SDRP Journal of Computational Chemistry & Molecular Modelling (ISSN: 2473-6260)

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

DOI: 10.25177/JCMP.2.1.5

RESEARCH

Author: Akori Esmel

September 2017

Received Date: 16h Aug 2017
Accepted Date: 24th Sep 2017
Published Date: 04th Oct 2017

Akori Esmel, [a] **Melalie Keita,** [a, c] **Eugene Megnassan,** [a, b, c]* **Beguemsi Toï,** [b] **Vladimir Frecer,** [c, d, e] **and Stanislav Miertus** [c, e, f]

a.Laboratoire de Physique Fondamentale et Appliquée, University of Abobo Adjamé (now University Nangui Abrogoua), Abidjan, Cote D'Ivoire
a.Laboratoire de Chimie Organique Structurale et Théorique, University of Cocody (now Félix Houphouët-Boigny), Abidjan , Côte D'Ivoire
a.ICS-UNIDO, Area Science Park, Padriciano 99, Trieste I-34012, Italy
b.Faculty of Pharmacy, Comenius University, Bratislava SK-83232, Slovakia
c.International Centre for Applied Research and Sustainable Technology, Bratislava SK-84104, Slovakia
d.Faculty of Natural Sciences, University of SS. Cyril and Methodius, Trnava SK-91701, Slovakia

CORRESPONDENCE AUTHOR
Eugène Megnassan
University of Abobo Adjamé (now University Nangui Abrogoua), UFR SFA, Laboratoire de Physique Fondamentale et Appliquée, 02 BP 801, Abidjan 02, Cote D'Ivoire
Email: megnase@yahoo.com

CONFLICTS OF INTEREST
There are no conflicts of interest for any of the authors.

ABSTRACT

We have carried out computational optimization of antiparasitic azadipeptide nitrile inhibitors (AZN) of falcipain-3 (FP3) of Plasmodium falciparum (Pf), a cysteine protease of the papain superfamily, using structure-based drug design and computer-assisted combinatorial chemistry. Three-dimensional (3D) models of complexes of inhibitor - FP3 for a training sets of published AZN analogs with experimentally determined inhibitory potencies were prepared by in situ modification of the crystal structure of PfFP3 inhibited by K11017 (Protein Data Bank entry 3BWK). We have used molecular mechanics, conformational searching and implicit solvation model to compute Gibbs free energies of inhibitor - FP3 receptor complex formation and built quantitative structure-activity relationships (QSAR) model by correlating the experimental inhibitory potencies with the computed binding affinities. The model was able to explain 97% of the FP3 inhibition data variation and was further validated with help of 3D-QSAR pharmacophore model generation (PH4). Structural information obtained from the 3D models of the AZN - FP3 complexes and the PH4 guided us in designing virtual combinatorial libraries of novel AZN analogs. Comparative analysis of the active site interactions directed us in the selection of building blocks used in the libraries. The initial virtual library was focused by means of computationally predicted oral bioavailability and subsequently in silico screened with the PH4 pharmacophore models to identify new AZN inhibitor candidates. Their inhibitory activities predicted by the QSAR model fall into the low nanomolar concentration range.

Keywords: Azadipeptide nitrile, FP3 inhibitors, *in silico* screening, Malaria, Molecular Modeling, Pharmacophore, QSAR model, Virtual Combinatorial Library.

INTRODUCTION

According to the World Health Organization (WHO) the year 2000 6th Millennium Development Goals (MDGs) to improve the state of the world by 2015, namely by reducing the number of malaria cases by 50% by 2015 [1] has been achieved. This "job half-done" picture is much more a statistical success than a great achievement. The other 50%, the one in the need of "malaria resistance overcoming" treatment, remains poised on a knife-edge. Indeed as resistance occurring against artemisinin combined therapy (ACT) has been reported [2,3], our world is kept in a crucial need of efficient and resistance overcoming treatment. This picture is worsened in sub-Saharan Africa where the most defeating and resistant parasite, *P. falciparum* (*Pf*) strains are prevalent.

The development of new orally bioavailable, resistance overcoming and low cost antimalarial is subordinated to deep constrain. First, address attractive new target; second, proceed through a rational drug design approach; third, screen large, diverse library of compounds and finally come out with an almost perfect pharmacokinetic profile and multi-target compound.

Proteases have been identified as eligible and interesting targets and gathered global funding to address HIV/AIDS, opening the gate to the same family enzymes for malaria and tuberculosis [4]. Further, comparative studies initiative has been carried out to get insight into S2' pocket fitting at *Pf* and HIV aspartic proteases' active sites for inhibition with compounds such as KNI -10006 containing a norstatine core [5]. Several *Pf* proteases: Aspartyl Proteases (plasmepsins I, II, IV-X and HAP), zinc metallopeptidases (Falcilysin), dipeptidylaminopeptidase 1 (DPAP1), aminopeptidases and Cysteine Proteases (falcipains: FP1, FP2A, FP2B, FP3), have been reported in enzyme inhibition studies [6,7]. Falcipains have drawn great interest due to their central role in the life cycle of *Pf* through hemoglobin degradation [8]. FP3, differently from FP2 looks more attractive since its inhibition is lethal to *Pf* [9,10]. Importantly, due to the difference in nature of *Pf* cysteine proteases and their human orthologues, *Pf*FP2-3 inhibitors will be selective over hFP making in this way *Pf*FP of great interest for antimalarial development [11].

This research study focuses on falcipain 3 (FP3) inhibition since no approved drug against FP-3 is available. Reversible and irreversible inhibitors of falcipains have been reported. Among them, reversible inhibitors, namely azadipeptides nitrile (AZN) [12] are of interest. AZNs structurally helped in FP2/3 pockets filling despite their poor ADME profile, are really potent against both Plasmodium chloroquine-sensitive (3D7), and chloroquine-resistant (Dd2) strains.

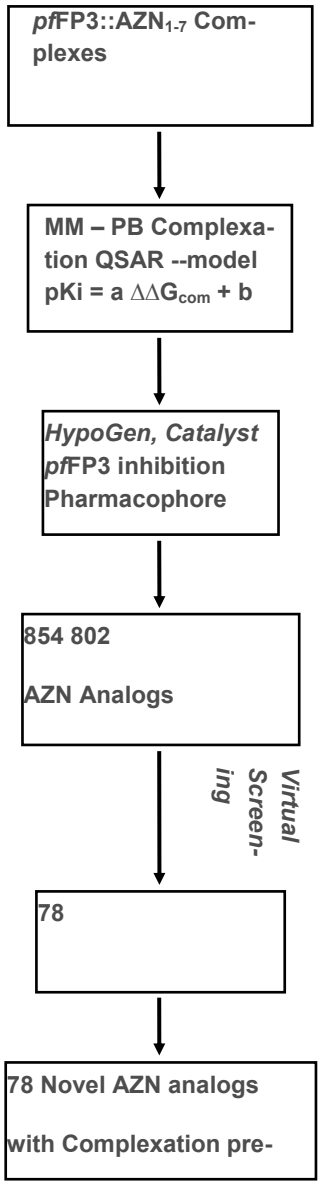
The availability of crystal structure of FP2/3 alone or FP2/3:Inhibitor complexes [13,14,15,16] opened the gate for knowledge and structure based approaches to widen the candidate population taking benefit from the large number of combinatorial libraries built in anti HIV design projects.

Very high FP3 inhibition increase is obtained from preferred interaction between the cyano group (electrophile) with the active site cysteine residue, precisely the thiol and the nitrogen atom in the P1 position of dipeptide nitriles [17,18,19]. For this reason we retain the peptidic AZN with the goal to identify best candidates through Quantitative Structure-Activity Relationship (QSAR) process and derive the FP3 inhibition pharmacophore which will further orientate the design of more potent non peptidic FP3 inhibitors.

In this work, we have built Hansch-type 'complexation' QSAR FP3 inhibition models in order to explain the biological activity variation versus the relative Gibbs free energy upon FP3: AZN complex formation's respectively according to Molecular Mechanics – Poisson-Boltzmann (MM-PB) *in silico* approach. The robustness of the built one descriptor QSAR models was confirmed by a five features 3D-QSAR pharmacophore models (PH4) [20], which were prepared with help of bound conformations of the training sets of AZNs. In the next step virtual combinatorial libraries of AZN analogues respectively have been built with the aim to virtually design more potent orally bioavailable *Pf*FP3 inhibitors. Starting from initial diversity libraries, computed ADME-related properties helped to identify subsets of predicted orally bioavailable AZN analogues to undergo screening by the PH4 pharmacophore models to yield best fit analogues. Based on the *in silico* complexation QSAR models we were able to predict the activities and select the best analogues with the highest predicted inhibitory potencies. Finally, computed ADME profiles of the best designed analogues were compared with those of the antiamalarials currently in use.

MATERIALS AND COMPUTATIONAL METHODS

The workflow describing the steps of the whole process of virtual design of novel AZN analogues is presented in scheme 1.



Scheme 1: workflow describing the multistep approach to virtually design novel AZN analogues with higher predicted potency against FP-3.

2.1 Training sets

Chemical structures and biological activities (IC_{50}^{exp}) of training and validation sets of azadipeptide nitrile (AZN) inhibitors of *pf*FP3 studied here were taken from literature [12]. The potencies of these compounds cover a broad range of half-maximal inhibitory concentrations ($110 \leq IC_{50}^{exp} \leq 50000$ nM), in order to allow construction of QSAR models. The training sets of AZN contained 7 inhibitors taken from the reference [12].

2.2 Model building by in situ modification

Three dimensional (3D) molecular models of enzyme-inhibitor complexes FP3-AZNx, free enzyme FP3 and free inhibitors AZNx were constructed from high-resolution (2.42 Å) X-rays crystal structure of the inhibitor K11017 (Mu-Leu-Hph-VSPh where VSPh: phenyl vinyl sulfone; Hph: homophenylalanyl; Mu: morpholino urea:) bound to the plasmodial FP3 (Protein Data Bank (PDB) [21] entry code 3BWK [20]) using Insight-II molecular modeling program [22].

The structures of FP-3 and enzyme-inhibitor (E:I) complexes were considered in the computations to be at the pH of 7 with neutral N- and C-terminal residues. All protonizable and ionizable residues were charged. All crystallographic water molecules were removed from the model. The inhibitors were built from the reference structure 3BWK [20] by in situ modification of derivatized groups in the molecular scaffold of the co-crystallized inhibitor K11017. An exhaustive conformational search was carried out over all rotatable bonds of the replacing function groups coupled with a careful gradual energy-minimization of the modified inhibitor and active site residues of the FP3 located in the close vicinity of the inhibitor (≤ 5 Å). This process helped to identify low-energy bound conformations of the modified inhibitor leading to various low-energy structures of the E:I complexes which were then carefully refined by minimization of the whole complex. This procedure has been successfully used to build viral, bacterial and protozoal enzyme-inhibitor complexes models and design of peptidomimetic, hydroxynaphthoic, thymidine, triclosan and pyrrolidine-based enzyme inhibitors [23,24,25,26,27,28,29,30,31].

2.3 Molecular mechanics

Modeling of inhibitors, FP3 and E:I complexes was carried out in all-atom representation using CFF91 force field [32] atomic and charge parameters. In the enzyme a dielectric constant of 4 was retained for all molecular mechanics (MM) calculations in order to take into account the dielectric shielding effect in proteins. Energy Minimizations of the E:I complexes, free E and I were carried out by relaxing the structures gradually, starting with the added hydrogen atoms, further with inhibitor heavy atoms, followed by residue side chains and concluded with protein backbone relaxation and alpha carbons. In all the geometry optimization process, a sufficient number of steepest descent followed by conjugate gradient iterative cycles were

used while the convergence criterion for the average gradient was set to 0.01 kcal×mol^{-1}×Å^{-1}.

2.4 Conformational search

Free inhibitor's conformations were obtained from their bound conformations in the E:I complexes by gradual relaxation to the nearest local energy minimum. Then a Monte Carlo (≤ 50000 iterations) low-energy conformations search over all rotatable bonds except those in the rings was performed using Discovery Studio 2.5 (DS 2.5) molecular modeling program [33]. Two hundred inhibitor unique conformations were generated after randomly varying torsion angles of the last accepted conformer by ±15 ° at 5000 K followed by subsequent energy minimization. During the minimization a dielectric constant e = 80 was used to approximately take account of for the dielectric screening effect of hydration. The conformer with the lowest total energy was selected and re-minimized at e = 4.

2.5 Solvation Gibbs free energies

The electrostatic component of solvation Gibbs free energy that includes also the effects of ionic strength via solving nonlinear Poisson-Boltzmann equation [29,30] was computed by the DelPhi module in Discovery Studio [33]. The program treats the solvent as a continuous medium of high dielectric constant (e_o = 80) and the solute as a cavity of low dielectric (e_i = 4) with boundaries linked to the solute's molecular surface, which encloses the solute's atomic charges. The program uses a finite difference method to numerically solve for the molecular electrostatic potential and reaction field around the solute. DelPhi calculations were carried out on a (235 × 235 × 235) cubic lattice grid for the E:I complexes and free E and (65 × 65 × 65) grid for the free I with full coulombic boundary conditions. Two subsequent focusing steps (starting at 50% and reaching 70%) led in both cases 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^{-3}, atomic partial charges and radii defined in the CFF force field parameter set [33] and a probe sphere radius of 1.4 Å were used. The electrostatic component of the Poisson Boltzmann solvation Gibbs free energy was calculated as the reaction field energy [34,35,36,37,38].

