Molecular docking and 3D-quantitative structure activity relationship analyses of peptidyl vinyl sulfones: *Plasmodium Falciparum* cysteine proteases inhibitors

**Cátia Teixeira · José R. B. Gomes ·
Thierry Couesnon · Paula Gomes**

Abstract Comparative molecular field analysis (CoM-FA) and comparative molecular similarity indices analysis (CoMSIA) based on three-dimensional quantitative structure–activity relationship (3D-QSAR) studies were conducted on a series (39 molecules) of peptidyl vinyl sulfone derivatives as potential *Plasmodium Falciparum* cysteine proteases inhibitors. Two different methods of alignment were employed: (i) a receptor-docked alignment derived from the structure-based docking algorithm GOLD and (ii) a ligand-based alignment using the structure of one of the ligands derived from a crystal structure from the PDB databank. The best predictions were obtained for the receptor-docked alignment with a CoMFA standard model ($q^2 = 0.696$ and $r^2 = 0.980$) and with CoMSIA combined electrostatic, and hydrophobic fields ($q^2 = 0.711$ and $r^2 = 0.992$). Both models were validated by a test set of nine compounds and gave satisfactory predictive r^2_{pred} values of 0.76 and 0.74, respectively. CoMFA and CoM-SIA contour maps were used to identify critical regions where any change in the steric, electrostatic, and hydrophobic fields may affect the inhibitory activity, and to highlight the key structural features required for biological activity. Moreover, the results obtained from 3D-QSAR analyses were superimposed on the *Plasmodium Falciparum* cysteine proteases active site and the main interactions were studied. The present work provides extremely useful guidelines for future structural modifications of this class of compounds towards the development of superior antimalarials.

Introduction

The history of malaria treatment is one of acquired drug resistance and toxic side effects. There is known, widespread resistance to once highly effective but now virtually useless antimalarials, like the well-known example of chloroquine [1–3]. The increasing resistance of malaria parasites is a key factor in the persistence of malaria as a global major health concern [4]. Therefore, novel, less toxic, and more specific inhibitors are urgently needed to control this infectious disease. Potential biochemical targets for antimalarial development have been identified after the unveiling of the *Plasmodium falciparum* (*Pf*) genome in 2002 [5, 6]. Among them, there is particular emphasis on proteases having key roles on the degradation of host's hemoglobin within the food vacuole of blood-stage parasites, as these depend on such process for their survival. Among such enzymes, the cysteine proteases, Falcipain-2 (FP2) and Falcipain-3 (FP3), are highly promising antimalarial drug targets [7–10].

Cysteine proteases were given this name due to the function of a catalytic cysteine. This amino acid catalyzes protein hydrolysis via nucleophilic attack to the carbonyl carbon of a susceptible bond. FP2 and FP3 are the best

C. Teixeira (✉) · P. Gomes
Centro de Investigação em Química da Universidade do Porto, Departamento de Química, Faculdade de Ciências, Universidade do Porto, R. Campo Alegre, 687, 4169-007 Porto, Portugal
e-mail: ca.teixeira@ua.pt

C. Teixeira · J. R. B. Gomes
CICECO, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

T. Couesnon
ITODYS, University Paris 7, CNRS UMR 7086, 15 rue Jean Antoine de Baif, 75205 Paris Cedex13, France

characterized cysteine proteases of the malaria parasite and are fairly typical papain-family (clan CA) cysteine proteases [7]. Over the past years the development of small molecule inhibitors against *Pf* falcipains has received a lot of attention by both the pharmaceutical industry and the medicinal chemistry community. Indeed, most studies in this area have shown that falcipain inhibitors prevent hemoglobin hydrolysis, block parasite development and cure murine malaria [11–13]. Moreover, the combination of high-throughput screening, molecular modeling methods and the availability of X-ray structures of falcipains has contributed to the discovery of potent inhibitors able to inactivate these *Pf* enzymes in a reversible or irreversible manner [7, 14].

In 1996, Rosenthal and colleagues evaluated the antimalarial effects of a series of vinyl sulfones as cysteine proteases inhibitors [12]. These organic compounds tend to covalently bind to the enzymes by an irreversible addition of the thiol group of the catalytic cysteine to the electrophilic vinyl sulfone moiety, which behaves as a Michael acceptor. The study revealed that some compounds of this family were shown to be nontoxic, active against *Pf* falcipains and exhibited antiparasitic activity. Several vinyl sulfones presented nanomolar activities and the compound Mu-Leu-hPh-VSPh (also named K11017, Fig. 1) was the one that most effectively inhibited falcipains, blocked hemoglobin degradation and parasite development. To increase both bioavailability and aqueous solubility, several structural replacements on K11017 were tested [15]. Although some compounds demonstrated better activities than the parent molecule, they also revealed limited utility as therapeutic agents due to their susceptibility to protease degradation and their poor absorption through cell membrane [14]. In an effort to well define the structure–activity relationships of this class of compounds, Shenai et al. recently disclosed a new series of 39 new vinyl sulfone, vinyl sulfonate ester, and vinyl sulfonamide cysteine protease inhibitors and determined their FP2 inhibitory activity [16].

This previous experimental structure–activity study of vinyl sulfones as FP inhibitors motivated us to engage in a complementary computational 3D-QSAR analysis, along with docking studies, of the same set of 39 peptidyl vinyl sulfone derivatives. This dataset was analyzed in an effort to: (i) derive a statistically significant and predictive model
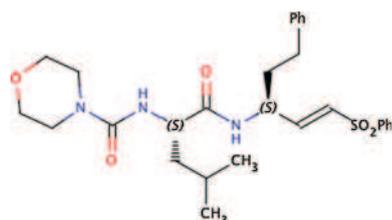
of peptidyl vinyl sulfones affinity towards the FP2 cysteine protease and (ii) identify the physicochemical properties that have a substantial effect on the binding affinity of those ligands. The knowledge of the protein structure is not a pre-requisite to perform QSAR analysis. However, combining docking and 3D-QSAR analysis is advantageous because it allows the direct visualization and interpretation of the results obtained by the derived models with respect to the protein environment. This information provided valuable insight into structure activity interpretations by revealing the interactions contributing positively or negatively to the binding affinity.

Computational methods

Inhibitor data set and ligand preparation

The set of 39 peptidyl vinyl sulfones (Fig. 2) with *Pf* FP2 inhibitory activity was identified from the literature [16]. All inhibitors were synthesized in the same laboratory and the activity data, reported as IC_{50}, were determined using the same assay protocol [12]. This aspect is crucial since homogeneity of experimental conditions is required to get reliable 3D-QSAR models. The IC_{50} value of each inhibitor was converted into pIC_{50} ($-\log IC_{50}$) in order to use the data as a dependent variable in the CoMFA and CoMSIA models. The structures, drawn with Marvin-Sketch [17], and inhibitory activities (IC_{50} and pIC_{50}) of the studied compounds are displayed in Table 1.

The 3D structures of the small organic molecules were built based on the docked structure of K11017 (ligand 14), within the FP2 catalytic site. The molecular structures of the inhibitors were generated employing the Tripos force field within the Sybyl X 1.3 molecular modeling program [18]. Charges were assigned using the Gasteiger-Marsili method. Energy minimization was performed using 20 simplex iterations followed by 1,000 steps of Powell

Fig. 1 Structure of vinyl sulfone, Mu-Leu-hPh-VSPh (or K11017)
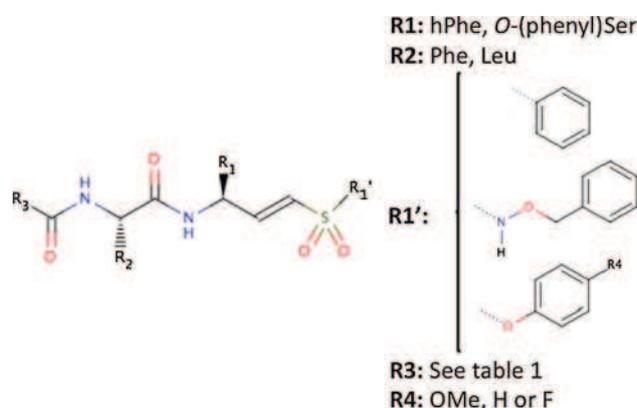
Fig. 2 General structure of peptidyl vinyl sulfone derivatives

minimization until achieving the gradient norm of 0.05 kcal/mol.

