

# In silico design of *Plasmodium falciparum* cysteine protease falcipain 2 inhibitors with favorable pharmacokinetic profile

## Abstract

We realized a virtual design of new azadipeptides nitriles (ADPNs) potential inhibitors of the cysteine protease falcipain 2 (FP2) protease inhibitors of *Plasmodium falciparum* (pf) by structure-based drug design (SBDD). From a series of 7 ADPNx with known FP2 inhibition potency, we constructed the FP2-ADPNx complexes by *in situ* modification of the X-rays crystal structure of FP2 in complex with epoxysuccinate E64 (pdb entry code: 3BPF). Out of then a one descriptor quantitative structure activity relationships (QSAR) model was built resulting in linear correlations between the gas phase computed enthalpy  $\Delta\Delta H_{MM}$  upon the FP2-ADPN complex formation and  $IC_{50}^{exp}$  (R square of 0.88; cross validated R square of 0.86; F-Test of 39.48). Thereafter taking into account the solvent effect and the loss of vibrational entropy of the inhibitor upon binding to the enzyme led to an improved QSAR model correlating the computed Gibbs free energy (GFE:  $\Delta\Delta G_{com}$ ) of FP2-ADPN complex formation and  $IC_{50}^{exp}$  (R square of 0.94; cross validated R square of 0.94; F-Test of 91.44). The estimated  $IC_{50}^{pred}$  from FP2 inhibition pharmacophore model derived from the QSAR model linearly correlates with  $IC_{50}^{exp}$  (R square of 0.99) bearing in this way structural inhibition information that served in the virtual screening of a combinatorial subset of a virtual library (VL) of more than 8000 ADPNs analogues. From the ADME focused VL, 68 best hit fit orally bioavailable analogues were selected and finally *in silico* evaluated with the GFE QSAR model to identify new powerful ADPNs with predicted  $IC_{50}$  reaching 0.5nM.

**Keywords:** malaria, FP2, QSAR, pharmacophore, virtual screening, ADME

Volume 7 Issue 3 - 2018

Yves Kily Herve Fagnidi,<sup>1</sup> Beguemi Toi,<sup>2</sup>  
 Eugene Megnassan,<sup>1,2,3</sup> Vladimir Frece,<sup>3,4,5</sup>  
 Stanislav Miertus<sup>3,5,6</sup>

<sup>1</sup>Laboratoire de Physique Fondamentale et Appliquée, University of Abobo Adjame (now Nangui Abrogoua), Côte d'Ivoire

<sup>2</sup>Laboratoire de Chimie Organique Structurale et Théorique, University of Cocody (now Felix Houphouët Boigny), Côte d'Ivoire

<sup>3</sup>ICS-UNIDO, Area Science Park, Italy

<sup>4</sup>Faculty of Pharmacy, Comenius University in Bratislava, Slovakia

<sup>5</sup>International Centre for Applied Research and Sustainable Technology, Slovakia

<sup>6</sup>Faculty of Natural Sciences, University of SS, Slovakia

**Correspondence:** Eugene Megnassan, Laboratoire de Physique Fondamentale et Appliquée, UFR SFA, University Nangui Abrogoua, 02 BP 801, Abidjan 02, Cote D'Ivoire, Tel +225-02-36-30-08, Fax +225-02-36-30-08, Email megnae@yahoo.com

**Received:** February 21, 2018 | **Published:** June 07, 2018

**Abbreviations:** FP2/FP3, falcipain 2/falcipain 3; pf, *plasmodium falciparum*; ADPN, azadipeptide nitrile; SBDD, structure-based drug design; E64, epoxysuccinate; PDB, protein data bank; QSAR, quantitative structure activity relationships;  $\Delta\Delta G_{com}$ , relative gibbs free energy change related to the enzyme-inhibitor complex formation;  $\Delta\Delta H_{MM}$ , relative enthalpic contribution to the Gibbs free energy change derived by molecular mechanics;  $\Delta\Delta T_{vib}$ , relative entropic contribution of the inhibitor to the Gibbs free energy;  $\Delta\Delta G_{sol}$ , the relative solvation Gibbs free energy contribution to the gibbs free energy change; GFE, gibbs free energy;  $IC_{50}^{exp}$ , experimental inhibition constant; VL, virtual library; ADME, adsorption distribution metabolism and excretion; ACTs, artemisin-based therapies; E, enzyme; DS, discovery studio (molecular modeling program); MOE, molecular operating environment (molecular modeling program); I, inhibitor; E:I, enzyme-inhibitor complex; SAR, structure-activity relationship; MM, molecular mechanics; CFF91, consistent force field 91; eint, interaction energy; 2D/3D, Two dimension/three dimension; RMSD, root mean square deviation; PH4, pharmacophoric hypotheses four features; Hypo, hypotheses; Et/Met/Ph, ethyl/methyl/phenyl; HB, hydrogen bond; vdW, van der waals; HOA, human oral absorption; nM, nanomolar

## Introduction

Malaria is endemic all over the world and causes many deaths among populations, especially in the tropics, mainly in sub-Saharan Africa. It is one of the largest parasitic infections in the world and its incidence is estimated at 212 million clinical cases with a mortality rate of 429 thousand people of which more than 303 thousand are children under 5 years or 70% of cases.<sup>1</sup>

This disease is caused by unicellular eukaryotic parasites of genus plasmodium. Of the five malaria parasites know to infect humans (*falciparum*, *vivax*, *malariae*, *ovale* and *knowlesi*), *plasmodium falciparum* is the most virulent.<sup>2</sup> One concern in the treatment of malaria is the spread of resistance to most available drugs. Chloroquine was one mainstay of antimalarial therapy. However, this drug is now inefficient against *Plasmodium falciparum* (Pf),<sup>3</sup> and recently clinical treatments failures have been reported for other antimalarial drugs, including artemisi-nin combined therapies (ACTs).<sup>4-6</sup> Since resistance to a wide range of antimalarial drugs is spreading throughout "at parasite infection risk" populations worldwide and a protective vaccine remains unavailable,<sup>7</sup> there is therefore urgent and imperative need to identify and develop novel classes of antimalarials. *Plasmodium falciparum* relies on hemoglobin hydrolysis to supply amino acids for protein synthesis and to maintain osmotic stability. Hemoglobin is transported to an acidic (pH 4-6) food vacuole, where it is hydrolyzed by multiple enzymes (E), including food vacuole cysteine, aspartic, and metalloproteases, before further action against hemoglobin fragments by a cytosolic aminopep-tidase.<sup>8,9</sup> Hemoglobin degradation is blocked by cysteine protease inhibitors (I), causing characteristic morphological abnormality in which the food vacuole is full of non-degraded hemoglobin blocking by this way parasite development.<sup>10,11</sup>

The Pf genome contains 33 open reading frames predic-ted to encode cysteine proteases, including 4 cathepsin L-like papain family proteases collectively known as falcipains.<sup>12</sup> Falcipain-2 (FP2) and falcipain-3 (FP3) are key papain-family (C1) Clan CA trophozoite cysteine proteases localized in the digestive food vacuole that cleave host native or denatured human hemoglobin. FP2, the most-expressed and best-studied enzyme among falcipains, is a promising

target for novel anti-malarial drugs development.<sup>13</sup> Disruption of FP2 leads to marked reduction in hemoglobin hydrolysis by trophozoites, although this phenotype is transient.<sup>14</sup> The main classes of FP2 inhibitors are peptides or peptido-mimetics bearing the most popular pharmacophores of cysteine protease inhibitors, such as vinyl sulfones, halomethyl ketones, and aldehydes. Furthermore, many other chemotypes have been identified as FP2 inhibitors, such as isoquinolines, thiosemicarbazones, and chalcones.<sup>15</sup> Also peptide nitriles have drawn increasing attention in recent years as the cyano group gives rise to preferred interaction with the active site thiol by the formation of an enzyme-bound thioimide that can result in potent enzyme inhibition.<sup>16–18</sup> The active site of FP2 is large and contains four pockets, S1, S'1, S2 and S3, each pocket accommodating one substituent, P1, P'1, P2 and P3 of peptide substrate. Thus, the scissile bond is localized between the residues P1' and P1 of the peptide and a cut that minimizes the energy of complex (protein-substrate) is obtained when the remaining substituents, P2 and P3, fit together perfectly with the remaining pockets S2 and S3. Previous structure-activity relationships (SAR) studies highlighted that FP2 and related enzymes, react preferentially with peptidyl ligands containing large hydrophobic side chain at P2 position and, have a clear preference for substrate that contain a leucine (Leu) residue at this position rather than phenylalanine (Phe) or valine (Val). In fact, the fragments in position P2 provides the basis for the specificity of the peptide inhibitors and the S2 subsite reflects the key determinant of selectivity of FP2.<sup>19</sup> Moreover, azadipeptide nitriles (ADPN) exhibit anti-malarial activity dependent on their structure and Reik Loser et al. reported the prominent role of the residue in P1 for development of this class of antimalarials.<sup>20</sup>

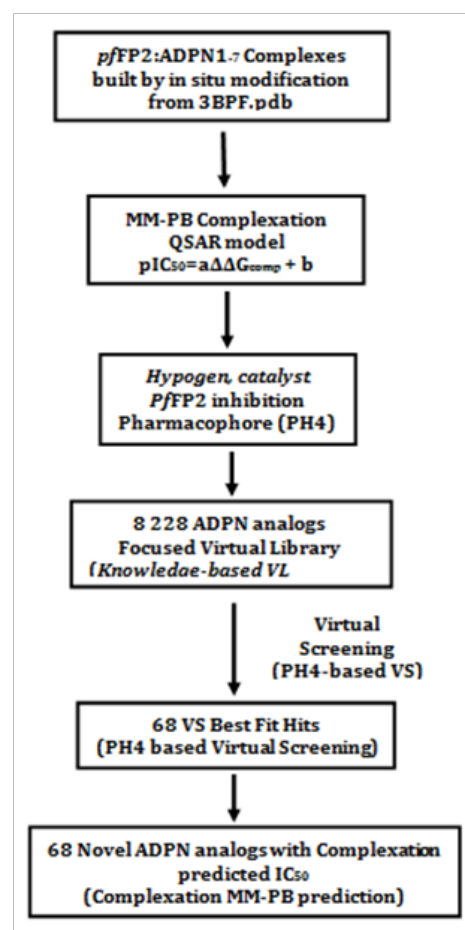
The reported ADPN inhibitors activity ( $IC_{50}$ ) range is very large: 59–50,001 nM.<sup>20</sup> The availability of X-ray crystal structure of FP2 in complex with epoxysuccinate E64 (pdb code 3BPF to resolution 2.9 Å), opened a gate to the structure-based design of novel potent antimalarial agents. In this work we start from a reduced series of 7 ADPN as a training set to construct a one descriptor (Gibbs free energy, GFE upon FP2-Inhibitor complex formation) QSAR model of FP2 inhibition, correlating GFE with experimental  $IC_{50}$ . Subsequently, a 3D-QSAR pharmacophore protocol was used to prepare a four-feature pharmacophore (PH4) model from FP2 active site bound ADPN conformations. This PH4's robustness due to the quality of FP2 inhibition structural information therein served to screen a virtual combinatorial library of ADPN analogues with the aim to design more potent orally bioavailable FP2 inhibitors.