2.6 Calculation of binding affinity and QSAR model

Inhibition constant (K_i) of a reversible inhibitor I is related to the standard Gibbs free energy (GFE) change upon formation of the enzyme-inhibitor (E:I) complexes (ΔG_{comp}) in a solvent. Thus prediction of K_i value from the complexation GFE as $\ln K_i = -\Delta G_{comp}/RT$, is achievable assuming the following equilibrium:

$$\{E\}_{aq} + \{I\}_{aq} \leftrightarrow \{E:I\}_{aq} \qquad (1)$$

where {}aq indicates solvated species. Half-maximal inhibitory concentration IC_{50} is for tight binding competitive inhibitors proportional to K_i:

$$IC_{50} = K_i \times (S/K_m + 1) + E/2 \qquad (2)$$

where S is the substrate concentration, K_m represents the Michaelis constant and E means the free enzyme concentration [39]. The standard GFE change of the reaction (1) can be derived by molecular simulations of the complex and the free reactants:

$$\Delta G_{comp} = G\{E:I\} - G\{E\} - G\{I\} \qquad (3)$$

In this work we approximate the exact values of standard GFE for larger systems such as enzyme-inhibitor complexes by the expression [23, 24, 25]:

$$G\{E:I\} \approx E_{MM}\{E:I\} + RT - TS_{trv}\{E:I\} + G_{solv}\{E:I\} \qquad (4)$$

where $E_{MM}\{E:I\}$ stands for MM total energy of the complex (including bonding and non-bonding contributions), $G_{solv}\{E:I\}$ is the solvation GFE and $TS_{trv}\{E:I\}$ is the entropic term:

$$TS_{trv}\{E:I\} = TS_{tran}\{E:I\} + TS_{rot}\{E:I\} + TS_{vib}\{E:I\} \qquad (5)$$

composed of the sum of contributions arising from translational, rotational and vibrational motions of E:I. Assuming that the *tran* and *rot* terms for the complex E:I and free enzyme E are approximately equal, we obtain:

$$\Delta G_{comp} \approx [E_{MM}\{E:I\} - E_{MM}\{E\} - E_{MM}\{I\}] + [G_{solv}\{E:I\} - G_{solv}\{E\} - G_{solv}\{I\}] + TS_{tran}\{I\} + TS_{rot}\{I\} - [TS_{vib}\{E:I\} - TS_{vib}\{E\} - TS_{vib}\{I\}] = \Delta H_{MM} + TS_{tran}\{I\} + TS_{rot}\{I\} - \Delta TS_{vib} + \Delta G_{solv} \qquad (6)$$

where $TS_{tran}\{I\}$ and $TS_{rot}\{I\}$ describe the translational and rotational entropy terms of the free inhibitor and $D TS_{vib}$ represents a simplified vibrational entropy change upon the complex formation: $D TS_{vib} = TS_{vib}\{I\}$ E - $TS_{vib}\{I\}$ [40,41]. In the same way $DD H_{MM}$ is the relative enthalpic contribution to the GFE change related to the intermolecular interactions in the enzyme-inhibitor complex derived by MM.

Relative changes in the complexation GFE of different inhibitors with respect to a reference inhibitor, I_{ref}, were computed assuming ideal gas behavior for the rotational and translational motions of the inhibitors:

$$\Delta\Delta G_{comp} = \Delta G_{comp}\ (I) - \Delta G_{comp}\ (I_{ref}) = \Delta\Delta H_{MM} - \Delta\Delta TS_{vib} + \Delta\Delta G_{solv} \qquad (7)$$

This evaluation of relative changes is preferable as a way to partial cancellation of errors originated from the approximate nature of the MM method, solvent and entropic effects description as well.

Quantitative structure-activity relationships (QSAR) model, in which a linear relationship between the computed relative GFE of the FP3-AZN complex formation $\Delta\Delta G_{comp}$ for the receptor structure and observed inhibitory potencies IC_{50}^{exp} specific to *pf*, is assumed according to eqs. (1) and (2):

$$pIC_{50}^{exp} = -\log_{10}IC_{50}^{exp} = a \times \Delta\Delta G_{comp} + b \qquad (8)$$

was prepared by linear regression for both training set using $\Delta\Delta G_{comp}$ quantities calculated via eq. (7), *a* and *b* are regression coefficients. These QSAR models (a target-specific scoring function) was then employed for prediction of inhibitory potencies (IC_{50}^{pred}) of newly designed and modeled AZN analogues (section 2.1).

2.7 Interaction energy

The MM interaction energy (E_{int}) protocol available in DS 2.5 [33] computes the non-bonded interactions (van der Waals and electrostatic terms) between enzyme residues and the inhibitor. The calculations were performed using CFF force field [32] with a dielectric constant of 4. The breakdown of E_{int} into active site residue contributions reveals the significance of individual interactions and allows a comparative analysis, which leads to identification of affinity enhancing and unfavorable AZN substitutions.

2.8 Pharmacophore generation

Pharmacophore modeling assumes that a set of structural features in a molecule is recognized at the receptor site and is responsible for biological activity of the compound. Bound conformations of inhibitors taken from E:I complexes were considered for constructing 3D-QSAR pharmacophore based on Catalyst HypoGen algorithm implemented in DS 2.5 [33]. The top scoring pharmacophore hypothesis was built up in three stages (constructive, subtractive and optimization step) from a set of the most active inhibitors while inactive ones served for definition of the excluded volume. During the pharmacophore generation, maximum number of five features allowed by the HypoGen algorithm was selected according to the AZN scaffold and substituents, namely: hydrophobic aromatic (HYdAr), hydrophobic aliphatic (HYd), hydrogen-bond donor, (HBD), hydrogen-bond acceptor (HBA) and ring aromatic (Ar) feature. Default values of adjustable parameters of the protocol were kept except the uncertainty on the activity and the minimum inter-feature distance, which were set to 1.25 and 0.5 Å (for small ligand). This parameter choice was intended to bring the uncertainty interval of experimental activity from a wide span áIC_{50}/3, 3×IC_{50}ñ to a relatively narrow one á4×IC_{50}/5, 5×IC_{50}/4ñ, taking thus account of the accuracy and homogeneity of the measured inhibitory activities since they are coming from the same laboratory [12]. During generation of 10 pharmacophores the number of missing features was set to 0. The best pharmacophore models were selected.

2.9 ADME 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 eighteen more descriptors related to adsorption, distribution, metabolism and excretion (ADME properties) of the inhibitors were computed by the QikProp program [42] based on the methods of Jorgensen [43,44,45]. 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) - the number of property descriptors that fall outside the range of values determined for 95 % of known drugs out of 24 selected descriptors computed by the QikProp [42], was used as an additional ADME-related compound selection criterion.

2.10 Virtual Combinatorial library generation

The analogue model building was performed with Molecular Operating Environment (MOE) program []. The library of analogues was enumerated by attaching R-groups (fragments, building blocks) onto AZN scaffold using the Quasar CombiDesign module of MOE [46]. Chemical reagents considered in this study were taken from the directories of chemicals available from the commercial suppliers of chemicals []. Each analogue was built as a neutral molecule in the MOE program [46] and its molecular geometry was refined by MM optimization using smart minimizer of Discovery Studio [33] with high convergence criteria (energy difference of 10^{-4} kcal×mol^{-1}, R.M.S. displacement of 10^{-5} Å) and a dielectric constant of 4 using class II consistent force field CFF [32] as described in the Molecular mechanics section 2.3.

2.11 ADME-based library focusing

Twenty four pharmacokinetic molecular descriptors available in QikProp [42], which characterize a wide spectrum of molecular properties as described in section 2.9. such as molecular mass, total solvent-accessible molecular surface, hydrophobic portion of the solvent-accessible molecular surface, total volume of molecule enclosed by solvent-accessible molecular surface, number of non-trivial non-hindered rotatable bonds, estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aque-

ous solution, estimated number of hydrogen bonds that would be accepted by the solute from water molecules, logarithm of partitioning coefficient between n-octanol and water phases, logarithm of predicted aqueous solubility, logarithm of predicted binding constant to human serum albumin, logarithm of predicted brain/blood partition coefficient, apparent Caco-2 cell membrane permeability in Boehringer-Ingelheim scale, number of likely metabolic reactions, percentage of human oral absorption in gastrointestinal tract, etc. Optimum ranges of the 24 descriptors were defined in terms of upper and lower bounds, and average values according to QikProp [42]. Drug likeness was used as ADME-related compound selection criterion. Only compounds with predicted drug likeness #stars equal to zero were selected for the focused library of drug like analogues.

2.12 Pharmacophore-based library focusing

The pharmacophore models (PH4) described in section 2.8. was derived from the bound conformations of AZNs at the active site of FP3. The enumerated and ADME-focused virtual library was further focused by using the ligand pharmacophore mapping protocol available of Discovery Studio [33]. Within this protocol, each generated conformer of the analogues was geometry optimized by means of the CFF forcefield 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 conformers were saved and clustered into ten conformational families according to their mutual r.m.s. deviations by means of Jarvis-Patrick complete linkage clustering method []. The best representative of each cluster was considered in the virtual screening of analogues.

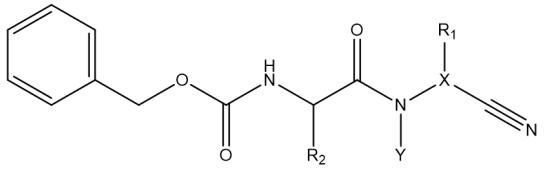
2.13 In silico screening

The conformer with the best mapping on the PH4 pharmacophore in each cluster of the focused library subset was selected for virtual screening by the complexation QSAR model. The GFE of E:I complex formation in a water DDG_{comp} was computed for each selected new analogue and then used for prediction of FP3 inhibitory potencies (IC_{50}^{pred}) of the focused virtual library of AZN and AZN analogues by inserting this parameter into the target-specific scoring function, eq. (9). The scoring function, which is specific for the FP3 receptor of *Pf*: $pIC_{50}^{pred}[FP3] = a \times DDG_{comp} + b$, was parameterized using the QSAR model described above, section 2.6

RESULTS AND DISCUSSION

In this study, a training set was selected from two homogeneous series of *pf*FP-3 inhibitors. The training set is composed of 7 Azadipeptide Nitrile (AZN), Table 1. Their chemical structures and experimental inhibitory concentrations IC_{50}^{exp} reported respectively by Reik Löser et al. [12], cover a concentration range sufficiently wide to serve well for building two reliable QSAR models of *pf*FP-3 inhibition.

Table 1. Training set (AZNx) of FP-3 inhibitors [12] used in the preparation of quantitative structure-activity relationships (QSAR) model of inhibitor binding.



Training Set	R_1	R_2	X	Y	IC_{50}^{exp} (nM)
AZN 1	$-CH_3$	$-CH_2$–Phe	N	$-CH_3$	18000
AZN 2	$-CH_3$	–*i*–But	N	$-CH_3$	1900
AZN 3	–Phe	$-CH_2$–Phe	N	$-CH_3$	1600
AZN 4	$-C_2H_4$–Phe	$-CH_2$–Phe	N	$-CH_3$	300
AZN 5	$-(CH_2)_4-CH_3$	$-CH_2$–Phe	N	$-CH_3$	110
AZN 6	$-CH_3$	$-CH_2$–Phe	C	–H	16000
AZN 7	$-CH_3$	$-CH_2$–Phe	C	$-CH_3$	50000

3.1 Quantitative Structure-Activity Relationships (QSAR) Model

The relative Gibbs free energy of the non-covalent enzyme-inhibitor (E:I) complex formation was computed for the FP3-AZNx complexes prepared by *in situ* modification of the inhibitor K11017 within the binding site of FP-3 of the refined crystal structure (PDB entry code 3BWK [16]), as described in the section 2.6. Table 3 and 4 lists computed values of complex formation Gibbs free energies ($\Delta\Delta G_{comp}$) and its components, Equation (7) for the training sets AZNs.

Since the $\Delta\Delta G_{comp}$ was computed in an approximate way, the relevance of the binding model is evaluated through a correlation with the experimental activity data (IC_{50}^{exp}, [12]) by linear regression, eq. (8). For this training set, a correlation equation obtained for the Gibbs free energy (GFE) of enzyme-inhibitor complex formation $\Delta\Delta G_{comp}$ is shown in Table 3 together with the relevant statistical data (see Figure 1 for the correlations plots). Relatively high values of the regression coefficient and the statistical significance Fischer F-test of the correlations involving $\Delta\Delta G_{comp}$, eq. (A), indicate that there is a strong relationship between the binding mode and the observed inhibitory potencies of the training set.