Molecular docking

As no experimental FP2 structure bound to a peptidyl vinyl sulfone inhibitor is available, docking calculations were conducted to predict the appropriate binding orientation of the ligands, into the FP2 catalytic site. For this process, we used the molecular docking program GOLD (Genetic Optimization for Ligand Docking) [19, 20] from the Cambridge Crystallographic Data Center.

The X-ray structure of FP2 with its bound inhibitor E64 (PDB code: 3BPF) [21] was used and prepared for docking

Table 1 Structures, experimental FP2 activities (IC_{50} and pIC_{50}), predicted activities (Calcd) and residuals (Res) by CoMFA and CoMSIA models of peptidyl vinyl sulfone derivatives. Compounds marked with * belong to the test set

	ID	R3	R5/R4	FP2 IC_{50} (nM)	pIC_{50}	CoMFA		CoMSIA	
						Calcd	Res	Calcd	Res
	1		CH_2	110	6.959	7.030	0.071	6.947	0.012
	2*		CH_2	120	6.921	7.067	0.146	7.235	0.314
	3		CH_2	71	7.149	7.084	0.065	7.065	0.084
	4		CH_2	120	6.921	6.774	0.147	6.984	0.063
	5		O	120	6.921	6.790	0.131	6.859	0.062
	6		O	56	7.252	7.247	0.005	7.256	0.004
	7		O	290	6.538	6.655	0.117	6.588	0.050
	8*		O	230	6.638	7.235	0.597	6.483	0.155
	9		O	130	6.886	6.889	0.003	6.859	0.027
	10			35	7.456	7.487	0.031	7.479	0.023
	11*			27	7.569	7.996	0.427	7.865	0.296
	12			8.7	8.060	7.936	0.124	7.900	0.160
	13			9.2	8.036	8.232	0.196	8.096	0.060
	14			3.5	8.456	8.456	0.000	8.392	0.064
	15			6.9	8.161	8.337	0.176	8.311	0.150
	16			9.9	8.004	7.941	0.063	8.053	0.049
	17			6.7	8.174	8.028	0.146	8.169	0.005
	18*			140	6.854	7.480	0.626	7.578	0.724

Table 1 continued

	ID	R3	R5/R4	FP2 IC_{50} (nM)	pIC_{50}	CoMFA		CoMSIA	
						Calcd	Res	Calcd	Res
	19*			2.3	8.638	8.091	0.547	7.985	0.653
	20			2.2	8.658	8.717	0.059	8.765	0.107
	21			25	7.602	7.711	0.109	7.594	0.008
	22			14	7.854	7.981	0.127	7.898	0.044
	23*			16	7.796	7.681	0.115	7.810	0.014
	24			3.6	8.444	8.460	0.016	8.375	0.069
	25			2.5	8.602	8.562	0.040	8.583	0.019
	26			2.2	8.658	8.733	0.075	8.643	0.015
	27			16	7.796	7.796	0.000	7.746	0.050
	28		OMe	0.7	9.155	9.198	0.043	9.153	0.002
	29*		OMe	1.9	8.721	8.862	0.141	8.798	0.077
	30		OMe	0.9	9.046	9.183	0.137	8.967	0.079
	31		F	0.7	9.155	9.045	0.110	9.114	0.041
	32		F	21	7.678	7.797	0.119	7.828	0.150
	33		F	0.9	9.046	8.809	0.237	9.055	0.009
	34*		H	0.7	9.155	8.918	0.237	8.943	0.212
	35		H	350	6.456	---	---	---	---
	36		H	0.7	9.155	9.109	0.046	9.131	0.024
	37		H	0.8	9.097	8.896	0.201	9.197	0.100
	38*		F	0.9	9.046	8.279	0.767	8.122	0.924
	39		OMe	0.9	9.046	9.076	0.030	8.956	0.090

with Sybyl X 1.3 software. The original ligand, ions and solvent molecules were removed and the protein was protonated to a pH = 5.5. The protein was then minimized with the AMBER program [22] by 500 steps of steepest descent followed by 2,000 steps of conjugate gradient to remove bad contacts using a generalized-Born solvent model. The biomolecular force field ff03 was used [23]. For docking studies the genetic operations were set to default parameters with a population size of 100

individuals. The selection pressure was set at 1.1 and default operator weights were used for crossover, mutation and migration, i.e., 95, 95 and 10, respectively. The binding site was defined as 15 Å radius from the catalytic amino acid Cys42 and the GoldScore function was used to rank the different binding poses.

Previously, the docking protocol was tested to verify if GOLD was able to reproduce the experimental data. For this validation, we selected the available complex of FP3 with K11017 inhibitor from the PDB databank (PDB code: 3BWK) [24]. The selection of FP3 structure (PDB code: 3BWK) to establish a docking protocol was based on the following reasons: (i) FP2 and FP3 are both cysteine proteases of *Plasmodium Falciparum* and contribute more or less equally to the digestion of hemoglobin in the food vacuole [21]; (ii) the superimposition of FP2 (3BPF, chainB) and FP3 (3BWK, chainA) using DaliLite server [25], matches 240 α-carbons with an RMSD of 1.1 Å, a Z score of 38.9 and a sequence of identity of 66% and (iii) FP3 X-ray structure (3BWK) give us access to the bioactive conformation of one of the peptidyl vinyl sulfones derivatives of this study (ligand 14). The X-ray FP3 structure presented a missing residue, the C-terminal GLU243, which was added with the module Biopolymer in Sybyl X 1.3 molecular modeling program. We followed exactly the same docking protocol as used above for FP2 whereas the binding site was defined as 15Å radius from the catalytic residue Cys52.

Ligand alignment method

The quality of the spatial alignment of the ligands is the most sensitive parameter in 3D-QSAR analyses since the quality and predictive ability of the model are dependent on the alignment method. In order to identify the most efficient alignment approach for this data set, two different procedures were employed: (i) a receptor-docked alignment derived from the structure-based docking algorithm GOLD and (ii) a ligand-based alignment using the structure of K11017 derived from the crystal structure 3BWK from the PDB databank. In the latter, the co-crystallized structure of K11017 was used as the basic skeleton to build the remaining compounds by modifying the required substitutions. Partial atomic charges were assigned with the Gasteiger-Marsili method and energy minimization of each molecule was then performed employing the Tripos standard force field while keeping the conformation of the common structure of peptidyl vinyl sulfones.

CoMFA analysis

The CoMFA fields (steric and electrostatic) were generated using C sp^3 atom with a +1 charge as probe and calculated using the standard Tripos force field with a distance-dependent dielectric constant. The cutoff was set to 30 kcal/mol. To form the basis for a statistical significant model, the method of partial least-squares (PLS) regression was used to analyze the inhibitors by correlating variations in their biological activities with variations in their interaction fields. The cross-validation analysis was performed using the leave-one-out (LOO) method wherein one of the compounds is removed from the data set and its activity predicted with the model derived from the remaining compounds. The optimum number of components, N, used in the model derivation was chosen from the analysis with the highest cross-validated coefficient, q^2, and lowest standard error of prediction, SEP. The column filtering was set at 2.0 kcal/mol to speed up the analysis and to reduce noise. Finally, the non-cross-validation analysis was performed to calculate conventional non cross-validated coefficient, r^2, using the optimum number of components obtained from the analysis above. In these analyses, a q^2 of 0 represents prediction based on random guessing, a q^2 of 1 represents perfectly accurate prediction and a negative q^2 represents prediction worse than random guessing. Generally, a minimum q^2 of 0.3 is recommended and a QSAR model with a value of $q^2 > 0.5$ is normally considered to possess significant predictive ability [26]. However, during the past few years, evidences have been indicative that q^2 appears to be necessary but not a sufficient condition for the model to have a high predictive power [27, 28]. This way, it has been proposed the validation on a sufficiently large test set (∼25–33% of total data set) to establish a reliable 3D-QSAR model [29]. Therefore, the panel comprising 39 compounds was split into a training set (30 compounds) and a test set (9 compounds, displayed with an asterisk in Table 1). The selection of these 9 compounds was made on the basis that the test set must represent structural diversity and a range of biological activities similar to that of the training set. A predictive r square (r^2_{pred}) value was then obtained with the following equation:

$$r^2_{pred} = (SD - PRESS)/SD$$

In the previous expression, SD is the sum of squared deviation between the biological activities and the mean activity of the test set molecules and PRESS represents the sum of squared deviations between the experimental and predicted activities of the test set compounds.

