J Comput Aided Mol Des (2011) 25:263–274
DOI 10.1007/s10822-011-9418-0

Structure-guided fragment-based *in silico* drug design of dengue protease inhibitors

Tim Knehans · Andreas Schüller · Danny N. Doan · Kassoum Nacro · Jeffrey Hill · Peter Güntert · M. S. Madhusudhan · Tanja Weil · Subhash G. Vasudevan

Received: 23 November 2010 / Accepted: 7 February 2011 / Published online: 23 February 2011

Abstract An *in silico* fragment-based drug design approach was devised and applied towards the identification of small molecule inhibitors of the dengue virus (DENV) NS2B-NS3 protease. Currently, no DENV protease co-crystal structure with bound inhibitor and fully formed substrate binding site is available. Therefore a homology model of DENV NS2B-NS3 protease was generated employing a multiple template spatial restraints method and used for structure-based design. A library of molecular fragments was derived from the ZINC screening database with help of the retrosynthetic combinatorial analysis procedure (RECAP). 150,000 molecular fragments were docked to the DENV protease homology model and the docking poses were rescored using a target-specific scoring function. High scoring fragments were assembled to small molecule candidates by an implicit linking cascade. The cascade included substructure searching and structural filters focusing on interactions with the S1 and S2 pockets of the protease. The chemical space adjacent to the promising candidates was further explored by neighborhood searching. A total of 23 compounds were tested experimentally and two compounds were discovered to inhibit dengue protease ($IC_{50} = 7.7$ μM and 37.9 μM, respectively) and the related West Nile virus protease ($IC_{50} = 6.3$ μM and 39.0 μM, respectively). This study demonstrates the successful application of a structure-guided fragment-based *in silico* drug design approach for dengue protease inhibitors providing straightforward hit generation using a combination of homology modeling, fragment docking, chemical similarity and structural filters.

Keywords Dengue virus · NS2B-NS3 protease · West Nile virus · Protease inhibitor · Fragment-based drug design (FBDD) · Homology modeling

Tim Knehans and Andreas Schüller have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9418-0) contains supplementary material, which is available to authorized users.

T. Knehans · P. Güntert
Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance and Frankfurt Institute for Advanced Studies, J.W. Goethe-University, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

T. Knehans · A. Schüller · D. N. Doan · S. G. Vasudevan (✉)
Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857, Singapore
e-mail: subhash.vasudevan@duke-nus.edu.sg

K. Nacro · J. Hill
Experimental Therapeutic Centre, Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, Nanos Level 3, Singapore 138669, Singapore

M. S. Madhusudhan
Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671, Singapore

T. Weil
Institute of Organic Chemistry III/Macromolecular Chemistry, University of Ulm, Albert-Einstein-Allee 11, 89069 Ulm, Germany

Introduction

The incidence of dengue has grown dramatically around the world in recent decades. Some 2.5 billion people—two fifths of the world's population—are now at risk from dengue [1]. The WHO currently estimates that 50 million dengue virus (DENV) infections occur worldwide every

year, including 500,000 cases of the more severe dengue hemorrhagic fever (DHF) and 22,000 deaths, mostly among children [2]. DENV belongs to the genus *Flavivirus* within the *Flaviviridae* family [3]. There are four closely related DENV serotypes that spread with its most common vector *Aedes aegypti,* an urban, diurnal mosquito species. So far there are no vaccines or an effective chemotherapy available against dengue.

The genome of DENV is organized in a ~ 11 kb single-strand plus-sense RNA which encodes three structural proteins (capsid protein C, membrane protein M, and envelope protein E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). The DENV genome is translated into a single polyprotein which is post-translationally processed via host proteases and the virus-encoded two-component protease NS2B-NS3pro. The NS3 protein contains an N-terminal protease domain (NS3pro, residues 1-179) and a C-terminal helicase domain (NS3hel). NS3pro is a classic trypsin-like serine endo-peptidase with a catalytic triad formed by His51, Asp75 and Ser135 [3]. The protease requires the hydrophilic part of the integral membrane protein NS2B (residues 49 to 95) for full enzymatic activity [4].