## Materials and methods

The complexation methodology has been described largely according to a procedure successfully used to elaborate one descriptor QSAR models of viral, bacterial and protozoal enzyme-inhibitor complexes and to design from them peptidomimetic, hydroxynaphthoic, triclosan, thymidine, pyrrolidine carboxamide derivatives and peptide inhibitors.<sup>21–31</sup> The workflow below in Figure 1 describes the steps of whole process of virtual design of novel ADPN analogues.

### Training sets

The chemical structures of the training set comprising the 7 ADPNs and the experimental biological activities ( $IC_{50}^{exp}$ ) studied here are taken from the literature.<sup>20</sup> The  $IC_{50}^{exp}$  values of these compounds cover a very wide concentration range (59 - 50,001 nM) in order to build a reliable QSAR model.<sup>20</sup>



**Figure 1** Multistep approach workflow to virtually design novel ADPN analogs with higher predicted potency against FP2.

### Model building and calculation of binding affinity

Simulations were carried out for the enzyme: inhibitor (FP2:ADPN) complexes, the free enzyme FP2 and the free ADPN inhibitor from the high-resolution 3BPF.pdb structure using Insight II molecular modeling program. Initially no molecules of crystallographic water were conserved in the molecular model, then the structures of the FP2 and FP2: ADPN complexes were at pH 7 with neutral N- and C-terminal and finally the protonable and ionizable residues of FP2 were loaded. Inhibitors were prepared from 3BPF.pdb by *in situ* modification of functional groups in the molecular scaffold of the endogenous E64 inhibitor. All rotatable bonds of the replacing fragments were subjected to an exhaustive conformational search coupled with careful gradual energy-minimization of the modified inhibitor and active-site FP2 residues located in the immediate vicinity (around 5 Å radius) in order to identify low-energy bounded conformations of the modified inhibitor. The resulting low-energy structures of the entire FP2:ADPN complexes were then carefully refined by energy-minimization procedure the stable structure. The complete description of the complexation binding affinity ( $\Delta G_{com}$ ) is described in ref.<sup>30</sup>:

$$\Delta\Delta G_{com} = \Delta G_{com}(I) - \Delta G_{com}(I_{ref}) = \Delta\Delta H_{MM} - \Delta\Delta TS_{vib} + \Delta\Delta G_{sol}(I).$$

$\Delta\Delta H_{MM}$  describes the relative enthalpic contribution to the GFE change related to the intermolecular interactions in E:I complex

derived by molecular mechanics (MM),  $\Delta\Delta G_{\text{sol}}$  and  $\Delta\Delta T S_{\text{vib}}$  represent, respectively, the relative solvation GFE and simplified relative vibrational entropy.

### Molecular mechanics (MM)

Simulations of the inhibitors and their complexes models were carried out with all-atom representation using atomic and charge parameters of the class II consistent force field CFF91.<sup>27</sup> A dielectric constant of 4 was used for all molecular mechanics (MM) calculations in order to take into account the dielectric shielding effect in proteins. Minimizations of the E:I complexes, free E and I were carried out by relaxing the structures gradually, starting with added hydrogen atoms, then with residue side chain heavy atoms and by the protein backbone relaxation. In all the geometry optimizations, a sufficient number of steepest descent and conjugate gradient iterative cycles were used with the convergence criterion for the average gradient set to 0.01 kcal/mol Å<sup>-1</sup>.

### Conformational search

Conformations of free inhibitors were obtained from the bound conformations in the binary E:I complexes by gradual relaxation to the nearest local energy minimum. Then a Monte Carlo search ( $\leq 50,000$  iterations) for low-energy conformations was performed over all rotatable bonds, except those in the rings, using Discovery Studio 2.5 (DS 2.5) molecular modeling program.<sup>32</sup> Two hundred unique inhibitor conformations were generated by randomly varying torsion angles of the last accepted conformer by  $\pm 15^\circ$  at 5000K followed by subsequent energy minimization with a dielectric constant  $\epsilon=80$  to approximate the dielectric screening effect of solvation. The conformer with the lowest total energy was selected and re-minimized at a dielectric constant of 4.

### Solvation Gibbs free energy

The electrostatic component of the solvation GFE which includes the effect of ionic strength via solving the nonlinear Poisson-Boltzmann (PB) equation<sup>33,34</sup> was computed by the DelPhi module of Discovery Studio.<sup>32</sup> Here the solvent is a high dielectric constant ( $\epsilon_0=80$ ) continuous medium and the solute a low dielectric constant ( $\epsilon_1=4$ ) charge distribution filling a cavity with boundaries linked to the solute's molecular surface. DelPhi numerically solves for the molecular electrostatic potential and reaction field around the solute using finite difference method on a (235×235×235) cubic lattice grid for the E:I complexes and free E and on (65×65×65) grid for the free I. Full coulombic boundary conditions were employed. Two subsequent focusing steps led to a similar final resolution of about 0.3 Å per grid unit at 70% filling of the grid by the solute. Physiological ionic strength of 0.145 mol·dm<sup>-3</sup>, atomic partial charges and radii defined in the CFF force field parameter set<sup>32</sup> and a probe sphere radius of 1.4 Å were used. The electrostatic component of the PB solvation GFE was calculated as the reaction field energy.<sup>26,29,34–36</sup>

### Interaction energy

The MM interaction energy ( $E_{\text{int}}$ ) calculation protocol available in Discovery Studio 2.5<sup>32</sup> was used to compute the non-bonded interactions (van der Waals and electrostatic interatomic potential terms) between two sets of atoms in the E:I complexes. All pairs of each FP2 residue–inhibitor interactions of the total enzyme–inhibitor  $E_{\text{int}}$  were evaluated using CFF force field parameters at  $\epsilon=4$ .<sup>32</sup> In particular,  $E_{\text{int}}$  breakdown into active-site residues contributions

shows in the detail the strength and level of the main individual interactions, allowing in this way a quantitative identification of residues with highest contribution to the ligand binding affinity and suggests structural modifications in terms of inhibitor molecular moieties which carries the key features responsible for favorable binding affinity or biological activity.<sup>27</sup>

### 3D-QSAR pharmacophore (PH4) generation

PH4 modeling assumes that a set of key structural features responsible for compound biological activity is recognized by the active site during receptor binding. In this work the PH4 was prepared by the 3D-QSAR pharmacophore protocol of Catalyst HypoGen algorithm.<sup>37</sup> implemented in DS 2.5.<sup>31</sup> Bound conformations of ADPNs taken from the refined E:I complexes were considered for the PH4 building. The top scoring PH4 hypothesis was prepared in three stages: constructive, subtractive and optimization step, from a set of most active ADPNs. The inactive compounds served for the definition of excluded volume. During the generation, five features available in the HypoGen algorithm were selected: hydrophobic aromatic (HYdAr), hydrophobic aliphatic (HYd), hydrogen-bond donor (HBD), hydrogen-bond acceptor (HBA), and ring aromatic (Ar) feature. Default values of the adjustable parameters were kept during the PH4 generation, except the uncertainty on the biological activity ( $IC_{50}$ ), which was reduced to 1.25 instead of 3. This modification of the uncertainty interval of  $IC_{50}$  from a wide span ( $IC_{50}/3$ ,  $3 \cdot IC_{50}$ ) to a relatively narrow one ( $4 \cdot IC_{50}/5$ ,  $5 \cdot IC_{50}/4$ ), is due to accuracy and homogeneity of the measured  $IC_{50}$  originating from the same laboratory.<sup>20</sup> The top ten PH4 were generated with the number of missing features set to 0. Finally, the best PH4 model was selected. Generally a PH4 model, as the one described here, can be used to estimate the  $pIC_{50}$  of new analogues on the basis of their mapping to its features. In this study, priority was given to the PH4 based screening of ADME focused virtual libraries (VLs).

### ADME-related properties

Properties that determine the pharmacokinetics profile of a compound, besides octanol/water partitioning coefficient, aqueous solubility, blood/brain partition coefficient, Caco-2 cell permeability, serum protein binding, number of likely metabolic reactions and other eighteen descriptors related to adsorption, distribution, metabolism and excretion (ADME properties) of the inhibitors were computed by the QikProp program<sup>38</sup> based on the methods of Jorgensen.<sup>39–41</sup> According to those methods, experimental results of more than 710 compounds among which about 500 drugs and related heterocycles were correlated with computed physicochemical descriptors resulting in an accurate prediction of molecule's pharmacokinetic profile. Drug likeness (#stars) is represented by the number of descriptors that exceed the range of values determined for 95% of known drugs out of 24 selected descriptors computed by the QikProp.<sup>38</sup> Drug-likeness was used as the global compound selection criterion related to ADME properties. The calculation of the selected ADME descriptors were performed from 3D structures of compounds considered. These descriptors were used to assess the pharmacokinetics profile of designed compounds and served also for the VL focusing.

