J Comput Aided Mol Des (2010) 24:195–212
DOI 10.1007/s10822-010-9326-8

Design, structure-based focusing and in silico screening of combinatorial library of peptidomimetic inhibitors of Dengue virus NS2B-NS3 protease

Vladimir Frecer · Stanislav Miertus

Received: 10 September 2009 / Accepted: 8 March 2010 / Published online: 21 March 2010

Abstract Serine protease activity of the NS3 protein of Dengue virus is an important target of antiviral agents that interfere with the viral polyprotein precursor processing catalyzed by the NS3 protease (NS3pro), which is important for the viral replication and maturation. Recent studies showed that substrate-based peptidomimetics carrying an electrophilic warhead inhibit the NS2B-NS3pro cofactor-protease complex with inhibition constants in the low micromolar concentration range when basic amino acid residues occupy P_1 and P_2 positions of the inhibitor, and an aldehyde warhead is attached to the P_1. We have used computer-assisted combinatorial techniques to design, focus using the NS2B-NS3pro receptor 3D structure, and in silico screen a virtual library of more than 9,200 peptidomimetic analogs targeted around the template inhibitor Bz-Nle-Lys-Arg-Arg-*H* (Bz—benzoyl) that are composed mainly of unusual amino acid residues in all positions P_1–P_4. The most promising virtual hits were analyzed in terms of computed enzyme-inhibitor interactions and Adsorption, Distribution, Metabolism and Excretion (ADME) related physico-chemical properties. Our study can direct the interest of medicinal chemists working on a next generation of antiviral chemotherapeutics against the Dengue Fever towards the explored subset of the chemical space that is predicted to contain peptide aldehydes with NS3pro inhibition potencies in nanomolar range which display ADME-related properties comparable to the training set inhibitors.

Keywords Dengue virus · NS2B-NS3 serine protease · Peptidomimetic inhibitors · QSAR model · Virtual library design · In silico screening · Target-specific scoring function

Abbreviations

ADME	Adsorption distribution metabolism and excretion
CFF91	Consistent class II force field
DEN2	Dengue virus serological type 2
DHF	Dengue hemorrhagic fever
MM	Molecular mechanics
NS2B	Nonstructural protein 2B (essential cofactor)
NS3	Nonstructural protein 3
NS3pro	Serine protease domain of the NS3 protein

Unusual amino acids

Abu	α-Aminobutyric acid
Bz	Benzoyl
Cit	Citrulline
hHis	Homo-histidine
N-MeArg	(*N*-methyl)-arginine
2Nal	β-(2-Naphthyl)-alanine
Nle	Norleucine
Nva	Norvaline
Orn	Ornithine
p(Py)Ala	β-(4-Pyridyl)-alanine
p(Ac)Phe	(4-*N*-Acetylamino)-phenylalanine
m(Am)Phe	(3-Amino)-phenylalanine
p(Am)Phe	(4-Amino)-phenylalanine

V. Frecer · S. Miertus (✉)
International Centre for Science and High Technology, UNIDO, AREA Science Park, Padriciano 99, 34012 Trieste, Italy
e-mail: stanislav.miertus@ics.trieste.it

V. Frecer
Cancer Research Institute, Slovak Academy of Sciences, 83391 Bratislava, Slovakia

V. Frecer
Laboratory of Molecular, Biostructural and Nanomaterial Modeling, Consortium AREA Science Park of Trieste, AREA Science Park, 34012 Trieste, Italy

p(Cl)Phe	(4-Chloro)-phenylalanine
p(CN)Phe	(4-Cyano)-phenylalanine
m(Gn)Phe	(3-Guanidino)-phenylalanine
p(Gn)Phe	(4-Guanidino)-phenylalanine
m(Im)Phe	(3-Imino)-phenylalanine
p(Im)Phe	(4-Imino)-phenylalanine
p(Ip)Phe	(4-Isopropyl)-phenylalanine
(Md)Phe	(3,4-Methylenedioxy)-phenylalanine
p(Hm)Phe	(4-Hydroxymethyl)-phenylalanine
p(Me)Phe	(4-Methyl)-phenylalanine
(dMo)Phe	(3,4-Dimethoxy)-phenylalanine
p(Ph)Phe	(4-Phenyl)-phenylalanine

Introduction

The Dengue virus belongs to the *flavivirus* genus of *flaviviridae*, which include lethal human pathogens such as Yellow Fever, West Nile, Japanese and Tick Borne Encephalitis viruses [1]. At present, there is no chemotherapy and no effective vaccines for most of the over 70 currently known members of the *flavivirus* genus, with the exception of Yellow Fever virus [2]. Dengue virus is endemic in tropical and sub-tropical regions with the abundance of the mosquito vectors of *Aedes* spp. causing almost one million infections each year and placing about three billion people at risk of contracting dengue fever or the more severe and potentially lethal dengue hemorrhagic fever (DHF) and dengue shock syndrome [3, 4]. Dengue virus consists of four distinct but closely related serological types (DEN1–4), the most prevalent one is the serological type 2 (DEN2, New Guinea strain). Recovery from infection by one Dengue serotype provides lifelong immunity against that virus but confers only partial and transient protection against subsequent infections by the other three. There is substantial evidence that sequential infection increases the risk of developing DHF [4].

Similar to other flaviviruses, the Dengue viruses are small, enveloped viruses containing a single positive-stranded RNA genome of 10,723 nucleotides, which encodes a polyprotein precursor of 3,391 amino acids. Dengue viruses replicate in the cytoplasm of infected cells, thus both the host cell and viral encoded proteases process the polyprotein precursor into the individual functional proteins consisting of three structural (capsid, membrane-associated, and envelope) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [5]. The viral protease (NS3pro) encoded within the 180 N-terminal amino acid residues of the multifunctional protein NS3 (618 a.a., 69 kDa, N-terminal third comprises protease and C-terminal two-thirds display nucleotide triphosphatase, RNA triphosphatase and helicase activity) [6–9]. The NS3pro recognizes dibasic residues at the P_1 and P_2 positions of substrate and cleaves the viral polyprotein at the junctions between the NS2A/NS2B, NS2B/NS3, NS3/NS4A and NS4B/NS5 and at internal sites within the capsid, NS2A, NS3 and NS4A [6–8]. The enzymatic activity of the NS3pro, a trypsin-like serine protease with a classical catalytic triad: His51, Asp75, and Ser135, is dependent on its interaction with the NS2B essential cofactor [10–12]. The NS3pro is vital for the post-translational proteolytic processing of the polyprotein precursor and is essential for viral replication and maturation of infectious Dengue virions. Therefore, the NS3pro is an attractive therapeutic target for the Dengue virus infection [6, 13].

Classical inhibitors of serine proteases display a limited effect on the Dengue NS3pro and only a few peptidic and non-peptidic inhibitors with moderate activities have been reported [12–19]. Two high affinity substrate-based tetrapeptides were recently identified through substrate profiling of the NS3pro [17]. Based on these peptides, Yin et al. synthesized and evaluated several peptidomimetic inhibitors containing electrophilic covalent warheads such as aldehydes, trifluoromethyl ketones, and boronic acids attached to the P_1 residues [20]. They also performed a structure–activity relationships study on tetrapeptides containing the aldehyde warhead to probe the specificity of the enzyme binding pockets of the NS2B-NS3pro complex including a truncated NS2B cofactor and observed that interactions of P_2 side chain are more important than the P_1, which can be effectively replaced by a non-basic side chain without noteworthy loss of binding affinity [21].

Recently, the crystal structures of the DEN2 NS3pro [10], full length NS3 [22], NS3pro with Bowman–Birk inhibitor [23], NS2B-NS3pro complex and West Nile virus NS2B-NS3pro complex with substrate-based inhibitor Bz-Nle-Lys-Arg-Arg-*H* containing the aldehyde warhead (-*H*) [24], have been reported. Structural analysis of the West Nile virus NS2B-NS3pro complex (which displays about 45% sequence identity and 80% sequence similarity to the DEN2 NS3pro) showed that the NS2B cofactor not only stabilizes both the N- and C-terminal barrels of the NS3pro by providing additional β-strands but the catalytically important C-terminal part of the NS2B (residues 82–85) completes the substrate-binding site [24]. The structure of this enzyme-inhibitor complex also revealed key interactions for the substrate recognition in the S_1 pocket where Asp129 located at the bottom of S_1 interacts with the positively charged side chain of arginine in the P_1 position of the inhibitor [24]. The S_2 pocket is dominated by the negative electrostatic potential originating from the backbone atoms of NS2B residues 82–84 that stabilizes the guanidinium group of arginine residue in the P_2 position, which is also within hydrogen bonding distance of NS2B Gly83 and NS3pro Asn152 [24]. The availability of structural data opened the possibility to achieve

improvements in selectivity and inhibitory activity of the peptidomimetic NS3pro inhibitors by rational structure-based approaches.

The concept of the peptidomimetics with a reversible electrophilic warhead as well as macrocyclic inhibitors has been successfully applied to the design of NS3-4A serine protease inhibitors of the hepatitis C virus (HCV) from the same *flaviviridae* family [25–32], which shares structural similarity with the DEN2 NS2B-NS3pro [14]. The HCV serine protease represents a validated but challenging drug target due to its shallow, hydrophobic, highly flexible binding pocket and helicase domain of the NS3 contributing to the protease inhibitor binding [25, 30]. For the Dengue virus, observed binding of the inhibitor D1 to the NS2B-NS3pro with an inhibition constant K_i^{exp} of 5.8 μM was retained also in the full-length NS3 protein with an almost identical K_i^{exp} of 7.0 μM [21]. Therefore, we may assume that the role of the helicase domain of the DEN2 NS3 protein on the inhibitory potency of protease inhibitor D1 and its peptidomimetic analogs is less significant than the effect observed in the HCV [26, 30].

In this study, we have explored extensively the chemical space of P_1, P_2, P_3, and P_4 positions of the peptidomimetic NS3pro inhibitors targeted around the substrate-based Bz-Nle-Lys-Arg-Arg-*H* template inhibitor D1 [20, 21] (Fig. 1). A virtual library of more than 9,200 analogs containing natural and especially unusual amino acids was designed, structure-based focused and in silico screened by computer-assisted combinatorial techniques [33–36] aiming at finding more potent and specific antiviral compounds. The three-dimensional structure of the DEN2 NS2B-NS3pro complex was employed to develop a QSAR model, parameterize a target-specific scoring function specific for the NS3pro target and select analogues which display the highest predicted binding to the NS3pro. We have derived a small highly focused combinatorial subset of peptidomimetics, which contains virtual hits that are predicted to inhibit DEN2 NS3pro in the nanomolar concentration range and display ADME-related properties comparable to the template inhibitor.

