_2 activity remains a potential target for the development of new drugs for the treatment of inflammatory diseases [3]. Nevertheless, the evidence now exists that more than one form of PLA_2 might be found [3]. The first one which corresponds to the low molecular weight (14 kDa) is a Ca^{2+}-dependent extra cellular PLA_2. It is found in mammalian pancreata, several snake venoms, human platelets, human placentas, and rheumatoid synovial fluids [4]. The second form exhibits a high-molecular weight (85 kDa) cytosolic PLA_2 [5]. The latter, a less-known form, will not be used in this bioinformatic study, even if some authors suggested that both the 14 kDa and 85 kDa PLA_2 may be

potential targets for the modulation of inflammatory diseases [3].

Within the secreted 14 kDa PLA_2, one enzyme is worth attention: the human non-pancreatic secretory PLA_2 (hnps-PLA_2) [6]. High levels of this enzyme have been observed in the synovial fluid of arthritic joints [7] and in the serum of patients with severe acute pancreatitis [8], septic shock [9] or with multiple injuries [10]. The hnps-PLA_2 inhibitors might lead to useful drug for diseases where elevated levels of this enzyme are found. Furthermore, hnps-PLA_2 has been crystallized with or without different ligands [11–13] and several chemical classes of synthetic and natural inhibitors are known [14]. However, optimization of the therapeutic properties of the hnps-PLA_2 inhibitors requires: (i) a better understanding of the inhibitor-protein interaction mechanism, and (ii) finding a strategy to predict the activity of new molecules. Approaches related to computational chemistry may help to resolve these two problems.

Previous computational studies [15–23], focusing on the inhibitors of this class of PLA_2, can be clearly subdivided into two groups according to the computational methods used: (i) protein-based studies, which deal with modeling ligand–receptor interactions through docking [15–18] or molecular dynamics (MD) [19–21], and (ii) quantitative structure–activity relationships (QSAR) analyses [22, 23], which do not require necessarily an a priori hypothesis about the receptor structure but imply a common biochemical mechanism. Thus, two types of QSAR analyses exist: the structure-based QSAR and, more recently, the protein-based QSAR [22].

Docking or MD studies [15–21] are computationally expensive and, therefore, they were applied only to a small number of molecules. While protein-based techniques usually provide accurate geometries of ligand–receptor complexes, they cannot easily be used for predicting binding affinity or inhibitory activity, due to the accumulated uncertainties related to the multiple calculations of the ligand–receptor interaction energy which are necessary to evaluate the free energy of the ligand–receptor complex. Nevertheless, some authors solved this problem in their series of compounds [22]. The applicability of the protein-based methods to drug design problems was recently reviewed by Kuntz et al. [24].

QSAR studies on hnps-PLA_2 inhibitors [22, 23] were performed using either particular descriptors [22], correlating the activity of 26 ligands with the enzyme/ligands binding energies, or comparative molecular field analysis (CoMFA) [23, 25] methods. The latter method aligns the studied molecules in a three-dimensional (3D) space. In conventional CoMFA studies, in agreement with the original study of Cramer et al. [25], the molecules are fitted to a reference one. This reference compound should be the most rigid or constrained among the active ones. In this case, this reference compound is supposed to correspond to the 'biologically active' conformation. However, in the case of flexible compounds, finding an appropriate alignment becomes a cumbersome task with a high probability of misleading results.

When a protein structure is known, applying protein-based methods can remove all ambiguity about the definition of the best active conformation. An ideal solution would be to dock all the available molecules to the receptor, this would create an incontestable 'protein-based' alignment for the CoMFA model. In practice, however, this may prove a time-consuming procedure. Moreover, in many cases, it would not completely remove the ambiguity related to the choice of the 'biologically active' conformation, since there can exist multiple hypotheses for the structure of the ligand–receptor complex.

This paper describes a second generation of CoMFA studies in which, thanks to the 'critical' mass of available experimental data, it became possible to dock all 188 indole hnps-PLA_2 inhibitors to the catalytic site of the hnps-PLA_2 in an unequivocal manner and with reasonable computational expenses. The docked inhibitor structures were then used as a 'protein-based' alignment to feed CoMFA procedure, showing that total docking allows to obtain a high quality CoMFA model.

2. Methods

2.1. Automated docking

2.1.1. Enzyme

Many sources of secreted PLA_2 have been used and crystallized: snake [26] and bee venoms [27]; bovine [28] or porcine [29] PLA_2. Since the inhibitory activity data used in the present study were measured on hnps-PLA_2, it was preferable to perform our docking study on this enzyme.

The crystal structures of hnps-PLA_2 were obtained from the Brookhaven Protein Data Bank (PDB). Sev-

eral hnps-PLA_2 proteins are presently available from PDB where the structure of the non-complexed hnps-PLA_2 have been determined in two crystal forms [11, 12] (PDB codes: 1POE [11] and 1BBC [12]). The structures of three hnps-PLA_2/ligand complexes had also been determined. The three ligands associated to the enzyme are (i) a phosphonate transition state analogue (TSA) (PDB code: 1POE) [11], (ii) a highly potent inhibitor (FPL67047XX) (PDB code: 1KVO) [13] and (iii) an acylamino phospholipid analog (PDB code: 1AYP) [30], respectively. To make the structures suitable for further modeling, hydrogen atoms were added to the four original PDB structures using the BIOPOLYMER module of SYBYL [31]. Then, the geometry of the protein was optimized using the Tripos force field [32] with a gradient convergence of 0.01 kcal mol^{-1} and with the Gasteiger–Hückel charges as electrostatic component.

To further exploit known crystal structures of ligand–enzyme complexes as a source of geometric constraints for the automated docking procedure, it was necessary to select the appropriate enzyme structure among the four available enzyme structures. In fact, after minimization, it was shown that: (i) the three ligands of any crystallized complexes showed the same mode of binding inside the catalytic site of the hnps-PLA_2 and (ii) that the presence of a ligand in the active site did not modify notably the position of the residues of the enzyme compared to their positions in the free form. Thus, any enzyme structures could be used for the automated docking procedure.

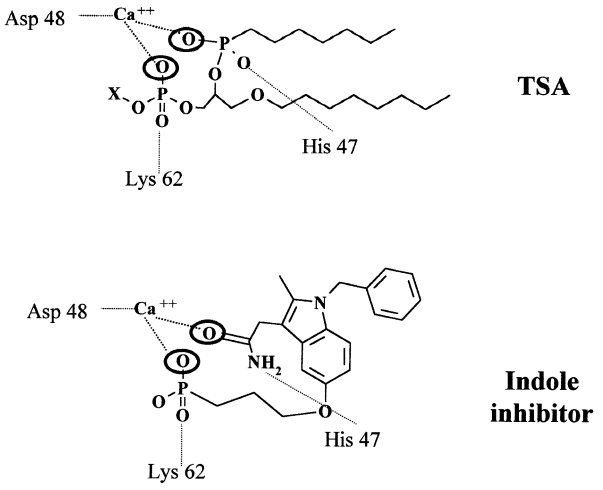


Figure 1. Schematic representation of inhibitor binding to hnps-PLA_2. Homology binding between the TSA and one indole inhibitor, the compound **110**. Atoms selected as anchor points, for the automated docking procedure, are circled with a bold line.

In addition, four compounds belonging to the indole inhibitor series were crystallized with the hnps-PLA_2 [33]. Nevertheless, these complexes are not available in PDB format. The authors of this publication [33] compared the common interactions between these new indole inhibitors/enzyme complexes and the phosphoryl TSA/enzyme complex from reference 11 (*figure 1*). The same three amino acid residues (His47, Asp48, and Lys62) and one calcium atom included in the enzyme structure were implicated in the interaction with a phosphoryl group and/or a carbonyl group of the indole inhibitors and the TSA. Hence, the phosphoryl group and the carbonyl group of the indole inhibitors were successively used as anchor points. Moreover, as indole inhibitor/enzyme complexes were compared with the TSA/enzyme complex, the selected enzyme structure used in this study came from the TSA/enzyme complex [11]. From the PDB file 1POE, the TSA was removed, the hydrogen atoms were added, and the geometry of the protein was optimized using the Tripos force field.

2.1.2. Ligands

Each of the 188 ligands was modeled using SYBYL 6.5 on a Silicon Graphics O_2 R10000 station connected to a Silicon Graphics Origin 200 R10000×2 server. The ligands were considered in their neutral form. The starting conformations were optimized by molecular mechanics algorithm using the Tripos force field. The lowest energy conformations were found by means of the SYBYL/SEARCH option and then used as initial conformations for docking.

2.1.3. Ligand–enzyme docking

The FIT ATOMS option of SYBYL allowed the placement of the phosphoryl group or carbonyl group of each indole inhibitor into the enzyme, prepared as mentioned above, at the position corresponding to the TSA phosphoryl group. The two oxygen atoms linked to the calcium atom of the enzyme were successively used as the anchor point, as shown in *figure 1*. The molecule was then rotated around the three-coordinate axis as previously described [34], with a step of 30°. For each orientation, a complete conformational search was made using the SYBYL/SYSTEMATIC SEARCH option. All rotatable bonds were relaxed and the rotation step was

also fixed at 30°. The total energy of the enzyme–ligand complex was calculated for each conformation generated. The geometry of the lowest energy conformations found for the enzyme–ligand complexes were then refined by further spatial rotations and conformational search with a step of 5° in a ±30° interval around the torsion values obtained. These refined enzyme–ligand complexes were optimized using the Tripos Force Field with a gradient convergence of 0.01 kcal mol^{-1} and with the Gasteiger–Hückel charges as electrostatic component. The dielectric constant was set to 1 and the van der Waals radius scale factor was fixed to 0.6 Å. The lowest energy geometries of the inhibitors were then used in the CoMFA analysis.