### Virtual combinatory library generation

The analogue model building was performed with Molecular Operating Environment (MOE) program.<sup>42</sup> The library of analogues was enumerated by attaching R-groups (fragments, building blocks)



onto ADPN scaffold using the Quasar CombiDesign module of MOE.<sup>42</sup> Reagents and chemicals considered in this paper were selected from the directories of chemicals available from the commercial sources. Each analogue was built as a neutral molecule in the MOE program,<sup>42</sup> its molecular geometry was refined by MM optimization through smart minimizer of DS<sup>32</sup> at high convergence criteria (threshold on energy difference of 10<sup>-4</sup> kcal/mol and root mean square deviation (RMSD) of 10<sup>-5</sup> Å), ε=4, using class II consistent force field CFF.<sup>43</sup>

### ADME-based library focusing

Twenty four pharmacokinetics-related molecular descriptors available in QikProp,<sup>38</sup> which characterize a wide spectrum of molecular properties as described in §2.8, Optimum ranges of these 24 descriptors were defined in terms of upper and lower bounds according to QikProp.<sup>38</sup> Only compounds with predicted drug-likeness (#stars, Section 2.8) equal to zero, were retained in the focused VL of drug-like ADPN analogues.

### Pharmacophore based library focusing

The PH4model described in Section 2.7 was derived from the bound conformations of ADPNs at the active-site of FP2. The enumerated and ADME-focused VL was further focused by using the ligand pharmacophore mapping protocol available of Discovery Studio.<sup>32</sup> Within this protocol, each generated conformer of the analogues was geometry optimized by means of the CFF force field for a maximum of 500 energy minimization steps and subsequently aligned and mapped to the PH4 model in order to select the top ranking overlaps. Twenty best-fitting inhibitor conformers were saved and clustered into 10 conformational families according to their mutual RMSD by Jarvis-Patrick complete linkage clustering method.<sup>44</sup> The best representative of each cluster was considered in the virtual screening of analogues. Only those analogues mapping to all four PH4 features were retained for the *in silico* screening.

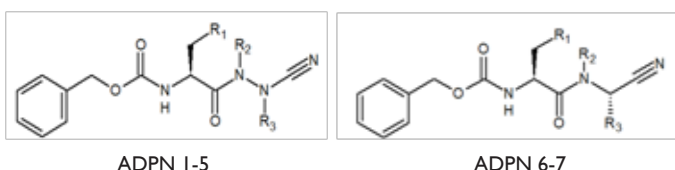
### In Silico screening

The conformer with the best mapping on the PH4 in each cluster of the focused library subset was selected for *in silico* screening by the complexation QSAR model. The relative GFE of E:I complex formation in water  $\Delta\Delta G_{\text{com}}$  was computed for each selected new ADPN analogue and then used for prediction of FP2 inhibitory potencies ( $IC_{50}^{\text{pre}}$ ) of the focused VL by inserting this parameter into the  $p/FP2$  receptor target-specific scoring function:  $pIC_{50}^{\text{pre}} = a \cdot \Delta\Delta G_{\text{com}} + b$  (2).

## Results

A series of 7 azadipeptide nitriles (ADPN) and their experimental activity ( $IC_{50}^{\text{exp}}$ ) from the same laboratory<sup>20</sup> were selected and presented in Table 1. Their  $IC_{50}^{\text{exp}}$  covering a very wide range of activity of 59 to 50 001 nM allowed to build a reliable QSAR model.

**Table 1** Training set of azadipeptide nitriles (ADPNx) inhibitors used for QSAR models preparation



Training set	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	IC <sub>50</sub> <sup>exp</sup> (nM)
ADPN1	-Ph	-CH <sub>3</sub>	-CH <sub>3</sub>	570
ADPN2	-i-Pr	-CH <sub>3</sub>	-CH <sub>3</sub>	59
ADPN3	-Ph	-CH <sub>3</sub>	-CH <sub>2</sub> -Ph	1000
ADPN4	-Ph	-CH <sub>3</sub>	-C <sub>2</sub> H <sub>5</sub> -Ph	430
ADPN5	-Ph	-CH <sub>3</sub>	-C <sub>5</sub> H <sub>11</sub>	310
ADPN6	-Ph	-H	-CH <sub>3</sub>	28 000
ADPN7	-Ph	-CH <sub>3</sub>	-CH <sub>3</sub>	50 001

The relative GFE upon (E:I) complex formation  $\Delta\Delta G_{\text{com}}$  was calculated for the FP2-DNAPx complexes as described in § Method. Table 2 shows GFE and its components, eq. (1).  $\Delta\Delta G_{\text{com}}$  reflects the mutual affinity between the enzyme and the inhibitor. Since it's calculated with an approximate approach the relevance of the binding model is evaluated by a linear regression, eq. (2) analysis which led to a linear correlation with experimental activity data ( $IC_{50}^{\text{exp}}$ ,<sup>20</sup>). Two correlation equations obtained for the GFE  $\Delta\Delta G_{\text{com}}$ , eq. (B) and its enthalpic component  $\Delta\Delta H_{\text{MM}}$ , eq. (A) are presented in Table 3 with the relevant statistical data. The relatively high values of the regression coefficient R<sup>2</sup> and the Fischer F-test of the correlation involving  $\Delta\Delta G_{\text{com}}$ , eq. (B), Table 3, indicate a strong relationship between the binding model and the experimental inhibitory potency  $IC_{50}^{\text{exp}}$  of the series of ADPN as plotted in Figure 2.

**Table 2** Complexation gibbs free energy (binding affinity) and its components for the training set of FP2 inhibitors ADPN 1-7

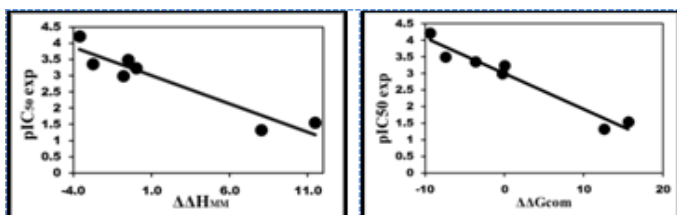
Training Set <sup>a</sup>	M <sub>w</sub> <sup>b</sup> (g/mol)	$\Delta\Delta H_{\text{MM}}$ <sup>c</sup> (kcal/mol)	$\Delta\Delta G_{\text{sol}}$ <sup>d</sup> (kcal/mol)	$\Delta\Delta TS_{\text{vib}}$ <sup>e</sup> (kcal/mol)	$\Delta\Delta G_{\text{com}}$ <sup>f</sup> (kcal/mol)	IC <sub>50</sub> <sup>exp</sup> g (μM)
ADPN1	366	0	0	0	0	570
ADPN2	332	-3.62	-1.43	4.33	-9.38	59
ADPN3	442	-0.8	0.53	0	-0.27	1000
ADPN4	456	-2.77	0.14	1.05	-3.68	430
ADPN5	422	-0.51	0.6	7.53	-7.44	310
ADPN6	351	11.46	2.61	-1.54	15.61	28000
ADPN7	365	8.03	3.37	-1.16	12.56	50 001

<sup>a</sup>For the chemical structures of the training set of inhibitors see Table 1

<sup>b</sup>M<sub>w</sub> is the molecular mass of the inhibitor; <sup>c</sup> $\Delta\Delta H_{\text{MM}}$  is the relative enthalpic contribution to the Gibbs free energy change related to the protease-inhibitor complex formation derived by molecular mechanics (MM):  $\Delta\Delta H_{\text{MM}} \cong [E_{\text{MM}}\{\text{PR:ADPNx}\} - E_{\text{MM}}\{\text{ADPNx}\}] - [E_{\text{MM}}\{\text{PR:ADPN1}\} - E_{\text{MM}}\{\text{ADPN1}\}]$ , ADPN1 is the reference inhibitor; <sup>d</sup> $\Delta\Delta G_{\text{sol}}$  is the relative solvation Gibbs free energy contribution to the Gibbs free energy change related to protease-inhibitor complex formation:  $\Delta\Delta G_{\text{sol}} = [G_{\text{sol}}\{\text{PR:ADPNx}\} - G_{\text{sol}}\{\text{ADPNx}\}] - [G_{\text{sol}}\{\text{PR:ADPN1}\} - G_{\text{sol}}\{\text{ADPN1}\}]$ ;

<sup>e</sup> $-\Delta\Delta TS_{\text{vib}}$  is the relative entropic contribution of the inhibitor to the Gibbs free energy related to protease-inhibitor complex formation:  $\Delta\Delta TS_{\text{vib}} = [TS_{\text{vib}}\{\text{ADPNx}\} - TS_{\text{vib}}\{\text{ADPN1}\}] - [TS_{\text{vib}}\{\text{PR:ADPNx}\} - TS_{\text{vib}}\{\text{PR:ADPN1}\}]$ ; <sup>f</sup> $\Delta\Delta G_{\text{com}}$  is the relative Gibbs free energy change related to the enzyme-inhibitor complex formation:  $\Delta\Delta G_{\text{com}} \cong \Delta\Delta H_{\text{MM}} + \Delta\Delta G_{\text{sol}} - \Delta\Delta TS_{\text{vib}}$ .