Results and discussion

Library design

Enzyme inhibition data from amino acid scan of peptidomimetic NS3pro inhibitors prepared by Yin et al. [21] revealed the preferences of the nonprime binding pockets S_1–S_4 of the DEN2 NS3pro for specific amino acids as: P_1—Arg, basic and aromatic residues, P_2—Arg, P_3—basic and polar residues, P_4—nonspecific (larger hydrophobic residues). Template inhibitor D1 (Fig. 1) containing an aldehyde warhead showed reversible competitive binding to the NS3pro with an inhibition constant K_i^{exp} of 5.8 μM and was almost equally active also in the full-length DEN2 NS3 protein [21]. Libraries of D1 analogs can be readily prepared by solid-phase parallel synthesis of peptide aldehydes from amino acid building blocks using Weinreb resin [21, 37, 38]. We have designed a virtual diversity library around the D1 template tetrapeptide considering larger selection of unusual amino acids corresponding to the observed NS3pro binding pocket preferences. Natural and unusual amino acids of suitable type available from the catalogues of suppliers of chemicals (Fig. 2), were selected to form the initial diversity library of the size: 21 (R_1) × 10 (R_2) × 11 (R_3) × 4 (R_4) = 9,240 analogs. This virtual library was generated by attaching of the R-groups (Fig. 2) onto the tetrapeptide aldehyde scaffold of D1 (Fig. 1) using the Cerius^2 program [39]. The range of the suitable residue properties was somewhat augmented to permit the use of diverse building blocks to explore a wider subset of the chemical space and propose new D1 analogs. Larger number of R_1-groups was chosen in line with the elevated variability of the P_1 residues, strict specificity of the S_2 pocket for arginine and the key role of P_1 and P_2

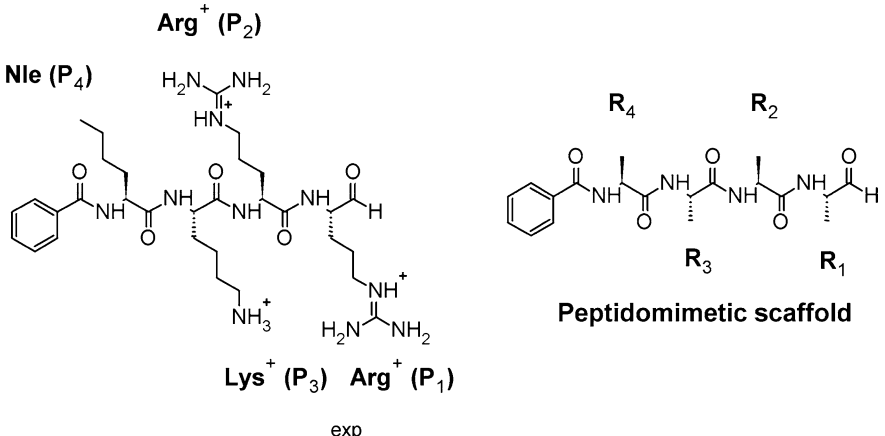


Fig. 1 Chemical structure of template inhibitor D1. Substrate-based inhibitor of Dengue virus NS3pro Bz-Nle-Lys-Arg-Arg-*H* with aldehyde electrophilic warhead [20, 21] and the peptidomimetic scaffold used in the virtual library generation with R-groups notation

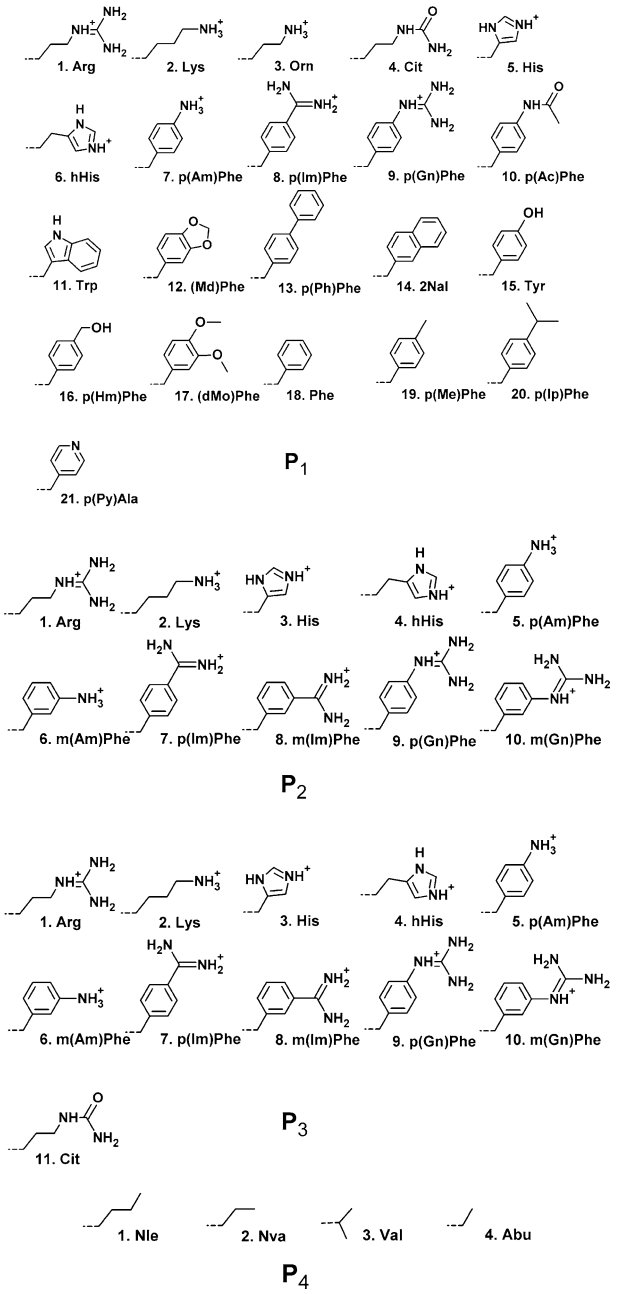


Fig. 2 R-groups (amino acid side chains) used in design of library of analogs of NS3pro inhibitor D1. For the R-groups numbering and residue notation see the scaffold on Fig. 1. Annotation: Abu—α-aminobutyric acid, Bz—benzoyl, Cit—citrulline, hHis—homo-histidine, *N*-MeArg—(*N*-methyl)-arginine, 2Nal—*β*-(2-naphthyl)-alanine, Nle—norleucine, Nva—norvaline, Orn—ornithine, *p*(Py)Ala—*β*-(4-pyridyl)-alanine, *p*(Ac)Phe—(4-*N*-acetylamino)-phenylalanine, *m*(Am)Phe—(3-amino)-phenylalanine, *p*(Am)Phe—(4-amino)-phenylalanine, *p*(Cl)Phe—(4-chloro)-phenylalanine, *p*(CN)Phe—(4-cyano)-phenylalanine, *m*(Gn)Phe—(3-guanidino)-phenylalanine, *p*(Gn)Phe—(4-guanidino)-phenylalanine, *m*(Im)Phe—(3-imino)-phenylalanine, *p*(Im)Phe—(4-imino)-phenylalanine, *p*(Ip)Phe—(4-isopropyl)-phenylalanine, (Md)Phe—(3,4-methylenedioxy)-phenylalanine, *p*(Hm)Phe—(4-hydroxymethyl)-phenylalanine, *p*(Me)Phe—(4-methyl)-phenylalanine, (dMo)Phe—(3,4-dimethoxy)-phenylalanine, *p*(Ph)Phe—(4-phenyl)-phenylalanine

residues manifested through potent inhibition of the NS3pro by dipeptide aldehydes [21].

To design a focused library of a reduced size and increased content of drug-like molecules, we have introduced a set of filters, which can help to select the most suitable analogs.

Analog-based focusing

In the next step, analog-based library focusing was applied to reduce the size of the generated library of D1 analogs. Based on the observed low specificity of the S_4 pocket of the NS3pro and relatively small variations of the inhibitory potency with respect to the P_4 residue replacement, it was decided to keep only analogs that contained the largest R_4-group, the Nle residue present in D1, and filter out analogs containing sub-optimal smaller R-groups: Nva, Val and Abu (Fig. 2). Thus a library subset of 2,310 analogs retaining full diversity in the R_1- to R_3-groups was chosen from the initial diversity library for a subsequent structure-based focusing step.

NS2B-NS3pro target and structure-based focusing

The 3D model of the DEN2 NS2B-NS3pro drug target including the essential cofactor was prepared by superposition of the crystal structure of Dengue virus NS2B-NS3pro available from the PDB (2FP7) [24] with its highly homologous counterpart (45% sequence identity, 80% sequence similarity) from the West Nile virus (Fig. 3), for which a complete crystal structure of the NS2B-NS3pro-D1 inhibitor complex (2FOM) exists [24] (Fig. 4). Refined 3D structure of the cofactor-protease-inhibitor complex was used in the structure-based focusing of the library (see the Methods section for details).

To select a small highly focused combinatorial subset of the virtual library of D1 analogs with the highest predicted binding affinities to the NS3pro, structure-based focusing and in silico screening procedures have been applied. Each analog was docked into the binding pocket of the NS2B-NS3pro represented by a binding site grid (Fig. 5) by employing the Monte Carlo procedure for conformers generation and ligand fitting algorithm of the LigFit module of Cerius^2 for the docking [40]. The docking procedure yielded 20 best binding conformers per analog, which were clustered into 10 conformational families depending on the mutual r.m.s. deviations [41]. In each cluster, the conformer with the highest docking score was selected for subsequent virtual screening.

QSAR model and target-specific scoring function

Target-specific scoring functions exhibit better predictivities on average than their generic counterparts [42]. To

```
CLUSTAL W2.0.10 sequence alignment:

2FP7_B          ---------------GDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFHTLWHTTKGAALMS 45
2FOM_B          AGVLWDVPSPPPVGKAELEDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMH 60
                              .:   *.***  :*:**  * ****  **.***:**.*:**.**
2FP7_B          GEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDEVQMIVVEPGKNVKNVQTKPGVFKTPE 105
2FOM_B          KGKRIEPSWADVKKDLISYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFKTNT 120
                   *::* *..**:* :.*** ***: :*:  :***::.:***** : ******:***
2FP7_B          GEIGAVTLDYPTGTSGSPIVDKNGDVIGLYGNGVIMPNGSYISAIVQGERMEEPAPAGFE 165
2FOM_B          GTIGAVSLDFSPGTSGSPIVDKKGKVVGLYGNGVVTRSGAYVSAIANTEKSIEDNPE-IE 179
                * ****:**:..**********:*.*:*******:  .*:*:***.: *:  *  *  :*
2FP7_B          PEMLRKK 172
2FOM_B          DDIFRK- 185
                 :::**

170 aligned residues, 91 * = 45% sequence identity, 144 .:* = 80% similarity
```

Fig. 3 Sequence alignment of NS3pro of Dengue and West Nile virus. ClustalW2 sequence alignment with EMBL—EBI server [54] displayed overall alignment score of 584, sequence identity of 45% and sequence similarity of 80% between the NS3pro of the two flaviviruses. 2FOM_B—Dengue virus (serotype 2) NS3pro (chain B), chain A is the NS2B cofactor [24], 2FP7_B—West Nile virus NS3pro (chain B), chain A is the NS2B cofactor [24]

parameterize a scoring function specific for the DEN2 NS2B-NS3pro target, we have correlated the scores implemented in the Cerius^2 program [39] with the experimental activities (K_i^{exp}) of a training set of peptidomimetic NS3pro inhibitors D1–D12 (Table 1) [20, 21]. The molecules of the training set were docked into the NS3pro receptor model using the LigFit algorithm of Cerius^2 [40]. Comparison of the docked analogs with the bound conformation of the D1 at the active site of NS3pro (Fig. 6) demonstrated that the binding mode for the training set inhibitors was very similar.

Various QSAR models relating the observed K_i^{exp}s to computed LigScore, LUDI, PMF and PLP scores [43–47], were tested. From the set of available scoring functions, the LUDI score led to the best fit of the experimental activities to the enzyme-inhibitor binding energies computed by means of a scoring function. The following QSAR equation was obtained by linear regression (Fig. 7):

$$\begin{aligned} pK_i &= -\log_{10} K_i[\text{DEN2 NS3pro}] \\ &= -3.50364 + 0.00624\,\text{LUDI} \end{aligned} \quad (1)$$