2.2. CoMFA

2.2.1. Data set

The 188 indole inhibitors and the corresponding biological data used in this study were selected from literature [35]. The molecular structures and hnps-PLA_2 inhibitory activity data for these 188 indole derivatives are summarized in *table I*. All the collected biological data (IC_{50}) were measured in vitro under the same experimental conditions [36].

These 188 compounds were divided into two subsets: 126 of them were used as a training set and 62 were used as a test set. In fact, every third compound was selected from the series and used in the test set. This choice raises the problem of the separation of a series of compounds into a training set and a test set. The order of the compounds found in *table I* is following an order related to the organic synthesis, i.e. to the structural filiation. Thus, *table I* is formed by the sum of chemical sub-series. Selecting every third compound corresponds in fact to making a selection inside each chemical subseries. This allows a good molecular diversity in the test set. Then, we decided to analyze more deeply the molecular diversity between the training set and the test set. In a recent paper, we explained how to use the Kohonen neural network (self-organizing map [SOM]) [37] for comparison of chemical databases [38]. The Kohonen SOM was therefore applied to investigate molecular diversity among the 126 compounds of the training set and the 62 indole inhibitors of the test set. For this purpose, a series of 60 2D molecular descriptors was used, including several steric, electrostatic and lipophilic descriptors, among which were 20 Kier–Hall molecular connectivity indices [39] with numbers of paths and vertices with degrees 1–4, Gutman [40] and Platt [41] indices, a series of information indices [42], as well as a set of physico-chemical parameters, such as molecular weight, molecular volume, molecular refractivity, octanol-water partition coefficient [43], and a set of electronegativity parameters established by Sanderson [44], mean, variance and maximum value of the electronegativity of atoms. The descriptors were normalized according to the formula:

$$X_{ij}^{n} = \frac{Xij - Xj, \min}{Xj, \max - Xj, \min},$$

where Xij and X_{ij}^{n} are the non-normalized and normalized j-th descriptor values for compound I; correspondingly, Xj,min and Xj,max are the minimum and maximum values for the j-th descriptor. The following parameters had been set: size of the map 10×10, the number of iteration steps for learning 20 000, the starting learning coefficient 0.9. Calculations were performed using proprietary software.

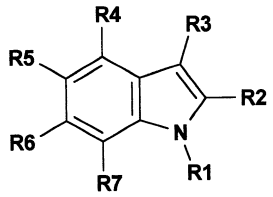
2.2.2. CoMFA method

A CoMFA study normally begins with searching a suitable alignment of the compounds under investigation by using a constrained reference compound. In the present study this problem was a priori solved by docking the compounds to the hnps-PLA_2 crystal structure.

Both the steric and electrostatic CoMFA fields were calculated using a sp^3 carbon atom as a probe, with a charge of +1 in the grid points around the molecules. The grid points were spaced 1.5 Å apart in all three dimensions. Partial atomic charges for the electrostatic field calculation were obtained by the MOPAC AM1 method [45]. The CoMFA region was chosen to include all the molecules within margins of at least 4.0 Å. The CoMFA region was defined as a cube with an edge of 24 Å. The field values were truncated at +30 kcal mol^{-1} for steric interactions and ±30 kcal mol^{-1} for electrostatic interactions.

The partial least squares method (PLS) [46] was used to linearly relate the CoMFA fields to the inhibitory activity values. The optimal number of PLS components was determined using the 'leave-one-out' and 'leave-some-out' (10 groups) cross-validation procedures [46]. The model quality is expressed in terms of Q^2 the cross-validated correlation coefficient, R^2 the conventional correlation coefficient, s the standard error and F the Fisher test.

Table I. In vitro inhibition of the hnps PLA_2 by 188 indole inhibitors.



cpd	R1	R2	R3	R4	R5	R6	R7	IC_{50} (μM)	Log $1/IC_{50}$
1	$C_6H_5CH_2$	H	CH_2CONH_2	H	OH	H	H	3.680	−0.57
2	$C_6H_5CH_2$	CH_3	CH_2CONH_2	OH	H	H	H	8.280	−0.92
3	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	OH	H	H	1.020	−0.01
4	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	OH	H	3.610	−0.56
5	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	OH	H	H	0.342	0.47
6	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	Allyl	OH	H	H	0.143	0.84
7	$C_6H_5CH_2$	$CH_3CH_2CH_2$	CH_2CONH_2	H	OH	H	H	34.000	−1.53
8	$C_6H_5CH_2$	Cyclopropyl	CH_2CONH_2	H	OH	H	H	0.118	0.93
9	$C_6H_5CH_2$	Br	CH_2CONH_2	H	OH	Cl	H	0.085	1.07
10	$C_6H_5CH_2$	CH_3S	$CH_2C\text{-}ONH_2$	H	OH	H	H	0.379	0.42
11	3-$ClC_6H_4CH_2$	CH_3	CH_2CONH_2	OH	H	H	H	5.880	−0.77
12	$C_6H_5CH_2$	CH_3	CH_2CONH_2	$OCH_2CO_2CH_3$	H	H	H	1.380	−0.14
13	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	$O(CH_2)_3CO_2CH_2CH_3$	H	0.450	0.35
14	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	$O(CH_2)_4CO_2CH_2CH_3$	H	0.306	0.51
15	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3CO_2CH_2CH_3$	H	H	1.270	−0.10
16	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CONH_2	H	H	H	0.124	0.91
17	$C_6H_5CH_2$	H	$COCONH_2$	H	$O(CH_2)_3CO_2H$	H	H	62.000	−1.79
18	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	Propyl	H	H	2.070	−0.32
19	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	H	$O(CH_2)_3CO_2H$	H	H	0.210	0.68
20	$C_6H_5SO_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	0.028	1.55
21	H	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	1.600	−0.20
22	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	CH_3	H	0.008	2.10
23	$C_6H_5CH_2$	H	$COCONH_2$	OCH_2CO_2H	H	H	H	0.096	1.02
24	$C_6H_5CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.011	1.96
25	2-(C_6H_5)$C_6H_4CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.006	2.22
26	3-(C_6H_5)$C_6H_4CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.009	2.05
27	4-(C_6H_5)$C_6H_4CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.043	1.37
28	4-$FC_6H_4CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.030	1.52
29	2,6-$Cl_2C_6H_3CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.006	2.22
30	1-naphthylCH_2	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.009	2.05
31	2-$ClC_6H_4CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.009	2.05
32	2-$CH_3C_6H_4CH_2$	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.015	1.82

Table I. (Continued)

cpd	R1	R2	R3	R4	R5	R6	R7	IC_{50} (μM)	Log $1/IC_{50}$
33	n-C_8H_{17}	CH_3	$COCONH_2$	OCH_2CO_2H	H	H	H	0.008	2.10
34	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	0.009	2.05
35	2-(C_6H_5) $C_6H_4CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	0.006	2.22
36	2-($C_6H_5CH_2$) $C_6H_4CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	0.004	2.40
37	3-$ClC_6H_4CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	0.007	2.15
38	C_6H_5CO	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	H	H	H	0.081	1.09
39	2-(C_6H_5) $C_6H_4CH_2$	n-propyl	$COCONH_2$	OCH_2CO_2H	H	H	H	0.082	1.09
40	$C_6H_5CH_2$	Cyclopropyl	$COCONH_2$	OCH_2CO_2H	H	H	H	0.028	1.55
41	2-(C_6H_5) $C_6H_4CH_2$	Cyclopropyl	$COCONH_2$	OCH_2CO_2H	H	H	H	0.006	2.22
42	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OCH_2CO_2H	allyl	H	H	1.620	−0.21
43	2-(C_6H_5) $C_6H_4CH_2$	CH_3	$COCONH_2$	$O(CH_2)_3CO_2H$	H	H	H	0.243	0.61
44	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	$OC(CH_3)_2CO_2H$	H	H	H	0.025	1.60
45	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OC(*R*) H($CH_2CH_2Ph)_2$ CO_2H	H	H	H	0.046	1.34
46	$C_6H_5CH_2$	CH_3	$COCONH_2$	OC(*R*)H(CH_3) CO_2H	H	H	H	0.011	1.96
47	$C_6H_5CH_2$	CH_3	$COCONH_2$	OC(*S*)H(CH_3)CO_2H	H	H	H	0.188	0.73
48	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OC(*S*)H(CH_3)CO_2H	H	H	H	0.141	0.85
49	$C_6H_5CH_2$	H	CH_2CONH_2	OCH_2CO_2H	H	H	H	0.321	0.49
50	$C_6H_5CH_2$	H	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.180	0.74
51	$C_6H_5CH_2$	CH_3	CH_2CONH_2	OCH_2CO_2H	H	H	H	0.052	1.28
52	$C_6H_5CH_2$	CH_3	CH_2CONH_2	$O(CH_2)_3CO_2H$	H	H	H	0.399	0.40
53	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	OCH_2CO_2H	H	H	1.790	−0.25
54	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_2CO_2H$	H	H	0.527	0.28
55	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.152	0.82
56	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_4CO_2H$	H	H	1.150	−0.06
57	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	(2-C_6H_4)CH_2	H	H	0.147	0.83
58	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	(3-C_6H_4)CH_2	H	H	1.900	−0.28
59	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	$O(CH_2)_3CO_2H$	H	12.210	−1.09
60	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	$O(CH_2)_4CO_2H$	H	7.960	−0.90
61	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	OCH_2CO_2H	H	H	H	0.024	1.62
62	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.189	0.72
63	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	(2-C_6H_4)CH_2	H	H	0.555	0.26
64	$C_6H_5CH_2$	Cyclopropyl	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.044	1.36
65	$C_6H_5CH_2$	Cl	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.077	1.11
66	$C_6H_5CH_2$	Br	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.073	1.14
67	$C_6H_5CH_2$	CH_3S	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.162	0.79
68	3-$ClC_6H_4CH_2$	CH_3CH_2	CH_2CONH_2	OCH_2CO_2H	H	H	H	0.039	1.41
69	CyclohexylCH_2	CH_3	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.654	0.18
70	$CH_3(CH_2)_9$	CH_3	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	H	0.683	0.17
71	$C_6H_5CH_2$	CH_3	$CH_2CO_2CH_3$	H	OCH_3	H	H	115.800	−2.06