Table 2. Complexation Gibbs free energy (binding affinity) and its components for the training set of FP3 inhibitors AZN**1**-**7**

Training Set [a]	M_w [b]	$\Delta\Delta H_{MM}$ [c]	$\Delta\Delta G_{solv}$ [d]	$\Delta\Delta TS_{vib}$ [e]	$\Delta\Delta G_{comp}$ [f]	IC_{50}^{exp} [g]
	(g·mol^{-1})	(kcal·mol^{-1})	(kcal·mol^{-1})	(kcal·mol^{-1})	(kcal·mol^{-1})	(nM)
AZN 1	366	0.00	0.00	0.00	0.00	18000
AZN 2	332	-1.67	1.56	3.54	-3.64	1900
AZN 3	443	-7.20	2.13	-1.10	-3.96	1600
AZN 4	457	-8.67	3.14	2.45	-7.98	300
AZN 5	423	-10.24	2.78	2.05	-9.51	110
AZN 6	351	-3.38	1.35	-0.55	-1.49	16000
AZN 7	365	0.24	0.57	-0.03	0.84	50000

[a] for the chemical structures of the training set of inhibitors see Table 1;
[b] M_w is the molecular mass of inhibitors;
[c] $\Delta\Delta H_{MM}$ is the relative enthalpic contribution to the Gibbs free energy change related to enzyme-inhibitor (E:I) complex formation derived by MM: $\Delta\Delta H_{MM}$ = $[E_{MM}\{E{:}I_x\} - E_{MM}\{I_x\}] - [E_{MM}\{E{:}I_{ref}\} - E_{MM}\{I_{ref}\}]$, I_{ref} is the reference inhibitor AZN **1**;
[d] $\Delta\Delta G_{solv}$ is the relative solvation Gibbs free energy contribution to the Gibbs free energy change of E:I complex formation: $\Delta\Delta G_{solv} = [G_{solv}\{E{:}I_x\} - G_{solv}\{I_x\}] - [G_{solv}\{E{:}I_{ref}\} - G_{solv}\{I_{ref}\}]$;
[e] $-\Delta\Delta TS_{vib}$ is the relative entropic contribution of inhibitor I_x to the Gibbs free energy related to E:I complex formation: $\Delta\Delta TS_{vib} = [\Delta\Delta TS_{vib}\{I_x\}_E - \Delta\Delta TS_{vib}\{I_x\}] - [\Delta\Delta TS_{vib}\{I_{ref}\}_E - \Delta\Delta TS_{vib}\{I_{ref}\}]$;
[f] $\Delta\Delta Gcomp$ is the relative Gibbs free energy change related to E:I_x complex formation: $\Delta\Delta Gcomp = \Delta\Delta H_{MM} + \Delta\Delta G_{solv} - \Delta\Delta TS_{vib}$;
[g] IC_{50}^{exp} is the experimental half-maximal inhibition concentration of FP3 inhibition obtained from reference [12].

The robustness of this one descriptor QSAR model is assessed through the components of GFE namely the enthalpic H_{MM}, solvation G_{solv} and the loss of vibrational entropy upon the AZN binding TS_{vib}. The relevance of the enthalpic contribution to GFE is well confirmed by the quality of the regression coefficient 0.81, the cross validated R^2_{xv} of 0.68 and the F-test of 21.54, indicating that in gas phase a large part (some 80%) of the variation of the IC_{50} is explained by that of H_{MM}. Adding to H_{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 vibrational entropy TS_{vib} to explain some 97% of the variation of 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. [49]. For this reason the lack of validation set (VS), due to the small number of provided AZNs 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 is based on the bound conformation of the AZNs (see figure 5 and table 6).

Therefore, the correlation equation (B) and computed $\Delta\Delta G_{comp}$ quantities can be used for prediction of inhibitory potencies IC_{50}^{pred} against *pf*FP-3 for novel AZN analogues (AZNA) respectively, provided that they share the same binding mode as their corresponding training set.

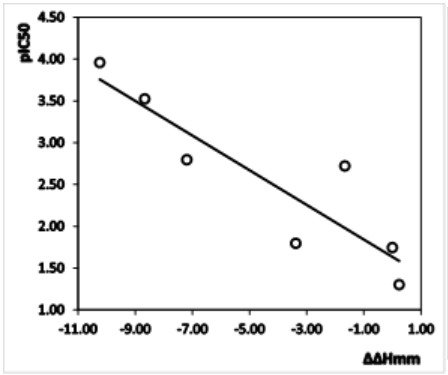
Table 3. Regression analysis of computed binding affinities $\Delta\Delta G_{comp}$, its component $\Delta\Delta H_{MM}$, and experimental half-maximal inhibitory concentrations $pIC_{50}^{exp} = -\log_{10}(IC_{50}^{exp}/10^6)$ [12] of AZNs towards *pf*FP3.

Statistical Data of Linear Regression	(A)	(B)
$pIC_{50}^{exp} = -0.20726 \times \Delta\Delta H_{MM} + 1.6330$ (A)		
$pIC_{50}^{exp} = -0.24871 \times \Delta\Delta G_{comp} + 1.6342$ (B)		
Number of compounds n	7	7
Squared correlation coefficient of regression R^2	0.81	0.97
LOO cross-validated squared correlation coefficient R^2_{xv}	0.68	0.96
Standard error of regression s	0.47	0.17
Statistical significance of regression, Fisher F-test	21.54	194.02
Level of statistical significance a	> 95 %	
Range of activities IC_{50}^{exp} [nM]	110-50000	

3.2 Binding mode of inhibitors

Beside the robustness of the QSAR model, the analysis of the interactions between AZNs and FP3 is expected to reveal key interactions justifying AZN::FP3 affinity such as hydrogen bonds (HBs), van der Waals (vdW), hydrophobic contacts, etc. As displayed in Figure 2, the binding mode of AZNs at FP3 active site of the best active AZN **5**, and AZN **4** in 2D and 3D is supported by the following interactions: π-π stacking with Tyr 93, HB with Trp 52, Gly 92 and hydrophobic contacts. All interactions involving the catalytic and other key

residues particularly Gly 92 are conserved. In order to verify whether other interesting interactions not displayed have to be taken into account in the description of AZN binding mode at FP3 active site for the design of new analogues, interaction energy (IE) between each active site residue and AZNx is computed. The peptidic structure of AZN shed light on the structural features for binding affinity and opened the gate to the design of non peptidic FP3 inhibitors taking benefit from S1 to S3 pockets filling. As displayed in Figure 4 the breakdown of IE diagram into each FP3 S1-3 pocket residue contribution for AZNs indicated a particular behavior of P1 comparatively to P2-3. The focus on S1 pocket comes from the fact that substitutions were mainly performed at P1 position of the AZNs during the SAR performed by Reik Löser et al. [12] (see table 1). This is also confirmed by the good correlation between P1 interaction energy with the FP3 and AZNs IC_{50} noticed in Figure 3. The major role of S1 pocket in the inhibition of FP3 is also confirmed by Falgun Shah et al. [50]. Unfortunately the lack in the training set of detailed SAR information involving other FP3 pockets justifies the low correlation between S2 and S3 contribution to IE with pIC_{50}. In such case the unique IE diagram analysis cannot guide precisely in the choice of the R-groups in S1, S2 and S3 pockets as we were able to do for designing thymine-like inhibitors of thymidine monophosphate kinase [47]. Usually in such case, a large and diverse combinatorial virtual library (VL) of AZNs has to be built with the purpose to screen it with our 3D pharmacophore (PH4) of FP3 inhibition, based on the complexation one descriptor QSAR model. This approach successfully has been used to design pyrrolidine carboxamide inhibitors of *Mycobaterium tuberculosis* InhA [31].



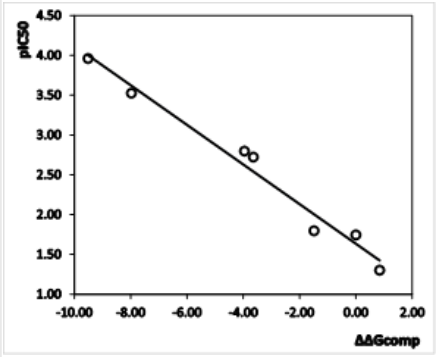
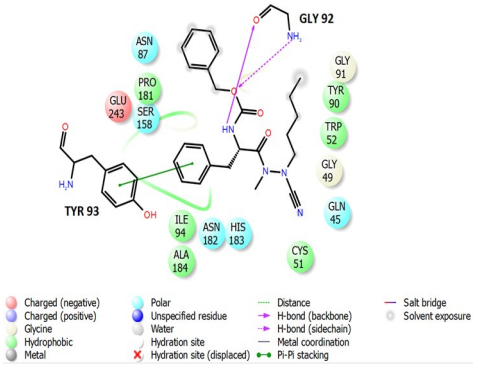
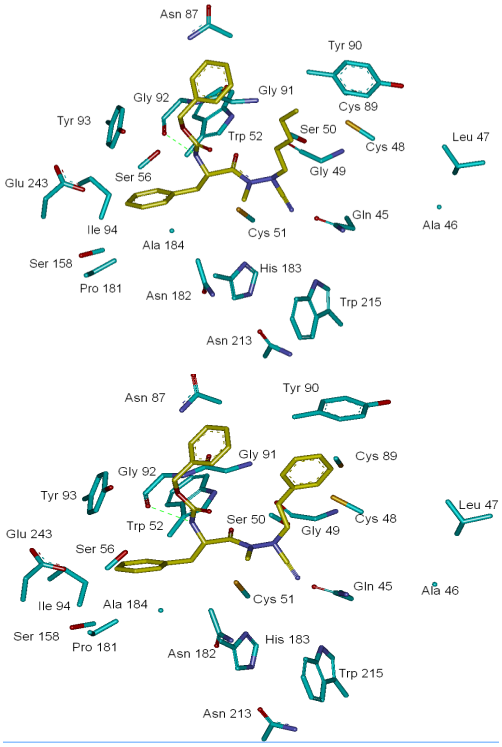


Figure 1. (Left) Plot of correlation equation between pIC_{50}^{exp} and the relative enthalpic contribution to the Gibbs free energy of *Pf*FP3-AZNx complex formation $\Delta\Delta H_{MM}$ (Eq. 6); (Right) Similar plot for relative complexation Gibbs free energies of the *Pf*FP3-AZNx complex formation $\Delta\Delta G_{comp}$ of the training set, all in kcal.mol^{-1}.





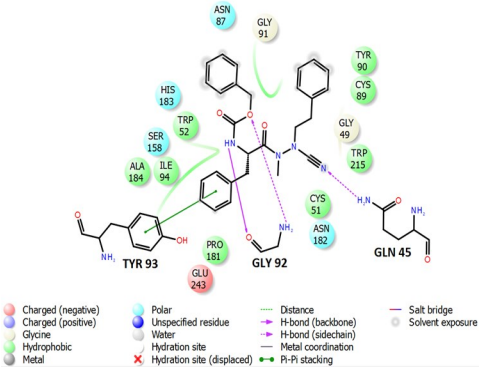


Figure 2. (Top) Left: 2D schematic interaction diagram of the most potent inhibitor AZN **5** at the active site of *Pf*FP3. Right. 3D structure of the active site with bound inhibitor AZN **5**. (Bottom) Left: 2D schematic interaction diagram of the most potent inhibitor AZN **4** at the active site of *Pf*FP3. Right. 3D structure of the active site with bound inhibitor AZN **4**.

Table 4. Interaction energy (IE) contribution between each part of inhibitor versus FP3.

Compd	Scaffold[a]	P1[b]	P2[c]	P3[d]	IE[e]	pIC_{50}
AZN 1	-5.06	-9.46	-16.38	-11.59	-42.49	1.74
AZN 2	-4.02	-9.78	-11.03	-11.57	-36.40	2.72
AZN 3	-3.27	-13.67	-15.12	-12.16	-44.22	2.80
AZN 4	-3.85	-16.46	-15.24	-11.20	-46.75	3.52
AZN 5	-3.73	-13.93	-15.11	-12.08	-44.85	3.96
AZN 6	-10.85	-1.83	-15.98	-11.00	-39.66	1.80
AZN 7	-12.57	-1.98	-16.30	-12.12	-42.97	1.30

[a] Scaffold contribution to total IE (kcal.mol^{-1});
[b] P1 (including nitrile fragment) contribution to total IE (kcal.mol^{-1});
[c] P2 contribution to total IE (kcal.mol^{-1});
[d] P3 contribution to total IE (kcal.mol^{-1});
[g] Total IE (kcal.mol^{-1}).