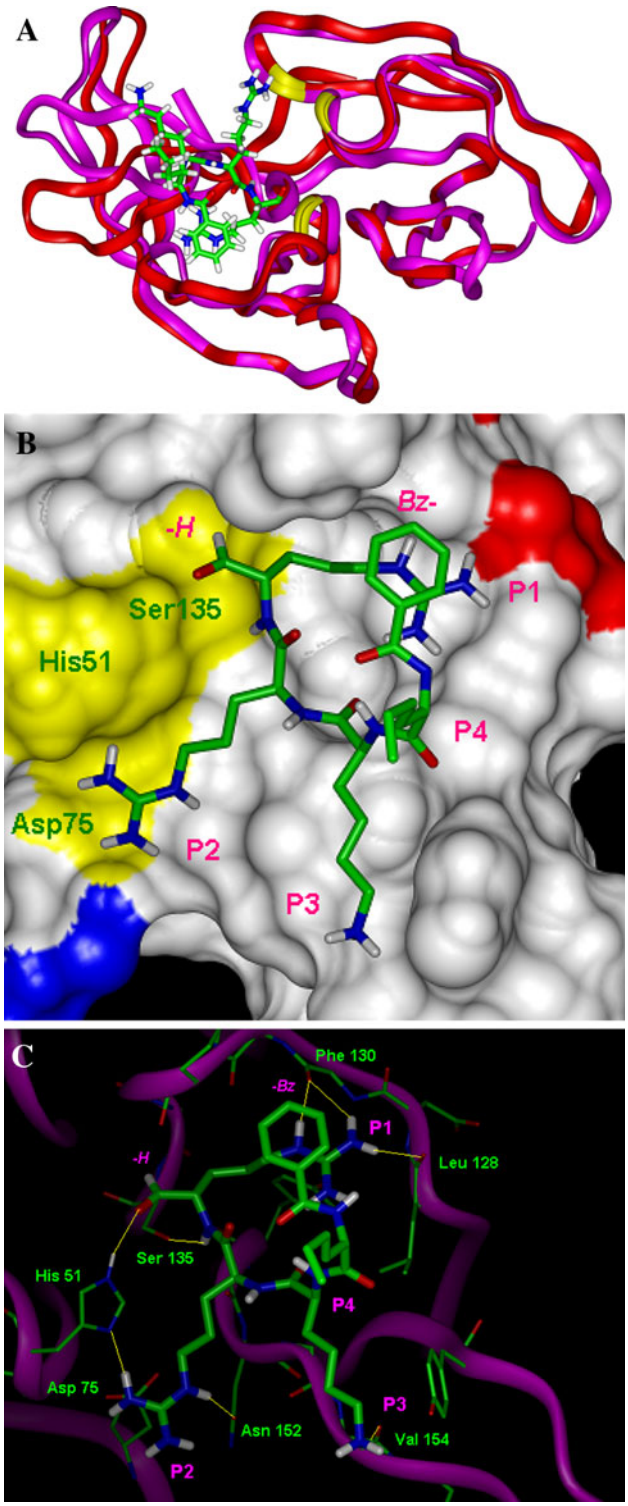
(number of samples $n = 12$, correlation coefficient $R^2 = 0.86$, leave-one-out cross validated correlation coefficient $R^2_{xv} = 0.79$, standard error $\varsigma = 0.13$, Fisher F-test = 58.8, statistical significance of the correlation $\alpha > 95\%$). The leave-one-out cross validated correlation coefficient R^2_{xv} of 0.79 indicates that major portion of the variance of the training set data was well described by this QSAR model. Empirical scoring functions, which estimate the free energy of binding of a ligand to a protein receptor with known 3D structure, tend to fall in general into two major classes emphasizing either hydrogen bonding and ionic interactions or van der Waals and hydrophobic interactions. As LUDI belongs the first class of functions, it was able to reproduce the binding of the training set of cationic peptidomimetic D1 analogs to the NS3pro better than the other tested scoring functions. The LUDI scoring function has been quite successful in correctly rank-ordering accurate ligand poses that are close to the crystal structure binding mode [44, 48, 49].

The quality of the QSAR model was also confirmed by predicting the NS3pro inhibition constants for a validation set of 3 analogs V1–V3 not included into the training set (Table 1) [21]. The ratio of the predicted activities derived from Eq. 1 and observed activities pK_i^{pre}/pK_i^{exp}, which yielded values close to 1, confirmed the predictive power of the QSAR model. Although the training and validation sets displayed somewhat limited variation in the R-groups space due to restricted availability of experimental data, prediction of inhibitory potencies by the trained target-specific scoring function, which slightly exceed the activity ranges of the training set, is still possible, as QSAR models using ligand–receptor binding affinity estimate (LUDI score) of docked analogs are less sensitive to size of the training set [33–36].

The parameterized scoring function, Eq. 1, which is specific for the DEN2 NS3pro, was subsequently used for in silico screening and ranking of the best conformers docked to the active site of the NS2B-NS3pro in the structure-based focusing step.

In silico screening

The best conformers of the analogs were screened in silico by using the LUDI score [43, 44]. The predicted NS3pro inhibitory activities K_i^{pre} of the analogs were calculated from the target-specific scoring function, Eq. 1. The analogs were then rank-ordered according to the estimated K_i^{pre} (Table 2). 3D structures of 16 analogs with the highest predicted potencies in low micromolar and nanomolar ranges (virtual hits) are shown on Fig. 8 and superposition of their bound conformations is given on Fig. 9. Relatively tight overlap of the poses of the high ranking analogs with the experimental structure of the template inhibitor D1 indicates that the LigFit docking algorithm of Cerius^2 [40]



◀**Fig. 4** Superposition of Dengue and West Nile virus NS3pros and inhibitor–enzyme interactions. **a** Superimposed structures of Dengue virus NS2B-NS3pro (2FOM—purple ribbon) and West Nile virus NS2B-NS3pro (2FP7—red ribbon) [24] with the Bz-Nle-Lys-Arg-Arg-*H* (D1) inhibitor at the active site in stick representation. The NS2B cofactors are not shown for better clarity. Positions of the catalytic residues His51, Asp75and Ser135 are shown as yellow ribbon segments. **b** Inhibitor D1 positioned over the molecular surface of active site of DEN2 NS2B-NS3pro. Surface coloring scheme: acidic residues—*red*, neutral—*white*, basic—*blue*. **c** Hydrogen bonds (*yellow lines*) of D1 with residues of the DEN2 NS2B-NS3pro active site in the fully relaxed structure of the complex. Peptide backbone of the NS3pro active site is displayed as a purple ribbon

generated inhibitor geometries close to the observed binding mode and the LUDI score correctly discriminated between realistic and artificial docking solutions. Chemical structures of the best designed analogs are presented on Fig. 10.

Inhibitor–protease interactions

The designed analog with the highest predicted inhibitory potency DA1: *Bz*-Nle-His-*m*(Gn)Phe-*p*(Gn)Phe-*H* (Figs. 10, 11), contains aromatic analogs of arginine at the P_1 (4-guanidino-phenylalanine) and P_2 (3-guanidino-phenylalanine) positions and a histidine at the P_3 position displaying a total net molecular charge Q_{tot} of 3 ė, similar to D1 (Table 2). The bound conformation of DA1 is stabilized by π–π stacking interactions of the benzoyl group and phenyl ring of the *p*(Gn)Phe-*H* and aromatic–aromatic interactions involving Tyr 150 and Pro 132 (Fig. 11). The guanidine group of *p*(Gn)Phe-*H* forms 3 hydrogen bonds with the backbone carbonyls of Leu 128, Asp 129 and Phe 130. The aldehyde oxygen of the warhead interacts with the catalytic serine residue Ser 135. The guanidine group of P_2 residue made by *m*(Gn)Phe forms an ion pair with the catalytic Asp 75 and a weak stacking interaction with the catalytic His 51. The histidine at P_3 position forms hydrogen bonds with the amide carbonyls of Gly 153 and Val 154 and an aromatic–aromatic interaction with Tyr 161. The overlap of the virtual hit DA1 and template inhibitor D1 at the catalytic site of the NS3pro is rather close (Fig. 12).

We have analyzed the inhibitor–enzyme interactions of 16 best virtual hits docked to the NS2B-NS3pro in terms of the individual R-groups contributions after careful molecular mechanics refinement of the resulting complexes (Table 3) (for the refinement details see "Methods"). The refinement by MM minimization partially compensated lost receptor flexibility that was neglected during the semi-flexible docking. After relaxation, most of the virtual hits displayed improved total inhibitor–enzyme non-bonding interaction energies (E_{int}) with the NS3pro with respect to DA1 ($K_i^{exp} = 5.8$ μM) [21]. The residue contribution analysis revealed that *p*(Gn)Phe and *p*(Am)Phe are suitable building blocks for the P_1 position (Fig. 2), *p*(Gn)Phe, *m*(Gn)Phe and *p*(Im)Phe for P_2 and His for the P_3 position of the peptidomimetic scaffold. The largest contribution to E_{int} arises from basic residues. In the virtual hits with 3 basic residues ($Q_{tot} = 3$ ė) the major stabilization

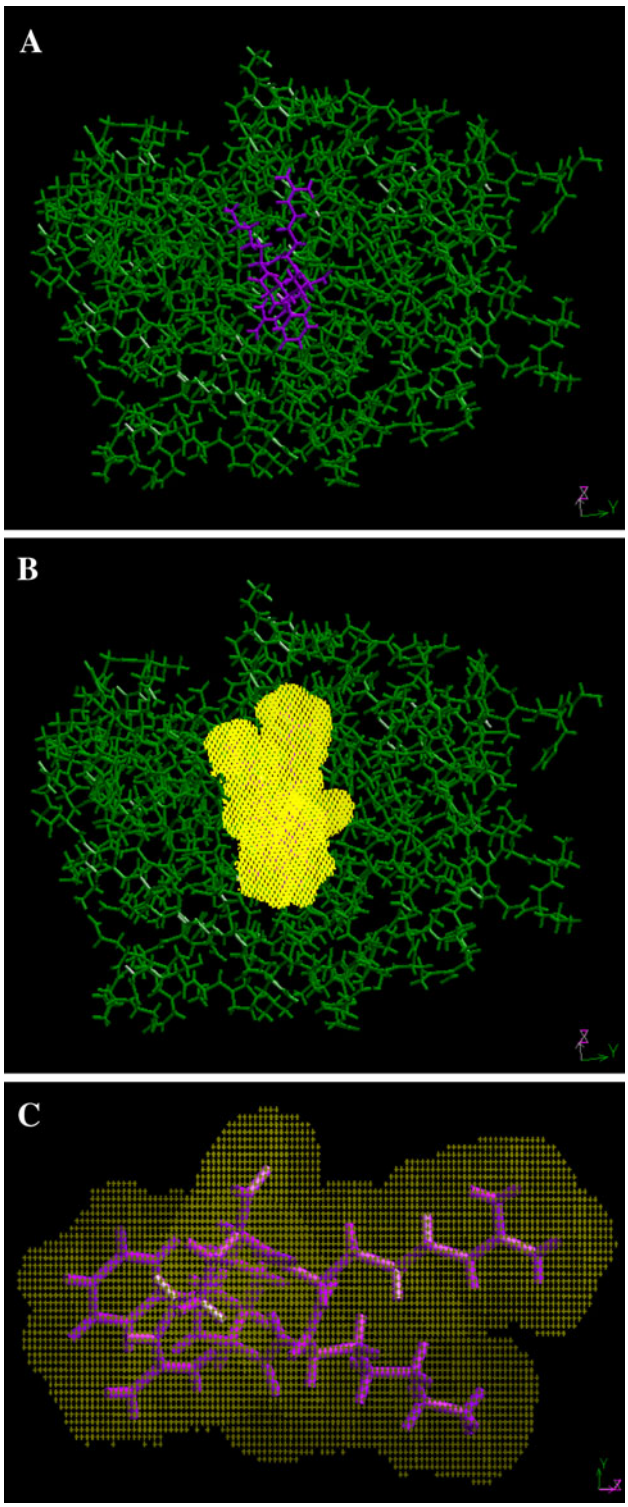


Fig. 5 Site model of active site of Dengue virus NS3pro with bound inhibitor D1. The site model (*yellow grid* with a spacing 0.25 Å per unit) was derived from the crystal structures of DEN2 NS2B-NS3pro and West Nile virus NS2B-NS3pro-D1 [24] using the LigFit module of Cerius^2 [39]. **a** The complex DEN2 NS2B-NS3pro (*green*)—D1 (*purple*) in stick representation. **b** The complex DEN2 NS2B-NS3pro (*green*) with site model (*yellow grid*). **c** Site model overlaying the template inhibitor D1 [20, 21] (in stick representation)

Table 1 Training and validation sets of tetrameric peptidomimetic inhibitors of NS2B-NS3pro used in the QSAR model of the inhibitory effect [20, 21] and parameterization of the target-specific LUDI scoring function [43, 44]