Table I. (Continued)

cpd	R1	R2	R3	R4	R5	R6	R7	IC_{50} (μM)	Log $1/IC_{50}$
72	$C_6H_5CH_2$	H	$CH_2CONHNH_2$	H	OCH_3	H	H	4.900	−0.69
73	2-(C_6H_5) $C_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	0.260	0.59
74	2-(C_6H_5) $C_6H_4CH_2$	CH_3CH_2	$CH_2CONHNH_2$	H	OCH_3	H	H	2.350	−0.37
75	2-(C_6H_5) $C_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	0.940	0.03
76	2-pyridylCH_2	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	27.180	−1.43
77	$CH_3(CH_2)_9$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	9.430	−0.97
78	$C_6H_5(CH_2)_3$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	16.700	−1.22
79	3-$ClC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	CO_2H	H	H	0.360	0.44
80	3-$NH_2C_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	4.320	−0.64
81	3-$OHC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	3.000	−0.48
82	C_6H_5	CH_3	$CH_2CONHNH_2$	H	H	H	H	38.200	−1.58
83	$C_6H_5CH_2$	Cl	$CH_2CONHNH_2$	H	OCH_3	H	H	0.390	0.41
84	$C_6H_5CH_2$	Br	$CH_2CONHNH_2$	H	OCH_3	H	H	0.430	0.37
85	$C_6H_5CH_2$	CH_3S	$CH_2CONHNH_2$	H	OCH_3	H	H	0.620	0.21
86	$C_6H_5CH_2$	CH_3SO	$CH_2CONHNH_2$	H	OCH_3	H	H	46.000	−1.66
87	$C_6H_5CH_2$	CH_3	$CH_2CONHNH_2$	H	H	H	H	2.360	−0.37
88	$C_6H_5CH_2$	CH_3	$CH_2CONHNH_2$	OCH_3	H	H	H	0.550	0.26
89	$C_6H_5CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	0.860	0.07
90	$C_6H_5CH_2$	CH_3	$CH_2CONHNH_2$	H	Br	H	H	1.020	−0.01
91	$C_6H_5CH_2$	CH_3	$CH_2CONHNH_2$	H	C_6H_5	H	H	0.860	0.07
92	$C_6H_5CH_2$	CH_3CH_2	$CH_2CONHNH_2$	H	OCH_3	H	H	0.420	0.38
93	$C_6H_5CH_2$	$CH_3CH_2CH_2$	$CH_2CONHNH_2$	H	OCH_3	H	H	70.500	−1.85
94	$C_6H_5CH_2$	Cyclopropyl	$CH_2CONHNH_2$	H	OCH_3	H	H	0.390	0.41
95	$C_6H_5CHCH_3$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	5.320	−0.73
96	$C_6H_5CHC_6H_5$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	40.200	−1.60
97	3-$CH_3OC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	2.120	−0.33
98	3-$C_6H_5CH_2OC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	0.610	0.21
99	3-$NO_2C_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	15.220	−1.18
100	2-$ClC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	1.710	−0.23
101	3-$ClC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	OCH_3	H	H	H	0.230	0.64
102	3-$ClC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	1.360	−0.13
103	4-$ClC_6H_4CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	9.280	−0.97
104	2,5-$Cl_2C_6H_3CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	9.390	−0.97
105	2,6-$Cl_2C_6H_3CH_2$	CH_3	$CH_2CONHNH_2$	H	OCH_3	H	H	0.640	0.19
106	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3PO(OCH_3)_2$	H	H	0.266	0.58
107	$C_6H_5CH_2$	CH_3	CH_2CO_2H	H	OCH_3	H	H	13.600	−1.13
108	$C_6H_5CH_2$	H	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.203	0.69
109	$C_6H_5CH_2$	CH_3	CH_2CONH_2	$OCH_2PO(OH)_2$	H	H	H	1.290	−0.11
110	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.057	1.24
111	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.023	1.64
112	$C_6H_5CH_2$	$CH_3CH_2CH_2$	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	6.130	−0.79

Table I. (Continued)

cpd	R1	R2	R3	R4	R5	R6	R7	IC_{50} (μM)	Log $1/IC_{50}$
113	$C_6H_5CH_2$	Cyclopropyl	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.074	1.13
114	$C_6H_5CH_2$	Br	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.033	1.48
115	3-$ClC_6H_4CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.016	1.80
116	2-(C_6H_5) $C_6H_4CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	H	H	0.022	1.66
117	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3PO(OH)$ OCH_3	H	H	0.040	1.40
118	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	SCH_2CO_2H	H	H	H	0.145	0.84
119	$C_6H_5CH_2$	CH_3	$CH(CH_3)$ $CONHNH_2$	H	OCH_3	H	H	32.640	−1.51
120	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_3SO_3H$	H	H	0.195	0.71
121	$C_6H_5CH_2$	CH_3	$CH(CH_3CH_2)$ $CONHNH_2$	H	OCH_3	H	H	24.200	−1.38
122	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3SO_3H$	H	H	0.050	1.30
123	$C_6H_5CH_2$	CH_3	$CH(CH_3(CH_2)_5)$ $CONHNH_2$	H	OCH_3	H	H	101.200	−2.01
124	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	H	$S(CH_2)_3CO_2H$	H	H	0.049	1.31
125	$C_6H_5CH_2$	H	CH_2CONH_2	H	OCH_3	H	H	3.700	−0.57
126	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	NH_2	H	H	1.610	−0.21
127	$C_6H_5CH_2$	CH_3	CH_2CONH_2	OCH_3	H	H	H	2.360	−0.37
128	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	OCH_3	H	H	0.840	0.08
129	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	OCH_3	H	H	0.260	0.59
130	$C_6H_5CH_2$	$CH_3CH_2CH_2$	CH_2CONH_2	H	OCH_3	H	H	38.100	−1.58
131	$C_6H_5CH_2$	Cyclopropyl	CH_2CONH_2	H	OCH_3	H	H	0.336	0.47
132	3-$ClC_6H_4CH_2$	CH_3	CH_2CONH_2	OCH_3	H	H	H	1.380	−0.14
133	3-$ClC_6H_4CH_2$	CH_3	CH_2CONH_2	H	OCH_3	H	H	0.910	0.04
134	$C_6H_5(CH_2)_3$	CH_3	CH_2CONH_2	H	OCH_3	H	H	11.130	−1.05
135	$C_6H_5CH_2$	Cl	CH_2CONH_2	H	OCH_3	H	H	0.250	0.60
136	$C_6H_5CH_2$	Br	CH_2CONH_2	H	OCH_3	H	H	0.230	0.64
137	$C_6H_5CH_2$	Br	CH_2CONH_2	H	OCH_3	Cl	H	0.120	0.92
138	$C_6H_5CH_2$	CH_3S	CH_2CONH_2	H	OCH_3	H	H	0.450	0.35
139	CyclohexylCH_2	CH_3	CH_2CONH_2	H	OCH_3	H	H	2.050	−0.31
140	2-PyridylCH_2	CH_3	CH_2CONH_2	H	OCH_3	H	H	21.360	−1.33
141	$CH_3(CH_2)_9$	CH_3	CH_2CONH_2	H	OCH_3	H	H	9.470	−0.98
142	C_6H_5CO	CH_3	CH_2CONH_2	H	OCH_3	H	H	7.210	−0.86
143	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	OCH_3	H	1.230	−0.09
144	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	NO_2	H	H	1.510	−0.18
145	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	OCH_3	H	H	H	1.180	−0.07
146	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_3$ 5-1H-tetrazoyl	H	H	2.450	−0.39
147	$C_6H_5CH_2$	CH_3	CH_2CONH_2	$OCH_2CONHNH_2$	H	H	H	0.422	0.37
148	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	$OCH_2CONHNH_2$	H	H	H	0.121	0.92
149	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	$O(CH_2)_3$ $CONHNH_2$	H	1.270	−0.10
150	$C_6H_5CH_2$	CH_3	CH_2CONH_2	OCH_2CONH_2	H	H	H	0.360	0.44
151	$C_6H_5CH_2$	CH_3	$COCONH_2$	H	OCH_3	H	H	1.230	−0.09
152	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	OCH_2CONH_2	H	H	H	0.154	0.81