CoMSIA analysis

CoMSIA similarity indices [30], between the compounds of interest and the probe atom, were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å, taking the same aligned molecules and the same lattice box that were used for CoMFA. A probe atom with radius 1.0 Å and +1.0 charge with hydrophobicity of +1.0 and

hydrogen bond donor and hydrogen bond acceptor properties of +1.0 was used to calculate steric, electrostatic, hydrophobic, donor and acceptor fields. Gaussian type distance dependence and the default value of the attenuation factor ($\alpha = 0.3$) were used.

Results and discussion

The experimental results cover a wide range of pIC_{50} from 6.456 to 9.155 (calculated from IC_{50} values in M units) giving us useful information to understand the different activity profiles of the molecules in our dataset.

Docking results

Docking accuracy was validated using the known X-ray structure of FP3 in complex with ligand 14. The ligand was docked into the binding site of FP3 and the docked conformation corresponding to the highest GoldScore value was selected as the most likely binding conformation. The RMSD between predicted and experimental poses of ligand 14 was found to be 1.36 Å, which was quite satisfactory considering the number of rotatable bonds of the molecule (Fig. 3).

Subsequent docking of ligand 14 in FP2 active site reproduced the group arrangements into S1–S3 substrate binding sites (Fig. 4a). Moreover, H-bonds of ligand 14 with Gly92, Asn182, His183, and Trp215 of FP3 binding site were reproduced in the FP2 active site: Gly83, Asn173, His174 and Trp206, respectively (Fig. 4b). Although the ligand adopt similar conformations in each complex, *i.e.*, same groups fit into the same substrate binding sites, in the FP2 ligand 14 complex the conserved phenyl sulfone group at the S1′ position is flipped about ∼90° out of the active site in relation to the FP3 complex.

This situation is analogous to that observed for the X-ray structure of rhodesain with the same ligand when compared with that of FP3 [24]. The authors attributed this event to the substitution of Ala166 in FP3 for the slightly bulky Met145 in rhodesain, which prevent the phenyl sulfone substituent from lying flat. Superimposition of the FP3 and FP2 complexes indicates that a similar substitution could be at the origin of this phenyl sulfone flip. In fact, the amino acid equivalent to Ala161 in FP3 is the slightly bulkier Val152 in FP2, which hinders the phenyl sulfone of the inhibitor to rest on the floor of the S1′ pocket (Fig. 5). However, we believe that this flipping may be transient, like for the complex of rhodesain and a vinyl sulfone where the P1′ moiety was modeled at half occupancy in both the "in" and "out" conformations [24].

The docking protocol was then applied to all the peptidyl vinyl sulfone derivatives in order to determine their likely binding conformations into the active site of FP2. All
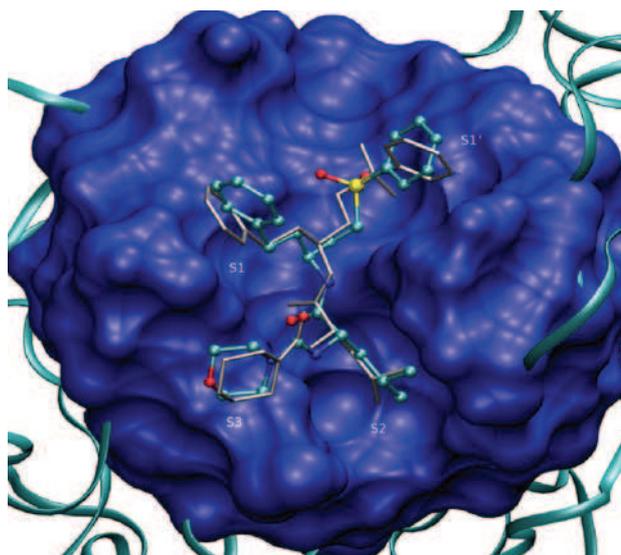
Fig. 3 Superimposition of co-crystallized pose (CPK representation) of ligand 14 and top-ranked docked binding mode (*silver*) into FP3 catalytic site. Substrate binding sites S1, S1′, S2 and S3 of FP3 are shown

of the inhibitors bonded to the active site of FP2 in a similar conformation to ligand 14 and the common chain of the structures superimposed rather well. Based on the set of binding conformations and their alignment, CoMFA and CoMSIA were performed.

CoMFA model and its predictive power

To explore the effect of different alignment methods on the predictive ability of the CoMFA model, separate CoMFA models were built for each alignment scheme.

Model 1 was derived from the docking-based alignment using the structure-based docking algorithm GOLD. To obtain a good superimposition of the molecules, the alignment was improved by manually selecting docking poses for ligands that showed deviations from it. The set of top-ranked conformations contained 7 compounds (5, 6, 13, 17, 22, 37 and 38) whose structural features did not superimpose well with the rest of the inhibitors. For these 7 ligands, conformations with lower scores that aligned with the majority of the other molecules were selected. An average decrease of 0.9 in the GOLD fitness score was observed between the top-ranked conformations and those presenting better alignment profile. These differences in the GoldScore can be considered as of minor importance. Figure 6a illustrates the resulting conformational alignment of the 30 compounds taken from the docked conformations. Model 2 was built from the ligand-based alignment, which is represented in Fig. 6b. The partial least square (PLS) statistics of CoMFA models 1 and 2 are summarized in Table 2.

Fig. 4 a Surface representation of the substrate binding sites of FP2 with the top-ranked docked conformation of ligand 14 and **b** *Ball* and *stick* representation showing important FP2 residues that establish hydrogen bonds with docked ligand 14 (thicker conformation)
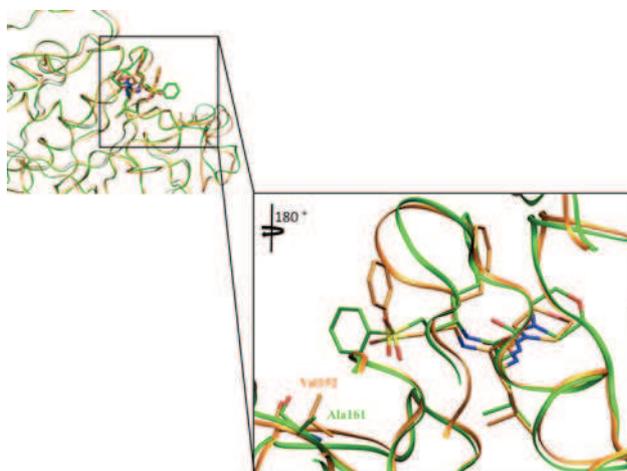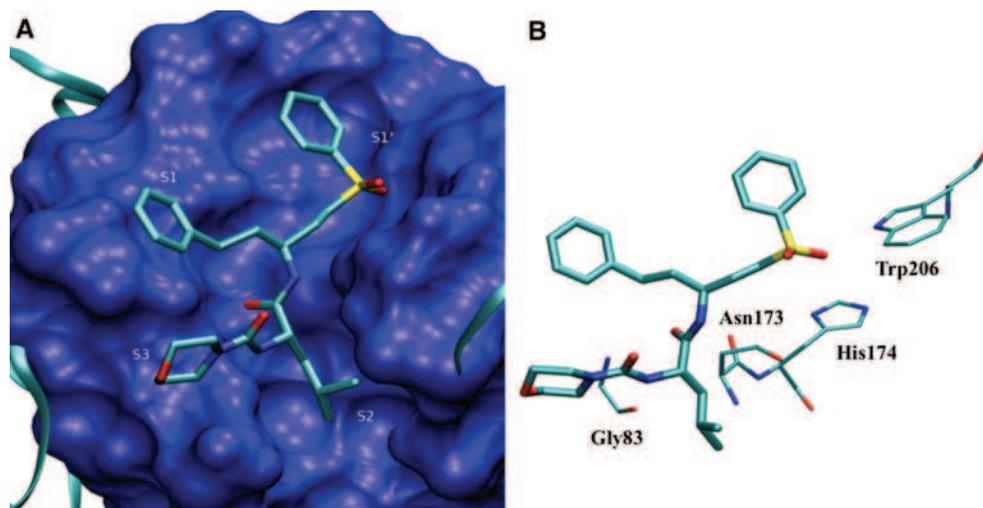
Fig. 5 Phenyl sulfone flipping in the FP2 ligand 14 complex: *ribbon* representation of FP3 (*green*) and FP2 (*orange*) are shown. Docked ligand 14, Val152 and Ala161 are colored as their respective enzyme partners