Training set	Inhibitor sequence[a] Bz-P_4-P_3-P_2-P_1-*H*[b]	Q_{tot}[c] [ė]	K_i^{exp}[d] [μM]
D1	Bz-Nle-Lys-Arg-Arg-*H*[e]	3	5.8
D2	Bz-Nle-Ala-Arg-Arg-*H*	2	22.1
D3	Bz-Ala-Lys-Arg-Arg-*H*	3	5.3
D4	Bz-Nle-Lys-Arg-Phe-*H*	2	15.9
D5	Bz-Nle-Phe-Arg-Arg-*H*	2	15.8
D6	Bz-Phe-Lys-Arg-Arg-*H*	3	6.8
D7	Bz-Nle-Lys-Arg-Lys-*H*	3	20.5
D8	Bz-Nle-Lys-Lys-Arg-*H*	3	41.3
D9	Bz-Nle-Pro-Arg-Arg-*H*	2	61.4
D10	Bz-Nle-Lys-Arg-*p*(CN)Phe-*H*	2	18.6
D11	Bz-Nle-Lys-Arg-*p*(Me)Phe-*H*	2	6.0
D12	Bz-Nle-Lys-Arg-*p*(Ph)Phe-*H*	2	11.6
Validation set	**Inhibitor sequence Bz-P_4-P_3-P_2-P_1-*H***	**Q_{tot} [ė]**	**pK_i^{pre}/pK_i^{exp}[f]**
V1	Bz-Nle-Lys-Arg-Trp-*H*	2	1.12
V2	Bz-Nle-Lys-Arg-*p*(Cl)Phe-*H*	2	0.94
V3	Bz-Nle-Lys-*N*-MeArg-Arg-*H*	3	1.21

[a] The position of R_1–R_4-groups and correspondence to P_1–P_4 residues see Fig. 1 where the wedged bonds indicate the attachment points of amino acid side chains

[b] -*H* means aldehyde warhead at P_1 C-terminus (Fig. 1) [20]

[c] Q_{tot}—Total molecular charge at neutral pH was determined according to the pK_a constants of amino acid side chains [59]

[d] K_i^{exp}—Experimental inhibition constants of the NS2B-NS3pro complex were obtained from Ref. [21]

[e] The notation of unusual amino acids see Fig. 2

[f] Ratio of predicted pK_i^{pre} and experimental pK_i^{exp} inhibition constants ($pK_i^{exp} = -\log_{10}K_i^{exp}$) of the validation set. pK_i^{pre} were computed from the QSAR Eq. 1

originates about equally from P_2 and P_1 residues, while in doubly charged analogs ($Q_{tot} = 2$ ė) and neutral R_1-groups the strongest contribution is donated by the P_2 residue.

Combinatorial subset selection

Over 250 analogs displayed activities predicted from the LUDI score lower than 15 μM. Relative frequency of occurrence of the individual R-groups was monitored in this group of the most potent analogs (Fig. 13). From Fig. 13 a clear preference for basic (Arg) and large aromatic (*p*(Ph)Phe) side chains in P_1, guanidine group containing building blocks (*m*(Gn)Phe and Arg) in P_2 and homo histidine in P_3 residue, is evident. The fragments that

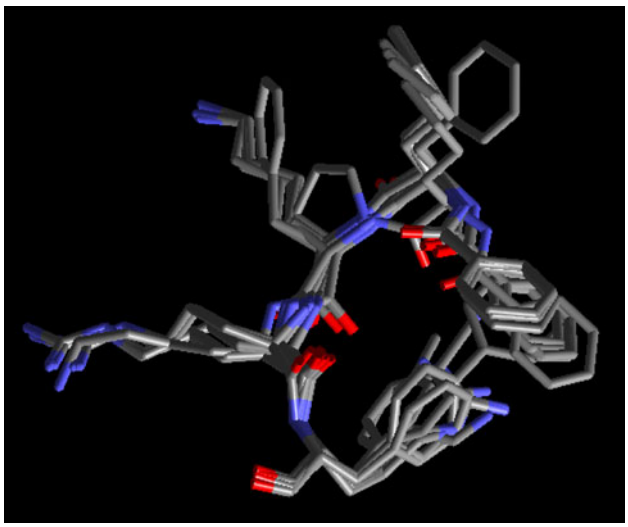


Fig. 6 Bound conformations of NS3pro inhibitors of the training set. Overlap of NS3pro inhibitors of the training set [20, 21] shown in *docked conformation* at the binding site of the NS2B-NS3pro target. Hydrogen atoms are omitted for a better clarity. Carbon atoms are shown in *grey color*, oxygens are *red*, and nitrogens *blue*

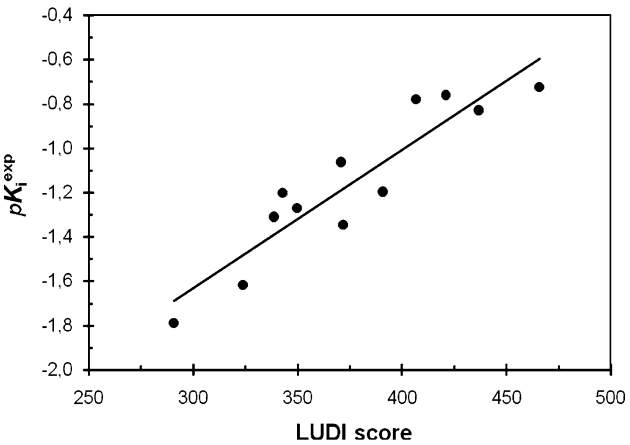


Fig. 7 Plot of QSAR regression equation. Regression equation derived for the training set of 12 NS3pro inhibitors [20, 21] correlates computed LUDI score with observed inhibition constants pK_i^{exp}. This equation was used as the scoring function specific for the NS2B-NS3pro target of the serological type 2 of Dengue virus

displayed the highest frequency of occurrence were selected to constitute a highly focused combinatorial subset of the initial diversity library Fig. 14). This subset displays greatly increased probability to contain potent D1 analogs with inhibition activities towards the DEN2 NS3pro in the nanomolar range. The size of the resulting combinatorial subset was narrowed down to only: 8 (R_1) × 4 (R_2) × 5 (R_3) × 1 (R_4) = 160 analogs, which permits its rapid synthesis and testing for the inhibitory activity.

ADME-related properties and hit prioritization

Numerous drugs at a late stage of pharmaceutical development and many lead compounds fail due to unsatisfactory pharmacokinetic properties. It is therefore essential to incorporate ADME properties prediction into the lead compound selection.

We have computed a number of physicochemical molecular properties (descriptors) relevant to the pharmacokinetic profile of compounds that can be used for prioritization of virtual hits selection for further development. A set of 13 ADME-related properties out of total of 36 descriptors calculated by the QikProp program [50] is given in Table 4. We have used an overall score—drug likeness parameter (#stars, Table 4), to assess the ADME profiles of the hits. The #stars parameter indicates the number of descriptors computed by QikProp that fall outside of the optimum range of values fulfilled by 95% of known drugs and can thus serve as an overall indicator of favorable ADME properties.

Ten out of the 16 best virtual hits with the highest predicted NS3pro inhibitory potencies are forecast to comply with the pharmacokinetic profiles of drug-like compounds equally or better than the template inhibitor D1 (Table 4). Therefore, we can expect that some of the virtual hits could be further developed into lead compounds. Although charged peptides presumably do not penetrate well cell membranes, the considered peptidomimetic NS3pro inhibitors may eventually have to be administered intravenously, which may still be an acceptable option, especially in cases of potentially lethal dengue hemorrhagic fever and dengue shock syndrome.

Conclusions

The presented study yielded a small highly focused virtual combinatorial subset of new analogs of the template tetrapeptide D1, selective inhibitor of *flavivirus* serine protease, which contains peptidomimetics composed of natural and mainly unusual basic amino acids, with increased predicted inhibitory potencies against the NS3pro of the Dengue virus (serotype 2). The predicted potency of the analog DA1 from this combinatorial subset falls into nanomolar range and is about 50-times lower than the K_i^{exp} value of the template inhibitor D1, the most potent one from the training set. The most promising analogs DA1, DA8 and DA12 are predicted to bind to the NS3pro with an improved affinity compared to the template tetrapeptide D1. The virtual hits are endowed with ADME-related properties comparable to those of the training set inhibitors. The proposed set of P_1 to P_4 residues optimized by structure-based focusing and virtual screening

Table 2 Virtual hits, peptidomimetic inhibitors of NS2B-NS3pro derived by structure-based focusing and virtual screening

Virtual hit	Sequence[a] Bz-P_4-P_3-P_2-P_1-*H*[b]	Q_{tot}[c] [ė]	LUDI[d]	K_i^{pre}[e] [μM]
DA1	Bz-Nle-His-*m*(Gn)Phe-*p*(Gn)Phe-*H*[f]	3	714.0	0.1
DA2	Bz-Nle-*p*(Am)Phe-*m*(Gn)Phe-*p*(Am)Phe-*H*	3	663.0	0.2
DA3	Bz-Nle-hHis-*m*(Gn)Phe-*p*(Gn)Phe-*H*	3	605.0	0.5
DA4	Bz-Nle-His-*m*(Gn)Phe-2Nal-*H*	2	573.0	0.8
DA5	Bz-Nle-His-*p*(Gn)Phe-Trp-*H*	2	572.0	0.9
DA6	Bz-Nle-*m*(Am)Phe-*m*(Im)Phe-(dMo)Phe-*H*	2	541.0	1.3
DA7	Bz-Nle-hHis-*p*(Gn)Phe-*p*(Ph)Phe-*H*	2	539.0	1.4
DA8	Bz-Nle-Arg-*m*(Gn)Phe-*p*(Am)Phe-*H*	3	535.0	1.5
DA9	Bz-Nle-Lys-Arg-Trp-*H*	2	531.0	1.6
DA10	Bz-Nle-His-*m*(Im)Phe-Arg-*H*	3	526.0	1.7
DA11	Bz-Nle-*p*(Am)Phe-*m*(Im)Phe-Arg-*H*	3	524.0	1.7
DA12	Bz-Nle-His-*p*(Gn)Phe-*p*(Gn)Phe-*H*	3	521.0	1.8
DA13	Bz-Nle-*p*(Am)Phe-*p*(Gn)Phe-Trp-*H*	2	521.0	1.8
DA14	Bz-Nle-*m*(Gn)Phe-*p*(Gn)Phe-*p*(Gn)Phe-*H*	3	520.0	1.8
DA15	Bz-Nle-hHis-*m*(Gn)Phe-Arg-*H*	3	517.0	1.9
DA16	Bz-Nle-*p*(Am)Phe-*m*(Gn)Phe-Arg-*H*	3	514.0	2.0

[a] The position of R_1–R_4-groups and correspondence to P_1–P_4 residues see Fig. 1, where the wedged bonds indicate the attachment points of amino acid side chains

[b] -*H* means aldehyde warhead at P_1 C-terminus (Fig. 1) [20]

[c] Q_{tot}—Total molecular charge at neutral pH was determined according to the pK_a constants of amino acid side chains [59]

[d] LUDI—Computed LUDI score (energy estimate type 3) [43, 44] computed using Cerius^2 program [39]

[e] K_i^{pre}—Predicted inhibition constants of virtual hits towards the DEN2 NS2B-NS3pro using the target-specific scoring function, QSAR Eq. 1: $K_i^{pre} = 10^{-(-3.50364+0.00624\ LUDI)}$, in μM. The experimental inhibition constants of the template inhibitor D1 towards the NS2B-NS3pro K_i^{exp} is equal to 5.8 μM [21]

[f] The notation of unusual amino acids see Fig. 2. For chemical structures of the virtual hits see Fig. 10

methods may represent the best choice of reagents independent of the type of electrophilic warhead used.

Nevertheless, the study can help to direct the interest of medicinal chemists working on the preparation of peptidomimetic anti-dengue chemotherapeutics to this particular subset of the chemical space, which is predicted to contain tetrapeptide aldehydes with elevated inhibitory potencies towards the Dengue virus serine protease.

Naturally, the predicted activities derived by this computational approach need to be verified by synthesis and inhibitory activity testing. Such activities are planned, and the results will be reported in due time.

Methods

Virtual library generation

The building of the library of peptidomimetic analogs was accomplished in the Cerius^2 program [39] using the class II consistent force field CFF91 [51] and Rappé and Goddard equilibrated charges [52]. The library of analogs was enumerated by attaching the R-groups (amino acid side chains) onto the benzoyl- and aldehyde-capped tetrapeptide backbone scaffold (Fig. 1) kept in its bound conformation [24] using the CombiChem module of Cerius^2 [39]. Natural and unusual amino acids considered in this study were taken from the directories of amino acids available from the commercial suppliers of chemicals [53]. Each peptidomimetic analog was built as a molecular ion containing neutral or protonated side chain in case of basic residues. Molecular geometry of the side chains was refined by molecular mechanics (MM) energy minimization by employing smart minimizer algorithm of the Cerius^2 [39] 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 2.