Table I. (Continued)

cpd	R1	R2	R3	R4	R5	R6	R7	IC_{50} (μM)	Log $1/IC_{50}$
153	$C_6H_5CH_2$	CH_3	$COCONH_2$	H	H	OCH_3	H	1.770	−0.25
154	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	H	$O(CH_2)_3CONH_2$	H	5.830	−0.77
155	$C_6H_5CH_2$	SCH_3	CH_2CONH_2	H	$O(CH_2)_3CONH_2$	H	H	0.093	1.03
156	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	OCH_3	H	H	H	0.550	0.26
157	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	NO_2	H	H	H	37.800	−1.58
158	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	H	OCH_3	H	H	2.560	−0.41
159	$C_6H_5CH_2$	CH_3	$CH(OH)CONH_2$	H	OCH_3	H	H	20.060	−1.30
160	$C_6H_5CH_2$	CH_3	$CH(OH)CONH_2$	H	H	OCH_3	H	120.900	−2.08
161	$C_6H_5CH_2$	CH_3CH_2	$CH(OH)CONH_2$	OCH_3	H	H	H	53.000	−1.72
162	$C_6H_5CH_2$	CH_3	$CH_2CH_2CONH_2$	H	OCH_3	H	H	28.100	−1.45
163	$C_6H_5CH_2$	CH_3	CH_2CSNH_2	H	OCH_3	H	H	74.100	−1.87
164	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	$NHCH_2CO_2H$	H	H	H	0.619	0.21
165	$C_6H_5CH_2$	CH_3CH_2	$COCONH_2$	$NHCH_2CO_2H$	H	H	H	1.000	0.00
166	$C_6H_5CH_2$	CH_3	CH_2CN_4H	H	OCH_3	H	H	84.700	−1.93
167	$C_6H_5CH_2$	CH_3	CH_2CNHNH_2	H	OCH_3	H	H	64.700	−1.81
168	$C_6H_5CH_2$	CH_3	$CH_2CONHCH_3$	H	OCH_3	H	H	141.800	−2.15
169	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$NHCH_2CH_2CO_2H$	H	H	3.370	−0.53
170	$C_6H_5CH_2$	CH_3	$CH_2CONHOCH_3$	H	OCH_3	H	H	40.000	−1.60
171	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$NH(CH_2)_3CO_2H$	H	H	1.830	−0.26
172	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$(CH_2)_3CO_2H$	H	H	0.155	0.81
173	$C_6H_5CH_2$	CH_3	$COCONH_2$	$CH=CHCO_2H$	H	H	H	46.000	−1.66
174	$C_6H_5CH_2$	CH_3	$COCONH_2$	$(CH_2)_2CO_2H$	H	H	H	0.145	0.84
175	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$S(CH_2)_2CO_2H$	H	H	0.059	1.23
176	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$S(CH_2)_3CO_2H$	H	H	0.023	1.64
177	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$S(CH_2)_3CO_2H$	CH_3	H	0.022	1.66
178	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$S(CH_2)_3CO_2H$	i-Pr	H	0.033	1.48
179	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	H	CH_3	0.075	1.12
180	$C_6H_5CH_2$	CH_3	CH_2CONH_2	H	$S(CH_2)_3CO_2H$	H	CH_3	0.051	1.29
181	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	i-Pr	H	0.040	1.40
182	$C_6H_5CH_2$	Br	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	Cl	H	0.020	1.70
183	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3CO_2H$	CH_3	H	0.033	1.48
184	$C_6H_5CH_2$	CH_3CH_2	CH_2CONH_2	H	$O(CH_2)_3PO(OH)_2$	CH_3	H	0.015	1.82
185	$C_6H_5CH_2$	CH_3	$CH_2CONHNH_2$	H	$O(CH_2)_3CO_2H$	H	H	1.020	−0.01
186	$C_6H_5CH_2$	CH_3	$CH_2CH_2CONH_2$	H	$O(CH_2)_3CO_2H$	H	H	6.930	−0.84
187	$C_6H_5CH_2$	CH_3	CH_2CO_2H	H	$O(CH_2)_3CO_2H$	H	H	25.900	−1.41
188	$3\text{-}ClC_6H_4CH_2$	CH_3CH_2	$CONH_2$	OCH_2CO_2H	H	H	H	16.800	−1.23

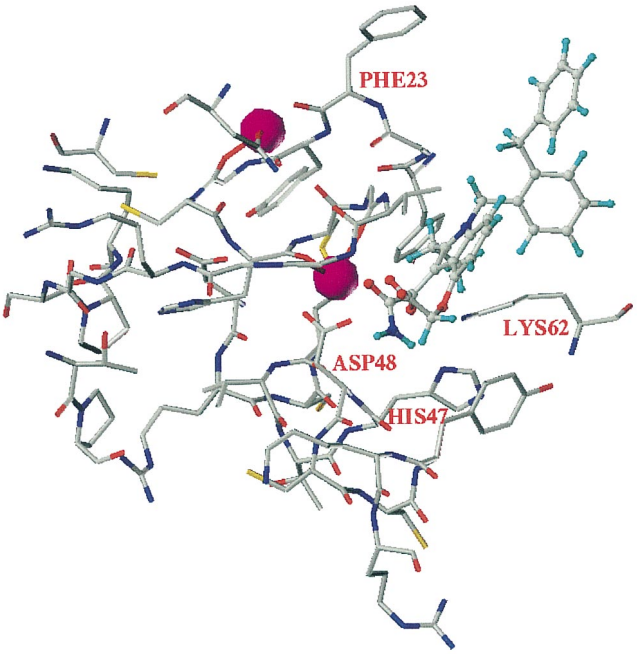


Figure 2. Enzyme/ligand complex obtained by automated docking between indole inhibitor **36** (ball and stick format) and hnps-PLA_2 (stick format).

3. Results and discussion

3.1. Docking

In the present section, a few key inhibitor structures that demonstrate the most characteristic ligand–protein interactions will be discussed. The attention will be focused more especially on compound **36**, the most active one in the series.

The biological activity data for these 188 hnps-PLA_2 are homogenously distributed on the activity scale, ranging from 4 nM for compound **36** up to 141.8 μM for compound **168**. Within this range, all the compounds were considered as being able to enter the active site of the enzyme and to interfere with it, via the same mechanism.

The automated docking procedure was firstly applied to the four compounds already crystallized with the enzyme and described by Schevitz et al. [33]. These four compounds, i.e. **55**, **107**, **110** and **128**, were docked in order to initiate our automated docking procedure. As mentioned above, the crystallographic data showed two interactions between two oxygen atoms of the ligand and the calcium atom of the protein. The automated docking procedure was applied with these two oxygen atoms as anchor points. Moreover, compound **110** possesses four oxygen atoms as possible anchor points, the three oxygen atoms of the phosphonate group and the oxygen atom of the amide group. Thus, for this compound, eight solutions were tested. Among these eight conformations, two identical solutions were in agreement with the crystallographic data of compound **110**. In these two conformations, the calcium atom is bonded to the oxygen amide group and to one oxygen of the phosphoryl group. The binding energy for these conformations is about -95 kcal mol^{-1} whereas the other conformations, unable to link twice to the calcium atom, have a lower binding energy (-30 to -60 kcal mol^{-1}). Compounds **55** and **128** present the same results. One remark concerns compound **107**. For the docking procedure of this compound, the calcium atom of the protein was removed, and then the carboxyl group directly interacts with the Asp48 residue [33].

This docking procedure was then applied for all the compounds in the series. *Figure 2* presents the most active compound in the series, compound **36**, docked to the hnps-PLA_2. This study revealed the three already described interactions with the enzyme, i.e. two interactions with the calcium atom and one hydrogen bond (HB) between residue His47 and the NH moiety of the amide group. In addition, another interaction was revealed between one hydrogen of the amide group and the carboxyl group of residue Asp48. This is a HB with a distance of about 1.8 Å between the acceptor and the donor. *Figure 3* shows this new interaction. This region is particularly sensitive to modifications because the substitution of a methyl group for one amide hydrogen decreases the activity about 170 times when considering, for example, compounds **128** ($CONH_2$) and **168** ($CONHCH_3$). The decrease in activity is probably due to electrostatic factors because no steric hindrance was revealed between the residues of the enzyme and the methyl group of compound **168**. Compound **89** ($CONHNH_2$) is in agreement with this fact. Some authors [47] already suggested the importance of electrostatic forces in the binding of secretory PLA_2 to substrate aggregates. Thus, electrostatic forces are very important in the substitutions occurring at R3, R4 and R5 on indole inhibitors.

The other positions (R1, R2, R6 and R7) on indole inhibitors consist in steric adjustments of the inhibitor to the protein. *Figures 4* and *5* show the active conformation of compound **36** compared to the conformations of the three ligands extracted from the crystallographic complexes. Compound **36** adopts the same alignment.

The three aromatic groups of compound **36** are fitted with the aromatic groups of the highly potent inhibitor from reference 13. Moreover, each ligand forms a loop and the two extremities of this loop end with the two oxygen atoms binding the calcium atom of the protein. The protein-based alignment of all the ligands allows for the definition of a volume with a size of about 10×10×15 Å. The R2 substitution is directed toward the center of the enzyme whereas the R6 and R7 substitutions are directed toward the exterior of the enzyme. The R2 position does not accept large substituents as shown by compounds **5** and **7**, **25** and **36** or **92** and **93**. The substitution of a n-propyl group for a methyl group decreases the activity 10–100 times. The direction toward the outside of the protein for the R6 and R7 substitutions explains the small variations in activity when large substituents are in these positions. Compounds **4**, **13** and **14** or **176**, **177** and **178** reflect this fact for position R6. The substitution of a large $O(CH_2)_3CO_2CH_2CH_3$ group in compound **13** for the R6 hydroxyl group of compound **4** does not modify the

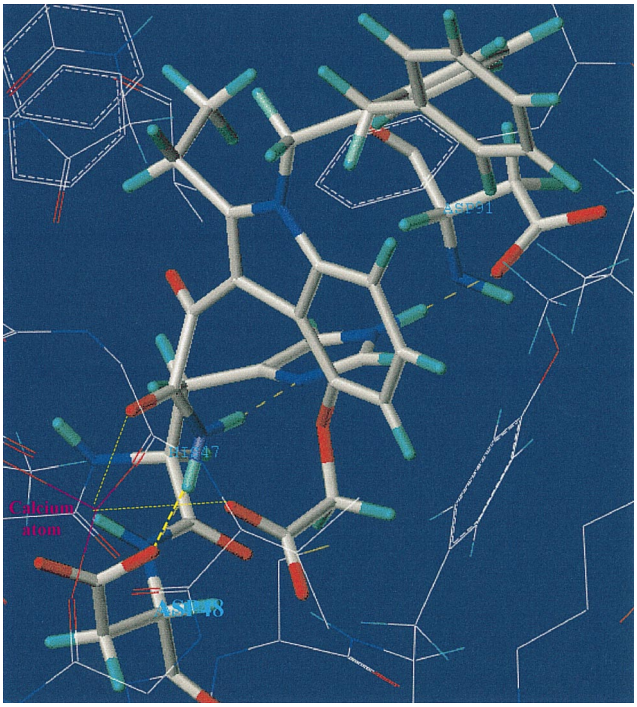


Figure 3. Hydrogen bond (HB) network, shown with a yellow dash line, binding indole inhibitor **36** to hnps-PLA_2. A new HB, indicated in bold, was found between the ligand and the enzyme residue Asp48.