<sup>g</sup>IC<sub>50</sub><sup>exp</sup> is the experimental inhibition constant obtained from.<sup>20</sup>

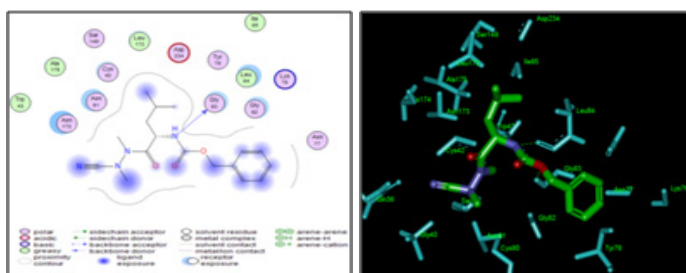


**Figure 2 (Left):** Plot of correlation equation between  $pIC_{50}$  and relative enthalpic contribution to the GFE (eq. (A))  $\Delta\Delta H_{MM}$ ; **(Right)** Similar plot for relative complexation GFE (eq. (B))  $\Delta\Delta G_{com}$  of the training set, all in kcal mol<sup>-1</sup>.

**Table 3** Regression analysis of computed binding affinities  $\Delta\Delta G_{com}$ , its enthalpic component  $\Delta\Delta H_{MM}$ , and their observed activities  $pIC_{50}^{exp} = -\log_{10} IC_{50}^{exp20}$  of azadipeptide nitriles (ADPNs)

Statistical Data of Linear Regression	(A)	(B)
$pIC_{50}^{exp} = -0.1738 \cdot \Delta\Delta H_{MM} + 3.1784$ (A)	-	-
$pIC_{50}^{exp} = -0.1067 \cdot \Delta\Delta G_{com} + 2.9995$ (B)	-	-
Number of compounds n	7	7
Square correlation coefficient regression $R^2$	0.88	0.94
LOO cross-validated Square correlation $R^2_{cv}$	0.86	0.93
Standard error of regression $\sigma$	0.388	0.264
Statistical significance of regression, Fisher F-test	39.48	91.44
Level of statistical significance $\alpha$	> 95 %	> 95 %
Range of activities $IC_{50}^{exp}$ [nM]	59-50 001	

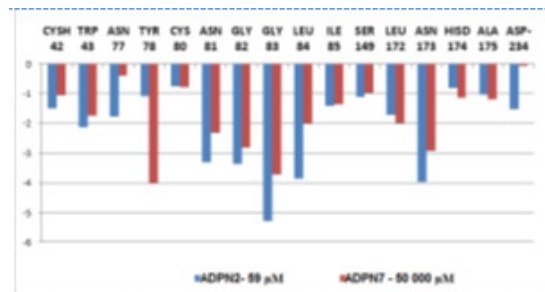
The ADPN's active conformation from this QSAR is revealed in Figure 3, (Figure 4) (Figure 5) (Table 4) for the most active ADPN2. Subsequently the enzyme-inhibitor intermolecular interaction energy breakdown to the active site pockets key residues are listed in Table 4 with their correlation with experimental activity ( $IC_{50}^{exp}$ ) plotted in Figure 4. A comparative individual residue contribution between the most active ADPN2 and the less active ADPN7 confirms the observed  $IC_{50}^{exp}$  trend.



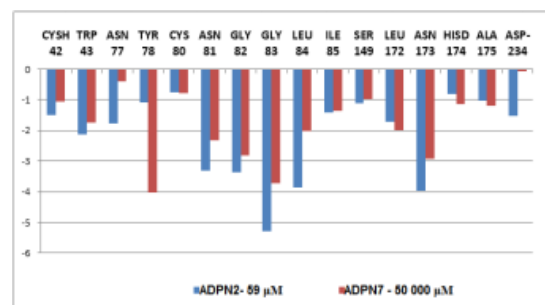
**Figure 3 (Left)** 2D schematic interaction diagram of the most potent inhibitor ADPN2 (Table 1) at pFP2 active-site; **(Right)** 3D structure of the active-site with bound inhibitor ADPN2.

The FP2 inhibitory pharmacophore (PH4) model elaborated from the QSAR model is presented in Table 5 & Table 6, Figure 6. The 3D-QSAR PH4 generation was carried out in three steps: the constructive, the subtractive and the optimization step (§ Method). During the constructive phase of HypoGen the most active ADPNs for which  $IC_{50}^{exp} \leq 2 \times 59$  nM, were automatically selected as the leads. Thus, the most active ADPN2 ( $IC_{50}^{exp} = 59$  nM) alone was used to generate the starting PH4 features and those matching this lead, were

retained. During the subsequent subtractive phase, features which were present in more than half of the inactive ADPNs were removed. The PH4 which contained all features were retained. None of the training set compounds was found to be inactive ( $IC_{50}^{exp} > 59 \times 10^3$  nM = 186574.3 nM). During the final optimization phase, the score of the PH4 hypotheses was improved. Hypotheses are scored via simulated annealing approach according to errors in the activity estimates from regression and complexity. At the optimization end, 10 best scoring unique hypotheses (Table 5) displaying four features were kept. The reliability of the generated PH4 models was then assessed using the calculated cost parameters ranging from 26.6 (Hypo1) to 29.9 (Hypo10). Their statistical data (costs, root-mean-square deviation RMSD,  $R^2$ ) are listed in Table 6,  $0.353 \leq RMSD \leq 0.801$  and  $0.98 \leq R^2 \leq 0.99$ . The PH4 hypo1, with the best RMSD and highest  $R^2$  was retained for further analysis; its regression equation  $pIC_{50}^{exp} = 0.9926 \cdot pIC_{50}^{pre} + 0.0527$  (plotted in Figure 6),  $R^2$  and  $R^2_{cv}$  greater than 0.9 and F-test of 1272.8 attest the predictive capacity of the PH4.



**Figure 4** Plot of the correlation between active site key residues CysH42, Trp43, Ser149, Asp234, Gly82, Gly83, P2 and PIP2 and S2 contribution to  $E_{int}$  of pFP2-ADPNx.



**Figure 5** Molecular mechanics intermolecular interaction energy  $E_{int}$  breakdown (in kcal.mol<sup>-1</sup>) to residue contributions for ADPN2 and ADPN7, Table 2.<sup>20</sup>

The Hypo1 fixed cost (26.2), lower than the null (127.0) by  $\Delta = 100.8$ , is a chief indicator of the PH4 model predictability ( $\Delta > 70$  corresponds to a probability higher than 90% that the model represents a valid correlation).<sup>16</sup> The difference  $\Delta \geq 97.1$  for the set of 10 hypotheses confirms high quality of the pharmacophore model. The best-selected hypothesis Hypo1 represents with a probability of 98% a PH4 model with a similar level of predictive power as the complexation GFE enzyme-inhibitor binding QSAR model.

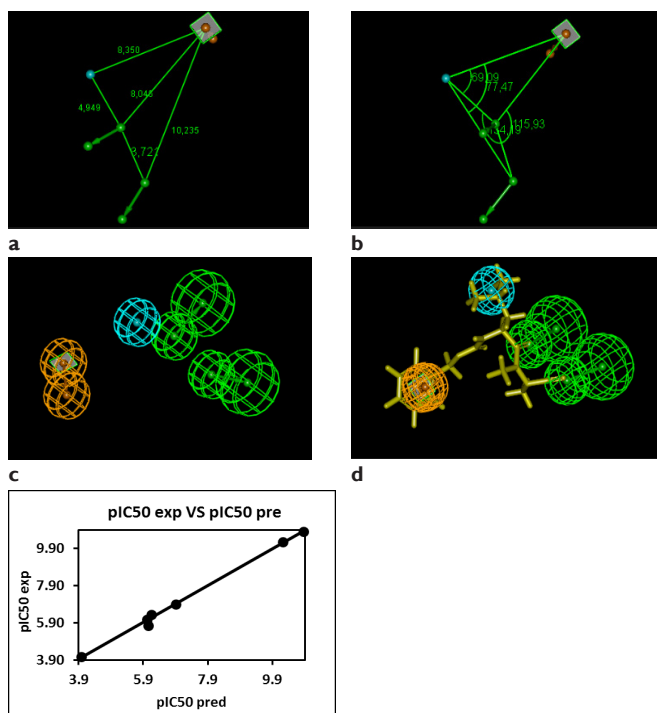
We have built a virtual library (VL) of new azadipeptide nitriles compounds with a variety of substitutions in positions P1, P2 and of the nitrogen adjacent to the cyano group with the goal to identify more potent orally bioavailable inhibitors of the pFP2. They have

been synthesized following a previously published protocol.<sup>12</sup> During the VL enumeration the R-groups listed in Table 7 were attached to positions R1–R3 of the ADPN scaffold to form a combinatorial library of the size:  $R1 \times R2 \times R3 = 44 \times 17 \times 11 = 8228$  analogues. In order to match the substitution pattern of the best training set inhibitor ADPN2 and to take into account the reported structural information about P1 and P2 position that is suitable for substitutions<sup>20</sup> and not excluded through the Lipinski's rule violation (molecular weight  $>500 \text{ g.mol}^{-1}$ ).<sup>17</sup> In search of more focused library with increased content of drug-like

and orally bioavailable molecules, we have introduced a set of filters and penalties, which can help to select smaller number of suitable ADPNs that will undergo *in silico* screening. The initial VL was thus filtered in an ADME-based focusing step to remove compounds with expected poor oral bioavailability and low drug-likeness. Only those with high predicted percentage of human oral absorption (HOA) in the gastrointestinal tract larger than 80%<sup>19,20</sup> and compounds satisfying the Lipinski's rule of five<sup>17</sup> computed for the entire virtual library using QikProp software,<sup>21</sup> were kept.