Model evaluation using the Leave-One-Out (LOO) cross-validation and external validation methods generated only moderate satisfactory results. By further checking the observed activities versus the predicted activities (data not shown) of the models, it was noted that one of the molecules (ligand 35) had predicted errors close to one unit for Model 1 and higher for Model 2. This indicates that the target value for this compound was badly estimated. Molecules can be outliers for various reasons such as (but not only) structural uniqueness compared to the other molecules, experimental data outside the range of the other molecules, or poor experimental data. In our case, we think it may be due to the large difference in inhibitory activity compared with other compounds that show only minor structural differences (ligands 28–39).

When ligand 35 was treated as outlier and excluded from the models construction, the new models (Model 3 and 4 derived from receptor-based and ligand-based alignment, respectively) were clearly improved. The partial least square (PLS) statistics of the two new CoMFA models are summarized in Table 3.

For Model 3, we obtained a cross-validated LOO q^2 of 0.696 (SEP = 0.478) with three principal components and a non cross-validated r^2 of 0.98 (SEE = 0.121). Model 4, derived from the ligand-based alignment, gave a slightly better value of LOO cross-validated q^2 of 0.784 (SEP = 0.418) with three principal components and a non cross-validated r^2 of 0.936 (SEE = 0.223). In order to use these models toward lead optimization, they must have reasonable extrapolative validity in addition to interpolative accuracy. Subsequently, an external validation was performed on the test set of 9 compounds outside the training set to evaluate their predictive ability. All the compounds were predicted well by the two models. Indeed, models 3 and 4 gave comparable good predictive r^2_{pred} values of 0.755 and 0.768, respectively.

LOO cross-validated value, q^2, was introduced to generate an initial measure of the accuracy of model interpolation. The study suggested that all the derived models had a good cross-validated correlation ($q^2 > 0.5$). Although the q^2 value of the receptor-based CoMFA model was lower than the ligand-based model, conventional r^2 and F value were higher, while the SEE was lower.

The different alignments also affected the field contributions of the two models: Model 3 showed a higher contribution of electrostatic field while we verified the opposite for Model 4. We consider that this may be due to the fact that the receptor-docked alignment may provide, in this case, a better representation for predictive intents, since the observed variability in the position of the

Fig. 6 Receptor-based (**a**) and ligand-based (**b**) alignments used to derive CoMFA models
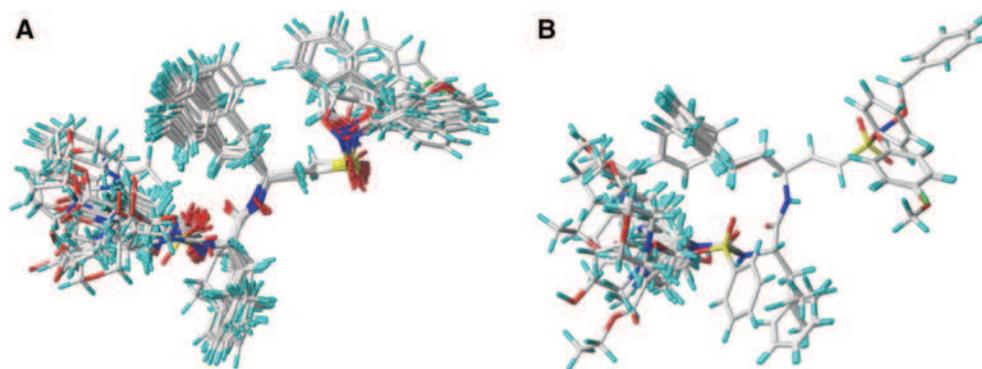
Table 2 Summary of CoMFA statistical results for the different alignment methods

QSAR parameter[a]	Model 1 (receptor-based alignment)	Model 2 (ligand-based alignment)
q^2	0.604	0.563
N	2	3
SEP	0.559	0.625
r^2	0.906	0.845
SEE	0.272	0.358
F test value	130.18	47.187
r^2_{pred}	0.68	0.63

[a] q^2 cross-validated coefficient, *N* number of components, *SEP* standard error of prediction, r^2 conventional non cross-validated coefficient, *SEE* standard error of estimation, r^2_{pred} predictive correlation coefficient

Table 3 Summary of CoMFA statistical results for models 3 and 4 (without outlier) derived from receptor and ligand-based alignment, respectively

| QSAR parameter[a]	Model 3 (receptor-based alignment)	Model 4 (ligand-based alignment)
q^2	0.696	0.784
N	3	3
SEP	0.478	0.418
r^2	0.980	0.936
SEE	0.121	0.223
F test value	427.861	121.26
r^2_{pred}	0.76	0.77
Field contribution (%)		
Steric	40	61
Electrostatic	60	39

[a] Labels as in Table 2

different ligands symbolizes the receptor's ability to differently accommodate (up to a point) compounds with substituents of varying size and electrostatic features. As a result, receptor-based alignment seems to provide additional information compared with the ligand-based method.

In addition, the cross-validated PLS analysis indicated that the CoMFA model derived from the receptor-based alignment is robust in predicting both the training and test set and was then selected for further structural analysis. For the selected model, the relative contributions of steric and electrostatic fields for CoMFA correspond to 40 and 60%, respectively. Electrostatic interactions of molecules from our dataset with active site of FP2 were found to be a slightly more influent factor for inhibitory activity. The predicted inhibitory activity (pIC_{50}) and the residual values for the training set and the test set, using the developed CoMFA Model 3, are described in Table 1, while the graphical plot between the experimental vs predicted FP2 inhibitory activity for both training and test set is shown in Fig. 7.

CoMSIA model and its predictive power

Taking the same alignment method, training and test set that were used to derive the best CoMFA model, various CoMSIA models were generated considering all possible combinations of CoMSIA field descriptors, i.e., steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D) and acceptor (A) fields. The partial least square (PLS) statistics of all the CoMSIA models are summarized in Table 4.

Based on the cross-validated and conventional correlation coefficients, q^2 and r^2, the PLS analyses yielded consistent results of high statistical significance for 11 models (highlighted in bold, Table 4). The q^2 and r^2 values of these models range from 0.651 to 0.754 and 0.956 to 0.998, respectively. However, on further validation of these eleven best CoMSIA models with respect to their ability to explain the activity variation among the test set of 9 compounds covering almost the same range of FP2 inhibitory activity, only models EH and SEH represented the best results with a predictive r^2_{pred} of 0.74 and 0.77, respectively. The combination of electrostatic and hydrophobic fields in CoMSIA gave the best results (Model EH), giving cross-validation correlation coefficient q^2 of 0.711, conventional correlation coefficient r^2 of 0.992 and
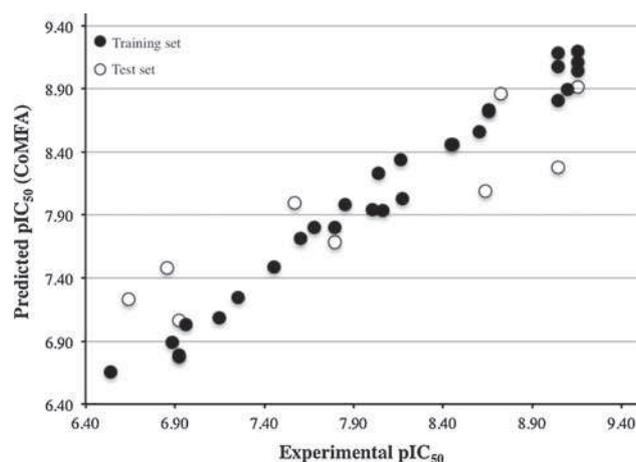
Fig. 7 CoMFA predicted versus experimental pIC$_{50}$ values. *Solid and open circles* represent predictions for the training and test sets, respectively

predictive correlation coefficient r^2_{pred} of 0.74. The other combination of steric, electrostatic and hydrophobic fields (Model SEH) in CoMSIA also gave statistically significant models, but exhibited relatively lower q^2 and r^2 values compared to Model EH. In terms of descriptors, the only difference between the two models is the contribution of the steric field in Model SEH.