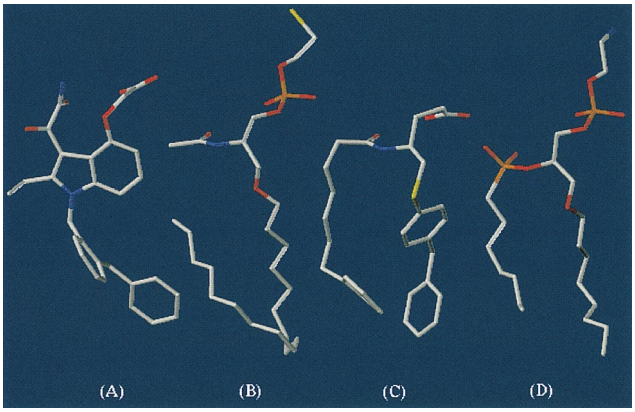


Figure 4. Active conformations of four ligands, A, B, C and D, of hnps-PLA_2. (A) Indole inhibitor **36** derived by our automated docking procedure; (B), (C) and (D) Three ligands extracted from the crystallographic complexes of references [30], [13] and [11], respectively.

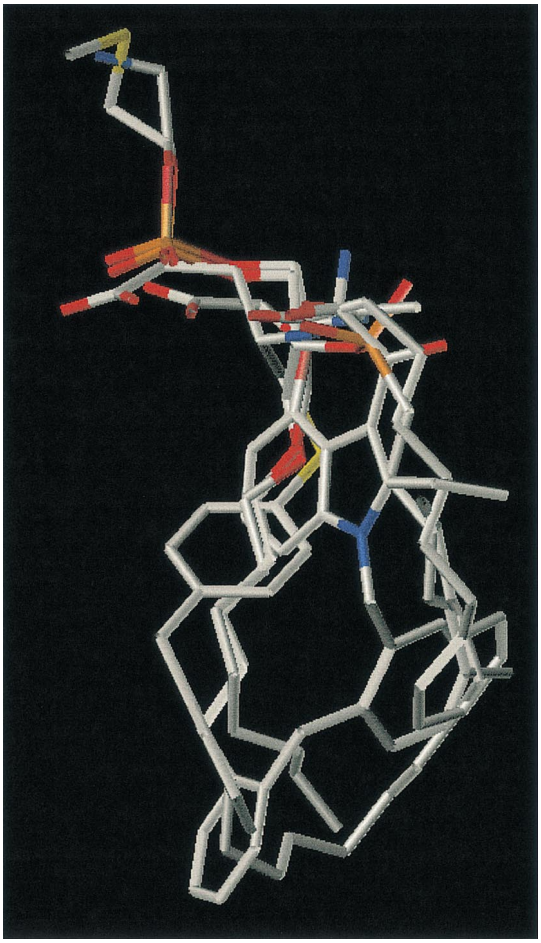


Figure 5. Protein-based alignment of the four ligands of hnps-PLA_2 presented in *figure 4*.

activity more than 10 times. The same remark can be made for compounds **176** and **180** about the R7 position.

As for the R1 position, it corresponds to a key position for governing the activity, especially due to steric effects. The substitution of a benzyl group in compounds **34** for a R1 hydrogen atom in compound **21** increases the activity more than 100 times. In addition, R1 substitutions of different alkyl groups also favor an increase in activity (see compounds **55** with **69** or **70** and compounds **24** with **33**). Finally, the different halogen substitutions in R1 do not considerably affect the activity (see compounds **2**, **11**, **34** and **37** ...).

3.2. CoMFA

The originality of the present CoMFA study is supported by the fact that each of the 188 indole inhibitors studied was aligned in an independent way. No reference compound was used and the geometries of the inhibitors were only determined by interactions with the biological receptor. The automated docking study showed that, despite the great flexibility of the molecules studied, the choice of the suitable active conformations was not ambiguous. Indeed, in most cases the conformational search yielded only a single conformation (or a single conformational family) for the ligand–protein complex whose energy was much lower than that of the other candidate complexes. Hence, it may be suggested once more that the active site of the protein is quite selective and does not permit the conformational mobility of the bonded inhibitor.

The protein-based alignment may have an impact on the nature of the structural information which we expect to obtain from the CoMFA analysis. A good statistic and predictive quality of the CoMFA model normally suggests that the molecules in the real biological system are aligned in accordance with the initial alignment of the CoMFA procedure. In the present case, as the alignment was prepared using the structure of the receptor as a template, the good quality of the CoMFA model would suggest the validity of the proposed model of inhibitor–enzyme interactions obtained by the automated docking procedure.

The separation of the series into a training set and a test set was realized as described in the methods section. The use of SOM for the analysis of the molecular diversity in both the training set and the test set, presented in *figure 6*, indicated that the selection of each third compound allowed the selection of a test set with a good level of molecular diversity. The 188 compounds occupied all the map and the test set was homogenously

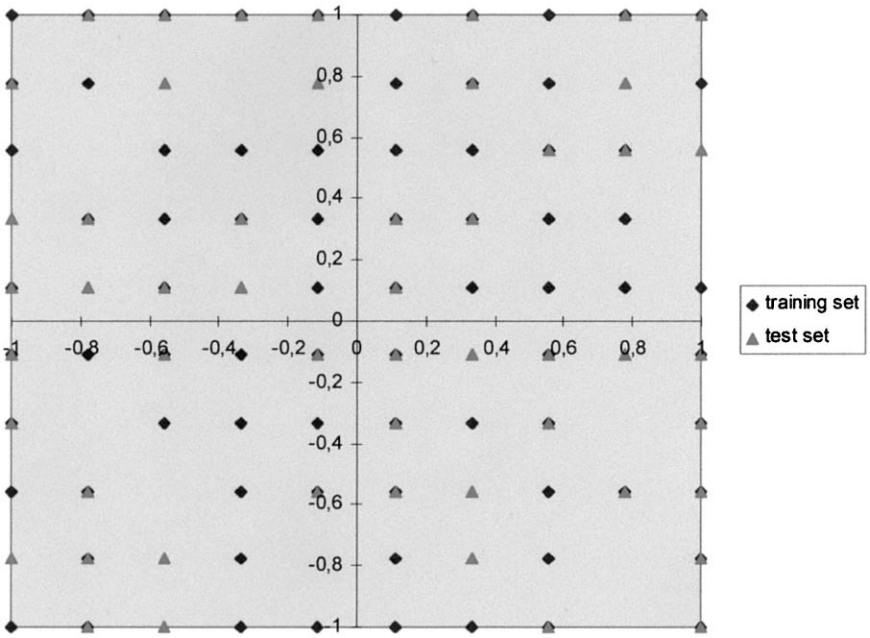


Figure 6. Evaluation of molecular diversity for the training and test sets dealing with 188 inhibitors of hnps-PLA_2. Distribution within the nodes of a self-organizing map (SOM).

Table II. Statistics and cross-validation results of the three CoMFA models.

		Steric	Electrostatic	Steric and electrostatic
N			126	
'leave-one-out'	n_{PC}	3	6	6
	Q^2	0.49	0.62	0.63
'leave-some-out' (10 groups)	n_{PC}	4	6	6
	Q^2	0.47	0.59	0.60
R^2		0.72	0.90	0.97
F		106	174	365
sd		0.63	0.39	0.22

distributed on the map amidst the training set. Moreover, some test set compounds fell into nodes not represented by the training set, e.g. compounds **21**, **30** or **99**. This would suggest that that these test set compounds possessed some structural particularities not found in the training set compounds. For example, only compound **21** has a hydrogen atom in the R1 position. In the other compounds, the R1 position corresponds to a bulky hydrophobic group. A good prediction of the activity for such compounds expresses the quality of the model and its extrapolation capabilities.

For a better understanding of the factors that underlie the activity, three different CoMFA models were derived: (i) a model with steric field only, (ii) a model with electrostatic field and (iii) a model taking both fields into account. The results of these analyses are presented in *table II*, which underlines the important role of the CoMFA electrostatic field compared to that of the steric field. This remark is in agreement with the previous discussion in the 'Docking' section and with reference 47 about the electrostatic forces in the ligand binding with the hnps-PLA_2. Indeed, the analysis including both fields shows that the relative contributions to the model are 37% for the steric field and 63% for the electrostatic one. Moreover, the statistic and predictive criteria for the analysis including steric field are significantly lower than the corresponding criteria for the other two analyses. A superiority of the electrostatic contribution means, however, that the process of the hnps-PLA_2 inhibition is mainly guided by electrostatic forces. Nevertheless, in the present study, the model comprising both steric and electrostatic fields was taken to plot the CoMFA statistic fields and to predict the hnps-PLA_2 inhibitory activity for new compounds.

Finally, a non-cross-validated PLS run was performed with six principal components in order to generate the final CoMFA plots. This value was selected as the addition of any new component adds less than 5% to the Q^2 coefficient. The CoMFA plots outline a statistic field expressing the relationship between the variation of the steric and electrostatic fields, and the variation of the biological activity. The values of the fields are calculated at each lattice intersection and are equal to the product of the descriptor coefficient by the corresponding standard deviation (STDEV*COEFF). Hence, an extremely low value of STDEV*COEFF indicates that the presence of the corresponding steric or electrostatic field is not desirable in this point because this causes a decrease in activity. New molecules should not contain fragments that generate such a field in the lattice intersections with low STDEV*COEFF. A high value of STDEV*COEFF means that the corresponding field is desirable in this point and that the presence of fragments producing such a field favors the activity. A STDEV*COEFF plot for the electrostatic field is shown in *figure 7* and a plot for the steric field is presented in *figure 8*.