**Table 4** Active site residue contribution to  $E_{\text{int}}$  (in kcal/mol)

Pockets	Residues	ADPN1	ADPN2	ADPN3	ADPN4	ADPN5	ADPN6	ADPN7
<b>pIC<sub>50</sub><sup>exp</sup></b>		3.24	4.22	3	3.36	3.50	1.55	1.33
	Gln36	-0.89	-0.24	-1.23	-1.03	-1.12	-0.13	-0.11
	Gly40	-1.52	-0.34	-2.39	-1.53	-1.96	-0.42	-0.52
	Ser41	-0.37	-0.09	-0.61	-0.40	-0.51	-0.06	-0.15
<b>S1</b>	Cys42	-1.69	-1.50	-1.42	-1.97	-1.64	-0.96	-1.06
	Trp43	-2.11	-2.13	-2.12	-2.44	-2.06	-1.77	-1.74
	Cys80	-1.16	-0.74	-1.46	-1.07	-1.24	-0.60	-0.77
	Asn81	-3.02	-3.29	-5.42	-3.18	-4.55	-1.96	-2.32
	<b>total</b>	<b>-10.71</b>	<b>-8.33</b>	<b>-14.65</b>	<b>-11.62</b>	<b>-13.08</b>	<b>-5.90</b>	<b>-6.67</b>
<b>S2</b>	Leu84	-2.44	-3.86	-2.28	-2.62	-2.40	-2.31	-2.02
	Ile85	-1.85	-1.42	-1.80	-1.93	-1.87	-1.48	-1.37
	Ser149	-1.02	-1.11	-0.96	-1.07	-0.99	-0.95	-0.92
	Leu172	-1.79	-1.71	-1.73	-1.97	-1.75	-1.97	-2.00
<b>S2</b>	Ala175	-1.89	-0.80	-1.96	-1.87	-1.94	-1.13	-1.20
	Asp-234	-1.29	-1.52	-1.05	-1.49	-1.17	-0.19	-0.07
	<b>total</b>	<b>-10.28</b>	<b>-10.42</b>	<b>-9.78</b>	<b>-10.95</b>	<b>-10.12</b>	<b>-8.07</b>	<b>-7.62</b>
<b>S'I</b>	Val150	-0.18	-0.10	-0.18	-0.19	-0.18	-0.20	-0.22
	Val152	-0.15	-0.10	-0.17	-0.22	-0.16	-0.10	-0.11
	Asn173	-3.45	-3.96	-3.53	-5.18	-3.56	-3.01	-2.92
	His174	-1.80	-0.80	-2.06	-1.86	-1.95	-1.19	-1.13
<b>S'I</b>	Trp206	-0.65	-0.23	-1.03	-0.61	-0.82	-0.18	-0.19
	Asn204	-0.08	-0.04	-0.11	-0.06	-0.09	-0.05	-0.05
	<b>total</b>	<b>-6.31</b>	<b>-5.23</b>	<b>-7.08</b>	<b>-8.12</b>	<b>-6.76</b>	<b>-4.73</b>	<b>-4.62</b>
<b>S3</b>	Lys+76	0.12	-0.34	0.09	0.11	0.11	-0.05	-0.07
	Asn77	-0.41	-1.77	-0.39	-0.44	-0.39	-0.37	-0.39
	Tyr78	-2.18	-1.08	-2.24	-2.02	-2.23	-4.13	-4.03
	Gly82	-3.03	-3.37	-3.24	-3.38	-3.19	-2.40	-2.80
<b>S3</b>	Gly83	-4.56	-5.29	-4.36	-4.58	-4.50	-3.44	-3.71
	<b>total</b>	<b>-10.06</b>	<b>-11.85</b>	<b>-10.14</b>	<b>-10.31</b>	<b>-10.20</b>	<b>-10.39</b>	<b>-11.00</b>



**Figure 6** Coordinates (a,b), features (c) and mapping (d) of the FP2 inhibitor pharmacophore with the best training set ADPN2 (yellow). The correlation plot of experimental vs. predicted inhibitory activity is displayed above. The features are colored blue for hydrophobic aliphatic (HYd), green for hydrogen-bond acceptor (HBA), purple for hydrogen-bond donor (HBD) and orange for Aromatic (Ar). The arrows represent the projection for donor and acceptor features.

**Table 5** Output parameters of 10 generated PH4 hypotheses for FP2 inhibitors after CatScramble validation procedure

Hypothesis	RMSD <sup>a</sup>	R2 <sup>b</sup>	Total costs <sup>c</sup>
Hypo1	0.353	0.99	26.868
Hypo2	0.509	0.99	27.216
Hypo3	0.528	0.99	27.275
Hypo4	0.393	0.99	27.332
Hypo5	0.597	0.99	27.53
Hypo6	0.745	0.98	28.722
Hypo7	0.658	0.99	28.831
Hypo8	0.74	0.98	28.95
Hypo9	0.792	0.98	29.908
Hypo10	0.801	0.98	29.941
Fixed Cost	0	1	26.281
Null Cost	5.619	0	127.077

<sup>a</sup>root mean square deviation (RMSD); <sup>b</sup>squared correlation coefficient; <sup>c</sup>overall cost parameter of the PH4

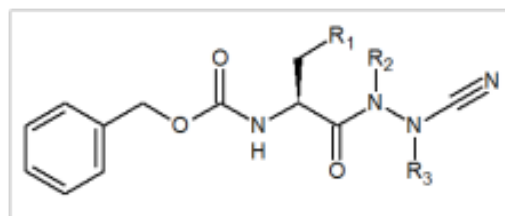
**Table 6** Regression analysis of  $pIC_{50}^{exp20}$  and computed  $pIC_{50}^{pred}$  of ADPNs toward pFP2

#### Statistical Data of Linear Regression for Hypo 1

$$pIC_{50}^{exp} = 0.9926 \times pIC_{50}^{pred} + 0.0527 \text{ (C)}$$

Number of compounds n	7
Square correlation coefficient regression R2	0.9
LOO cross-validated Square correlation R2xv	0.9
Standard error of regression $\sigma$	0.169
Statistical significance of regression, Fisher F-test	1272
Level of statistical significance $\alpha$	>95 %
Range of activities $IC_{50}^{exp}$ [nM]	59-50 001

**Table 7** R-groups (fragments, building blocks, substituents) used in the design of the diversity library of ADPN analogs



#### R-groups<sup>a,b</sup>

1. Cyclopentane	28. 3-Me Pyridine
2. Me-C <sub>6</sub> H <sub>11</sub>	29. 4-Et Pyridine
3. Et-Cyclopentane	30. 4-Me Pyrimidine
4. 2-Me-2H-imidazole	31. 4-Et Pyrimidine
5. Cyclohexene	32. i-but
6. 3-Et-Cyclopentane	33. -CH <sub>2</sub> CCF
7. Et-C <sub>6</sub> H <sub>11</sub>	34. -CH <sub>2</sub> CCCl
8. Et-Ph	35. -CHCCH <sub>2</sub>
9. Ethylfurane	36. -CH <sub>2</sub> CONH <sub>2</sub>
10. 4-Me-tetrahydro-2H-pyran	37. -C <sub>2</sub> H <sub>5</sub> COO <sup>-</sup>
11. 2,2-diMe-butane	38. -CH <sub>2</sub> CCBr
12. 3-ethylfurane	39. -C <sub>2</sub> H <sub>5</sub> CHO
13. 1-Et-4-Me-C <sub>6</sub> H <sub>11</sub>	40. -CHO
14. 1-Et-3-Me-C <sub>6</sub> H <sub>11</sub>	41. -C <sub>2</sub> H <sub>5</sub> OH
15. 1-Et-2-Me-C <sub>6</sub> H <sub>11</sub>	42. -COO <sup>-</sup>
16. 1-Et-4-Me-Ph	43. -OH
17. 1-Et-3-Me-Ph	44. -F
18. 1-Et-2-Me-Ph	45. -Cl
19. 5-Et-2-Me-piperidine	46. -Br
20. 2-Et-5-Me-piperidine	47. -I
21. 3-Et-5-Me-piperidine	48. -CH <sub>3</sub>
22. 2-Et-4-Me-piperidine	49. -CH <sub>2</sub> F
23. 3-Et-4-Me-piperidine	50. -CH <sub>2</sub> Cl
24. 3-Et-3-Me-piperidine	51. -CH <sub>2</sub> Br
25. 4-Et-tetrahydro-2H-pyran	52. -CH <sub>2</sub> I
26. 2-Et-tetrahydrofuran	
27. 2,3-diMe-tetrahydrofuran	
	53. -CF <sub>3</sub>
	54. -CCl <sub>3</sub>
	55. -CBr <sub>3</sub>
	56. -Cl <sub>3</sub>
	57. -NO <sub>2</sub>
	58. -NO
	59. -OCH <sub>3</sub>
	60. -C <sub>3</sub> H <sub>7</sub> CHO
	61. -C <sub>2</sub> H <sub>5</sub> OH
	62*. -CH <sub>2</sub> COO <sup>-</sup>
	63. -C <sub>2</sub> H <sub>5</sub> OH
	64*. -CONH <sub>2</sub>
	65. -C <sub>2</sub> H <sub>5</sub> CONH <sub>2</sub>
	66. -C <sub>2</sub> H <sub>5</sub> F
	67. -C <sub>2</sub> H <sub>5</sub> Cl
	68. -C <sub>2</sub> H <sub>5</sub> Br
	69. -C <sub>2</sub> H <sub>5</sub> I
	70. 2-Me-C <sub>5</sub> H <sub>11</sub>

<sup>a</sup>fragments 1– 42 and 62\* and 64\* were also used in R1-groups; fragments 43–59 were used in R2-group and fragments 60–70 were used in R3-groups; <sup>b</sup>dashed bonds indicates the attachment points of individual fragments.