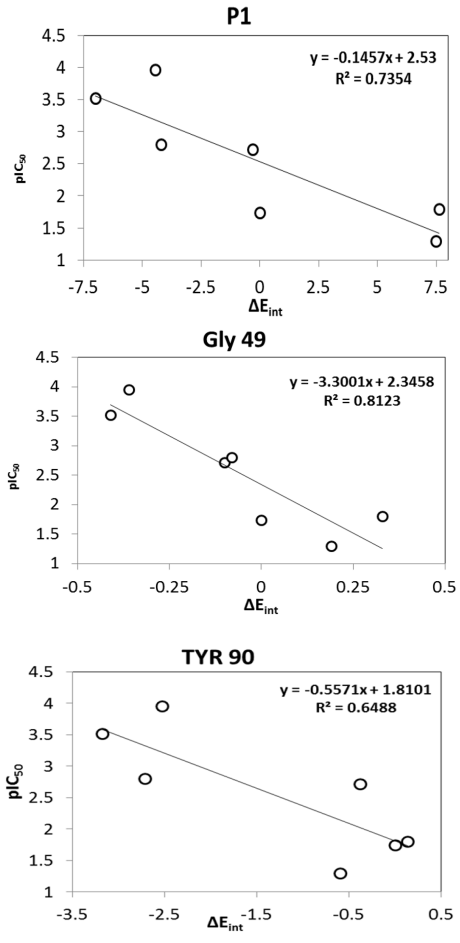


Figure 3. Plot of correlation between pIC_{50}^{exp} and P1, Gly 49 and Tyr 90 contribution to the intermolecular interaction energy of *Pf*FP3-AZNx

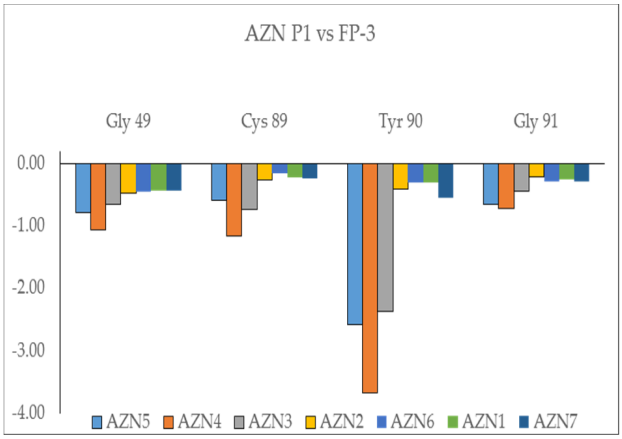


Figure 4. Breakdown of FP3-AZN interaction energy into each active site residue contribution at S1 pocket.

3.3 Ligand-Based 3D-QSAR Pharmacophore Model of Inhibitory Activity

The 3D-QSAR pharmacophore generation process followed three main steps: (i) the constructive step, (ii) the subtractive step, and (iii) the optimization step. The constructive phase of HypoGen automatically selected the most active compounds for which $IC_{50}^{exp} \leq 1.25 \times 110$ nM as leads. Thus, the most active AZN **5** (IC_{50}^{exp} = 110 nM) alone was used to generate the starting PH4 features. Only those features were retained which matched this lead. Next, in the subtractive phase inactive compounds with are used to remove those pharmacophoric features that were present in more than 50 % of these compounds, while pharmacophores which contained all features were retained. None of the training set compounds was found to be inactive ($IC_{50}^{exp} > 110 \times 10^{3.5}$ nM = 347 850.5 nM). During the final optimization phase, the score of the pharmacophoric hypotheses was improved. Hypotheses are scored according to errors in activity estimates from regression and complexity via a simulated annealing approach. At the end of the optimization the top scoring 10 unique pharmacophoric hypotheses (see Table 5) were kept, all displaying four features.

The generated pharmacophore models were then assessed for their reliability based on the calculated cost parameters. The overall costs ranged from 36.7 (Hypo **1**) to 136.9 (Hypo **10**). The relatively small gap between the highest and lowest cost parameter corresponds well with the homogeneity of the generated hypotheses and the consistency of the training set. For this PH4 model the fixed cost (19.8) is lower than the null cost (321.0) by a difference $\Delta = 301.2$. This difference is a major quality indicator of the PH4 predictability ($\Delta > 70$ corresponds to an excellent chance or a probability higher than 90% that the model represents a true correlation) [33]. To be statistically significant the hypotheses have to be as close as possible to the

fixed cost and as far as possible from the null cost. For the set of 10 hypotheses, the difference $\Delta \geq 184.2$ attests the high quality of the pharmacophore model. The standard indicators such as the root-mean-square deviations (RMSD) between the hypotheses ranged from 2.083 to 5.717 and the squared correlation coefficient (R^2) falls to an interval from 0.98 to 0.79. The first PH4 hypothesis with the best RMSD and R^2 was retained for further analysis. The statistical data for the set of hypotheses (costs, RMSD, R^2) are listed in Table 6.

The geometry of the Hypo **1** pharmacophore of FP3 inhibition is displayed on Figure 5. The regression equation for pIC_{50}^{exp} vs. pIC_{50}^{pred} estimated from Hypo **1**: $pIC_{50}^{exp} = 0.9926 \times pIC_{50}^{pred} + 0.0187$ (see Table 6 for its statistical data) is also plotted in Figure 5. The regression equation coefficient is close to one and the intercept close to zero indicating that the ratio of predicted and observed activities ($pIC_{50}^{pred}/pIC_{50}^{exp}$) for the training is relatively close to one (table 7) except for AZN1. Moreover, the statistical data in table 6 such as R^2 and R^2_{xv} greater than 0.9 and a significance F-test of 96.73 document substantial predictive power of this regression for the best PH4 model.

Table 5. Output parameters of 10 generated PH4 pharmacophoric hypotheses for FP-3 inhibitors after CatScramble validation procedure.

Hypothesis	RMSD [a]	R^2 [b]	Total Costs [c]
Hypo 1	2.083	0.98	36.68
Hypo 2	4.187	0.90	82.04
Hypo 3	4.171	0.90	85.54
Hypo 4	4.817	0.86	102.82
Hypo 5	5.186	0.83	114.84
Hypo 6	5.131	0.84	116.00
Hypo 7	5.491	0.81	128.12
Hypo 8	5.602	0.80	131.37
Hypo 9	5.663	0.80	132.35
Hypo 10	5.717	0.79	136.86
Fixed Cost	0.0	1.0	19.82
Null Cost	9.391	0.0	321.04

[a] root mean square deviation; [b] squared correlation coefficient; [c] overall cost parameter of the PH4 pharmacophore.

Table 6. Regression analysis of experimental half-maximal inhibitory concentrations pIC50exp [12] and computed half-maximal inhibitory concentrations pIC50pred of AZNs towards pfFP-3.

Statistical Data of Linear Regression for HYPO 1	
$pIC_{50}^{exp} = 0.9926 \cdot pIC_{50}^{pred} + 0.0187$	
Number of compounds n	7
Squared correlation coefficient of regression R^2	0.95
LOO cross-validated squared correlation coefficient R^2_{xv}	0.91
Standard error of regression s	0.24
Statistical significance of regression, Fisher F-test	96.73
Level of statistical significance a	>95
Range of activities IC_{50}^{exp} [nM]	110-50000

Table 7. Computed the ratio of predicted and observed activities ($pIC_{50}^{pred}/pIC_{50}^{exp}$) for the training set

Compounds	pIC_{50}^{exp}	pIC_{50}^{pred}	$pIC_{50}^{exp}/ pIC_{50}^{pred}$
AZN 1	1.74	2.10	0.83
AZN 2	2.72	2.38	1.14
AZN 3	2.80	2.69	1.04
AZN 4	3.52	3.44	1.02
AZN 5	3.96	4.11	0.96
AZN 6	1.80	1.82	0.98
AZN 7	1.30	1.30	1.00

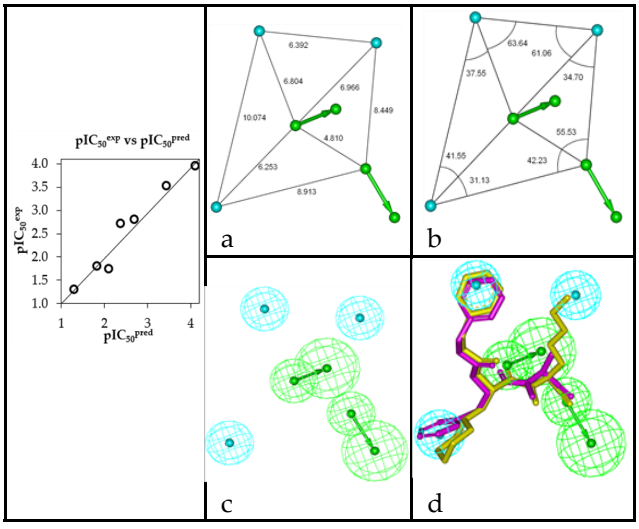


Figure 5. (a,b) Coordinates of the Hypo **1** pharmacophore of *Pf*FP3 inhibiton, (c) features of the pharmacophore of *Pf*FP3 inhibition and (d) pharmacophore mapping with AZN **1** (IC_{50}^{exp}=18000 nM) (purple) and the best fit hit **21-70-166** (IC_{50}^{pred}=1.1 nM) (yellow). The correlation plot of experimental vs. predicted inhibitory activity is displayed at the left. The features are colored blue for hydrophobic aliphatic (HYd), green for hydrogen-bond acceptor (HBA) and purple for hydrogen-bond donor (HBD). The arrows represent the projection for donor and acceptor features.

The randomization validation of the PH4 model was carried out by the CatScramble algorithm of the Catalyst for 49 random runs corresponding to a 98 % confidence level. This procedure created 10 valid hypotheses for each run. However, none of them was as predictive as the Hypo **10**, the hypothesis with the highest cost of the ten best hypotheses generated and shown in Table 5. Thus there is a 98% probability that the best selected hypothesis Hypo **1** represents a pharmacophore model for inhibitory activity of AZNs with a similar level of predictive power as the complexation QSAR model, which relies on the 3D structures of the FP3-AZNx complexes and computed Gibbs free energies of enzyme-inhibitor binding DDG_{comp}. The main information provided by the PH4 is the relative coordinates of the three hydrophobic features (Figure 5 a, b) for an accurate filling of the active site pockets. Since it has been built from the active conformation of AZNs, this PH4 is suitable for an efficient screening of AZN analogues virtual library.

Therefore we have performed computational design and selection of new AZN analogues with increased inhibition potencies against FP3 of *Pf*. The design strategy relied on the mapping to the hydrophobic features included in the best PH4 pharmacophore model at the position of P1, P2 and P3 substitutions (PH4 hypothesis Hypo **1** in Figure 5).

3.4. Library design and ADME focusing

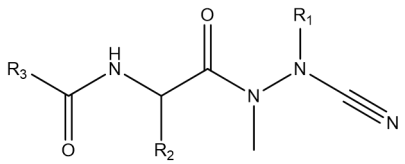
We have built a virtual libraries of new azadipeptide nitriles compounds with a variety of substitutions in free positions (P1, P2 and P3) with the goal to identify more potent orally bioavailable inhibitors of the *pf*FP3. During the virtual library enumeration the R-groups listed in Table 8 were attached in positions R^1 to R^3 of the AZN scaffold to form a combinatorial library of the size: $R^1 \times R^2 \times R^3 = 167 \times 167 \times 52 = 1{,}450{,}228$ AZNA.

To design a more focused library of a reduced size and 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 AZNs which can undergo *in silico* screening. The initial virtual library was then filtered in an ADME-based focusing step to remove compounds with expected poor oral bioavailability and low drug-like character. Only analogues with high predicted percentage of human oral absorption in the gastrointestinal tract HOA larger than 80% [43, 44] and compounds satisfying the Lipinski's rule of five [51] computed for the entire virtual library using QikProp software [42], were kept. This focusing has reduced the size of the initial library to 819,642 AZNA less than 60% its original number size.

3.5. In silico screening of AZNs virtual library

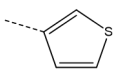
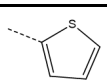
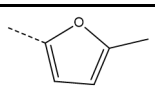
The library of AZN analogues was further screened for molecular structures matching to the 3D-QSAR PH4 pharmacophore model Hypo **1** of FP3 inhibition. From the reduced libraries few thousands of AZNA mapped to at least 2 features, 150 AZNA of which mapped to 5 features of the pharmacophore. Out of then, only 78 AZNA best fitting analogues (PH4 hits) have been retained and submitted to virtual evaluation with the help of the complexation QSAR model: their Gibbs free energy (GFE) upon complex formation with *pf*FP3 was computed along with its component and their predicted half-maximal inhibitory concentration IC_{50}^{pred} was estimated with the correlation equation (B) (Table 3). All the results are given in Table 9.

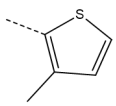
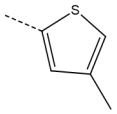
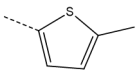
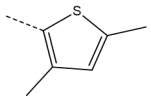
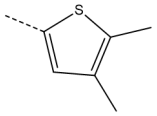
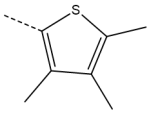
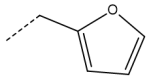
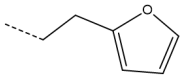
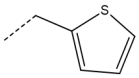
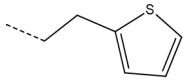
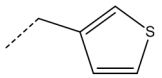
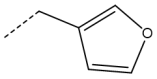
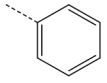
Table 8. R-groups (fragments, building blocks, substituents) used in the design of the initial diversity library of azadipeptide nitriles analogues (AZNA).