Mittal et al. previously concluded that the CoMFA steric field calculated using the steeper Lennard-Jones potential tends to perform better than CoMSIA steric field calculated with Gaussian version, which has a slower and smoother decrease [31]. Indeed, this is true when comparing predictive ability of CoMSIA Model SE with the related selected CoMFA model (Model 3). Although the values of q^2 and r^2 are similar for those CoMSIA and CoMFA models, CoMFA Model 3 clearly demonstrates higher predictive correlation coefficient, r^2_{pred} of 0.76 compared

Table 4 Summary of partial least-squares statistics for CoMSIA models

CoMSIA model[b]	q^2	N	SEP	r^2	SEE	F test value	Field contribution (%)	r^2_{pred}
EA	**0.754**	**8**	**0.480**	**0.991**	**0.092**	**275.391**	**68 (E), 32 (A)**	**0.48**
SH	0.564	1	0.550	0.737	0.427	75.848	31 (S), 69 (H)	
EH	**0.711**	**7**	**0.508**	**0.992**	**0.084**	**378.244**	**61 (E), 39 (H)**	**0.74**
SA	**0.672**	**10**	**0.585**	**0.993**	**0.085**	**257.12**	**44 (S), 56 (A)**	**0.42**
SD	0.464	2	0.623	0.789	0.390	48.583	33 (S), 67 (D)	
SE	**0.697**	**5**	**0.499**	**0.979**	**0.132**	**210.748**	**29 (S), 71 (E)**	**0.68**
HA	0.686	2	0.475	0.895	0.275	110.754	57 (H), 43 (A)	
ED	**0.660**	**8**	**0.565**	**0.983**	**0.125**	**146.627**	**72 (E), 28 (D)**	**0.58**
DA	0.629	2	0.517	0.824	0.356	60.883	56 (D), 44(A)	
HD	0.555	2	0.566	0.803	0.377	52.832	50 (H), 50 (D)	
SEA	**0.709**	**8**	**0.522**	**0.994**	**0.075**	**418.005**	**16 (S), 58 (E), 26 (A)**	**0.52**
SEH	**0.670**	**4**	**0.507**	**0.963**	**0.171**	**154.676**	**18 (S), 52 (E), 30 (H)**	**0.77**
EHA	**0.729**	**10**	**0.531**	**0.998**	**0.046**	**877.078**	**46 (E), 33 (H), 21 (A)**	**0.53**
SHA	0.668	2	0.492	0.901	0.267	118.591	23 (S), 42 (H), 35 (A)	
SED	0.645	7	0.563	0.986	0.111	213.176	21 (S), 54 (E), 25 (D)	
EDA	**0.712**	**9**	**0.533**	**0.991**	**0.096**	**226.141**	**53 (E), 21 (D), 26 (A)**	**0.42**
EHD	**0.651**	**7**	**0.558**	**0.989**	**0.098**	**275.324**	**49 (E), 31 (H), 20 (D)**	**0.67**
SHD	0.536	1	0.567	0.718	0.443	68.729	14 (S), 45 (H), 41 (D)	
SDA	0.630	2	0.516	0.868	0.308	85.784	23 (S), 42 (D), 35 (A)	
HDA	0.651	2	0.502	0.865	0.312	83.571	40 (H), 35 (D), 25 (A)	
SEHA	**0.702**	**3**	**0.473**	**0.956**	**0.181**	**181.435**	**10 (S), 42 (E), 27 (H), 21 (A)**	**0.53**
SEHD	0.628	7	0.576	0.988	0.105	237.846	11 (S), 45 (E), 26 (H), 18 (D)	
SEDA	0.667	2	0.49	0.903	0.264	121.512	13 (S), 40 (E), 27 (D), 20 (A)	
EHDA	0.678	2	0.482	0.899	0.269	116.065	33 (E), 26 (H), 23 (D), 18 (A)	
SHDA	0.629	2	0.517	0.873	0.303	89.394	13 (S), 34 (H), 30 (D), 23 (A)	
SEHDA	0.659	2	0.496	0.898	0.271	114.759	10 (S), 30 (E), 23 (H), 21 (D), 16 (A)	

[a] Labels as in Table 2

[b] E electrostatic, S steric, H hydrophobic, D donor, A acceptor

Models with high statistical significance are highlighted in bold

with 0.68 for CoMSIA SE model. Hence, we assumed that the lower values of q^2 and r^2 for CoMSIA Model SEH might be similar due to the use of steric molecular field data as one of the descriptors.

In terms of the statistical results, the CoMSIA EH model was found to be the best, except for predictive correlation coefficient r^2_{pred} that, still, remains statistically significant. In addition, the CoMSIA contour map for steric field, issued from Model SEH, also presented similar sized polyhedra than CoMFA steric contour maps at the same regions (data not shown). Hence, the Model EH of CoMSIA was chosen as a complementary source of information of the previously selected CoMFA model and was used for final analysis. The observed and calculated activity values for the training and test set molecules are given in Table 1, and the plots of the predicted versus the actual activity values for the training set and test set are shown in Fig. 8.

Graphical interpretation of the CoMFA and CoMSIA results

The greatest advantage of CoMFA and CoMSIA is that the field effect on the compounds property can be viewed as 3D contour plots. In our case, these contour plots are useful to: (i) identify critical regions where any change in the steric, electrostatic, and hydrophobic fields may affect the inhibitory activity, and (ii) highlight the key structural features required for FP2 inhibitory activity. The CoMSIA electrostatic contour maps were found to be approximately identical to the corresponding CoMFA contour maps. Therefore, the graphical interpretation of the CoMSIA results will only focus on the CoMSIA hydrophobic contour maps since electrostatic contour maps will be discussed from the CoMFA analysis.
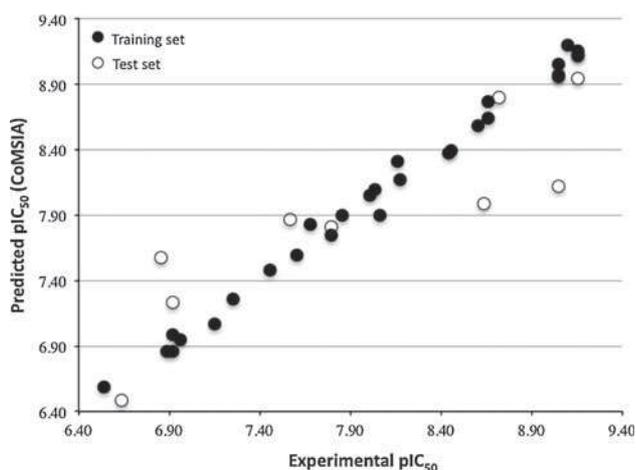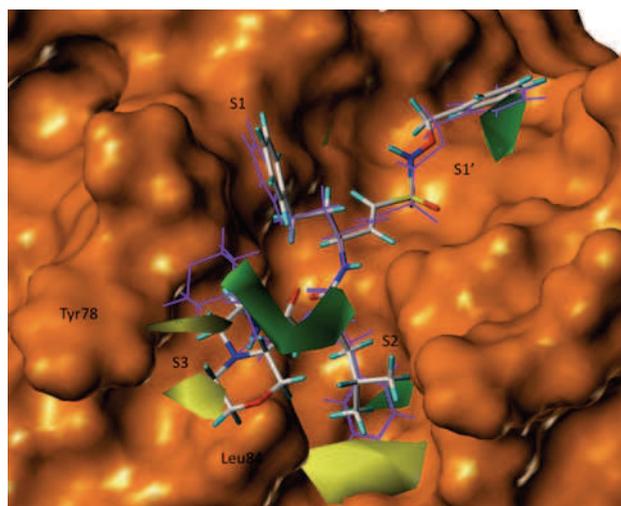
CoMFA steric contour maps

Figure 9 illustrates the contours of the steric fields, showing in green and in yellow the favored and unfavored bulky groups, respectively. In order to better understand the steric field contribution, their corresponding contour maps were projected on the FP2 active site surface.