NS2B-NS3pro target

To prepare a 3D model of the Dengue virus serotype 2 NS2B-NS3pro-D1 complex as drug target we have first completed sequence alignment of the of NS3pro of Dengue virus (2FP7) [24] and West Nile virus, for which a

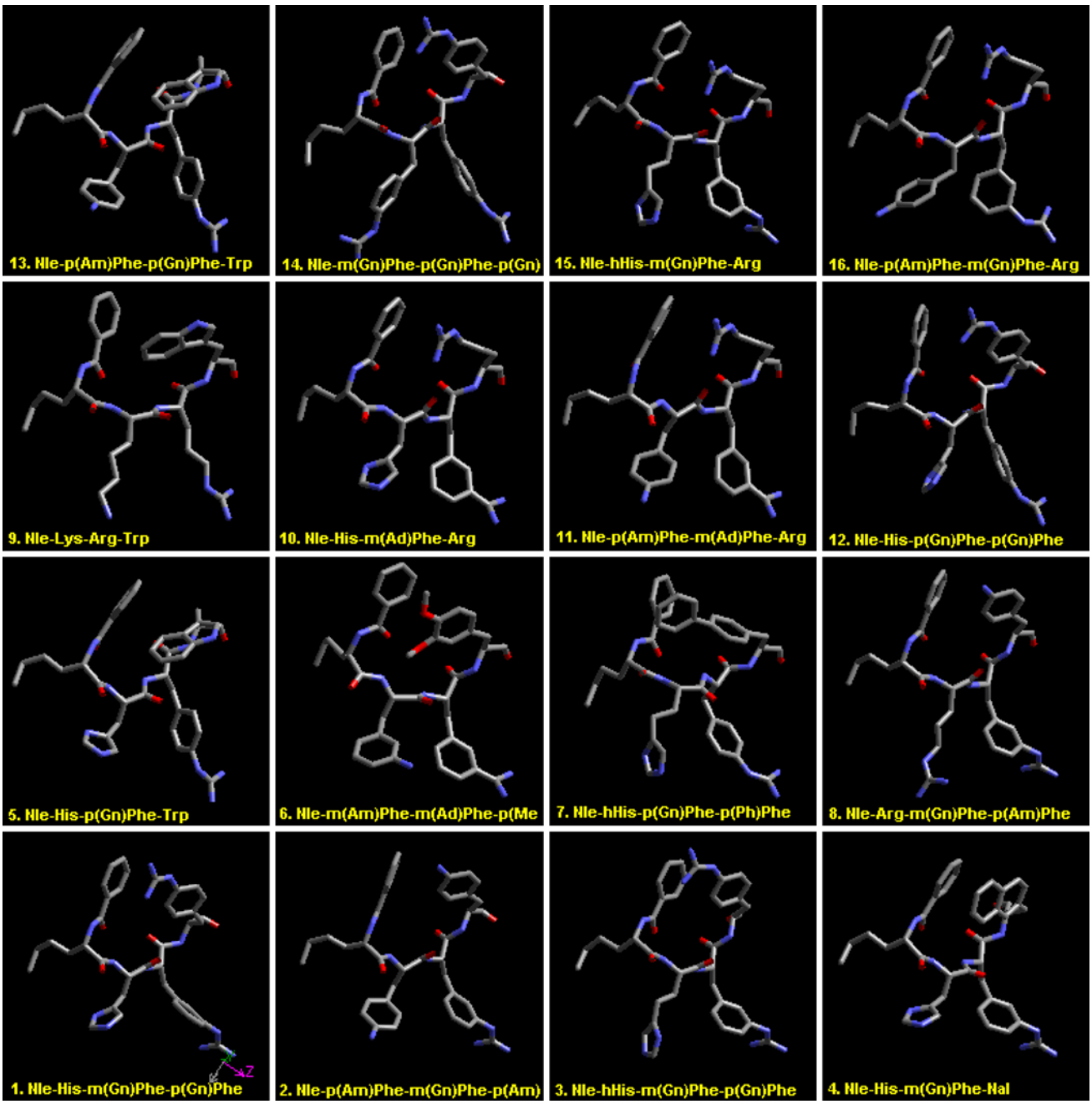


Fig. 8 3D-structures of 16 best virtual hits rank-ordered by scoring function. 3D models in stick rendering are shown without hydrogen atoms for better clarity. The atom coloring scheme: *gray*—carbon, *blue*—nitrogen, and *red*—oxygen

complete crystal structure of the protease-cofactor-inhibitor complex is known (2FOM) [24]. ClustalW2 sequence alignment carried out with help of the EMBL-EBI server, using its default settings (gap open penalty of 10.0, gap extension penalty 0.2, and Gonnet 250 substitution scoring matrix) [54], displayed an overall alignment score of 584, sequence identity of 45% and sequence similarity of 80% over 170 aligned NS3pro residues of the two flaviviruses (Fig. 3). A high degree of sequence similarity permitted to accomplish automatic structural alignment over structurally conserved regions of the NS3pro of Dengue and West Nile viruses (PDB entry codes 2FOM and 2FP7) [24] by means of Homology module of Insight-II molecular modeling package [55]. A very good overlap of the backbone atoms of the enzymes including the protease active site (r.m.s.d. of 0.865 Å over backbone atoms measured in 6

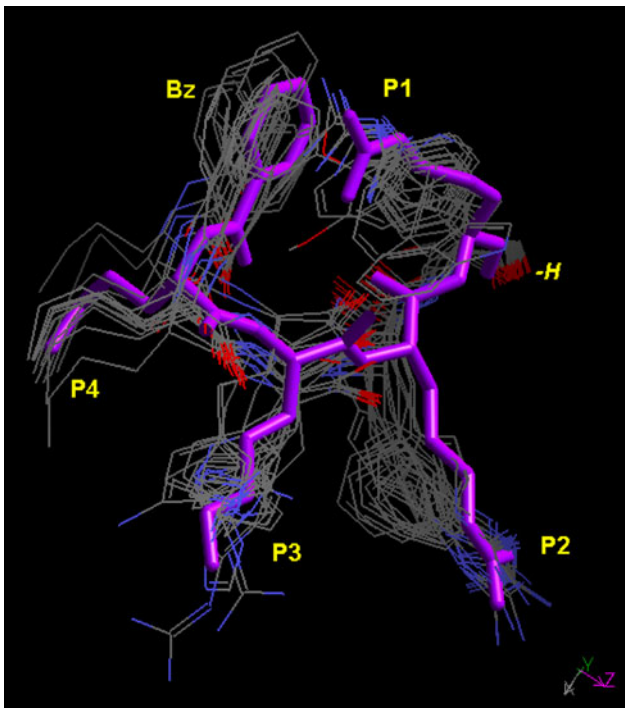


Fig. 9 Superposition of 3D structures of best virtual hits and crystal structure of NS3pro inhibitor D1. Heavy atoms of bound conformation of 16 best virtual hits shown in line representation are superimposed over the template inhibitor D1 displayed in stick rendering in *purple color*. Positions of the side chains P_1–P_4, benzoyl cap (Bz), and aldehyde warhead (*-H*) are indicated

blocks containing 66.7% of residues in 128 pairs) allowed to place the D1 inhibitor Bz-Nle-Lys-Arg-Arg-*H* present in the West Nile virus NS2B-NS3pro-D1 complex into the active site of the Dengue virus NS2B-NS3pro target assuming the same binding mode of the inhibitor in both similar proteases (Fig. 4). Then the DEN2 NS2B-NS3pro-D1 complex was carefully refined by MM energy-minimization using all-atom representation and CFF91 force field [51]. A non-bonding interaction distance cut-off set to 15 Å was employed and dielectric constant of 2 was used to take into account dielectric shielding effects in proteins. Minimization of the complex was carried out by relaxing the structure gradually, starting with hydrogen atoms, then proceeding with all inhibitor atoms, followed by protease and cofactor residue side chains and concluding the relaxation with freeing of all atoms including the enzyme and cofactor backbone. In the geometry optimization, a sufficient number of conjugate gradient iterative cycles, with a convergence criterion set to the average gradient of 0.01 kcal mol^{-1} $Å^{-1}$, has been used.

Structure-based library focusing

The 3D structure of the DEN2 NS2B-NS3pro-D1 complex was used as the target receptor for docking of virtual library of peptidomimetic analogs. The binding site of NS2B-NS3pro receptor was mapped onto a 3D energy grid of the size (114 × 94 × 114) points with a resolution of 0.25 Å per unit after augmenting it by 5 Å in all directions and subsequently enumerated with a non-bonding interaction cut-off distance of 15 Å using the LigFit algorithm of the Cerius^2 [40]. The shape and size of the ligand binding site was defined from the bound inhibitor D1. The binding site model was enlarged in areas selected around the P_1–P_4 side chain headgroups by one grid layer. Subsequently, the analogs were docked into the binding site model as flexible molecules by generating conformers of each analog via Monte Carlo randomization of the dihedral angles (10^4 steps). The generated conformers were fitted to the site model via flexible fit algorithm by comparing principal moments of inertia of the site and the analog after 100 rigid body minimization steps of 4 ligand orientations and 1,000 steps of final minimization [40]. Subsequently, the ligands were energy minimized at binding site of the NS2B-NS3pro with a maximum number of iterations set to 3,000 steps. The docking score (ligand–receptor interaction energy) was computed as non-bonding MM energy term using the CFF91 force field [51] for each generated conformer. An energy penalty of 100 kcal mol^{-1} was applied to each atom placed outside of the grid to favor the realistic ligand poses. During the docking score calculation a grid representation of rigid NS2B-NS3pro receptor, a cut-off distance of 15 Å applied to the non-bonded interactions and a dielectric constant of 2, were used. 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 [41]. The best representative of each cluster was considered in the virtual screening of the analogs.

QSAR model

A training set of 12 substrate-based peptidomimetic inhibitors containing the aldehyde warhead D1–D12 prepared by Yin et al. [20, 21] were evaluated for inhibitory potency (K_i^{exp}) in an enzyme inhibition assay against viral NS3 protease DEN2 CF40 · NS3pro, a truncated Dengue 2 NS3 enzyme fused via flexible linker to a 47 amino acid region of NS2B [21]. Molecular models of the inhibitors were docked to the binding site grid of the NS2B-NS3pro receptor model using the LigFit docking procedure [40] of Cerius^2 [39]. Various scores implemented in the Cerius^2 (such as LigScore, LUDI, PMF, and PLP [40, 43–47]) were computed for the clustered conformations of the inhibitors and several QSAR models, which relate the K_i^{exp} to the computed scores, were prepared by linear regression analysis encoded in the QSAR module of the Cerius^2 [39]. The regression equation $pK_i = f(\text{LUDI})$, which correlates