R^1, R^2 and R^3-Groups [a,b]

1	–F	2.	–Cl	3.	–Br
4.	$–CH_3$	5.	$–SCH_3$	6.	$–CH_2F$
7.	$–CH_2Cl$	8.	$–CH_2Br$	9.	$–C_2H_5$
10.	$–(CH_2)_2F$	11.	$–(CH_2)_2Cl$	12.	$–(CH_2)_2Br$
13.	$–(CH_2)_2CH_3$	14.	$–(CH_2)_3F$	15.	$–(CH_2)_3Cl$
16.	$–(CH_2)_3Br$	17.	*n*-butyl	18.	$–(CH_2)_4F$

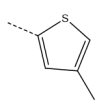
19.	$-(CH_2)_4Cl$	20.	$-(CH_2)_4Br$	21.	*n*-pentyl
22.	$-(CH_2)_5F$	23.	$-(CH_2)_5Cl$	24.	$-(CH_2)_5Br$
25.	$-(CH_2)_5CH_3$	26.	$-(CH_2)_6F$	27.	$-(CH_2)_6Cl$
28.	$-(CH_2)_6Br$	29.	*i*-prop	30.	*i*-butyl
31.	$-(CH_2)_2CH(CH_3)_2$	32.	$-(CH_2)_3CH(CH_3)_2$	33.	$-(CH_2)_4CH(CH_3)_2$
34.	$-(CH_2)_5CH(CH_3)_2$	35.	$-CH(CH_3)_3$	36.	$-CH_2CH(CH_3)_3$
37.	$-(CH_2)_2CH(CH_3)_3$	38.	$-(CH_2)_3CH(CH_3)_3$	39.	$-(CH_2)_4CH(CH_3)_3$
40.	$-(CH_2)_5CH(CH_3)_3$	41.	sec-butyl	42.	$-CH_2CH(CH_3)C_2H_5$
43.	$-(CH_2)_2CH(CH_3)C_2H_5$	44.	$-(CH_2)_3CH(CH_3)C_2H_5$	45.	$-(CH_2)_4CH(CH_3)C_2H_5$
46.	$-(CH_2)_5CH(CH_3)C_2H_5$	47.	$-CH(C_2H_5)C_2H_5$	48.	$-CH_2CH(C_2H_5)C_2H_5$
49.	$-(CH_2)_2CH(C_2H_5)C_2H_5$	50.	$-(CH_2)_3CH(C_2H_5)C_2H_5$	51.	cyclopropyl
52.	$-CH_2$–cyclopropyl	53.	$-(CH_2)_2$–cyclopropyl	54.	$-(CH_2)_3$–cyclopropyl
55.	$-(CH_2)_4$–cyclopropyl	56.	$-(CH_2)_5$–cyclopropyl	57.	cyclobutyl
58.	$-CH_2$–cyclobutyl	59.	$-(CH_2)_2$–cyclobutyl	60.	$-(CH_2)_3$–cyclobutyl
61.	$-(CH_2)_4$–cyclobutyl	62.	$-(CH_2)_5$–cyclobutyl	63.	cyclopentyl
64.	$-CH_2$–cyclopentyl	65.	$-(CH_2)_2$–cyclopentyl	66.	$-(CH_2)_3$–cyclopentyl
67.	$-(CH_2)_4$–cyclopentyl	68.	$-(CH_2)_5$–cyclopentyl	69.	cyclohexyl
70.	$-CH_2$–cyclohexyl	71.	$-(CH_2)_2$–cyclohexyl	72.	$-(CH_2)_3$–cyclohexyl
73.	$-(CH_2)_4$–cyclohexyl	74.	$-(CH_2)_5$–cyclohexyl	75.	4-methylcyclohexyl
76.	cyclopropenyl	77.	2-methyl-2-cyclopropenyl	78.	2,3-dimethyl-2-cyclopropenyl
79.	1-methyl-2-cyclopropenyl	80.	1,2-dimethyl-2-cyclopropenyl	81.	1,2,3-trimethyl-2-cyclopropenyl
82.	(2-cyclopropenyl)methyl	83.	(2-cyclopropenyl)ethyl	84.	(2-cyclopropenyl)n-propyl
85.	(2-cyclopropenyl)n-butyl	86.	2-methyl-1,3-cyclopropadienyl	87.	1,3-cyclopropadienyl
88.	3-methyl-1,3-cyclopropadienyl	89.	2,3-dimethyl-1,3-cyclopropadienyl	90.	2,4-dimethyl-1,3-cyclopropadienyl
91.	2,3,4-trimethyl-1,3-cyclopropadienyl	92.	(1,3-cyclopropadienyl)methyl	93.	(1,3-cyclopropadienyl)ethyl
94.	(1,3-cyclopropadienyl)n-propyl	95.	(1,3-cyclopropadienyl)n-butyl	96.	2-methylcyclopentyl
97.	3-methylcyclopentyl	98.	2,2-dimethylcyclopentyl	99.	3,3-dimethylcyclopentyl
100.	2-oxacyclopentyl	101.	(2-oxacyclopentyl)methyl	102.	(2-oxacyclopentyl)ethyl
103.	(2-oxacyclopentyl)n-propyl	104.	(2-oxacyclopentyl)n-butyl	105.	(2-oxacyclopentyl)n-pentyl
106.	3-oxacyclopentyl	107.	(3-oxacyclopentyl)methyl	108.	(3-oxacyclopentyl)ethyl
109.	(3-oxacyclopentyl)n-propyl	110.	(3-oxacyclopentyl)n-butyl	111.	2-oxacyclohexyl
112.	3-oxacyclohexyl	113.	4-oxacyclohexyl	114.	2-cyclopentenyl
115.	3-furanyl	116.		117.	2-furanyl
118.		119.		120.	4-methyl-2-oxapenta-3,5-dienyl

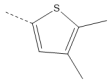
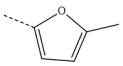
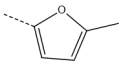
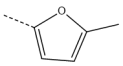
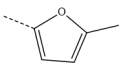
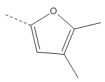
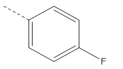
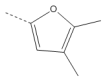
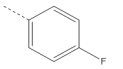
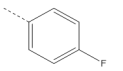
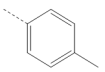
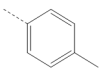
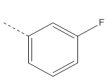
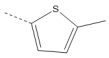
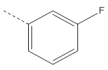
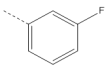
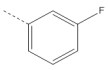
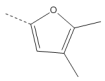
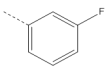
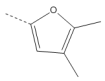
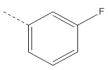
121.	5-methyl-2-oxapenta-3,5-dienyl	122.	3,5-dimethyl-2-oxapenta-3,5-dienyl	123.	3,4-dimethyl-2-oxapenta-3,5-dienyl
124.	3,4,5-trimethyl-2-oxapenta-3,5-dienyl	125.		126.	
127.		128.		129.	
130.		131.		132.	
133.		134.		135.	
136.		137.		138.	p-fluorophenyl
139.	p-chlorophenyl	140.	p-bromophenyl	141.	p-methylphenyl
142.	m-fluorophenyl	143.	m-chlorophenyl	144.	m-bromophenyl
145.	m-methylphenyl	146.	o-fluorophenyl	147.	o-chlorophenyl
148.	o-bromophenyl	149.	o-methylphenyl	150.	p-hydroxyphenyl
151.	p-methyloxyphenyl	152.	p-hydroxybenzyl	153.	p-methyloxybenzyl
154.	benzyl	155.	p-fluorobenzyl	156.	p-chlorobenzyl
157.	p-bromobenzyl	158.	p-methylbenzyl	159.	3,5-dimethylbenzyl

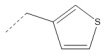
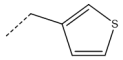
[a] fragments **1**–**167** were used in R^1 to R^2-groups; fragments **115-119**, **137-158** and **164**–**193** were used in R^3-group; [b] dashed bonds indicates the attachment points of individual fragments.

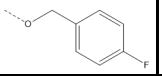
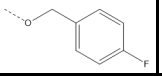
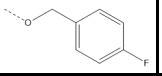
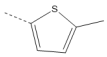
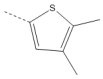
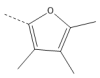
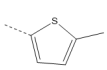
The selected libraries subset then underwent virtual screening by means of their corresponding PH4 pharmacophore models of AZN inhibitory activity towards *pf*FP3.

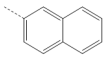
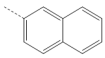
Table 9. Complexation Gibbs free energies and their components for the top 78 scoring virtually designed analogue AZNA. The analogue numbering concatenates the index of each substituent R1 to R3 with the substituent numbers taken from Table 8.

AZNA	Substituents Substituents Substituents			M_w [a]	$\Delta\Delta H_{MM}$ [b]	$\Delta\Delta G_{solv}$ [c]	$\Delta\Delta TS_{vib}$ [d]	$\Delta\Delta G_{comp}$ [e]	IC_{50}^{pred} [f]
	R^1	R^2	R^3						
AZN 1 (4-154-166)	$-CH_3$	$-CH_2$Phe	$-OCH_2$Phe	366.0	0	0	0	0	18000.00 [g]
22-45-115	$-(CH_2)_5F$	$-(CH_2)_4CH(CH_3)C_2H_5$		422.5	-7.11	2.31	7.04	-11.85	26.30
26-149-117	$-(CH_2)_6F$	o-methylphenyl		479,3	-5.89	0.74	-1.91	-3.25	3618.15
126-49-117		$-(CH_2)_2CH(C_2H_5)C_2H_5$		416.5	-7.62	1.35	0.64	-6.91	443.84

129-33-117		–$(CH_2)_4CH(CH_3)_2$		430.6	-8.99	0.76	1.92	-10.16	69.16
26-30-119	–$(CH_2)_6F$	*i*-butyl		394.5	-3.39	2.39	5.72	-6.72	496.13
26-33-119	–$(CH_2)_6F$	–$(CH_2)_4CH(CH_3)_2$		436.6	-4.36	2.99	8.03	-9.40	106.50
126-33-119		–$(CH_2)_4CH(CH_3)_2$		430.6	-8.82	0.67	1.82	-9.97	77.05
126-49-119		–$(CH_2)_2CH(C_2H_5)C_2H_5$		430.6	-8.01	3.42	2.11	-6.70	499.60
123-30-138		*i*-butyl		400.5	-6.08	2.16	1.68	-5.61	934.742
123-33-138		–$(CH_2)_4CH(CH_3)_2$		422.8	-11.12	1.63	3.19	-12.68	16.30
126-49-138		–$(CH_2)_2CH(C_2H_5)C_2H_5$		444.6	-8.44	1.96	1.67	-8.15	218.00
26-30-141	–$(CH_2)_6F$	*i*-butyl		404.5	-1.85	-1.79	6.40	-10.04	73.83
26-33-141	–$(CH_2)_6F$	–$(CH_2)_4CH(CH_3)_2$		446.6	-8.19	1.46	9.02	-15.75	2.81
25-30-142	–(CH2)5CH3	*i*-butyl		390.5	-3.36	2.86	5.36	-5.86	810.44
25-127-142	–$(CH_2)_5CH_3$			430.5	-8.26	2.39	-2.82	-3.05	4058.20
25-145-142	–$(CH_2)_5CH_3$	m-methylphenyl		459.0	-8.35	2.00	0.33	-6.68	506.36
26-145-142	–$(CH_2)_6F$	m-methylphenyl		462.9	-8.84	2.06	-1.94	-4.84	1455.03
123-30-142		*i*-butyl		400.5	-6.47	2.11	1.67	-6.03	735.27
123-33-142		–$(CH_2)_4CH(CH_3)_2$		442.5	-11.42	1.75	3.27	-12.94	14.08
22-120-147	–$(CH_2)_5F$		o-chlorophenyl	434.9	-7.61	1.81	-2.39	-3.41	3297.05
21-9-166	*n*-pentyl	–C_2H_5	–OCH_2Phe	360.5	-0.53	1.22	3.92	-3.23	3644.94
21-36-166	*n*-pentyl	–$CH_2CH(CH_3)_3$	–OCH_2Phe	402.	-5.35	-	7.88	-	5.91