Two large sterically favorable contours are reported. The first is located near the R_3 position while the second is situated aside the R_1' substituents. These two contours suggest that there is a requisite for bulky groups in this region for potent FP2 inhibitory activity. Indeed, those two zones are highly exposed to solvent (Fig. 9) with enough space to accommodate bulky substituents that would enhance the interactions between the ligand and protein surface.

One sterically unfavorable contour is localized at the R_2 group, which is docked into the S2 pocket of FP2 catalytic site. This unfavorable contour represents the limitation of S2 cavity (Fig. 9) and suggests that the occupation of this area by a bulky group would have a negative effect on the FP2 inhibitory activity. It is also associated to this FP2 cavity, a small green contour region, which indicates that certain bulkiness is favored. Thus, these two contour maps at the S2 pocket reveal that the substituent could be slightly bulky in depth but not in length.

Indeed, a comparison between molecules that change from Leu to hPhe shows that bulky groups (in length) are not preferred at R_2 as can be seen by the activities of compounds 1, 2, 3 and 4 which are lower than those of compounds 22, 21, 24 and 23. We also detect two smaller yellow regions near the S3 pocket, corresponding to

Fig. 8 CoMSIA predicted versus experimental pIC$_{50}$ values. *Solid and open circles* represent predictions for the training and test sets, respectively


Fig. 9 CoMFA steric maps projected on the Connolly molecular surface of FP2 substrate binding sites (*orange*). Sterically favorable and unfavorable areas are shown in green and *yellow*, respectively. Molecule 1 is displayed in *purple* while molecule 22 is colored by atom type

possible steric compression of bulky groups in this area by FP2 residues Tyr78 and Leu84 (Fig. 9).

CoMFA electrostatic contour maps

In the CoMFA electrostatic contours (Fig. 10), the introduction of electronegative substituents in red regions may increase the inhibitory activity while in blue regions decrease the affinity. In order to facilitate electrostatic fields' analysis, corresponding contour maps were projected with some FP2 binding pocket residues.

Two relatively large blue-coloured regions are observed over R_2 backbone and R_1 side-chain. This region shows an area where electropositive charged groups within this glycine rich region of the binding site enhance FP2 inhibitory activity. Docking calculations permitted to verify that, indeed, the –NH groups of the peptidic bonds form hydrogen bonds with the carbonyl groups of FP2 residues: Gly83, Gly82, and Asn81. However, compounds presenting *O*-(phenyl)Ser at R_1 position present modest activity against FP2 due to the disadvantage of repulsive interactions between the oxygen atom and the carbonyl backbone of Gly40 (e.g. compounds 2 e 7). Indeed, when we explore the maps around compound 2 and compare them to those around compound 7, it is clear that the oxygen atom of R_1 substituent falls in a blue region where an increase in
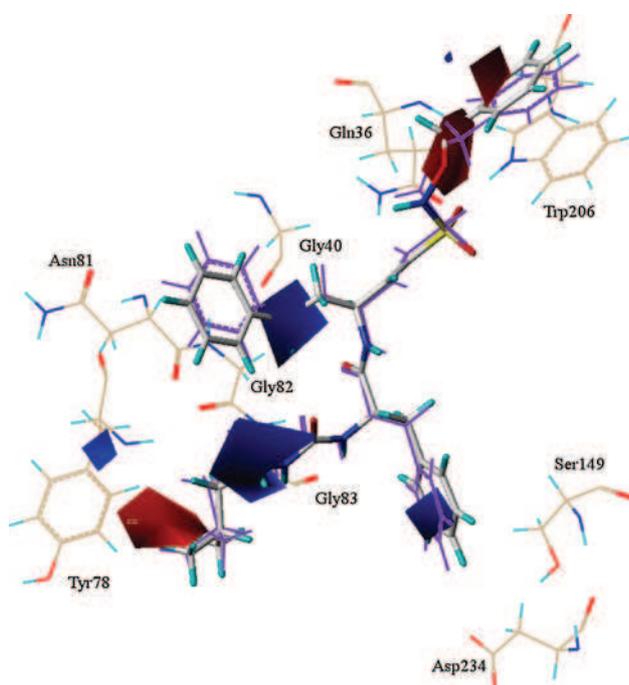
positive charge would enhance the activity of the compounds (Fig. 10).

A third smaller blue contour is noticed at the bottom of the S2 cavity. This can be explained by the fact that almost all S2 residues are hydrophobic with the exception of two polar amino acids, Ser159 and Asp234, that are located at the end of the cavity. The R_2 substituent of our dataset compounds toggles between Leu and hPhe. Although both residues are hydrophobic, Leu shows an electron density lower than that of hPhe aromatic ring, hence reducing the repulsion with the electron-rich groups of the two S2 polar residues.

Near the SO_2–R_1' group, a red-coloured region is noticed and represents a zone where electronegative charged groups improve the activity. Docking calculations of the dataset showed that electronegative fragments at this position form electrostatic interactions with Gln36 and Trp206 side-chains. This is apparent when comparing molecules that were identical except for R_1' substituent, which exhibit a general rank order of FP2 inhibitory activity: sulfonate esters > sulfonamides > sulfones (e.g. ligands 11, 24 and 37). Furthermore, an electron-donating group at this position would increase electron density of R_1' aryl ring and, consequently, enhance π–π interactions with Trp206.

Furthermore, a second small red-coloured region is obtained over the aryl ring of R_1' substituent. This suggests that the more electron density is delocalized to this ring, the more π–π interactions established between side-chain of Trp206 and the ligand are enhanced. Thus, we hypothesize that electron donor groups in *para*-position of vinyl sulfonate ester aryl ring are preferred to delocalize electron density to phenyl and, hence, would increase compound's activity. Although FP2 activities for vinyl sulfonate esters are quite similar, this can be observed with compounds 29 and 32. Indeed, the compounds differ only in the *p*-substitution of the aryl ring and the one bearing the strongest electron-donating group, compound 29 with *p*-OCH_3, has higher inhibitory activity compared to molecule 32 that has a *p*-fluoro substituent.

Close to Tyr78 in the S3 cavity, a third red zone is noted indicating that electron rich groups at this position would be preferred for FP2 inhibition. We believe that, this way, π–π interactions with S3 Tyr78 would be favored. Figure 11 shows electrostatic contours around compounds 8 and 9 that differ only in the R_3 group. We can see that the aromatic moiety of compound 9 establishing stacking contact with Tyr78 and, consequently, presents a better FP2 inhibitory activity than compound 8, which does not form any sort of π-interactions with the same residue.