From the set of 3023 analogs derived from Lipinski test, 93 of them were mapped to 4 features of the pharmacophore. Out of then, only 68 best fitting analogues (PH4 hits) have been retained and submitted to screening with help of the complexation QSAR model. Computed

GFE of complex formation with FP2 of *Pf* and its component as well as predicted half-maximal inhibitory concentrations  $IC_{50}^{pred}$  estimated from the correlation Equation (B), Table 3. For the majority of new ADPN analogues, the estimated inhibitory potencies shown in Table 8 are better than that for the most active training set compound ADPN2 ( $IC_{50}^{exp}=59nM$ ).<sup>20</sup>

In order to identify which substituents (Table 7) lead to new inhibitor candidates with the highest predicted potencies towards the FP2 of *Pf*, we have prepared histograms of the frequency of occurrence

of R1, R2 and R3 groups in the 68 PH4 best fitting hit ADPNs selected from the focused virtual library shown in Table 8 (Figure 7). From the histograms in Figure 7 it comes out that R1 groups numbered

11(20),13(4),14(4),15(4),19(4),20(4),21(4),22(4) and 23(4) are almost equally represented with the highest occurrence in the ADPNs subset. The R2 groups contain preferentially 42(25), 48(20), 49(4) and 50(5), 36(3) while the R3 groups include chiefly fragments 60(26), 61(28) and 70(12).

**Table 8** Complexation Gibbs free energy and its components for the 68 virtually designed ADPN

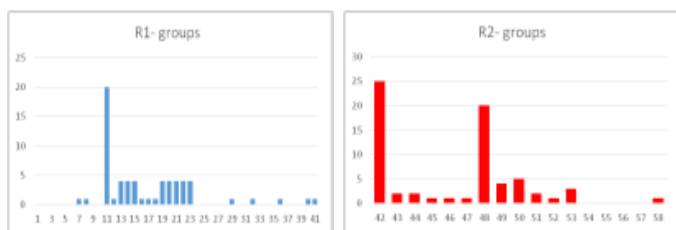
Analogs designed	$M_w^a$	$\Delta\Delta H_{MM}^b$	$\Delta\Delta G_{sol}^c$	$\Delta\Delta TS_{vib}^d$	$\Delta\Delta G_{com}^e$	$IC_{50}^{pre f}$
<b>ADPN2</b>	332	0	0	0	0	59 <sup>g</sup>
12-50-60	474.9	-12.41	-0.54	2.17	-15.12	23.63
7-59-60	472.2	-11.11	-1.15	8.53	-20.79	5.80
10-49-32	476.5	-9.39	1.50	7.9	-15.79	20.02
8-58-61	495.5	-6.74	-1.18	7.75	-15.67	20.62
32-50-61	467.0	-6.44	-0.42	11.2	-18.06	11.40
41-49-29	469.5	-11.16	-1.28	1.95	-14.39	28.32
36-50-61	467.9	-8.87	1.29	7.57	-15.15	23.46
38-59-60	479.3	-7.89	-3.17	3.73	-14.79	26.68
40-53-61	458.4	-11.39	0.74	5.12	-15.77	20.12
11-59-70	446.5	-4.44	-2.38	16.83	-23.65	2.85
11-43-70	460.6	-6.72	-2.11	13.10	-21.93	4.37
11-44-70	462.6	-5.54	-2.69	13.26	-21.49	4.87
11-45-70	479.0	-6.19	-1.79	11.58	-19.56	7.80
11-46-70	523.5	-5.9	-1.86	12.16	-19.92	7.19
11-47-70	570.5	-4.07	-1.88	11.08	-17.03	14.72
11-48-70	458.6	-5.31	-1.90	18.17	-25.38	1.86
11-49-70	476.6	-6.31	-1.69	17.49	-25.49	1.81
11-50-70	493.0	-5.85	-1.90	16.15	-23.90	2.68
11-51-70	537.5	-7.13	-1.71	16.68	-25.52	1.79
11-52-70	584.5	1.06	-1.42	15.83	-16.19	18.13
11-53-70	512.6	-10.34	-1.77	15.57	-27.68	1.05
11-53-61	514.5	-11.14	-0.71	11.12	-22.97	3.38
11-59-61	476.6	-4.89	-0.93	16.35	-22.17	4.12
11-43-61	462.5	-4.24	-1.39	13.58	-19.21	8.58
11-44-61	464.5	-2.87	-1.37	11.08	-15.32	22.49
11-48-61	460.6	-5.17	-2.10	15.88	-23.15	3.23
11-49-61	478.6	-5.89	-0.16	11.84	-17.89	11.89
11-50-60	495.0	-4.48	-0.65	10.99	-16.12	18.44
11-51-60	539.5	-9.84	0.36	9.21	-18.69	9.76
13-59-60	486.6	-12.06	-1.19	10.29	-23.54	2.93
14-59-60	486.6	-11.00	-2.03	9.98	-23.01	3.34
15-59-60	486.6	-10.44	-1.24	10.07	-21.75	4.57
19-59-60	480.5	-9.17	-0.91	5.25	-15.33	22.43



Table Continued

Analogs designed	M <sub>w</sub> <sup>a</sup>	$\Delta\Delta H_{MM}$ <sup>b</sup>	$\Delta\Delta G_{sol}$ <sup>c</sup>	$\Delta\Delta TS_{vib}$ <sup>d</sup>	$\Delta\Delta G_{com}$ <sup>e</sup>	IC <sub>50</sub> <sup>pre f</sup>
<b>ADPN2</b>	332	0	0	0	0	59 <sup>g</sup>
17-59-60	480.5	-11.03	-0.31	3.66	-15.00	24.35
18-59-60	480.5	-9.74	-0.80	6.49	-17.03	14.72
19-59-60	487.5	-12.49	-1.36	9.66	-23.51	2.95
20-59-60	487.5	-13.36	-0.72	10.56	-24.64	2.23
21-59-60	487.5	-10.79	-1.40	12.41	-24.60	2.25
22-59-60	487.5	-14.06	-1.42	10.52	-26.00	1.59
23-59-60	487.5	-11.66	-0.53	10.39	-22.58	3.72
24-59-60	487.5	-12.82	-1.79	11.48	-26.09	1.56
13-48-60	470.6	-13.21	-1.15	12.89	-27.25	1.17
14-48-60	470.6	-12.64	-1.66	10.36	-24.66	2.22
15-48-60	470.6	-11.88	-0.61	13.92	-26.41	1.44
19-48-60	471.5	-13.80	-0.50	8.61	-22.91	3.43
20-48-60	471.5	-14.53	-0.05	7.11	-21.69	4.64
21-48-60	471.5	-12.92	-1.91	9.94	-24.77	2.16
22-48-60	471.5	-14.02	-1.13	8.57	-23.72	2.80
23-48-60	471.5	-12.13	-0.06	11.47	-23.66	2.84
24-48-60	471.5	-14.00	-1.53	9.47	-25.00	2.04
13-48-61	500.6	-11.36	-0.21	15.43	-27.00	1.24
14-48-61	500.6	-0.14	-1.11	16.76	-18.01	11.55
<b>15-48-61</b>	<b>500.6</b>	<b>-9.93</b>	<b>-1.60</b>	<b>19.20</b>	<b>-30.73</b>	<b>0.5</b>
<b>19-48-61</b>	<b>501.6</b>	<b>-11.85</b>	<b>-1.61</b>	<b>15.61</b>	<b>-29.07</b>	<b>0.74</b>
<b>20-48-61</b>	<b>501.6</b>	<b>-12.56</b>	<b>0.21</b>	<b>16.26</b>	<b>-28.61</b>	<b>0.83</b>
21-48-61	501.6	-8.26	-0.54	16.80	-25.60	1.76
<b>22-48-61</b>	<b>501.6</b>	<b>-10.18</b>	<b>-1.27</b>	<b>19.21</b>	<b>-30.66</b>	<b>0.5</b>
23-48-61	501.6	-6.69	-0.19	16.48	-23.36	2.92
<b>24-48-61</b>	<b>501.6</b>	<b>-13.19</b>	<b>-0.50</b>	<b>16.96</b>	<b>-30.65</b>	<b>0.5</b>
<b>13-59-61</b>	<b>516.6</b>	<b>-9.85</b>	<b>-0.75</b>	<b>17.56</b>	<b>-28.16</b>	<b>0.93</b>
<b>14-59-61</b>	<b>516.6</b>	<b>-9.06</b>	<b>-1.55</b>	<b>18.31</b>	<b>-28.92</b>	<b>0.77</b>
15-59-61	516.6	-8.59	-0.50	17.38	-26.47	1.42
19-59-61	517.6	-10.55	-0.63	15.94	-27.12	1.20
20-59-61	517.6	-10.55	-0.25	15.30	-26.10	1.55
21-59-61	517.6	-9.99	-1.27	15.61	-26.87	1.28
<b>22-59-61</b>	<b>517.6</b>	<b>-13.19</b>	<b>-1.29</b>	<b>15.01</b>	<b>-29.49</b>	<b>0.67</b>
23-59-61	517.6	-9.11	-1.14	17.60	-27.85	1.00
24-59-61	517.6	-9.26	-1.39	15.92	-26.57	1.38

<sup>a</sup>MW (g/mol) is molecular mass of the inhibitor; <sup>b</sup> $\Delta\Delta H_{MM}$  (kcal/mol) is the relative enthalpic contribution to the Gibbs free energy change related to the FP2-ADPN complex formation  $\Delta\Delta G_{com}$  (for details see footnote of Table 2); <sup>c</sup> $\Delta\Delta G_{sol}$  (kcal/mol) is the relative solvation Gibbs free energy contribution to  $\Delta\Delta G_{com}$ ; <sup>d</sup> $\Delta\Delta TS_{vib}$  (kcal/mol) is the relative entropic (vibrational) contribution to  $\Delta\Delta G_{com}$ ; <sup>e</sup> $\Delta\Delta G_{com}$  (kcal/mol) is the relative Gibbs free energy change related to the enzyme-inhibitor FP2-ADPN complex formation  $\Delta\Delta G_{com} = -\Delta\Delta H_{MM} + \Delta\Delta G_{sol} + \Delta\Delta TS_{vib}$ ; <sup>f</sup>IC<sub>50</sub><sup>pre</sup> (μM) is the predicted half-maximal inhibitory concentration of ADPNx towards FP2 of Pf calculated from  $\Delta\Delta G_{com}$  using correlation Equation (B), Table 3; <sup>g</sup>IC<sub>50</sub><sup>exp</sup> is given for the reference inhibitor ADPN2 instead of IC<sub>50</sub><sup>pre</sup> (M).