21-69-166	*n*-pentyl	cyclohexyl	–OCH_2Phe	414.	-6.48	0.87	7.80	-	10.79
21-70-166	*n*-pentyl	–CH_2–cyclohexyl	–OCH_2Phe	428.6	-10.97	1.37	7.83	-17.43	1.07
21-75-166	*n*-pentyl	4-methylcyclohexyl	–OCH_2Phe	428.6	-7.83	1.90	10.37	-16.30	2.05
21-158-166	*n*-pentyl	p-methylbenzyl	–OCH_2Phe	436.6	-10.86	-1.37	4.37	-16.59	1.73
21-160-166	*n*-pentyl	phenylethyl	–OCH_2Phe	436.	-7.23	2.77	3.35	-7.81	265.59
22-20-166	–$(CH_2)_5F$	–$(CH_2)_4Br$	–OCH_2Phe	501.	-9.47	2.73	3.34	-	72.27
25-20-166	–$(CH_2)_5CH_3$	–$(CH_2)_4Br$	–OCH_2Phe	481.	-8.28	2.32	5.51	-	32.67
32-30-166	–(CH2)3CH(CH3)2	*i*-butyl	–OCH_2Phe	402.	-6.65	8.48	7.90	-6.07	718.12
32-154-166	–$(CH_2)_3CH(CH_3)_2$	benzyl	–OCH_2Phe	436.	-9.75	5.45	5.30	-9.60	95.31
32-158-166	–$(CH_2)_3CH(CH_3)_2$	p-methylbenzyl	–OCH_2Phe	450.6	-12.85	1.19	5.39	-17.05	1.34
38-30-166	–(CH2)3CH(CH3)3	*i*-butyl	–OCH_2Phe	416.	-3.90	3.42	8.83	-9.32	111.72
38-154-166	–$(CH_2)_3CH(CH_3)_3$	benzyl	–OCH_2Phe	450.	-9.14	0.35	6.50	-	3.66
66-30-166	–(CH_2)3-cyclopentyl	*i*-butyl	–OCH_2Phe	428.	-0.47	4.48	10.0	-6.07	716.80
66-135-166	–$(CH_2)_3$cyclopentyl		–OCH_2Phe	468.6	-2.22	0.24	7.32	-9.30	112.97
71-114-166	–$(CH_2)_2$–cyclohexyl	2-cyclopentenyl	–OCH_2Phe	438.	0.38	0.14	6.64	-6.12	696.75
71-135-166	–$(CH_2)_2$–cyclohexyl		–OCH_2Phe	468.6	-0.05	-0.21	6.20	-6.46	575.35
72-30-166	–$(CH_2)_3$–cyclohexyl	*i*-butyl	–OCH_2Phe	442.6	-3.28	7.80	11.16	-6.65	516.46
72-36-166	–$(CH_2)_3$–cyclohexyl	–$CH_2CH(CH_3)_3$	–OCH_2Phe	456.6	-4.46	3.72	12.72	-13.46	10.43
72-154-166	–$(CH_2)_3$–cyclohexyl	benzyl	–OCH_2Phe	476.	-9.23	2.26	8.34	-	3.62
160-36-166	phenylethyl	–$CH_2CH(CH_3)_3$	–OCH_2Phe	436.6	-4.93	-0.36	7.69	-12.97	13.77
160-69-166	phenylethyl	cyclohexyl	–OCH_2Phe	448.	-5.50	1.41	4.99	-9.07	128.53
160-70-166	phenylethyl	–CH_2–cyclohexyl	–OCH_2Phe	462.6	-10.58	1.39	7.56	-16.74	1.59
160-75-166	phenylethyl	4-methylcyclohexyl	–OCH_2Phe	462.	-7.60	1.85	4.41	-	69.38
160-154-166	phenylethyl	benzyl	–OCH_2Phe	470.	-8.95	2.16	2.33	-9.12	124.93
160-158-166	phenylethyl	p-methylbenzyl	–OCH_2Phe	470.6	-11.70	-0.43	3.36	-15.49	3.26
160-160-166	Phenylethyl	phenylethyl	–OCH_2Phe	470.	-6.98	3.77	1.05	-4.26	2020.69
164-70-166	phenyloxymethyl	–CH_2–cyclohexyl	–OCH_2Phe	464.6	-12.28	0.33	4.25	-16.19	2.18
164-154-166	phenyloxymethyl	benzyl	–OCH_2Phe	458.5	-11.92	2.14	-0.15	-9.63	93.33
164-158-166	phenyloxymethyl	p-methylbenzyl	–OCH_2Phe	472.5	-14.30	-1.79	0.45	-16.54	1.79

165-154-166	p-methyl phenyloxymethyl	benzyl	–OCH_2Phe	472.5	-11.14	2.28	-0.06	-8.81	149.70
165-158-166	p-methyl phenyloxymethyl	p-methylbenzyl	–OCH_2Phe	486.6	-14.31	-1.48	-0.85	-14.94	4.46
19-22-167	–$(CH_2)_4$Cl	–(CH2)5F		458.9	-8.61	3.28	2.85	-8.17	215.92
19-23-167	–$(CH_2)_4$Cl	–(CH2)5Cl		475.4	-9.30	2.85	2.27	-8.72	157.05
20-22-167	–$(CH_2)_4$Br	–(CH2)5F		503.4	-8.61	3.08	2.37	-7.90	251.81
19-22-168	–$(CH_2)_4$Cl	–$(CH_2)_5$F	–OCH_2Phe-(p)	475.	-9.01	3.57	2.06	-7.50	316.01
19-23-168	–$(CH_2)_4$Cl	–$(CH_2)_5$Cl	–OCH_2Phe-(p)	491.	-9.49	3.25	1.76	-8.00	237.69
21-23-169	*n*-pentyl	–$(CH_2)_5$Cl	–OCH_2Phe-(p)	515.	-9.32	3.31	3.54	-9.55	97.86
19-22-174	–$(CH_2)_4$Cl	–$(CH_2)_5$F		411.9	-8.44	3.75	0.97	-5.67	905.38
19-23-174	–$(CH_2)_4$Cl	–$(CH_2)_5$Cl		428.4	-8.37	3.06	0.26	-5.57	955.96
19-24-174	–$(CH_2)_4$Cl	–$(CH_2)_5$CBr		472.8	-8.07	2.50	-0.31	-5.26	1138.96
20-23-174	–$(CH_2)_4$Br	–$(CH_2)_5$Cl		472.8	-7.39	2.83	-0.47	-4.09	2232.38
21-127-174	*n*-pentyl			399.5	-5.17	1.34	-1.14	-2.69	4978.67
21-129-174	*n*-pentyl			413.5	-6.39	2.48	-0.01	-3.90	2482.25
25-124-174	–$(CH_2)_5CH_3$			425.5	-6.84	1.80	2.94	-7.98	240.26
25-127-174	–$(CH_2)_5CH_3$			413.5	-6.73	1.32	-1.85	-3.56	3016.43
25-130-174	–$(CH_2)_5CH_3$			441.6	-5.82	1.95	3.95	-7.81	264.99
25-138-174	–$(CH_2)_5CH_3$	p-fluorophenyl		407.5	-6.44	1.17	3.40	-8.67	161.94
25-145-174	–$(CH_2)_5CH_3$	m-methylphenyl		427.9	-7.58	1.16	-0.72	-5.70	888.44
25-149-174	–$(CH_2)_5CH_3$	o-methylphenyl		472.4	-6.55	1.27	-1.15	-4.13	2181.02

67-30-174	–(CH2)4–cyclopentyl	*i*-butyl		413.6	-0.96	2.64	9.22	-7.55	308.25
67-54-174	$–(CH_2)_4–$cyclopentyl	$–(CH_2)_3–$cyclopropyl		439.6	-7.77	1.98	8.78	-14.58	5.51
120-33-174	4-methyl-2-oxapenta-3,5-dienyl	$–(CH_2)_4CH(CH_3)_2$		411.5	-8.54	0.96	5.02	-12.61	17.02
120-49-174	4-methyl-2-oxapenta-3,5-dienyl	$–(CH_2)_2CH(C_2H_5)C_2H_5$		411.5	-6.28	1.74	5.80	-10.33	62.53
120-60-174	4-methyl-2-oxapenta-3,5-dienyl	$–(CH_2)_3–$cyclobutyl		409.5	-8.94	1.04	3.02	-10.92	44.54
18-30-181	–(CH2)4F	*i*-butyl		412.5	-4.85	3.50	4.18	-5.53	976.82
18-138-181	$–(CH_2)_4F$	p-fluorophenyl		446.5	-9.88	2.96	0.36	-7.28	358.97

[a] M_w is molecular mass of the inhibitor;
[b] DDH_{MM} is the relative enthalpic contribution to the Gibbs free energy change related to the FP3-AZN complex formation DDG_{comp} (for details see footnote of Table 2);
[c] DDG_{solv} is the relative solvation Gibbs free energy contribution to DDG_{comp};
[d] DDTS_{vib} is the relative entropic (vibrational) contribution to DDG_{comp};
[e] DDG_{comp} is the relative Gibbs free energy change related to the enzyme-inhibitor FP3-AZN complex formation DD*G*comp @ DDH_{MM} + DD*G*solv - DDTS_{vib}.
[f] IC_{50}^{pred} is the predicted inhibition constant towards *pf*FP3 calculated from DDG_{comp} using correlation equation (B), Table 3;
[g] IC_{50}^{exp} is given for the reference inhibitor AZN1 instead of IC_{50}^{pred}.

For 33 AZNA out of the 78, the estimated inhibitory potency shown in Table 9 is better than that for the most active training set compound AZN **5** (IC_{50}^{exp} = 110 nM [12]). In fact, the best designed AZN analogues **21-70-166** (1.1 nM) display predicted inhibitory potency about 100 times higher than the AZN **5** (21-154-166).

3.5 Analysis of new inhibitors

To identify which substituents on R-positions of AZN scaffold (Table 8) lead to new inhibitor candidates with the highest predicted potencies towards the FP-3 of *Pf*, we have prepared histograms of the absolute frequency of occurrence of R^1 to R^3 groups in the 78 PH4 best fit hits AZNAs selected from the focused virtual library shown in Table 9 (Figures 6).

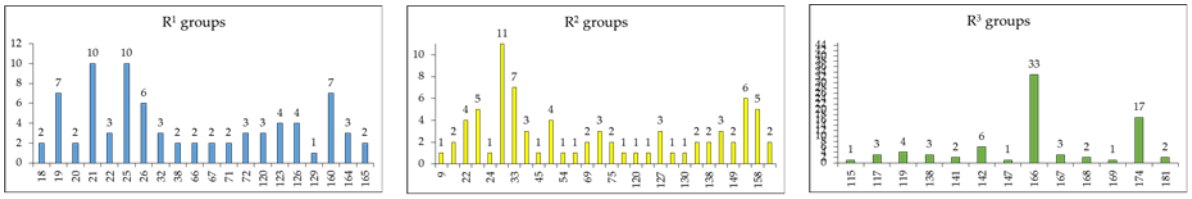


Figure 6. Histograms of absolute frequency of occurrence of individual R^1, R^2, R^3-groups in the 78 best-selected analogues mapping to the five features of the PH4 pharmacophore hypothesis Hypo **1** (for fragment's structure and number see Table 8).

From the histograms in Figure 6 it comes out that R^1 groups numbered **19** (7), **21** (10), **25** (10), **26** (6), and **160** (7) are almost equally represented with the highest occurrence in the AZNs subset. The R^2 groups contain preferentially **22** (4), **23** (5), **30** (11), **33** (7), **36** (3), **49** (4), **70** (3), **127** (3), **145** (3), **154** (6) and **158** (5) while the R^3 groups include chiefly fragments **142** (6), **166** (33), and **174** (17). The top scoring virtual hits are AZN analogues: **21-70-166** (1.1 nM), **32-158-166** (1.3 nM), **160-70-166** (1.6 nM), **21-158-166** (1.7 nM) **164-158-166** (1.8 nM).

Figure 7 displays the 3D interaction depiction of the best designed AZNA **21-70-166** at FP3 active site (right). The Connolly surface of the active site (left) shows the conserved pentyl (P1) and the P3 benzyloxy of the AZN **5**. The cyclohexyl substituted to the benzene ring in P2 resulted in a high increase of potency due to a better filling of the lipophilic S2 pocket and a stabilizing subsequent hydrophobic contact. This structural information is confirmed by the reported experimental results on FP2 and FP3 pyrimidinecarbonitrile inhibitors [] attesting the quality of the FP3 inhibition PH4 and its ability identify novel FP3 inhibitors.

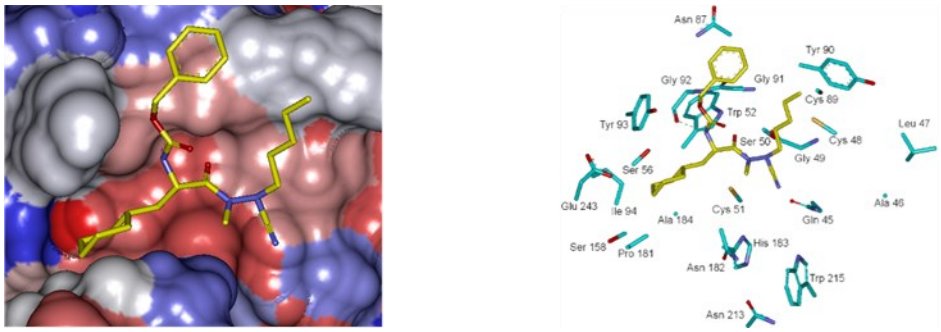


Figure 7. **(**Left): Connolly surface of the active site of *Pf* FP-3 with bound predicted most active AZN inhibitor **21-70-166**. The binding site surface is colored according to residue hydrophobicity: red - hydrophobic, blue - hydrophilic and white - intermediate. (Right): Close up of the best virtual hit **21-70-166** at the active site of FP-3. Carbon atoms of interacting residues's side chains are colored light green and those of ligand are shown in yellow color.