Compound **115** for *figure 7* and compound **36** for *figure 8* illustrate the main features of the CoMFA plots. Some of the colored regions on the plots mark essential ligand–protein interactions. In *figure 7*, the major electrostatic variations are registered in the R3, R4 and R5 positions. For instance, two red spheres present in the region between two phosphoryl oxygen atoms of the inhibitor suggest that there is always a negatively charged inhibitor group in these regions. The first negatively charged group is able to form an electrostatic bond with the calcium atom whereas the second oxygen atom is directed toward Lys62. The distance between this oxygen atom and the cation group of Lys62 is 5.12 Å. This result is in agreement with Schevitz et al. [33] (5 Å), which suggested the presence of a water molecule interacting between them. The blue and largest volume covering the R3 amine group explains the importance of HB donors with the residues His47 and Asp48. *Figure 8* expresses that structural variations at the R6 and R7 positions do not significantly influence the steric field.

This result is in agreement with the docking study, being the positions oriented toward the outside of the protein. Two yellow volumes beside the R3 amino group indicate the position of His47 and Asp48, where no bulky substitution is possible. On the contrary, the region around the R4 and R5 positions is favorable to the bulky substituents. Substitutions exhibit unfavorable steric effect near the R2 position around the ethyl group. Nevertheless, a green volume placed in the axes of the indole plane seems favorable to the activity. It will be interesting to further investigate this R2 position because this position looks very critical. Yellow zones around R1 position correspond to the presence of residues of the protein. This explains why some compounds, such as compound **96**, having substituents in this zone are not active.

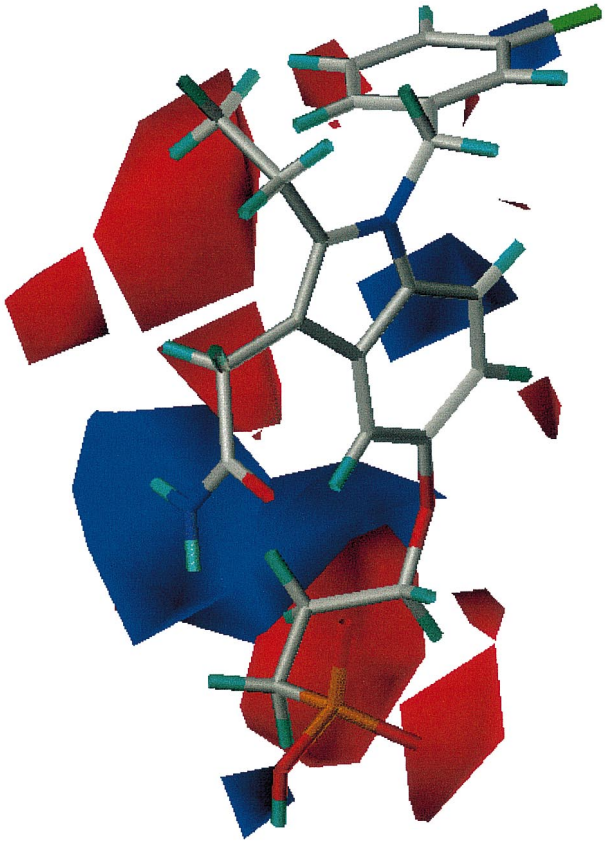


Figure 7. CoMFA electrostatic STDEV*COEFF field plot. Increasing negative charge inside red regions and increasing positive charge in blue regions favor the inhibitory activity. Compound **115** is shown.

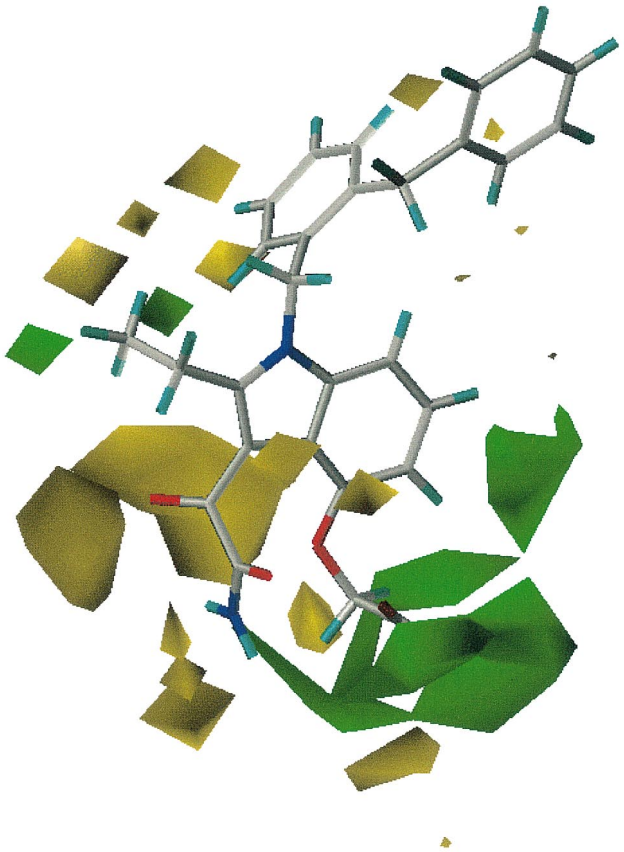


Figure 8. CoMFA steric STDEV*COEFF field plot. Increasing bulk inside green regions and removing bulk from yellow regions favor the inhibitory activity. Compound **36** is shown.

3.3. Predictive aspect of the CoMFA model

After validating our model by means of cross-validation, the next step of the investigation consisted in applying the model to a test set of inhibitors independently from the training set with which the model was established.

Compound numbers in the test set are referring to *table I*; predicted and real activity values with the corresponding deviations for the above 62 indole inhibitors are shown in *table III*. *Figure 9* illustrates how tightly the predicted values are correlated with the actual activity values ($R = 0.90$). It means that the relative inhibitory capacities are correctly predicted for the whole series of 62 molecules. The parameter that is widely used to estimate the quality of test set predictions, i.e. PRESS/SSY [48], has a good value of 0.18, which indicates no important deviations of the predicted

Table III. Predicted activity of the 62 hnps PLA_2 indole inhibitors with the 3D QSAR model.

Compound	Experimental log $1/IC_{50}$	Predicted log $1/IC_{50}$	Δ (Exp-Pred)
3	−0.01	−0.33	0.32
6	0.84	0.94	−0.10
9	1.07	0.28	0.79
12	−0.14	0.56	−0.70
15	−0.10	0.29	−0.39
18	−0.32	−1.02	0.70
21	−0.20	0.44	−0.64
24	1.96	1.72	0.24
27	1.37	1.48	−0.11
30	2.05	1.91	0.14
33	2.10	1.18	0.92
36	2.40	1.95	0.45
39	1.09	0.30	0.79
42	−0.21	−0.79	0.58
45	1.34	1.80	−0.46
48	0.85	0.97	−0.12
51	1.28	1.18	0.10
54	0.28	−0.24	0.52
57	0.83	0.16	0.67
60	−0.90	−0.80	−0.10
63	0.26	0.68	−0.42
66	1.14	0.95	0.19
69	0.18	0.69	−0.51
72	−0.69	−0.84	0.15
75	0.03	−0.56	0.59
78	−1.22	−0.26	−0.96
81	−0.48	−0.67	0.19
84	0.37	0.14	0.23
87	−0.37	−0.60	0.23
90	−0.01	−0.50	0.49
93	−1.85	−1.24	−0.61
96	−1.60	−1.06	−0.54
99	−1.18	−0.59	−0.59
102	−0.13	−0.59	0.46
105	0.19	−0.64	0.83
108	0.69	1.09	−0.40
111	1.64	1.59	0.05
114	1.48	1.22	0.26
117	1.40	1.26	0.14
120	0.71	0.25	0.46
123	−2.01	−1.44	−0.57
126	−0.21	−0.15	−0.06
129	0.59	0.12	0.47
132	−0.14	−0.17	0.03
135	0.60	0.22	0.38
138	0.35	0.32	0.03
141	−0.98	−0.15	−0.83
144	−0.18	0.04	−0.22
147	0.37	−0.24	0.61
150	0.44	0.19	0.25
153	−0.25	−0.72	0.47
156	0.26	−0.05	0.31
159	−1.30	−1.50	0.20
162	−1.45	−1.34	−0.11
165	0.00	0.45	−0.45
168	−2.15	−1.65	−0.50
171	−0.26	0.28	−0.54
174	0.84	0.60	0.24
177	1.66	1.75	−0.09
180	1.29	1.48	−0.19
183	1.48	1.27	0.22
186	−0.84	−1.42	0.58

values from the actual ones. Clementi and Wold suggested that: 'In a reasonable QSAR model, PRESS/SSY should be smaller than 0.4' [48].

Two conclusions may be derived from this part of the study. First, the tight correlation between the predicted and experimentally observed values in this study suggests that the present model is able to provide reliable predictions of the hnps-PLA_2 inhibitory capacities in a set of new inhibitors. The good correlation between predicted and observed values also indicates a good predictive capacity of the present CoMFA model. The second conclusion is that the good statistical parameters of the CoMFA model and its good predictive capacities allow to validate the inhibitor conformations selected by our docking procedure.

3.4. Global CoMFA model and hnps-PLA_2 pharmacophore

A new CoMFA model, including all the 188 hnpsPLA_2 indole inhibitors, was elaborated. The statistical results of this model and the plot showing the calculated versus experimental activity data are indicated in *table IV* and *figure 10*, respectively. These good results, validating the protein-based alignment of the inhibitors, were used to define the pharmacophore of the hnps-PLA_2.