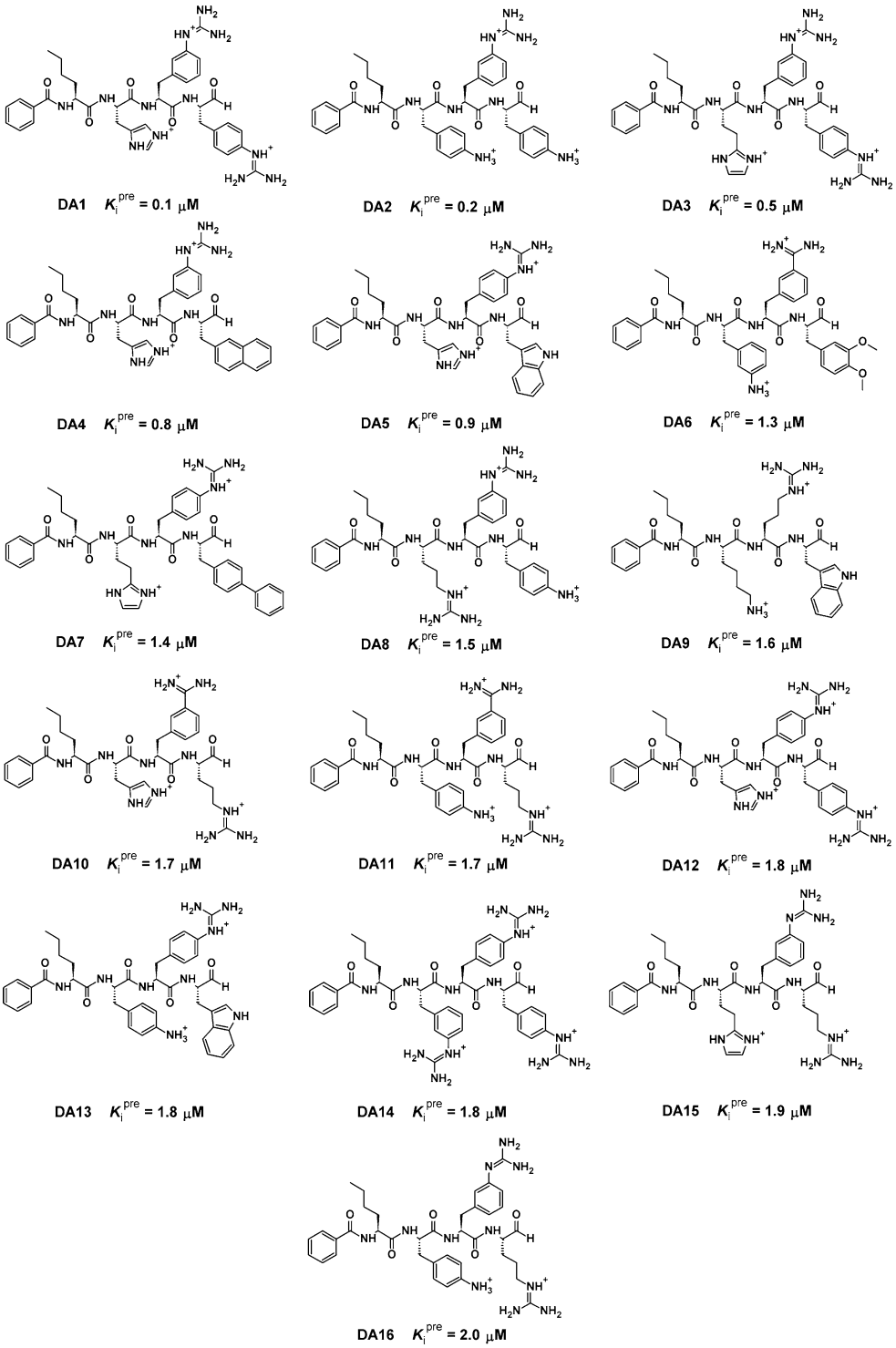


Fig. 10 Chemical structures of 16 best virtual hits with predicted inhibition constants towards the DEN2 NS3pro. Inhibition constants K_i^{pre} of DA1–DA16 were predicted from the regression Eq. 1

the computed LUDI score (energy estimate type 3) with the experimental inhibitory constants of the DEN2 NS3pro [20, 21], displayed the highest statistical significance. The LUDI score is defined as a sum of five ligand–receptor interaction contributions deriving from ideal hydrogen bonds, perturbed ionic interactions, lipophilic protein–ligand contact surface, contributions due to the freezing of internal degrees of freedom and a constant term representing the loss of translational and rotational entropy of the ligand [43, 44]. The statistical accuracy of this QSAR

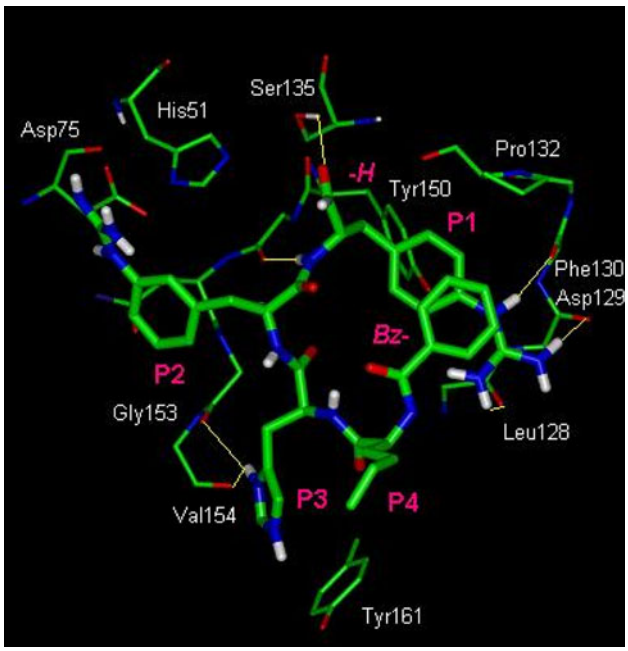


Fig. 11 Interactions of the best virtual hit DA1 with active site residues of NS3pro. Heavy atoms and polar hydrogens of DA1 are shown in stick representation bound to the active site of NS2B-NS3pro. Coloring scheme: carbon—*green*, oxygen—*red*, nitrogen—*blue*, hydrogen—*white*. Only selected residues of the NS3pro active site are shown for better clarity. Hydrogen atoms of NS3pro are omitted. Hydrogen bonds between DA1 and residues of NS3pro are displayed as *yellow lines*

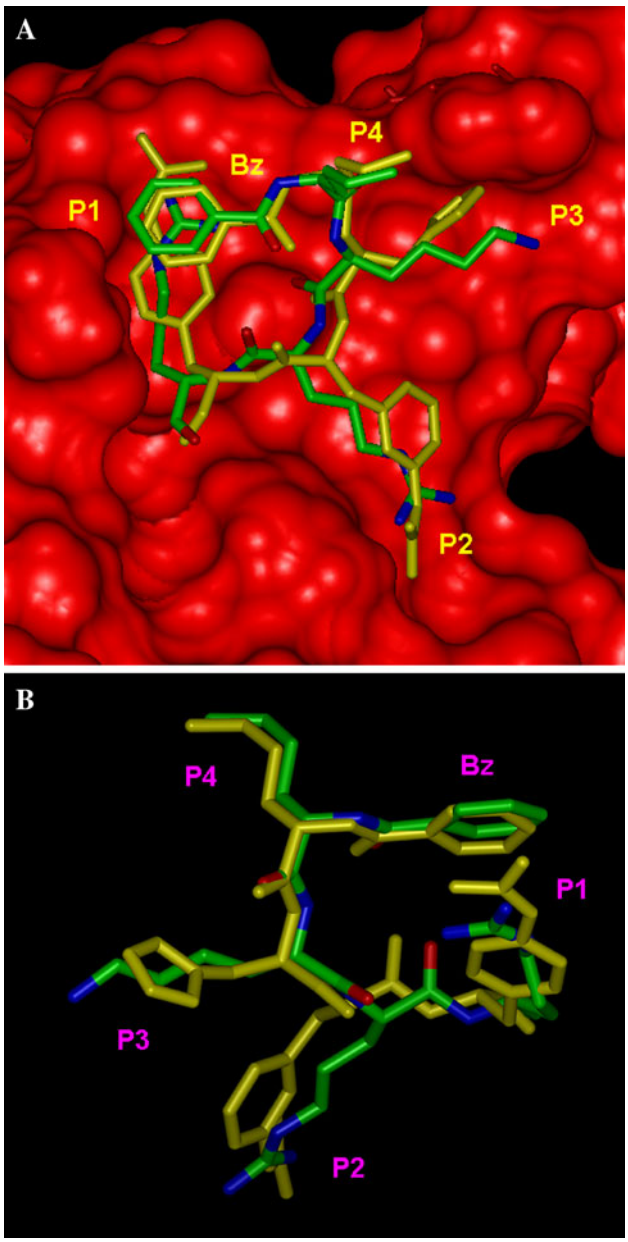


Fig. 12 3D structure of best virtual hit DA1. **a** Heavy atoms of virtual hit DA1: Bz-Nle-His-*m*(Gn)Phe-*p*(Gn)Phe-*H* (*yellow*) in bound conformation are shown in stick representation superimposed with D1 template inhibitor (carbons—*green*, nitrogens—*blue*, oxygens—*red*) over the molecular surface of the catalytic site of DEN2 NS3pro (*red*). Hydrogen atoms are omitted for better clarity. **b** Side view of superimposed DA1 and D1

model was verified by leave-one-out cross validation method. The predictive power of the Eq. 1, which was then used as the target-specific scoring function for in silico screening of the designed virtual library of peptidomimetics, was also verified by applying it to a validation set of 3 similar NS3pro inhibitors V1–V3 with known K_i^{exp} values, which were not included into the training set. The ratio of predicted activities K_i^{pre} obtained from the regression Eq. 1 and observed K_i^{exp} was used to evaluate the performance of the QSAR model.

In silico screening

The best member in each of the 10 clusters obtained from the ligand docking was selected for virtual screening by means of the LUDI scoring function (energy estimate type 3) [43, 44]. The LUDI function was selected since it performed best in the QSAR model of the NS3pro inhibitory potencies of the training and validation sets of inhibitors [20, 21]. The LUDI score was used for prediction of NS3pro inhibitory potencies (K_i^{pre}) of the structure-based focused virtual library by employing this parameter as a target-specific scoring function. The scoring function, specific for the NS3pro of the Dengue virus serotype 2: pK_i^{pre}[DEN2] = a + b LUDI, was parameterized using the QSAR model described above.

Receptor flexibility and inhibitor–enzyme interaction energy partitioning

The effect of the receptor flexibility was introduced partially by full relaxation of the NS2B-NS3pro-inhibitor

Table 3 Partitioning of NS2B-NS3pro-inhibitor interaction energy into contributions from individual residues (R-groups) for the 16 best virtual hits

Virtual hit	Q_{tot}[a] [ė]	E_{int}[b] [kcal mol^{-1}]					
		Bz	P_4 R_4-group	P_3 R_3-group	P_2 R_2-group	P_1-*H* R_1-group	Bz-P_4-P_3-P_2-P_1-*H* Total
DA1[c]	3	−4.4	−3.2	−62.1	−88.0	−95.2	−252.9
DA2	3	−7.3	−4.0	−52.0	−87.6	−97.4	−248.3
DA3	3	−4.5	−2.5	−60.4	−88.5	−94.5	−250.3
DA4	2	−5.1	−2.5	−65.7	−88.1	−24.6	−186.1
DA5	2	−6.8	−10.0	−49.7	−86.4	−28.9	−181.6
DA6	2	−4.9	−2.1	−62.0	−91.3	−31.2	−191.5
DA7	2	−2.8	−2.4	−58.3	−96.2	−25.3	−185.0
DA8	3	−7.1	−4.4	−58.6	−94.6	−101.4	−266.1
DA9	2	−2.5	−3.0	−54.0	−90.6	−23.7	−173.8
DA10	3	−6.3	−3.2	−57.9	−92.8	−88.6	−248.8
DA11	3	−5.0	−4.1	−53.1	−94.6	−87.4	−244.1
DA12	3	−3.9	−3.3	−58.2	−93.7	−101.6	−255.5
DA13	2	−8.1	−9.0	−47.5	−86.3	−27.0	−178.0
DA14	3	−4.0	−3.1	−49.9	−94.4	−96.0	−247.4
DA15	3	−5.1	−1.7	−60.6	−93.2	−89.5	−250.1
DA16	3	−5.1	−4.1	−52.6	−92.5	−87.0	−241.3
D1[d]	3	−4.3	−3.1	−56.0	−94.4	−87.3	−245.1

[a] Q_{tot}—Total molecular charge of analog at neutral pH. Total molecular charge of the NS2B-NS3pro receptor is equal to −5 ė

[b] E_{int}—Enzyme-inhibitor non-bonding interaction energy (1-6-9 potential) as defined in CFF91 force field [51]

[c] Chemical structures of the virtual hits see Fig. 10

[d] Reference inhibitor (Fig. 1) [20, 21]