**Figure 7** Histograms of frequency of occurrence of individual R1, R2-groups in the 68 best-selected analogues mapping to the four features of the PH4 pharmacophore hypothesis Hypo I (for fragments structures Table 7).

## Discussion

From a reduced series of 7 ADPN as training set a one descriptor (FP2-ADPN1-7 complex formation GFE:  $\Delta\Delta G_{\text{com}}$ ) QSAR model of FP2 inhibition ( $\text{pIC}_{50} = f(\Delta\Delta G_{\text{com}})$ ) was built revealing the ADPN active conformation. Then a 3D-QSAR pharmacophore (PH4) model of FP2 inhibition of bound ADPN conformations was elaborated. The PH4's predictability served to screen a virtual combinatorial library of ADPN analogues with the aim to design more potent orally bioavailable FP2 inhibitors.

The most important findings are first the likeness of the constructed active conformation of the ADPN1-7 from the complexation QSAR model and second the accuracy of the subsequent PH4's  $\text{IC}_{50}$  estimation as plotted in Figure 6. Overall this confirms the structural hypothesis according to which the inhibitory activity of structure-based designed similar ADPNs sharing the binding mode of the training set ADPN1-7 will be predicted suitably by the QSAR regression equation (B) and cross checked by the PH4 regression equation (C).

## QSAR model

The robustness of this one descriptor QSAR model is assessed through the components of GFE ( $\Delta\Delta G_{\text{com}}$ ) namely the enthalpic  $\Delta\Delta H_{\text{MM}}$ , solvation  $\Delta\Delta G_{\text{sol}}$  and the loss of vibrational entropy upon the ADPN binding  $\Delta\Delta TS_{\text{vib}}$ . The relevance of the enthalpic contribution to GFE is well confirmed by the quality of the regression indicating that in gas phase a large part (some 88%) of the variation of the  $\text{IC}_{50}$  is explained by that of  $H_{\text{MM}}$ . Adding to  $H_{\text{MM}}$  the solvation contribution in order to come closer to the biological medium, kept the level of strong relationship between the experimental data and the simulation results. Finally the likeliness of the model is increased by the loss of the inhibitor vibrational entropy  $TS_{\text{vib}}$  to explain some 94% of the variation of  $\text{IC}_{50}$  by that of GFE. This last contribution is one of the most reliable indicators of the predictive power of the QSAR model as reported by Freire et al.<sup>45</sup> For this reason the lack of validation set (VS), due to the small number of provided ADPNs experimental values is not detrimental, because the validation of the QSAR model is much more performed through the high predictive quality of the PH4 model derived from it than the VS, provided that it's based on the bound conformation of the ADPNs (Figure 6 & Table 6). Therefore, the correlation equation (B) and computed  $\Delta\Delta G_{\text{comp}}^{\text{pred}}$  quantities can be used for prediction of inhibitory potencies  $\text{IC}_{50}^{\text{pred}}$  against *p*/FP2 for novel ADPN analogues respectively, since they share the same binding mode as their corresponding training set.

## Binding mode of inhibitors

Beside the robustness of the QSAR model, the analysis of the

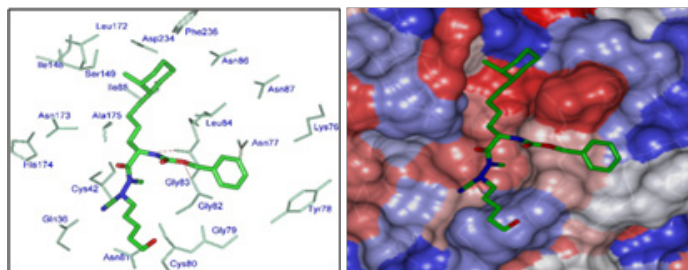
interactions between ADPNs and FP2 is expected to reveal key interactions justifying ADPNx: FP2 affinity such as hydrogen bonds (HBs), van der Waals (vdW), hydrophobic contacts, etc. As displayed in Figure 3, the binding mode of ADPNs at FP2 active site of the best active ADPN 2 in 2D and 3D is supported by the following interactions: HB with Gly83 and Leu84 and hydrophobic contacts. In order to verify whether other interesting interactions not displayed have to be taken into account in the description of ADPN binding mode at FP2 active site for the design of new analogues, interaction energy ( $E_{\text{int}}$ ) between each active site residue and ADPNx is computed (Table 4). The peptidyl structure of ADPN shed light on the structural features for binding affinity and opened the gate to the design of FP2 inhibitors taking benefit from S1 to S3 pockets filling. In this way key residues from these pockets contribution to the global interaction energy FP2-ADPNx correlate with their inhibitory potencies as displayed in Figure 4. S2 pocket filling exemplifies the overall impact for FP2 inhibition as reported<sup>13</sup> previously. Oppositely Figure 5 displaying comparative contribution of each active site residue to  $E_{\text{int}}$  for the most active training set ADPN2 and the less active ADPN7 confirm the observed  $\text{IC}_{50}^{\text{exp}}$  trend but cannot justify the large gap in their inhibitory potencies. However in a recent work we succeeded in justifying the observed 37.5% jump in experimental biological activity between methylphosphonic arginine and hydroxamic acid derivative *Pf*Leucyl aminopeptidase (*p*/A-M17) inhibitors by the enzyme active site residues contribution to  $E_{\text{int}}$  at a level of 35%.<sup>46</sup> Therefore the essential structural information in the design of novel potent ADPN analogues will not be derived from  $E_{\text{int}}$  but from a more predictive descriptor. They will be selected, according to the workflow (Figure 1), by virtual screening from a diverse Virtual Library of analogues with the S2 hydrophobic contact as the central structural requirement displayed by the Pharmacophore model of FP2 inhibition provided by the one descriptor (GFE) QSAR model (Table 3) (Figure 2).

## Analysis of new inhibitors from *in Silico* screening

The top scoring virtual hits are ADPN analogues: 15-48-61 (0.5nM), 22-48-61 (0.5nM), 24-48-61 (0.5nM), 19-48-61 (0.74nM), 22-59-61 (0.67nM), 14-59-61 (0.77nM), 20-48-61 (0.83nM) and 13-59-61 (0.9 nM) display predicted potency about one hundred times that of the best training set ADPN2. Albeit this exaggeratedly optimistic result, it suggests alternative suitable sidechains (different from current amino acid ones) for pockets filling. As we can see from Figure 8 the 3D interaction depiction (Top) of one of the best designed analogues 15-48-61 with FP2 and the Connolly surface of the binding site (Bottom), the n-pentyl (P1) of the training set is extended to n-pentyl hydroxyl while the benzyloxy is conserved in P3 for all the designed analogues mimicking the training set P3. The lipophilic S2 pocket accommodates a bulkier 1-ethyl-2-Me- $\text{C}_6\text{H}_{11}$  (15) in place of the cyclohexyl moiety leading to an increase of hydrophobic contact equivalent to a better stabilization and greater affinity. These results are in agreement with reported structural information from experimental Structure - Activity Relationship (SAR) on FP2 and FP3 pyrimidine-carbonitrile inhibitors<sup>47</sup> and Complexation QSAR based *in silico* design of dipeptide nitriles inhibitors of FP3;<sup>31</sup> they attest the accuracy of the FP2 pharmacophore derived from the one descriptor complexation QSAR model of FP2 inhibition by ADPNs.

The ADME-related properties described in Section 3.9 were computed. The values for the best active designed ADPNs were compared with those computed for drugs used for treatment of malaria or currently undergoing clinical trials. Most of the best designed

analogues ADPNs are predicted to display good level of druglikeness as well as high human oral absorption in the gastrointestinal tract (HOA).



**Figure 8 (Left)** Close up of virtual hit 15-48-61 at the active-site of FP2. **(Right)** Connolly surface of the active-site of FP2 of *Pf* with bound predicted most active ADPN inhibitor 15-48-61. The binding site surface is colored according to residue hydrophobicity: red—hydrophobic, blue—hydrophilic, and white—intermediate.

The structure-based design of *p*/FP2 inhibitors has been attempted with various scaffolds. The main strategy remains the increase of hydrophobic enzyme – ligand contact.<sup>48</sup> Due to the lack of quantitative accurate structural information about the *p*/FP2 active site subpockets efforts were oriented towards SAR studies of peptides and non peptidic scaffolds<sup>49,50</sup> but so far no integrated approach combining 3D-QSAR, PH4 and *in silico* screening has been reported and useful structural information is provided here.

## Conclusion

Assessment to FP2 inhibition key structural information either from crystal structure of FP2-ADPNx complex or *in situ* modification of an existing FP2-Generic Inhibitor as in this work is a reliable way to design potent, orally bioavailable and favorable pharmacokinetic ADPN antimalarials. From the short training set of Reik Löser et al.<sup>20</sup> despite the low diversity of the substitutions at P1, P2 and P3, a one descriptor QSAR model correlating the computed Gibbs free energy upon complex formation with the experimental enzyme inhibitory potency shed light on the main interactions impacting activity. Deriving a 3D QSAR Pharmacophore (PH4) model of FP2 inhibition from active conformations of ADPNs at the enzyme active site taking into account the inhibitor loss of entropy upon binding increased the robustness of the PH4 in order to screen virtual libraries (VL) of ADPNs analogues. The design strategy based on the hydrophobic contact at S2 pocket in the best 3D QSAR PH4 model generated since the P3 Aromatic moiety was conserved for all ADPNs. The PH4 selected VL subset of ADPN analogues of 68 was crossed with the GFE computation of predicted inhibitory potencies (Table 3, Equation B). The best designed analogues 15-48-61 (0.5nM), 22-48-61 (0.5nM), 24-48-61 (0.5nM), 19-48-61 (0.74nM), 22-59-61 (0.67nM), 14-59-61 (0.77nM), 20-48-61 (0.83nM) and 13-59-61 (0.93nM) in the low nanomolar range (Table 8) are recommended for synthesis and activity evaluation to check the consistency of the structural information provided in order to guide in the future discovery of non-peptide or peptidomimetic FP2 inhibitors.