Fig. 10 CoMFA electrostatic field contours shown in *red* (electronegative substituents favored) and *blue* (electropositive substituents favored) colors. Some FP2 binding pocket residues are represented in *tinny lines* while ligands 2 and 7 are colored by atom type and *purple*, respectively
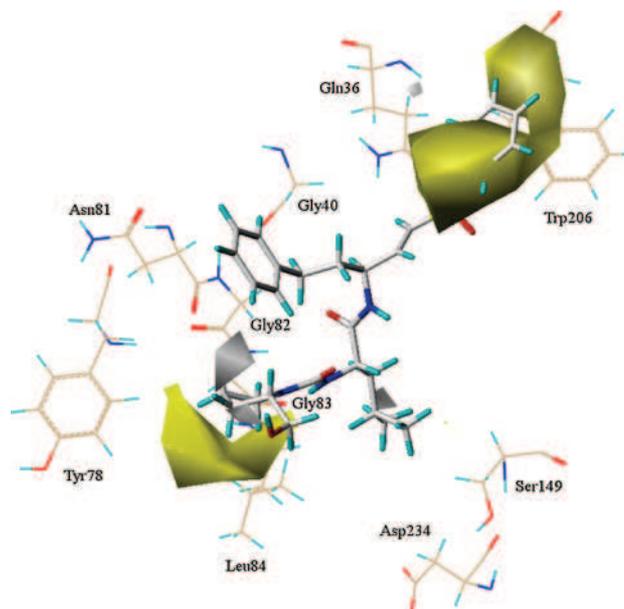
CoMSIA hydrophobic contour maps

Regarding CoMSIA hydrophobic contour maps, yellow regions indicate that hydrophobic groups would be favored
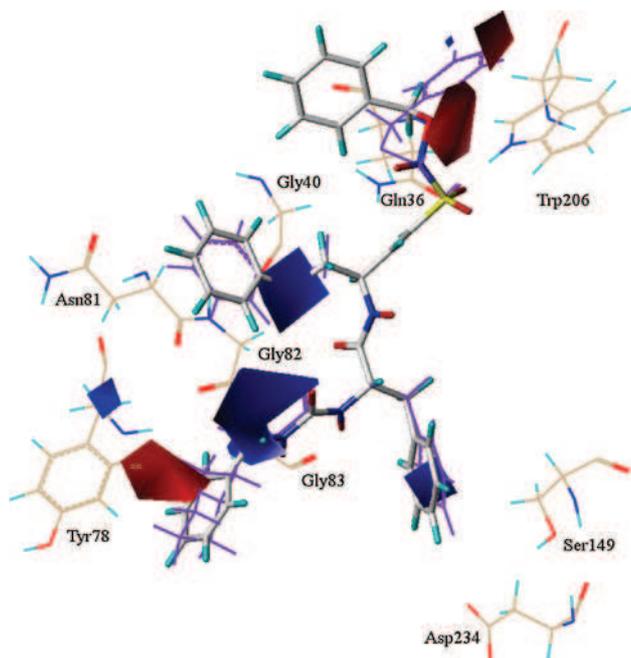
Fig. 11 CoMFA electrostatic field contours shown in *red* (electronegative substituents favored) and *blue* (electropositive substituents favored) colors. Some FP2 binding pocket residues are represented in *tinny lines* while ligands 8 and 9 are colored by atom type and *purple*, respectively


Fig. 12 CoMSIA hydrophobic field contours shown in *yellow* (hydrophobic substituents favored) and *grey* (hydrophilic substituents favored) colors. Some FP2 binding pocket residues and ligand 37 are represented in *tinny lines* and *capped sticks*, respectively

at these positions while grey regions suggest that hydrophilic groups would be preferred. We observed two large yellow regions at S1′ and S3 pockets, and one small grey region also close to the S3 cavity. The mapping of the hydrophobic contours into the active site of FP2 (Fig. 12) further explained the CoMSIA hydrophobic maps. The small grey region is observed at the solvent accessible area of S3 cavity and indicates that hydrophilic groups would be favorable at this position. This region also corresponds to a blue electrostatic contour where electropositive groups are preferred to establish hydrogen bonds with the backbone residues of this glycine rich region.

The yellow site existing near S1′ pocket coincides with FP2 residue Trp206, suggesting that π–π stacking interactions between this residue and compound's aryl moiety are an important factor to increase FP2 inhibitory activity. Regarding π–π interactions that our dataset compounds establish with this highly conserved residue in the S1′ subsite of clan CA cysteine proteases, it is noteworthy that we observed a preferential placement of the aryl moiety relative to Trp206 depending on the R$_1$′ substituent. Figure 13 represents the docking conformations of the compounds that have sulfonate ester (A), sulfonamide (B) and sulfone (C) as R$_1$′ substituent. It is clear that sulfonate esters present a high reproducibility in aryl moiety positioning relative to Trp206, establishing π–π stacking interactions with this aromatic amino acid. Regarding sulfonamides, we also verified that

many of these compounds place the aryl moiety in such a way that π–π interactions with Trp206 are best favored. Still, sulfonamides have greater flexibility in positioning R$_1$′ substituent compared with sulfonate esters. In relation to sulfones, we observed for all the compounds the same phenyl sulfone "flip out" discussed in section "Docking results". In this "flip out" positioning, the aromatic moiety of sulfones does not establish stacking interactions with Trp206. Based on these results, we infer that sulfonamides flexibility at R$_1$′ position may be due to a similar flipping phenomenon. However, we reinforce that this flipping may be transient like the one observed for the complex of rhodesain with a vinyl sulfone where the R$_1$′ moiety was modeled at half occupancy in both the "flip in" and "flip out" conformations [24]. Since the sulfonamides have more degrees of freedom and hence higher flexibility, it is likely that these compounds have higher occupancy in the "flip in" conformation than sulfones. Regarding sulfonate esters, R$_1$′ group flexibility clearly favors the accommodation of the aryl ring into S1′ subsite and thus the formation of π–π interactions with Trp206. We believe that R$_1$′ flexibility is also related to the general rank order of FP2 inhibitory activity observed when comparing compounds that only differ on R$_1$′ substituent, i.e., sulfonate esters > sulfonamide > sulfones.

The second hydrophobic zone, at the S3 cavity, is detected between Tyr78 and Leu84. Compounds that orient hydrophobic substituents into this region (e.g. compound 26) would have enhanced FP2 inhibitory activity comparatively to molecules that do not accommodate any

Fig. 13 Docked conformations of dataset compounds with sulfonate esters (**a**), sulfonamide (**b**) and sulfone (**c**) as R_1' substituent
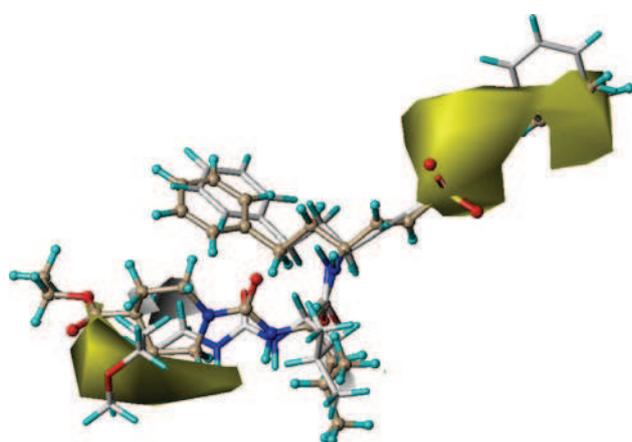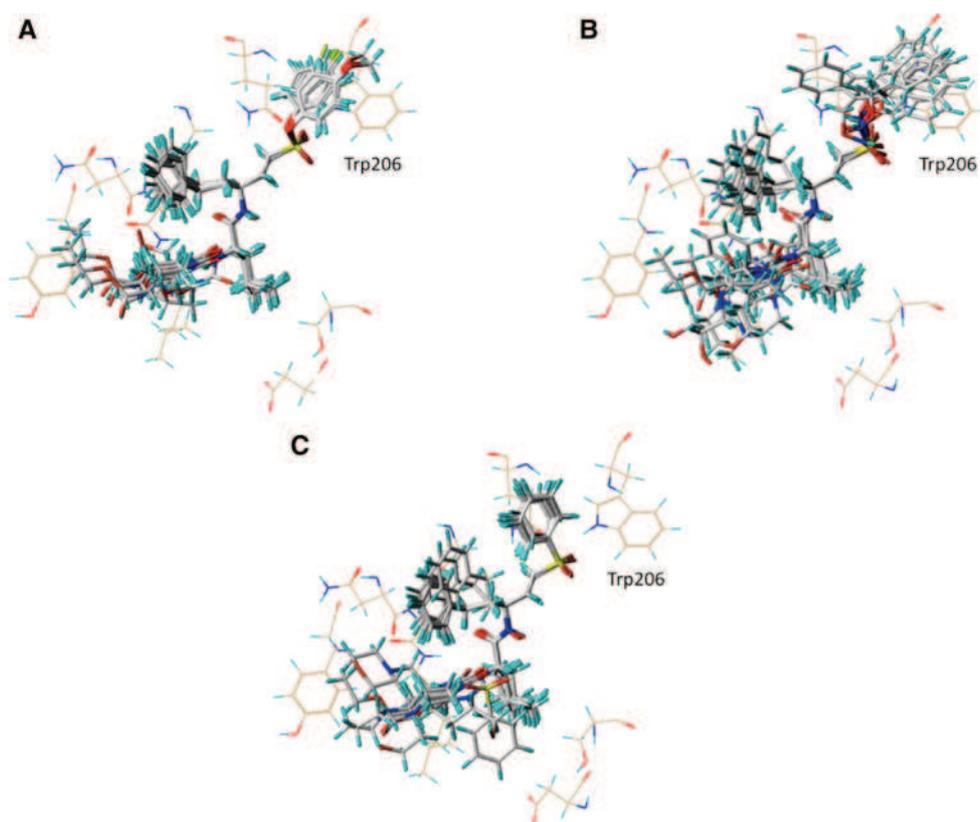
Fig. 14 Representation of the CoMSIA hydrophobic contour maps (code color: *yellow*, hydrophobic substituents favored; *grey*, hydrophilic substituents favored) projected on the docked conformations of compounds 22 and 26 colored by atom type and *beige*, respectively