3.6 ADME profiles of designed AZNs

ADME-related properties fully described in section 2.9 were computed. The values for the best active designed AZNs are compared with those computed for drugs used for treatment of malaria alone or in Artemisinin combined therapy (ACT) or currently undergoing clinical trials, Table 10. The best designed analogues all display #stars equal to zero meaning that the optimal "value range of drug likeness" descriptor was violated. One of the main requirements about new antimalarials, as stated by WHO, is their oral bioavailability. The last column of the table displays a high level drug likeness descriptor: the percentage of human oral absorption in gastrointestinal tract (HOA). None of the designed AZN analogues is outside the range of good oral bioavailability (<25% - poor, >80% high). In the same way most of currently in use antimalarials (see the lowest part of Table 9) display a percentage greater than 80%. Interestingly the #stars is equal to 0 for the designed AZN analogues oppositely to those of most of the ACT antimalarials.

Table 10. Predicted ADME-related properties of the best designed AZN analogues and known anti-malaria agents either in clinical use or currently undergoing clinical testing, as computed by QikProp [42].

Ana-logues [a]	#star [b]	Mw [c] [g.mol^{-1}]	Smol [d] [Å^2]	Smol,hfo [e] [Å^2]	Vmol [f] [Å^3]	RotB [g]	HBdon [h]	HBacc [i]	logPo/w [j]	logSwat [k]	logKHSA [l]	logB/B [m]	BIPcaco [n] [nm.s^{-1}]	*#meta* [o]	IC_{50}^{pre}	HOA[q]	%HOA[r]
26-30-141	0	405	761.0	488.0	1415.4	12	0.25	6.8	5.0	7.5	0.6	-1.12	1055.3	2	73.83	3	100.0
21-36-166	0	403	725.3	438.3	1387.8	12	0.3	6.8	4.5	7.8	0.5	-1.4	634.7	2	5.9	3	100.0
21-69-166	0	415	744.5	462.6	1407.2	11	0.3	6.8	4.6	8.0	0.6	-1.3	660.8	2	10.8	3	100.0
21-70-166	0	429	776.7	484.4	1468.6	12	0.3	6.8	5.0	7.9	0.7	-1.5	576.7	2	1.1	3	100.0

21-75-166	0	429	770.8	485.8	1465.5	11	0.3	6.8	5.0	8.1	0.7	-1.4	668.2	2	2.1	3	94.0
21-158-166	0	437	782.3	366.5	1474.2	12	0.3	6.8	5.2	8.8	0.7	-1.5	597.5	4	1.7	3	94.1
32-158-166	0	451	808.7	387.0	1531.2	12	0.3	6.8	5.6	8.9	0.9	-1.4	709.1	4	1.3	1	100.0
38-154-166	0	451	766.1	308.3	1491.2	12	0.3	6.8	5.4	9.0	0.8	-1.4	615.9	3	3.7	3	95.3
72-36-166	0	457	778.3	506.5	1532.8	12	0.3	6.8	5.5	7.8	0.9	-1.4	691.1	2	10.4	3	96.8
160-36-166	0	437	738.6	308.7	1441.9	11	0.3	6.8	5.1	9.0	0.7	-1.2	737.2	3	13.8	3	95.1
160-70-166	0	463	802.1	354.6	1531.0	11	0.3	6.8	5.6	9.3	0.9	-1.4	637.5	3	1.6	1	100.0
160-158-166	0	471	798.7	242.5	1526.3	11	0.3	6.8	5.8	10.0	0.9	-1.3	726.4	5	3.3	1	100.0
164-70-166	0	465	784.7	342.6	1497.0	11	0.3	7.5	5.0	9.9	0.7	-1.4	590.7	3	2.2	3	93.0
164-158-166	0	473	790.0	227.2	1499.7	11	0.3	7.5	5.2	10.8	0.7	-1.4	622.9	5	1.8	3	94.4
165-158-166	0	487	814.1	315.9	1555.6	11	0.3	7.5	5.5	10.4	0.8	-1.4	682.3	6	4.5	1	97.0
120-33-174	0	412	776.4	402.2	1410.6	11	0.3	8.3	4.0	10.2	0.2	-1.5	574.9	4	17.0	3	100.0
pyrimethamine	0	249	467.4	115.5	778.4	4	4	3.0	1.8	-2.8	-0.3	-0.6	556.5	1	-	3	86.7
Dapsone	1	236	431.6	0.0	687.9	2	0	7.0	-0.4	-0.5	-1.3	-0.9	289.1	0	-	2	68.8
Trimethoprim	0	272	500.2	223.9	835.9	5	0	6.5	0.6	-1.5	-0.9	-1.2	282.8	3	-	3	74.3
Quinine	0	324	522.0	301.0	990.1	5	1	5.5	3.3	-2.9	0.1	0.2	628.3	4	-	3	96.3
Chloroquine	1	294	594.1	188.9	982.9	6	0	3.0	4.6	-5.3	0.4	-0.1	3718.1	0	-	3	100.0
Amodiaquine	1	334	603.2	131.7	1018.7	6	0	5.0	3.6	-4.4	0.0	-0.4	1689.0	0	-	3	100.0
Mefloquine	2	362	533.1	0.0	925.1	2	0	4.0	4.1	-4.9	0.1	0.5	2903.1	0	-	3	100.0
Primaquine	0	259	528.1	242.7	909.6	7	3	3.8	2.0	-2.5	-0.1	-0.2	371.3	6	-	3	84.9
Pamaquine	0	316	654.8	443.4	1148.1	9	1	4.8	4.0	-3.8	0.4	0.2	1475.2	5	-	3	100.0
Sulfametopyra-zine	1	268	473.4	77.9	773.3	4	0	9.0	-1.0	0.2	-1.7	-1.3	195.8	1	-	2	61.9
Tetracycline	5	422	604.5	173.1	1111.8	2	0	16.0	-3.4	1.0	-2.5	-2.6	6.8	5	-	1	21.8
Quinacrine	0	370	680.5	268.8	1163.6	7	0	3.5	5.6	-6.5	0.8	-0.1	4435.7	1	-	1	100.0
Atovaquone	0	367	620.6	136.9	1099.8	1	1	4.8	4.1	-6.0	0.6	-0.4	917.5	3	-	3	100.0
Proguanil	1	238	478.2	125.3	768.6	6	0	6.0	1.1	-1.5	-1.0	-0.7	834.6	0	-	3	85.6
Clindamycin	0	425	721.5	534.2	1307.3	10	4	11.8	2.0	-2.3	-0.8	-0.7	139.2	6	-	3	77.1
Halofantrine	*5*	*470*	*785.4*	*160.2*	*1351.8*	*5*	*0*	*3.0*	*7.6*	*-9.9*	*1.5*	*0.2*	*2844.1*	*0*	-	*1*	*100.0*
Sulfadoxine	1	296	510.6	152.3	849.5	5	0	9.5	-0.8	-0.1	-1.7	-1.4	213.4	2	-	2	64.0
Hydroxychloro-quine	1	310	609.5	119.5	1006.5	6	0	5.0	3.4	-4.5	-0.1	-0.7	1023.7	0	-	3	100.0
Bulaquine	0	370	560.2	360.2	1097.8	9	1	5.8	3.6	-3.0	0.1	-0.4	3099.7	7	-	3	100.0
Lumefantrine	5	497	819.1	160.7	1437.5	7	0	3.0	8.3	-10.0	1.7	0.2	4337.2	0	-	1	100.0
Artemether	1	298	490.6	465.5	901.7	1	0	5.7	2.3	-2.4	-0.3	0.3	5729.0	0	-	3	100.0
Artesunate	0	384	644.1	465.1	1155.8	4	1	8.0	2.5	-4.4	-0.1	-1.4	50.4	2	-	3	72.0
Arteether	1	312	531.1	506.0	970.2	2	0	5.7	2.7	-3.0	-0.2	0.2	5731.8	0	-	3	100.0
Dihydroartemis-inine	1	284	477.4	395.7	864.6	1	1	5.7	1.8	-2.9	-0.1	-0.1	1664.9	0	-	3	95.4
Doxycycline	4	422	602.2	174.0	1104.2	2	0	17.3	-4.0	1.7	-2.9	-2.5	9.2	4	-	1	20.8
Artemisinin	0	282	456.6	380.6	848.4	0	0	5.3	1.7	-2.1	-0.3	0.001	1886	1	-	3	95.8

[a] best designed AZN analogues, Table 9;
[b] drug likeness, number of property descriptors (from 24 out of the full list of 49 descriptors of QikProp, ver. 3.7, release 14) that fall outside of the range of values for 95% of known drugs;
[c] molecular weight in g.mol^{-1} (range for 95% of drugs: 130 - 725 g.mol^{-1}) [43];
[d] total solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 300 - 1000 Å^2);
[e] hydrophobic portion of the solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 0 - 750 Å^2);
[f] total volume of molecule enclosed by solvent-accessible molecular surface, in Å^3 (probe radius 1.4 Å) (range for 95% of drugs: 500 – 2000 Å^3);
[g] number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds (range for 95% of drugs: 0 - 15);
[h] estimated number of hydrogen bonds that would be donated by the so lute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer (range for 95% of drugs: 0.0 - 6.0);
[i] estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are average es taken over a number of configurations, so they can be non-integer (range for 95% of drugs: 2.0 - 20.0);
[j] logarithm of partitioning coefficient between n-octanol and water phas es (range for 95% of drugs: -2 - 6.5);
[k] logarithm of predicted aqueous solubility, log S. S in mol.dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid (range for 95% of drugs: -6.0 - 0.5);
[l] logarithm of predicted binding constant to human serum albumin (range for 95% of drugs: -1.5 - 1.5);
[m] logarithm of predicted brain/blood partition coefficient. Note: QikProp predictions are for orally delivered drugs so, for example, dopamine and sero tonin are CNS negative because they are too polar to cross the blood-brain barrier (range for 95% of drugs: -3.0 - 1.2);
[n] predicted apparent Caco-2 cell membrane permeability in Boehringer-Ingelheim scale, in [nm/s] (range for 95% of drugs: < 25 poor, > 500 great);
[o] number of likely metabolic reactions (range for 95% of drugs: 1 - 8);
[p] predicted inhibition constants IC_{50}^{pre} (nM). IC_{50}^{pre} was predicted from computed DDG_{comp} using the regression equation shown in Table 3;
[q] human oral absorption (1 - low, 2 - medium, 3 - high);
[r] percentage of human oral absorption in gastrointestinal tract (<25% - poor, >80% high);
[(*)] star indicating that the property descriptor value falls outside the range of values for 95% of known drugs.

4 Conclusions

Structural information from the crystal structure of FP3 -K11017 complex guided us during elaboration of reliable QSAR models of non-covalent inhibition of the FP3 of *P. falciparum* by azadipeptide nitrile (AZN) inhibitors, which correlated computed Gibbs free energies of complex formation with observed inhibitory potencies [12]. In addition to this QSAR model, we have derived a 3D QSAR pharmacophore models for AZN inhibitors. Analysis of interactions between the FP3 and AZNs in the enzyme active site directed us in our effort to design an initial diversity virtual combinatorial library of new AZN analogues with multiple substitutions. The design strategy was based predominantly on the presence of the hydrophobic features included in the best PH4 pharmacophore models at the P1 to P3 positions of AZNs. The focused library filtered by a set of ADME-related descriptors and screened by matching of the analogues to the PH4 pharmacophore permitted selection of a library subset of orally bioavailable AZN. The subset of 78 best virtual hits was submitted to GFE computation of predicted inhibitory potencies by the complexation QSAR model derived from the training set. The best analogues reached predicted activities in the low nanomolar concentration range. The best designed AZN analogues, **21-70-166** (1.1 nM), **32-158-166** (1.3 nM), **160-70-166** (1.6 nM), **21-158-166** (1.7 nM) **164-158-166** (1.8 nM), Table 10, are recommended for synthesis and subsequent activity evaluation in FP3 inhibition assays and may lead to a discovery of novel potent orally bioavailable antimalarial. Usually the investigations of the active site of an enzyme start with the analysis its interactions with peptides inhibitors for identification of the best pharmacophore features which will guide the subsequent design and synthesis of peptidomimetic and non peptidic inhibitors []. The same approach, initiated by the reported AZN potencies [12], successfully provided in this work helpful information particularly for P1 and P2 position despite the small level of diversity of the training set. On the whole for FP3, the structural information provided by the PH4 especially the coordinates of the “pocket centered” hydrophobic features will be of crucial help for the design of non-peptide inhibitors. The report of a FP3-AZN complex crystal structure for one of the AZN best analogues we suggested for synthesis or three more AZN experimental evaluation with substitutions at P3 will help in the assessment to valuable additional structural information at the level of the one we obtained here from P2 with our model. In this way we’ll afford better orientation to the design of non peptidic falcipains inhibitors. Moreover our PH4 is in improvement in order to reach the selectivity over human cysteine proteases; we’ll report the results in due course.