## Acknowledgements

This work was also supported by the Slovak Research and Development Agency grants APVV-14-0294 and APVV-15-0111”.

## Conflict of Interest

The authors declare there is no conflict of interest in relation with the work presented herein.

## References

1. World Malaria Report 2016: Summary. Geneva: World Health Organization; 2017.
2. Vythilingam I, Noor Azian Y, T Huat, et al. *Plasmodium Knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. *Parasit vector*. 2006;1(1):26
3. Hyde JE. Drug-resistant malaria—an insight. *FEBSJ*. 2007;274(18):4688–4698.
4. Wongsrichanalai C, Meshnick SR. Declining artesunate-mefloquine efficacy against falciparum malaria on the Cambodia–Thailand border. *Emerg Infect Dis*. 2008;14(5):716–719.
5. Dondorp AM, Nosten F, Yi P, et al. Artemisinin resistance in *Plasmodium falciparum* malaria. *Eng J Med*. 2009;361(15):455–467.
6. Garcia Linares GE, Rodriguez JB. Current Status and Progresses Made in Malaria Chemotherapy. *Curr Med Chem*. 2007;14(3): 289–314.
7. Juliane Wunderlich, Petra Rohrbach, John Pius Dalton. The malaria digestive vacuole. *Front Biosci*. 2012;4:1424–1448.
8. Francis SE, Sullivan DJ, Goldberg DE. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu Rev Microbiol*. 1997;51:97–123.
9. Rosenthal PJ. *Protease Inhibitors*. In: Antiparasitic Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery. New York: Springer; 2001:325–345.
10. Rosenthal PJ, McKerrow JH, Aikawa M, et al. A malarial cysteine proteinase is necessary for hemoglobin degradation by *Plasmodium falciparum*. *J Clin Invest*. 1988;82(5):1560–1566.
11. Gamboa de Dominguez N D, Rosenthal PJ. Cysteine proteinase inhibitors block early steps in hemoglobin degradation by cultured malaria parasites. *Blood*. 1996;87:4448–4454.
12. Iain D Kerr, Ji H Lee, Kailash C Pandey, et al. Structures of Falcipain-2 and Falcipain-3 Bound to Small Molecule Inhibitors: Implications for Substrate Specificity. *J Med Chem*. 2009;52(3):852–857.
13. Boris D Bekono, Fidele Ntie-Kang, Luc COO, et al. Targeting cysteine proteases from *Plasmodium falciparum*: A general overview, rational drug design and computational approaches for drug discovery. *Curr Drug Targets*. 201819(5):501–526.
14. Rosenthal PJ. Cysteine proteases of malaria parasites. *Int J Parasitol*. 2004;34(13–14):1489–1499.
15. Ettari R, Bova F, Zappala M, et al. Falcipain-2 inhibitors. *Medicinal Research Reviews*. 2010;30(1):136–167.
16. Moon JB, Coleman RS, Hanzlik RP. *J Am Chem Soc*. 1986;108:1350.
17. Ward YD, Thomson DS, Frye LL, et al. Design and synthesis of dipeptide nitriles as reversible and potent Cathepsin S inhibitors. *J Med Chem*. 2003;45(25):5471–5482.
18. Löser R, Schilling K, Dimmig E, et al. Interaction of papain-like cysteine proteases with dipeptide-derived nitriles. *J Med Chem*. 2005;48(24):7688–7707.
19. Shenai BR, Lee BJ, Alvarez-Hernandez A, et al. Structure-activity relationships for inhibition of cysteine protease activity and development

- of *Plasmodium falciparum* by peptidyl vinyl sulfones. *Antimicrob Agents Chemother*. 2003;47(1):154–160.
20. Reik Löser, Jiri Gut, Philip J. et al. Antimalarial activity of azadipeptide nitriles. *Bioorganic & Medicinal Chemistry Letters*. 2010;20(1):252–255.
21. Freceer V, Kabelac M, De Nardi P, et al. Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus. *J Mol Graph Model*. 2004;22(3):209–220.
22. Freceer V, Jedinak A, Tossi A, et al. Structure Based Design of Inhibitors of Aspartic Protease of HIV-1. *Letters in Drug Design & Discovery*. 2005;2(8):638–646.
23. Freceer V, Berti F, Benedetti F, et al. Design of peptidomimetic inhibitors of aspartic protease of HIV-1 containing –Phe Psi Pro– core and displaying favourable ADME-related properties. *J Mol Graph Model*. 2008;27(3):376–387.
24. Dali B, Keita M, Megnassan E, et al. Insight into selectivity of peptidomimetic inhibitors with modified statine core for plasmepsin II of *Plasmodium falciparum* over human cathepsin D. *Chem Biol Drug Des*. 2012;79(4):411–430.
25. Megnassan E, Keita M, Bieri C, et al. Design of novel dihydroxynaphthoic acid inhibitors of *Plasmodium falciparum* lactate dehydrogenase. *Med Chem*. 2012;8(5):970–984.
26. Owono Owono LC, Keita M, Megnassan E, et al. Design of Thymidine Analogues Targeting Thymidilate Kinase of *Mycobacterium tuberculosis*. *Tuberc Res Treat*. 2013;670836.
27. Keita M, Kumar A, Dali B, et al. Quantitative structure–activity relationships and design of thymine-like inhibitors of thymidine monophosphate kinase of *Mycobacterium tuberculosis* with favourable pharmacokinetic profiles. *RSC Advances*. 2014;4(99):55853–55866.
28. Freceer V, Seneci P, Miertus S. Computer-assisted combinatorial design of bicyclic thymidine analogs as inhibitors of *Mycobacterium tuberculosis* thymidine monophosphate kinase. *J Comput Aided Mol Des*. 2011;25(1):31–49.
29. Owono Owono LC, Ntie-Kang F, Keita M, et al. Virtually Designed Triclosan-Based Inhibitors of Enoyl–Acyl Carrier Protein Reductase of *Mycobacterium tuberculosis* and of *Plasmodium falciparum*. *Mol Inform*. 2015;34(5):292–307.
30. Kouassi AF, Kone M, Keita M, et al. Computer-Aided Design of Orally Bioavailable Pyrrolidine Carboxamide Inhibitors of Enoyl–Acyl Carrier Protein Reductase of *Mycobacterium tuberculosis* with Favorable Pharmacokinetic Profiles. *Int J Mol Sci*. 2015;16(12):29744–29771.
31. Esmel A, Keita M, Megnassan E, et al. Insight into binding mode of nitrile inhibitors of *Plasmodium falciparum* Falcipain-3, QSAR and Pharmacophore models, virtual design of new analogues with favorable pharmacokinetic profiles. *J Comput Chem Molec Model*. 2017;2–1:1–21.
32. *Discovery Studio Molecular Modeling and Simulation Software*. USA: San Diego; 2009.
33. Gilson MK, Honig B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. *J Comput Aided Mol Des*. 1991;5(1):5–20.
34. Rocchia W, Sridharan S, Nicholls A, et al. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J Comput Chem*. 2002;23(1):128–137.
35. Böttcher CJF, Bordewijk P. *Theory of electric polarization*. 2nd edition. The Netherlands: Elsevier; 1973.
36. Miertuš S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics*. 1981;55(1):117–129.
37. Li H, Sutter J, Hoffmann R. *Pharmacophore Perception, Development and Use in Drug Design*. In: Güner OF, editor. USA: CA; 2000:171–189.
38. *Schrodinger Software*. New York: Schrodinger, USA; 2014.
39. Duffy EM, Jorgensen WL. Prediction of properties from simulations: Free energies of solvation in hexadecane, octanol, and water. *J Am Chem Soc*. 2000;122:2878–2888.
40. Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett*. 2000;10(11):1155–1158.
41. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. *Adv Drug Deliv Rev*. 2002;54(3):355–366.
42. [https://www.chemcomp.com/MOE-Molecular\\_Operating\\_Environment.htm](https://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm)
43. Maple JR, Hwang MJ, Stockfish TP, et al. Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J Comput Chem*. 1994;15(2):62–182.
44. Willett P. *Molecular Similarity in Drug Design*. Glasgow, UK: Chapman and Hall; 1994:110–137.
45. Freire E. Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today*. 2008;13(19–20):869–874.
46. N'Guessan H, Megnassan E. *In silico* Design of Phosphonic Arginine and Hydroxamic Acid Inhibitors of *Plasmodium falciparum* M17 Leucyl Aminopeptidase with Favorable Pharmacokinetic Profile. *J Drug Design Med Chem*. 2017;3(6):98–125.
47. Cotereon JM, Catterick D, Castro J, et al. Falcipain Inhibitors: Optimization Studies of the 2–Pyrimidine–carbonitrile Lead Series. *J Med Chem*. 2010;53(16):6129–6152.
48. Tiwari HK, Kumar P, Kumar K, et al. *In Vitro* Antimalarial Evaluation of Piperidine- and Piperazine-Based Chalcones: Inhibition of Falcipain-2 and Plasmepsin II Hemoglobins Activities from *Plasmodium falciparum*. *Chemistry Select*. 2017;2(25):7684–7690.
49. Himangini, Dharam Pal Pathak, Vidushi Sharma, et al. Designing novel inhibitors against falcipain-2 of *Plasmodium falciparum*. *Bioorg Med Chem Lett*. 28(9):1566–1569.
50. David J Weldon, Falgun Shah, Amar Chittiboyina, et al. Synthesis, biological evaluation, hydration site thermodynamics, and chemical reactivity analysis of  $\alpha$ -keto substituted peptidomimetics for the inhibition of *Plasmodium falciparum*. *Bioorg Med Chem Lett*. 24:1274–1279.