Two regions, implicated in the structure–activity relationships, were identified on these indole inhibitors. The R3, R4 and R5 positions on the indole group correspond to the hydrophilic moiety which is essentially implicated in electrostatic forces with the residues of the protein, whereas R1, R2 for a major contribution and R6, R7 for a minor contribution are implicated in hydrophobic interactions with the protein. The aromatic group in the R1 position of compound **36** is surrounded by four hydrophobic residues: Val3, Leu19, Phe23 and Phe63.

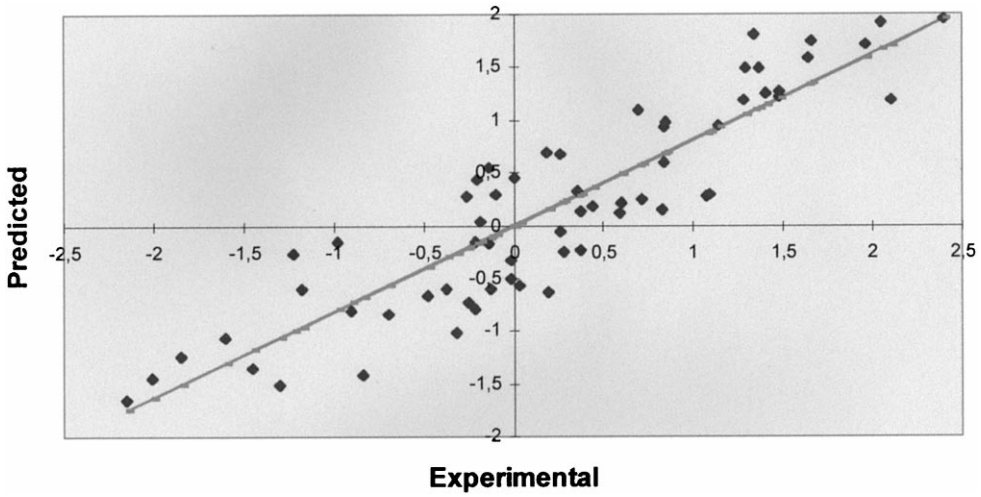


Figure 9. Experimental versus predicted inhibitory activity values on the 62 hnps-PLA_2 indole inhibitors. The corresponding structures are indicated in *table III*.

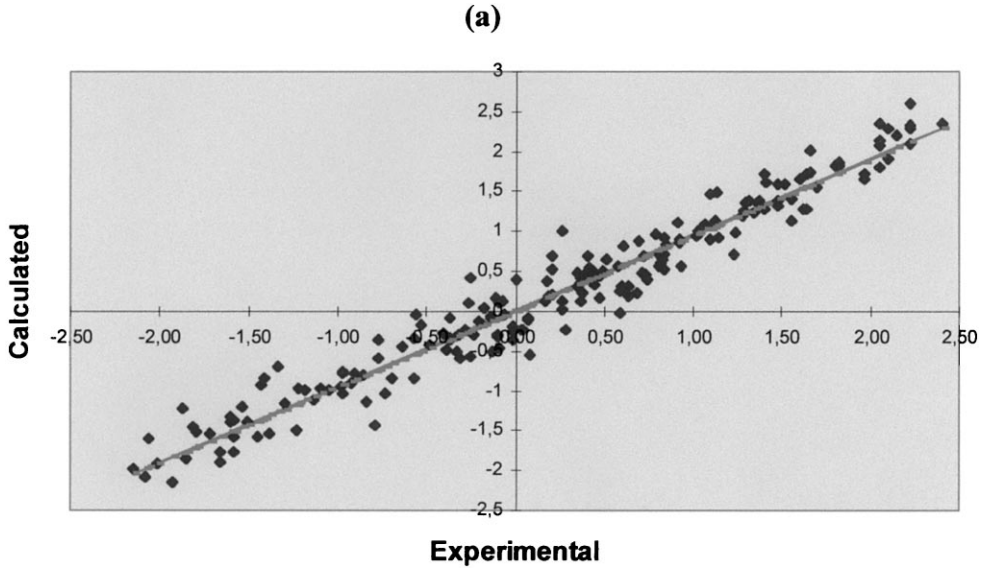


Figure 10. Plot of the experimental against calculated inhibitory activity values for the CoMFA model including the 188 hnps-PLA_2 indole inhibitors.

The pharmacophore taking docking data and CoMFA data into account is presented in *figure 11*. The protein-based alignment of four ligands of hnps-PLA_2, compound **36** and the three ligands extracted from the crystallographic complexes [11, 13, 30], is shown. The pharmacophore elements are represented in black, red and blue for the aromatic rings, the HB acceptors and the HB donors, respectively. The distances between each of the pharmacophore elements are indicated with dotted lines. Circles A, B and C indicate the acceptable variations of the HB acceptors in the pharmacophore elements. The radii of the circles are, respectively, 0.7, 0.6 and 0.5 Å. The two little spheres D and E around hydrogen atoms indicate the small variations found in all the docked indole inhibitors. These small variations underline the importance of these points. Volume F, described by two ligands not included in the model and nevertheless in agreement with the CoMFA plots, ac-

cepts substitutions. Volume G is also able to accept some substitutions (seen also through the CoMFA plots), but these substitutions must be inferior to 2 or 3 Å. The R2 methyl, ethyl or cyclopropyl groups are able to fit the pocket around volume G, but the propyl group contributes to the decrease in activity (see compounds **5** and **7** or **92** and **93**), because its only possible solution will be to adopt the alkyl chain position inside sphere G, *figure 11*. Nevertheless, this R2 propyl conformation is sterically unfavorable due to the presence of the R1 substituent.

Table IV. Statistical results of the global CoMFA model including the 188 hnps PLA_2 indole inhibitors.

	Steric and electrostatic fields
N	188
n_{PC}	8
Q^2	0.75
R^2	0.95
F	421
sd	0.26

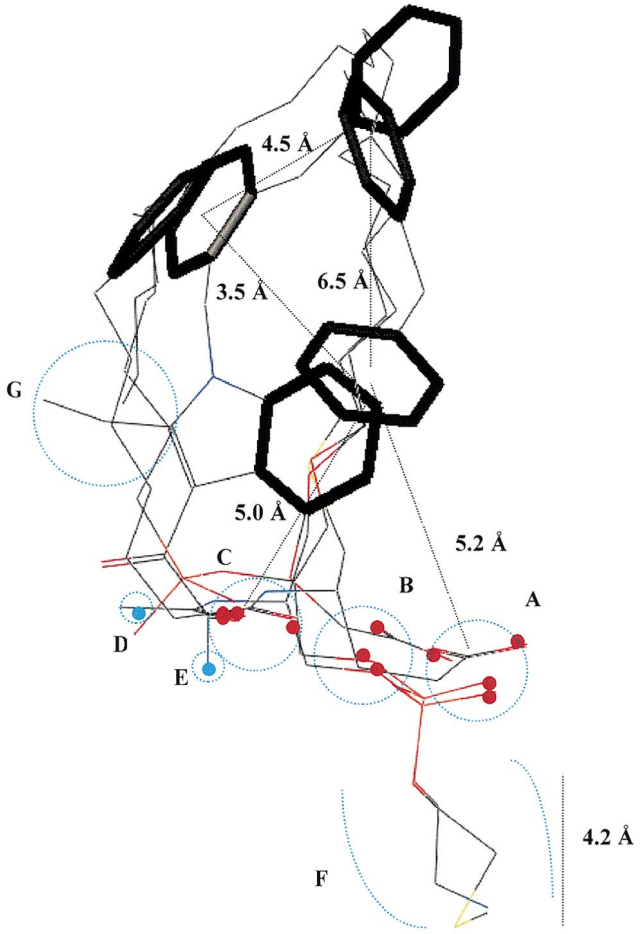


Figure 11. Pharmacophore of the hnps-PLA_2. The pharmacophore elements are shown in black, red and blue for aromatic rings, hydrogen bond (HB) acceptors and HB donors, respectively.

4. Conclusion

An automated docking study was performed on a series of 188 competitive hnps-PLA_2 inhibitors with the potential of being drugs against inflammatory diseases. The protein spatial constraints derived from the crystallographic data were initially imposed in order to reduce the search space. In view of these constraints, a reliable docking was performed for all the 188 molecules studied with reasonable computational expenses.

The present docking is in agreement with some previous studies [33] on the existence of interactions with the enzyme, especially with the calcium atom or with His47. A new interaction between the NH group of R1 indole acetamide or glyoxamide groups and residue Asp48 was suggested. It was also shown that inhibitors, as well as natural substrate phospholipides, exhibit two main parts: (i) a hydrophobic moiety with alkyl groups and/or aromatic rings and (ii) a hydrophilic moiety creating HB interactions with residues of the PLA_2 active site.

The docking data were then used to perform a CoMFA. The 188 hnps-PLA_2 inhibitors were divided into two sets, one with 126 compounds for the model, the other with 62 compounds for the validation of the model. The CoMFA model with the protein-based alignment has a good predictive capacity according to the cross-validation test ($Q^2 = 0.63$). The CoMFA plots are in agreement with the docking results and could be used efficiently for the design of new effective inhibitors. The inhibitory capacities were correctly predicted for the 62 molecules of the test set. Finally, the pharmacophore of hnps-PLA_2, in agreement with the complete 3D QSAR model and the docking study, was elaborated.

This robust and predictive model of the hnps-PLA_2 inhibitory activity might further be used for the selection of new active molecules and would also bear witness to the validity of the conformations selected by the docking study. Work is underway to develop a

virtual screening of a large database of natural substances on this modeled hnps-PLA_2.

Acknowledgements

We thank the French Ministère de l'Education Nationale de la Recherche et de la Technologie and the European Commission for their support within the Eureka Program Cosmactiv EU-1748.