Acknowledgement

This work was also supported by the Slovak Research and Development Agency grants APVV-14-0294 and APVV-14-0113"

References

1. Millennium development Goals, http://www.undp.org/content/undp/fr/home/mdgoverview/mdg_goals/mdg6
2. Wongsrichanalai, C., & Meshnick, S. R. (2008). Declining Artesunate-Mefloquine Efficacy against Falciparum Malaria on the Cambodia–Thailand Border. *Emerging Infectious Diseases, 14*(5), 716-719. doi: 10.3201/eid1405.071601.
3. O'Brien, C., Henrich, P.P., Passi, N., and Fidock, D.A. (2011). Recent clinical and molecular insights into emerging artemisinin resistance in *Plasmodium falciparum*. *Curr Opin Infect Dis.* 24, 570-577. DOI: 10.1097/QCO.0b013e32834cd3ed.
4. O'Neill, H. G., Mzilahowa, T., de Deus, N., Njenga, S. M., Mmbaga, E. J., & Kariuki, T. M. (2013). Evaluation of the European Foundation Initiative into African Research in Neglected Tropical Diseases by the African Fellows. *PLOS Neglected Tropical Diseases, 7*(3), e2019. doi: 10.1371/journal.pntd.0002019.
5. Kiso, A., Hidaka, K., Kimura, T., Hayashi, Y., Nezami, A., Freire, E., & Kiso, Y. (2004). Search for substrate-based inhibitors fitting the S2′ space of malarial aspartic

protease plasmepsin II. *Journal of Peptide Science, 10* (11), 641-647. doi: 10.1002/psc.609.
6. Perez, B., Teixeira, C., Gomes, J. R. B., & Gomes, P. (2013). Development of Plasmodium falciparum Protease Inhibitors in the Past Decade (2002–2012). *Current Medicinal Chemistry, 20*(25), 3049-3068. doi: http://dx.doi.org/10.2174/0929867311320250003.
7. Kolla, V. K., Prasad, R., Sayyad, Z., Atul, Shah, A. Y., Allanki, A. D., Sijwali, P. S. (2015). Independent amino acid residues in the S2 pocket of falcipain-3 determine its specificity for P2 residues in substrates. *Molecular and Biochemical Parasitology, 202*(2), 11-22. doi: http://dx.doi.org/10.1016/j.molbiopara.2015.09.005.
8. Boris, D. B., Fidele, N.-K., Luc, C. O. O., & Eugene, M. (2016). Targeting Cysteine Proteases from Plasmodium falciparum: A General Overview, Rational Drug Design and Computational Approaches for Drug Discovery. *Current Drug Targets, 17*, 1-26. doi: http://dx.doi.org/10.2174/1389450117666161221122432.
9. Kesharwani, R. K., Singh, D. V., & Misra, K. (2013). Computation-based virtual screening for designing novel antimalarial drugs by targeting falcipain-III: a structure-based drug designing approach. *J Vector Borne Dis, 50*(2), 93-102.
10. Teixeira, C., Gomes, J. R., & Gomes, P. (2011). Falcipains, Plasmodium falciparum cysteine proteases as key drug targets against malaria. *Curr Med Chem, 18*(10), 1555-1572.
11. Verissimo, E., Berry, N., Gibbons, P., Cristiano, M. L. S., Rosenthal, P. J., Gut, J., O'Neill, P. M. (2008). Design and synthesis of novel 2-pyridone peptidomimetic falcipain 2/3 inhibitors. *Bioorganic & Medicinal Chemistry Letters, 18*(14), 4210-4214. doi: http://dx.doi.org/10.1016/j.bmcl.2008.05.068.
12. Löser, R., Gut, J., Rosenthal, P. J., Frizler, M., Gütschow, M., & Andrews, K. T. (2010). Antimalarial activity of azadipeptide nitriles. *Bioorganic & Medicinal Chemistry Letters, 20*(1), 252-255. doi: http://dx.doi.org/10.1016/j.bmcl.2009.10.122.
13. Hogg, T., Nagarajan, K., Herzberg, S., Chen, L., Shen, X., Jiang, H., Schmidt, C. L. (2006). Structural and functional characterization of Falcipain-2, a hemoglobinase from the malarial parasite Plasmodium falciparum. *J Biol Chem, 281*(35), 25425-25437. doi: 10.1074/jbc.M603776200.
14. Hansen, G., Heitmann, A., Witt, T., Li, H., Jiang, H., Shen, X., Hilgenfeld, R. (2011). Structural Basis for the Regulation of Cysteine-Protease Activity by a New Class of Protease Inhibitors in Plasmodium. *Structure, 19*(7), 919-929. doi: http://dx.doi.org/10.1016/j.str.2011.03.025.
15. Wang, S. X., Pandey, K. C., Somoza, J. R., Sijwali, P. S., Kortemme, T., Brinen, L. S., McKerrow, J. H. (2006). Structural basis for unique mechanisms of folding and hemoglobin binding by a malarial protease. *Proc Natl Acad Sci U S A, 103*(31), 11503-11508. doi: 10.1073/pnas.0600489103.
16. Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., Brinen, L. S. (2009). Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. *Journal of Biological Chemistry, 284*(38), 25697-25703.
17. Moon, J. B.; Coleman, R. S.; Hanzlik, R.-P. *J. Am. Chem. Soc.* 1986, 108, 1350.
18. Ward, Y. D.; Thomson, D. S.; Frye, L. L.; Cywin, C. L.; Morwick, T.; Emmanuel, M. J.; Zindell, R.; McNeil, D.; Bekkali, Y.; Girardot, M.; Hrapchak, M.; DeTuri, M.; Crane, K.; White, D.; Pav, S.; Wang, Y.; Hao, M.-H.; Grygon, C. A.; Labadia, M. E.; Freeman, D. M.; Davidson, W.; Hopkins, J. L.; Brown, M. L.; Spero, D. M. *J. Med. Chem.* 2002, 45, 5471.
19. Löser, R.; Schilling, K.; Dimmig, E.; Gütschow, M. *J. Med. Chem.* 2005, 48, 7688.
20. Lin, S.-K. (2000). Pharmacophore Perception, Development and Use in Drug Design. Edited by Osman F. Güner. *Molecules, 5*(7). doi: 10.3390/50700987.
21. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.
22. Accelrys. (2005). Insight-II and discover molecular modeling and simulation package (Version 2005). New York: Accelrys.
23. Owono Owono, L. C., Keita, M., Megnassan, E., Frecer, V., & Miertus, S. (2013). Design of Thymidine Analogues Targeting Thymidilate Kinase of Mycobacterium tuberculosis. *Tuberculosis Research and Treatment, 2013*, 670836. doi: 10.1155/2013/670836.
24. Frecer, V., Miertus, S., Tossi, A., & Romeo, D. (1998). Rational design of inhibitors for drug-resistant HIV-1 aspartic protease mutants. *Drug Des Discov, 15*(4), 211-231.
25. Frecer, V., & Miertus, S. (2002). Interactions of ligands with macromolecules: Rational design of specific inhibitors of aspartic protease of HIV-1. *Macromolecular Chemistry and Physics, 203*(10-11), 1650-1657. doi: 10.1002/1521-3935(200207)203:10/11<1650::AID-MACP1650>3.0.CO;2-E
26. Frecer, V., Berti, F., Benedetti, F., & Miertus, S. (2008). 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, 27*(3), 376-387. doi: 10.1016/j.jmgm.2008.06.006.
27. Dali, B., Keita, M., Megnassan, E., Frecer, V., & Miertus, S. (2012). Insight into selectivity of peptidomimetic inhibitors with modified statine core for plasmepsin II of *Plasmodium falciparum* over human cathepsin D. *Chem Biol Drug Des, 79*(4), 411-430. doi: 10.1111/j.1747-0285.2011.01276.x.
28. Megnassan, E., Keita, M., Bieri, C., Esmel, A., Frecer, V., & Miertus, S. (2012). Design of novel dihydroxynaphthoic acid inhibitors of *Plasmodium falciparum* lactate dehydrogenase. *Med Chem, 8*(5), 970-984.
29. Keita, M., Kumar, A., Dali, B., Megnassan, E., Siddiqi, M. I., Frecer, V., & Miertus, S. (2014). Quantitative structure-activity relationships and design of thymine-like inhibitors of thymidine monophosphate kinase of *Mycobacterium tuberculosis* with favourable pharmacokinetic profiles. *RSC Advances, 4*(99), 55853-55866. doi: 10.1039/C4RA06917J.
30. Owono Owono, L. C., Ntie-Kang, F., Keita, M., Megnassan, E., Frecer, V., & Miertus, S. (2015). Virtually Designed Triclosan-Based Inhibitors of Enoyl-Acyl Carrier Protein Reductase of *Mycobacterium tuberculosis* and of Plasmodium falciparum. *Molecular Informatics, 34*(5), 292-307. doi: 10.1002/minf.201400141.
31. Kouassi, A. F., Kone, M., Keita, M., Esmel, A., Megnassan, E., N'Guessan, Y. T., Miertus, S. (2015).

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, 16*(12), 29744-29771. doi: 10.3390/ijms161226196.
32. Maple, J. R., Hwang, M. J., Stockfisch, T. P., Dinur, U., Waldman, M., Ewig, C. S., & Hagler, A. T. (1994). Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. *Journal of Computational Chemistry, 15*(2), 162-182. doi: 10.1002/jcc.540150207.
33. Accelrys. (2009). Discovery Studio Modeling Environment (Version 2.5).
34. Gilson, M. K., & Honig, B. (1991). The inclusion of electrostatic hydration energies in molecular mechanics calculations. *J Comput Aided Mol Des, 5*(1), 5-20.
35. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., & Honig, B. (2002). 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, 23*(1), 128-137. doi: 10.1002/jcc.1161.
36. Böttcher, C. J. F. (1973). HISTORICAL INTRODUCTION *Theory of Electric Polarization (Second Edition)* (pp. 1-8). Amsterdam: Elsevier.
37. Miertuš, S., Scrocco, E., & Tomasi, J. (1981). Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chemical Physics, 55*(1), 117-129. doi:http://dx.doi.org/10.1016/0301-0104(81)85090-2.
38. Frecer, V., & Miertuš, S. (1992). Polarizable continuum model of solvation for biopolymers. *International Journal of Quantum Chemistry, 42*(5), 1449-1468. doi: 10.1002/qua.560420520.
39. Copeland, R. A., Lombardo, D., Giannaras, J., & Decicco, C. P. (1995). Estimating KI values for tight binding inhibitors from dose-response plots. *Bioorganic & Medicinal Chemistry Letters, 5*(17), 1947-1952. doi: http://dx.doi.org/10.1016/0960-894X(95)00330-V.
40. Fischer, S., Smith, J. C., & Verma, C. S. (2001). Dissecting the Vibrational Entropy Change on Protein/Ligand Binding: Burial of a Water Molecule in Bovine Pancreatic Trypsin Inhibitor. *The Journal of Physical Chemistry B, 105*(33), 8050-8055. doi: 10.1021/jp0120920.
41. Schwarzl, S. M., Tschopp, T. B., Smith, J. C., & Fischer, S. (2002). Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? *J Comput Chem, 23*(12), 1143-1149. doi: 10.1002/jcc.10112.
42. Schrodinger. (2014). QikProp (Version 3.7, release 14). New York: Schrodinger.
43. Duffy, E. M., & Jorgensen, W. L. (2000). Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *Journal of the American Chemical Society, 122*(12), 2878-2888. doi: 10.1021/ja993663t.
44. Jorgensen, W. L., & Duffy, E. M. (2000). Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett, 10*(11), 1155-1158.
45. Jorgensen, W. L., & Duffy, E. M. (2002). Prediction of drug solubility from structure. *Adv Drug Deliv Rev, 54* (3), 355-366.
46. Chemical Computing Group. (2014). Molecular Operating Environment (MOE) (Version 2014).
47. Available Chemicals Directory, Version 95.1, MDL Information Systems, San Leandro, CA.
48. Dean, P. M. (1994). *Molecular simylarity in drug Design*: Springer.
49. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug discovery today*. 13(19-20):869-874.
50. Shah, F., Gut, J., Legac, J., Shivakumar, D., Sherman, W., Rosenthal, P. J., & Avery, M. A. (2012). Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure–Activity Relationships, Hydration Site Thermodynamics, and Reactivity Analysis. *Journal of Chemical Information and Modeling, 52*(3), 696-710. doi: 10.1021/ci2005516.
51. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews, 46*(1–3), 3-26. doi: http://dx.doi.org/10.1016/S0169-409X(00)00129-0.
52. Cotereon JM, Catterick D, Castro J, et al. (2010), Falcipain Inhibitors: Optimization Studies of the 2-Pyrimidinecarbonitrile Lead Series. *J Med Chem.* 53, 6129–6152.
53. Frecer V, Kabelac M, De Nardi P, et al. (2004), Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus. *J Mol Graph Mod.* 22(3):209-220.