hydrophobic part of R_3 group in this zone (e.g. compound 22), as shown in Fig. 14.

Conclusions

A set of 39 peptidyl vinyl sulfone derivatives was investigated to relate their FP2 inhibitory activity (IC_{50} values)

with molecular structure. Using CoMFA and CoMSIA techniques, stable and predictive 3D-QSAR models with acceptable q^2 values were developed. The best predictions were obtained for the receptor-docked alignment method and the predictive ability of these models was verified by leave-one-out and external validation methods. These results indicate that the structural alignment of high affinity binding poses obtained from molecular docking simulations comprises biologically active conformations of the peptidyl vinyl sulfone derivatives, confirming the validity and usefulness of the alignment. Besides the effective use of docking as an alignment method, the docking analysis also provided a qualitative representation of ligand and enzyme interactions, which are complementary with CoMFA and CoMSIA maps, and was then helpful in characterizing fundamental structural features required for biological activity. Some of the main features observed are:

- Bulky groups near R_3 and R_1' positions, which are zones highly exposed to the solvent, would enhance the interactions between the ligand and protein surface and, thus, the FP2 inhibitory activity.
- One sterically unfavorable contour is localized at the R_2 group, matching the limitation of the binding pocket, and suggests that bulky groups at this position tend to decrease biological activity. In addition, two additional smaller unfavorable regions near S3 pocket are detected,

indicating possible steric constraints for bulky groups in this area, due to FP2 residues Tyr78 and Leu84.

- Electropositive groups at R_2 backbone and R_1 side-chain are preferred to improve FP2 inhibitory activity. Docking analysis permitted to verify that the –NH groups of ligands peptidic bonds form hydrogen bonds with the backbone of FP2 residues: Gly83, Gly82, Asn81 and Gly40.
- Electronegative groups near the SO_2–R_1' group tend to increase biological activity by forming electrostatic interactions with Gln36 and Trp206 side-chains.
- A second small region, where electronegatively charged groups enhance FP2 inhibitory activity, is obtained over the aryl ring of R_1' substituent. Analysis of the docked conformations suggested that the more electron density is delocalized to this ring, the more π–π interactions established between the side-chain of Trp206 and the ligand are improved.
- Compounds that orient any hydrophobic part of R_3 group towards Tyr78 and Leu84 would exhibit enhanced FP2 inhibitory activity.
- Based on docking analysis, R_1' flexibility was related to biological activity depending on whether or not it favors the positioning of the aryl ring in such a way it can establish hydrophobic interactions with Trp206. The general rank order of FP2 inhibitory activity observed when comparing compounds that only differ on R_1' substituent was: sulfonate esters > sulfonamide > sulfones.

In summary, the present work is the first study based on 3D-QSAR and docking simulations for peptidyl vinyl sulfone derivatives as *Pf* falcipain inhibitors. The physicochemical meaning of the descriptors of the proposed models and the characterization of some ligand–protein interactions provide us valuable guidelines for future structural modifications of this class of compounds towards the design of potent antimalarials.

References

1. Das MK, Lumb V, Mittra P, Singh SS, Dash AP, Sharma YD (2010) J Antimicrob Chemother 65(6):1258
2. Martin RE, Kirk K (2004) Mol Biol Evol 21(10):1938
3. Wiesner J, Ortmann R, Jomaa H, Schlitzer M (2003) Angewandte Chemie 42(43):5274
4. Sachs J, Malaney P (2002) Nature 415(6872):680
5. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Nature 419(6906):498
6. Padmanaban G (2003) Adv Biochem Eng/Biotechnol 84:123
7. Teixeira C, Gomes JR, Gomes P (2011) Curr Med Chem 18(10):1555
8. Rosenthal PJ, Sijwali PS, Singh A, Shenai BR (2002) Curr Pharm Des 8(18):1659
9. Rosenthal PJ (2004) Int J Parasitol 34(13–14):1489
10. Subramanian S, Hardt M, Choe Y, Niles RK, Johansen EB, Legac J, Gut J, Kerr ID, Craik CS, Rosenthal PJ (2009) PloS one 4(4):e5156
11. Batra S, Sabnis YA, Rosenthal PJ, Avery MA (2003) Bioorg Med Chem 11(10):2293
12. Rosenthal PJ, Olson JE, Lee GK, Palmer JT, Klaus JL, Rasnick D (1996) Antimicrob Agents Chemother 40(7):1600
13. Singh A, Rosenthal PJ (2001) Antimicrob Agents Chemother 45(3):949
14. Ettari R, Bova F, Zappala M, Grasso S, Micale N (2010) Med Res Rev 30(1):136
15. Olson JE, Lee GK, Semenov A, Rosenthal PJ (1999) Bioorg Med Chem 7(4):633
16. Shenai BR, Lee BJ, Alvarez-Hernandez A, Chong PY, Emal CD, Neitz RJ, Roush WR, Rosenthal PJ (2003) Antimicrob Agents Chemother 47(1):154
17. MarvinSketch 5.2.2, 2009, ChemAxon (http://www.chemaxon.com)
18. Sybyl X 1.3, Tripos Software, St. Louis, USA
19. Jones G, Willett P, Glen RC (1995) J Mol Biol 245(1):43
20. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267(3):727
21. Kerr ID, Lee JH, Pandey KC, Harrison A, Sajid M, Rosenthal PJ, Brinen LS (2009) J Med Chem 52(3):852
22. Case DA, Darden TA, Cheatham TE, Simmerling CLI, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry KF, Wong KF, Paesani F, Vanicek F, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER 10. University of California, San Francisco
23. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) J Comput Chem 24(16):1999
24. Kerr ID, Lee JH, Farady CJ, Marion R, Rickert M, Sajid M, Pandey KC, Caffrey CR, Legac J, Hansell E, McKerrow JH, Craik CS, Rosenthal PJ, Brinen LS (2009) J Biol Chem 284(38): 25697
25. Holm L, Park J (2000) Bioinformatics 16(6):566
26. Perkins R, Fang H, Tong W, Welsh WJ (2003) Environ Toxicol Chem/SETAC 22(8):1666
27. Golbraikh A, Tropsha A (2002) J Mol Graph Model 20(4):269
28. Saxena AK, Prathipati P (2003) SAR QSAR Environ Res 14(5–6):433
29. Gramatica P (2004) Evaluation of different statistical approaches for the validation of quantitative structure-activity relationships. Final report for JRC Contract ECVA-CCR.496576-Z. European Chemicals Bureau, Joint Research Centre, European Commission, Ispra, Italy
30. Bohm M, Sturzebecher J, Klebe G (1999) J Med Chem 42(3):458
31. Mittal RR, McKinnon RA, Sorich MJ (2008) J Mol Model 14(1):59