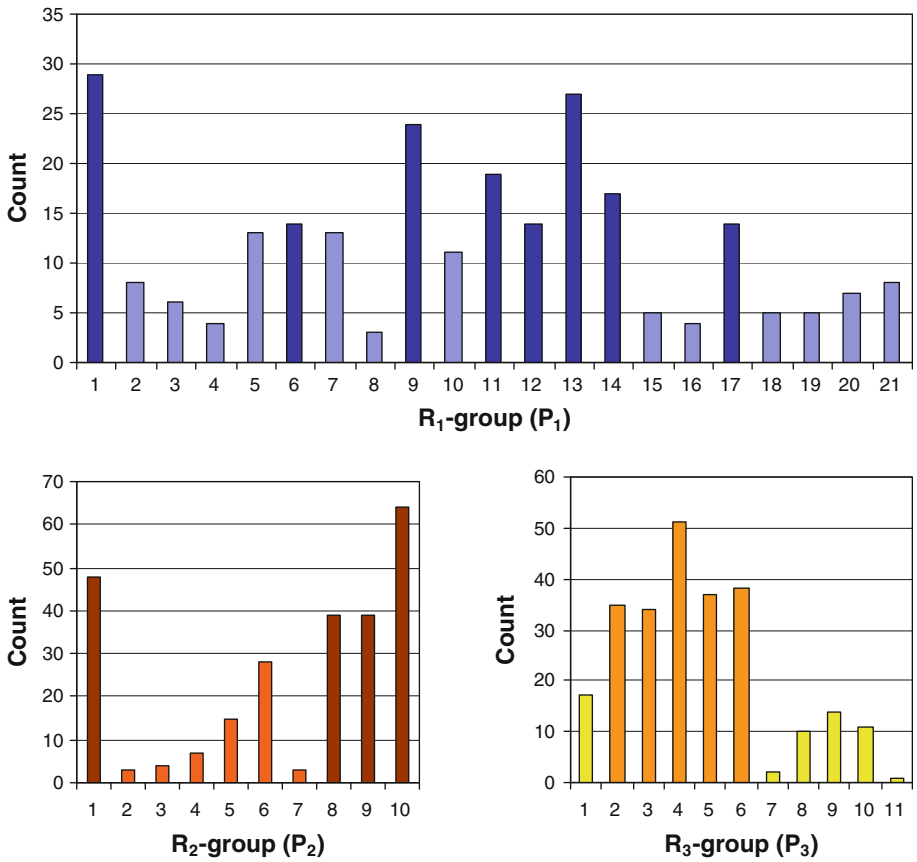


Fig. 13 Frequency of occurrence of R-groups in 250 best D1 analogs. Count of R_1 to R_3-groups (fragments) in the best 250 analogs for which K_i^{pre} towards DEN2 NS3pro ≤15 μM (for the numbering of fragments see Fig. 2). R_4-group contains -Nle- as the optimal fragment. The most frequently occurring fragments were selected to form a combinatorial subset of the size: 8 (R_1) × 4 (R_2) × 5 (R_3) × 1 (R_4) = 160 analogs

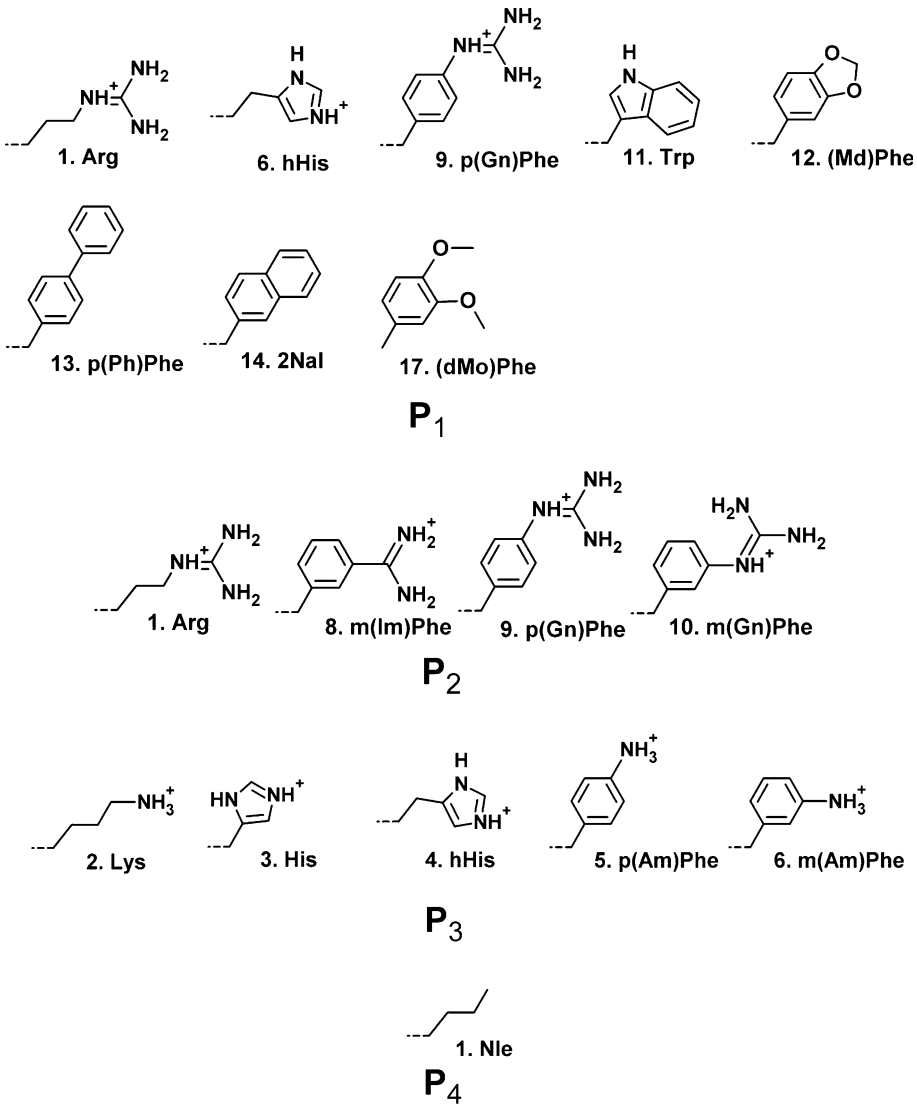


Fig. 14 Selected R-groups forming highly focused combinatorial subset. R_1 to R_3-groups were selected based on the frequency of occurrence of fragments in the 250 best D1 analogs (Fig. 13). R_4-group is represented by the best fragment from the training set—Nle

complexes of 16 best scoring virtual hits obtained by semi-flexible docking from the LigFit algorithm of Cerius^2 [39]. The inhibitor–enzyme interaction energy decomposed to contributions of P_1–P_4 residue was computed for fully minimized NS2B-NS3pro-DA complexes containing 16 best virtual hits (DA1–DA16) at active site. Interaction energy was computed as the 1-6-9 non-bonding pairwise interatomic interaction potential that includes coulombic, dispersion and repulsion terms, as defined in the CFF91 force field [51] using the Insight-II modeling package [55]. Original net atomic charges, van der Waals and repulsion atomic parameters of CFF91 were used together with a dielectric constant of 2 to account partially for the dielectric shielding effect within the target protein.

ADME-related properties

Prediction of descriptors related to adsorption, distribution, metabolism and excretion (ADME) of the analogs was carried out by QikProp program [50] based on the method of Jorgensen [56–58]. The QikProp program computes properties such as octanol/water partitioning coefficient, aqueous solubility, brain/blood partition coefficient, Caco-2 cell permeability, serum protein binding, number of likely metabolic reactions, and others. Drug likeness (#stars)—the number of property descriptors from the full list of 36 descriptors computed by the program, that fall outside the optimum range of values determined for 95% of known drugs, was used as an additional compound selection filter.

Table 4 Predicted ADME-related properties of the 16 best virtual hits computed by QikProp program [50] used to prioritize the hit selection

Virtual hits[a]	#stars[b]	M_w[c] [g mol^{-1}]	S_{mol}[d] [Å^2]	$S_{mol,hfob}$[e] [Å^2]	V_{mol}[f] [Å^3]	RotB[g]	HB_{don}[h]	HB_{acc}[i]	log $P_{o/w}$[j]	log S_{wat}[k] [g dm^{-3}]	log K_{HSA}[l]	log B/B[m]	BIP_{caco}[n] [nm s^{-1}]	#metab[o]	pK_i^{pre}[p] [μM]
DA1[q]	10	765	1,104	288	2,251	22	10	15	2.0	−2.9	−0.6	−5.8	0.04	8	0.1
DA2	11	748	1,150	285	2,319	22	9	13	3.5	−4.2	−0.1	−5.7	0.06	9	0.2
DA3	10	779	1,101	304	2,262	23	10	15	2.0	−2.7	−0.6	−6.0	0.03	7	0.5
DA4	8	758	1,056	286	2,193	20	6	13	4.0	−3.7	0.1	−4.3	1.83	7	0.8
DA5	9	747	1,030	266	2,156	20	7	13	3.4	−3.1	−0.1	−4.4	1.09	7	0.9
DA6	10	778	1,150	393	2,336	24	6	14	4.3	−3.0	0.1	−5.0	0.11	11	1.3
DA7	9	798	1,106	287	2,332	22	6	13	4.8	−4.2	0.4	−4.6	1.52	8	1.4
DA8	12	742	1,114	338	2,228	24	11	14	1.4	−1.3	−0.8	−6.8	0.00	7	1.5
DA9	8	690	1,090	428	2,166	24	8	12	2.4	−0.6	−0.3	−5.4	0.00	6	1.6
DA10	9	702	1,014	343	2,074	24	9	14	1.6	1.2	−0.9	−5.0	0.01	6	1.7
DA11	12	727	1,118	355	2,226	25	10	14	2.3	−0.1	−0.7	−6.1	0.00	7	1.7
DA12	14	765	1,137	293	2,277	22	10	15	2.1	−3.4	−0.5	−6.2	0.02	8	1.8
DA13	12	772	1,064	260	2,228	21	8	12	3.7	−3.4	0.0	−4.8	0.53	9	1.8
DA14	12	832	1,147	264	2,384	24	13	15	1.4	−2.5	−0.7	−7.2	0.00	8	1.8
DA15	10	731	1,086	363	2,189	24	10	15	1.6	−0.9	−0.8	−5.9	0.01	6	1.9
DA16	10	742	1,039	325	2,176	24	11	14	1.6	−0.1	−0.8	−5.5	0.02	7	2.0
D1[r]	10	660	1,012	485	2,045	27	11	14	−0.1	2.0	−1.3	−5.5	0.00	6	5.8[s]

[a] The chemical structures of the virtual hits see Fig. 10

[b] Drug likeness—number of property descriptors (from the full list of 36 descriptors of QikProp, ver. 2.0 [50]) that fall outside the range of values for 95% of known drugs are marked by *

[c] Molecular weight, in Da (range for 95% of drugs: 130–725 Da) [50]

[d] Total solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 300–1,000 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: 500–2,000 Å^3)

[g] Number of rotatable bonds (range for 95% of drugs: 0–15)

[h] Number of hydrogen bonds donated by the molecule (range for 95% of drugs: 0–6)

[i] Number of hydrogen bonds accepted by the molecule (range for 95% of drugs: 2–20)

[j] Logarithm of partitioning coefficient between *n*-octanol and water phases (range for 95% of drugs: −2 to 6)

[k] Logarithm of aqueous solubility (range for 95% of drugs: −6.0 to 0.5)

[l] Logarithm of predicted binding constant to human serum albumin (range for 95% of drugs: −1.5 to 1.2)

[m] Logarithm of predicted blood/brain barrier partition coefficient (range for 95% of drugs: −3.0 to 1.0)

[n] Predicted apparent Caco-2 cell membrane permeability in Boehringer-Ingelheim scale, in [nm s^{-1}] (range for 95% of drugs: <5 low, >100 high)

[o] Number of likely metabolic reactions (range for 95% of drugs: 0–15)

[p] Predicted pK_i^{pre} were estimated from the QSAR Eq. 1 and the computed LUDI score

[q] The QikProp calculation the neutral (nonprotonated) forms of DA1–DA16 were considered

[r] For comparison we list also the ADME-related properties of the reference inhibitor D1

[s] Experimental activity of D1 [21]

Acknowledgments This work has been done within the ICS-UNIDO global program on Rational Drug Design and Discovery. The institutional support to this work is gratefully acknowledged.