References

[1] Bernard P., Berthon J.-Y., Chrétien J.R., Curr. Opin. Drug Disc. Devel. 2 (1999) 213–223.

[2] Irvine R.F., Biochem. J. 204 (1982) 3–10.

[3] Mayer R.J., Marshall L.A., FASEB J. 7 (1993) 339–348.

[4] Dennis E.A., Rhee S.G., Billah M.M., Hannun Y.A., FASEB J. 5 (1991) 2068–2077.

[5] Clark J.D., Milona N., Knopf J.L., Proc. Natl. Acad. Sci. USA 87 (1990) 7708–7712.

[6] Kramer R.M., Hession C., Johansen B., Hayes G., McGray P., Chow E.P., Tizard R., Pepinski R.B., J. Biol. Chem. 264 (1989) 5768–5775.

[7] (a) Hara S., Kudo I., Chang H.W., Matsuta K.J., Biochem. 105 (1989) 395–399. (b) Lai C.-Y., Wada K., Biochem. Biophys. Res. Commun. 157 (1988) 488–493. (c) Bomalaski J.S., Lawton P., Browning J.L., J. Immunol. 146 (1991) 3904–3910.

[8] Gronroos J.M., Nevalainen T.J., Digestion 52 (1992) 232–236.

[9] (a) Green J.A., Smith G.M., Buchta R., Lee R., Ho K.Y., Rajkovic I.A., Scott K.F., Inflammation 15 (1991) 355–367. (b) Santos A.A. et al., Ann. Surg. 219 (1994) 183–192. (c) Vadas P., Pruzanski W., Circ. Shock 39 (1993) 160–167.

[10] Uhl W., Buchler M., Nevalainen T.J., Deller A., Beger H.G., Trauma J. 30 (1990) 1285–1290.

[11] Scott D.L., White S.P., Browning J.L., Rosa J.J., Gelb M.H., Sigler P.B., Science 254 (1991) 1007–1010.

[12] Wery J.P., Schevitz R.W., Clawson D.K., Bobbitt J.L., Dow E.R., Gamboa G., Goodson T.J.R., Hermann R.B., Kramer R.M., McClure D.B., et al., Nature 352 (1991) 79–82.

[13] Cha S.S., Lee D., Adams J., Kurdyla J.T., Jones C.S., Marshall L.A., Bolognese B., Abdel-Meguid S.S., Oh B.H., J. Med. Chem. 39 (1996) 3878–3881.

[14] (a) Teshiro I., Matsutani S., Shirahase K., Fujii Y., Yoshida T., Tanaka K., OhtaniM., J. Med. Chem. 39 (1996) 5183–5191. (b) Baba A., Kawamura N., Makino H., Ohta Y., Taketomi S., Sohda T., J. Med. Chem. 39 (1996) 5176–5182. (c) De Rosa M., Giordano S., Scettri A., Sodano G., Soriente A., Garcia Pastor P., Alcaraz M.J., Paya M., J. Med. Chem. 41 (1998) 3232–3238.

[15] Noel J.P., Bingman C.A., Deng T.L., Dupureur C.M., Hamilton K.J., Jiang R.T., Kwak J.G., Sekharudu C., Sundaralingam M., Tsai M.D., Biochemistry 30 (1991) 11801–11811.

[16] (a) Ortiz A.R., Pisabarro M.T., Gallego J., Gago F., Biochemistry 31 (1992) 2887–2896. (b) Pisabarro M.T., Ortiz A.R., Palomer A., Cabre F., Garcia L., Wade R.C., Gago F., Mauleon D., Carganico G., J. Med. Chem. 37 (1994) 337–341.

[17] Bennion C., Connoly S., Gensmantel N.P., Hallam C., Jackson C.G., Primrose W.U., Roberts G.C., Robinson D.H., Slaich P.K., J. Med. Chem. 35 (1992) 2939–2951.

[18] Sessions R.B., Dauber-Osguthorpe P., Campbell M.M., Osguthorpe D.J., Proteins 14 (1992) 45–64.

[19] Tomoo K., Yamane A., Ishida T., Fujii S., Ikeda K., Iwama S., Katsumura S., Sumiya S., Miyagawa H., Kitamura K., Biochim. Biophys. Acta 1340 (1997) 178–186.

[20] Thunnissen M.M., Kalk K.H., Drenth J., Dijkstra B.W., J. Mol. Biol. 216 (1990) 425–439.

[21] Hariprasad V., Kulkarni V.M., J. Mol. Recognit. 9 (1996) 95–102.

[22] Ortiz A.R., Pisabarro M.T., Gago F., Wade R.C., J. Med. Chem. 38 (1995) 2681–2691. Published erratum appears in additions and corrections, J. Med. Chem. 40 (1997) 4168.

[23] Ortiz A.R., Pastor M., Palomer A., Cruciani G., Gago F., Wade R.C., J. Med. Chem. 40 (1997) 1136–1148.

[24] Kuntz I.D., Meng E.C., Shoichet B.K., Acc. Chem. Res. 27 (1994) 117–123.

[25] Cramer R.D. III, Patterson D.E., Bunce J.D., J. Amer. Chem. Soc. 110 (1988) 5959–5967.

[26] White S.P., Scott D.L., Otwinowski Z., Gelb M.H., Sigler P.B., Science 250 (1990) 1560–1563.

[27] Scott D.L., Otwinowski Z., Gelb M.H., Sigler P.B., Science 250 (1990) 1563–1566.

[28] Noel J.P., Bingman C.A., Deng T.L., Dupureur C.M., Hamilton K.J., Jiang R.T., Kwak J.G., Sekharudu C., Sundaralingam M., Tsai M.D., Biochemistry 30 (1991) 11801–11811.

[29] Thunnissen M.M., Ab E., Kalk K.H., Drenth J., Dijkstra B.W., Kuipers O.P., Dijkman R., de Haas G.H., Verheij H.M., Nature 347 (1990) 689–691.

[30] Oh B.-H. (31 July 95, PDB code: 1AYP). To be published.

[31] SYBYL is available from Tripos Associates, 1699 South Hanley Road, St Louis, MO 63144.

[32] Clark M., Cramer R.D. III, Opdenbosch N.V., J. Comput. Chem. 10 (1989) 982–1012.

[33] Schevitz R.W., Bach N.J., Carlson D.G., Chirgadze N.Y., Clawson D.K., Dillard R.D., Draheim S.E., Hartley L.W., Jones N.D., Mihelich E.D., Olkowski J.L., Snyder D.W., Sommers C., Wery J.-P., Nat. Struct. Biol. 2 (1995) 458–465.

[34] Bernard P., Kireev D.B., Chretien J.R., Fortier P.-L., Coppet L., J. Comput.-Aided Mol. Des. 13 (1999) 355–371.

[35] (a) Dillard R.D., Bach N.J., Draheim S.E., Berry D.R., Carlson D.G., Chirgadze N.Y., Clawson D.K., Hartley L.W., Johnson L.M., Jones N.D., McKinney E.R., Mihelich E.D., Olkowski J.L., Schevitz R.W., Smith A.C., Snyder D.W., Sommers C.D., Wery J.-P., J. Med. Chem. 39 (1996) 5119–5136. (b) Dillard R.D., Bach N.J., Draheim S.E., Berry D.R., Carlson D.G., Chirgadze N.Y., Clawson D.K., Hartley L.W., Johnson L.M., Jones N.D., McKinney E.R., Mihelich E.D., Olkowski J.L., Schevitz R.W., Smith A.C., Snyder D.W., Sommers C.D., Wery J.-P., J. Med. Chem. 39 (1996) 5137–5158. (c) Draheim S.E., Bach N.J., Dillard R.D., Berry D.R., Carlson D.G. Chirgadze N.Y., Clawson D.K., Hartley L.W., Johnson L.M., Jones N.D., McKinney E.R., Mihelich E.D., Olkowski J.L., Schevitz R.W., Smith A.C., Snyder D.W., Sommers C.D., Wery J.-P., J. Med. Chem. 39 (1996) 5159–5175.

[36] Reynolds L.J., Hughes L.L., Dennis E.A., Anal. Biochem. 204 (1992) 190–197.

[37] Kohonen T. (Ed.), Self-Organization and Associative Memory, Springer-Verlag, Berlin, 1988.

[38] Bernard P., Golbraikh A., Kireev D.B., Chrétien J.R., Rozhkova N., Analusis 26 (1998) 333–341.

[39] Kier L.B., Hall L.H. (Eds.), Molecular Connectivity in Structure-Activity Analysis, John Wiley and Sons, New York, 1986.

[40] Gutman I., Ruscic B., Trinajstic N., Wilcox Jr C.F., J. Chem. Phys. 62 (1975) 3339–3405.

[41] Sabljic A., in: Karcher W., Devillers J. (Eds.), Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and toxicology, Kluwer Academic Publishers, Dordrecht, 1990, pp. 61–82.

[42] Sabljic A., in: Karcher W., Devillers J. (Eds.), Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Kluwer Academic Publishers, Dordrecht, 1990, pp. 83–103.

[43] Hansch C., Leo A. (Eds.), Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley Interscience Publications, 1979, p. 13.

[44] Sanderson R.T. (Ed.), Chemical Bonds and Bond Energy, Academic Press, New York, 1976.

[45] Dewar M.J.S., Zoebich E.G., Healy E.F., Stewart J.J.P., J. Amer. Chem. Soc. 107 (1985) 3902–3909.

[46] Wold S., Eriksson L., in: Waterbeemd H. (Ed.), Chemometric Methods in Molecular Design, VCH, Weinheim, 1995, pp. 309–318.

[47] Scott D.L., Mandel A.M., Sigler P.B., Honig B., Biophys. J. 67 (1994) 493–504.

[48] Clementi S., Wold S., in: Waterbeemd H. (Ed.), Chemometric Methods in Molecular Design, VCH, Weinheim, 1995, pp. 324–325.