References

1. Rice CM (1996) In: Fields BN, Knipe DM, Howley PM (eds) Fields virology. Lippincott-Raven, Philadelphia, pp 931–959
2. Gosh D, Basu A (2008) Present perspectives on flaviviral chemotherapy. Drug Discov Today 13:619–624
3. Mackenzie JS, Gubler DJ, Petersen LR (2004) Emerging flaviviruses: the spread and resurgence of *Japanese encephalitis*, West Nile and dengue viruses. Nat Med 10(12 suppl):S98–S109
4. Dengue and Dengue Haemorrhagic Fever (2009) WHO fact sheet no. 117. World Health Organization, Geneva. Retrieved from http://http://www.who.int/mediacentre/factsheets/fs117/en/ 1/9/2009
5. Chambers TJ, Hahn CS, Galler R, Rice CM (1990) Flavivirus genome organization, expression, and replication. Annu Rev Microbiol 44:649–688
6. Falgout B, Pethel M, Zhang YM, Lai CJ (1991) Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J Virol 65:2467–2475
7. Chambers TJ, Weir RC, Grakoui A, McCourt DW, Bazan JF, Fletterick RJ, Rice CM (1990) Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc Natl Acad Sci USA 87:8898–8902
8. Preugschat F, Yao CW, Strauss JH (1990) In vitro processing of dengue virus type 2 nonstructural proteins NS2A, NS2B, and NS3. J Virol 64:4364–7434
9. Li H, Clum S, You S, Ebner KE, Padmanabhan R (1999) The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. J Virol 73:3108–3116
10. Murthy HM, Clum S, Padmanabhan R (1999) Dengue virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects. J Biol Chem 274:5573–5580
11. Clum S, Ebner KE, Padmanabhan R (1997) Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. J Biol Chem 272:30715–30723
12. Leung D, Schroder K, White H, Fang NX, Stoermer MJ, Abbenante G, Martin JL, Young PR, Fairlie DP (2001) Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. J Biol Chem 276:45762–45771
13. Zhang L, Mohan PM, Padmanabhan R (1992) Processing and localization of Dengue virus type 2 polyprotein precursor NS3-NS4A-NS4B-NS5. J Virol 66:7549–7554
14. Chanprapaph S, Saparpakorn P, Sangma C, Niyomrattanakit P, Hannongbua S, Angsuthanasombat C, Katzenmeier G (2005) Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites. Biochem Biophys Res Commun 330:1237–1246
15. Ganesh VK, Muller N, Judge K, Luan CH, Padmanabhan R, Murthy KH (2005) Identification and characterization of non-substrate based inhibitors of the essential dengue and West Nile virus proteases. Bioorg Med Chem 13:257–264
16. Ekonomiuk D, Su XC, Ozawa K, Bodenreider C, Lim SP, Yin Z, Keller TH, Beer D, Patel V, Otting G, Caflisch A, Huang D (2009) Discovery of a non-peptidic inhibitor of west nile virus NS3 protease by high-throughput docking. PLoS Negl Trop Dis 3:e356
17. Li J, Lim SP, Beer D, Patel V, Wen D, Tumanut C, Tully DC, Williams JA, Jiricek J, Priestle JP, Harris JL, Vasudevan SG (2005) Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries. J Biol Chem 280:28766–28774
18. Wahab HA, Yusof R, Rahman NA (2007) A search for vaccines and therapeutic for dengue: a review. Curr Comput Aided Drug Des 3:341–352
19. Othman R, Kiat TS, Khalid N, Yusof R, Newhouse EI, Newhouse JS, Alam M, Rahman NA (2008) Docking of noncompetitive inhibitors into dengue virus type 2 protease: understanding the interactions with allosteric binding sites. J Chem Inf Model 48:1582–1591
20. Yin Z, Patel SJ, Wang WL, Wang G, Chan WL, Rao KR, Alam J, Jeyaraj DA, Ngew X, Patel V, Beer D, Lim SP, Vasudevan SG, Keller TH (2006) Peptide inhibitors of Dengue virus NS3 protease. Part 1: warhead. Bioorg Med Chem Lett 16:36–39
21. Yin Z, Patel SJ, Wang WL, Chan WL, Ranga Rao KR, Wang G, Ngew X, Patel V, Beer D, Knox JE, Ma NL, Ehrhardt C, Lim SP, Vasudevan SG, Keller TH (2006) Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. Bioorg Med Chem Lett 16:40–43
22. Luo D, Xu T, Hunke C, Grüber G, Vasudevan SG, Lescar J (2008) Crystal structure of the NS3 protease-helicase from dengue virus. J Virol 82:173–183
23. Murthy HM, Judge K, DeLucas L, Padmanabhan R (2000) Crystal structure of Dengue virus NS3 protease in complex with a Bowman–Birk inhibitor: implications for flaviviral polyprotein processing and drug design. J Mol Biol 301:759–767
24. Erbel P, Schiering N, D'Arcy A, Renatus M, Kroemer M, Lim SP, Yin Z, Keller TH, Vasudevan SG, Hommel U (2006) Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. Nat Struct Mol Biol 13:372–373
25. Llinàs-Brunet M, Bailey M, Fazal G, Ghiro E, Gorys V, Goulet S, Halmos T, Maurice R, Poirier M, Poupart M-A, Rancourt Thibeault JD, Wernic D, Lamarre D (2000) Highly potent and selective peptide-based inhibitors of the hepatitis C virus serine protease: towards smaller inhibitors. Bioorg Med Chem Lett 10:2267–2270
26. Johansson A, Hubatsch I, Åkerblom E, Lindeberg G, Winiwarter S, Danielson UH, Hallberg A (2001) Inhibition of hepatitis C virus NS3 protease activity by product-based peptides is dependent on helicase domain. Bioorg Med Chem Lett 11:203–206
27. Perni RB, Britt SD, Court JC, Courtney LF, Deininger DD, Farmer LJ, Gates CA, Harbeson SL, Kim JL, Landro JA, Levin RB, Luong Y-P, O'Malley ET, Pitlik J, Rao BG, Schairer WC, Thomson JA, Tung RD, Van Drie JH, Wei Y (2003) Inhibitors of hepatitis C virus NS3.4A protease 1. Non-charged tetrapeptide variants. Bioorg Med Chem Lett 13:4059–4063
28. Perni RB, Pitlik J, Britt SD, Court JJ, Courtney LF, Deininger DD, Farmer LF, Gates CA, Harbeson SL, Levin RB, Lin C, Lin K, Moon Y-C, Luong Y-P, O'Malley ET, Rao BG, Thomson JA, Tung RD, Van Drie JH, Wei Y (2004) Inhibitors of hepatitis C virus NS3.4A protease 2. Warhead SAR and optimization. Bioorg Med Chem Lett 14:1441–1446
29. Frecer V, Kabelac M, De Nardi P, Pricl S, Miertus S (2004) Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus. J Mol Graph Model 22:209–220
30. Liverton NJ, Holloway MK, McCauley JA, Rudd MT, Butcher JW, Carroll SS, DiMuzio J, Fandozzi C, Gilbert KF, Mao S-S, McIntyre CJ, Nguyen KT, Romano JJ, Stahlhut M, Wan B-L, Olsen DB, Vacca JP (2008) Molecular modeling based approach to potent P2–P4 macrocyclic inhibitors of hepatitis C NS3/4A protease. J Am Chem Soc 130:4607–4609

31. Boloor A, Hanway D, Joshi M, Winn DT, Mendez G, Walls M, Wei P, Qian F, Zhang X, Zhang Y, Hepperle ME, Li X, Campbell DA, Betancort JM (2009) Synthesis and antiviral activity of HCV NS3/4A peptidomimetic boronic acid inhibitors. Bioorg Med Chem Lett 19:5708–5711
32. Liverton NJ, Carroll SS, Dimuzio J, Fandozzi C, Graham DJ, Hazuda D, Holloway MK, Ludmerer SW, McCauley JA, McIntyre CJ, Olsen DB, Rudd MT, Stahlhut M, Vacca JP (2010) MK-7009, a potent and selective inhibitor of hepatitis C virus NS3/4A protease. Antimicrob Agents Chemother 54:305–311
33. Frecer V, Burello E, Miertus S (2005) Combinatorial design of nonsymmetrical cyclic urea inhibitors of aspartic protease of HIV-1. Bioorg Med Chem 13:5492–5501
34. 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 27:376–387
35. Frecer V, Megnassan E, Miertus S (2009) Design and in silico screening of combinatorial library of antimalarial analogs of triclosan inhibiting *Plasmodium falciparum* enoyl-acyl carrier protein reductase. Eur J Med Chem 44:3009–3019
36. Rungrotmongkol T, Frecer V, De-Eknamkul W, Hannongbua S, Miertus S (2009) Design of oseltamivir analogs inhibiting neuraminidase of avian influenza virus H5N1. Antiviral Res 82:51–58
37. Bajusz S, Szell E, Bagdy D, Barabas E, Horvath G, Dioszegi M, Fittler Z, Szabo G, Juhasz A, Tomori E (1990) Highly active and selective anticoagulants: D-Phe-Pro-Arg-H, a free tripeptide aldehyde prone to spontaneous inactivation, and its stable *N*-methyl derivative, D-MePhe-Pro-Arg-H. J Med Chem 33: 1729–1735
38. Fehrentz J-A, Paris M, Heitz A, Velek J, Liu C-F, Winternitz F, Martinez J (1995) Improved solid phase synthesis of C-terminal peptide aldehydes. Tetrahedron Lett 36:7871–7874
39. Cerius^2 Life Sciences software, version 4.6 (2002) Accelrys, San Diego, CA
40. Peters KP, Fauck J, Frommel C (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. J Mol Biol 256:201–213
41. Jarvis RA, Patrick EA (1973) Clustering using a similarity measure based on shared near neighbors. IEEE Trans Comput C22:1025–1034
42. Seifert MH (2009) Targeted scoring functions for virtual screening. Drug Discov Today 14:562–569
43. Böhm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8:243–256
44. Böhm HJ (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. J Comput Aided Mol Des 12:309–323
45. Muegge I, Martin YC (1999) A general and fast scoring function for protein–ligand interactions: a simplified potential approach. J Med Chem 42:791–804
46. Muegge I, Martin YC, Hajduk PJ, Fesik SW (1999) Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. J Med Chem 42:2498–2503
47. Verkhivker GM, Appelt K, Freer ST, Villafranca JE (1995) Empirical free energy calculations of ligand–protein crystallographic complexes. I. Knowledge-based ligand–protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. Protein Eng 8:677–691
48. Stahl M (2000) In: Böhm HJ, Schneider G (eds) Virtual screening for bioactive molecules. Wiley-VCH, Weinheim, pp 229–264
49. Jain AN (2004) Virtual screening in lead discovery and optimization. Curr Opin Drug Discov Devel 7:396–403
50. QikProp, version 2.0.005 (2002) Schrödinger LLC, New York, NY
51. Maple JR, Hwang MJ, Stockfish TP, Dinur U, Waldman M, Ewing CS, Hagler AT (1994) Derivation of class II force fields. I. Methodology and quantum force field for the alkyl functional group and alkane molecules. J Comput Chem 15:162–182
52. Rappé AK, Goddard WA III (1991) Charge equilibration for molecular dynamics simulations. J Phys Chem 95:3358–3363
53. Available Chemicals Directory, version 3.0 (2008) Symyx Technologies Inc., Santa Clara, CA
54. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. Bioinformatics 23:2947–2948
55. Insight-II Life Sciences software, version 2005 (2005) Accelrys, San Diego, CA
56. Duffy EM, Jorgensen WL (2000) Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. J Am Chem Soc 122:2878–2888
57. Jorgensen WL, Duffy EM (2000) Prediction of drug solubility from Monte Carlo simulations. Bioorg Med Chem Lett 10:1155–1158
58. Jorgensen WL, Duffy EM (2002) Prediction of drug solubility from structure. Adv Drug Deliv Rev 54:355–366
59. Voet D, Voet JG (1995) Biochemistry, 2nd edn. Wiley, New York