DENV protease shows high sequence (49% identity) and structural similarity (Table 1) to the protease from West Nile virus (WNV), a closely related flavivirus member. Several crystal structures of DENV/WNV NS2B-NS3pro have been determined either co-crystallized with an inhibitor (PDB ids: 2ijo [5]; 2fp7 [6]; 3e90 [7]; all WNV) or in absence of an inhibitor (PDB ids: 2ggv (WNV) [5]; 3l6p, 3lkw (DENV1) [8]; 2fom (DENV2) [6]; 2vbc (DENV4) [9]; 2whx (DENV4) [4]. According to these three-dimensional (3D) structures, the conformation of NS2B-NS3pro can assume an "open form" or a "closed form". In the closed conformation the NS2B peptide wraps around NS3pro like a belt and its C-terminal β-strand is associated with the C-terminal β-barrel of NS3pro forming part of the substrate binding site [3, 6]. In the open conformation, in contrast, this C-terminal β-strand of NS2B is dissociated from NS3 [3, 6]. A closed-form DENV protease structure has not yet been determined experimentally. Solution-phase NMR analyses indicated that the closed form is predominant in WNV regardless of the presence of an inhibitor while in DENV the open form is predominant when no inhibitor is present [10].

Table 1 RMSD matrix of WNV and DENV protease template structures used for homology modeling

	2fp7	3e90	2ijo	2fom	2vbc	Wichapong et. al
Model	0.330	0.388	0.380	0.820	0.650	0.350
2fp7 (WNV)		0.380	0.300	0.760	0.720	0.003
3e90 (WNV)			0.400	0.780	0.760	0.381
2ijo (WNV)				0.650	0.790	0.280
2fom (DENV2)					0.770	0.800
2vbc (DENV4)						0.754

Pair-wise root mean squared deviations (RMSD) of equivalent non-hydrogen atoms (in Å) of the generated homology model, the template structures (denoted by their PDB id), and a previously published homology model by Wichapong et al. [51]. The particularly low RMSD of Wichapong's homology model with the WNV structure 2fp7 [6] is likely due to their single template approach. In contrary, our model was based on five separate template structures

Drug discovery efforts have so far yielded several peptide and small molecules inhibitors with activities in the sub- to low micromolar ranges [11–26]. Tri- and tetra-peptide aldehyde inhibitors have been reported for WNV protease in the submicromolar range and for DENV protease in the low micromolar range [12–15]. The preferred amino acids in positions P1 and P2 (nomenclature of Schechter and Berger, 1967 [27]) are arginine and lysine indicating the role of electrostatics in the inhibitor-protease interaction. Only a few small molecule inhibitors have so far been identified utilizing both HTS and rational drug design approaches. These can be grouped into charged [17–19] and uncharged compounds [11, 16, 22, 24].

Here, we demonstrate the successful application of a structure-guided fragment-based *in silico* drug design approach of DENV2 protease inhibitors. Our approach provides straightforward hit generation using a combination of homology modeling, fragment docking, chemical similarity and structural filters. Ekonomiuk et al. reported the discovery of three cationic small molecule inhibitors for WNV protease by structure-guided fragment-based approaches [17, 18]. However, in the absence of any published closed-form DENV protease crystal structure and in contrast to related studies in the literature, the structure-based design of this study is based on a DENV2 NS2B-NS3pro homology model.

Materials and methods

Sequence alignment and homology modeling

All homology models were generated with the software MODELLER (version 9.7; Andrej Sali Laboratories, University of California, San Fransisco, CA, USA) [28]. The sequence alignment was performed in MODELLER with the SALIGN routine [29]. NS2B and NS3 template sequences were aligned separately and subsequently merged manually to yield the final NS2B-NS3pro sequence alignment (Fig. 1a, PDB ids: 2fom (DEN2) [6]; 2vbc (DEN4) [9]; 3e90 (WNV) [7]; 2fp7 (WNV) [6]; 2ijo (WNV) [5]). The model was based on the sequence of 2fom and spatial restraints were derived from all five

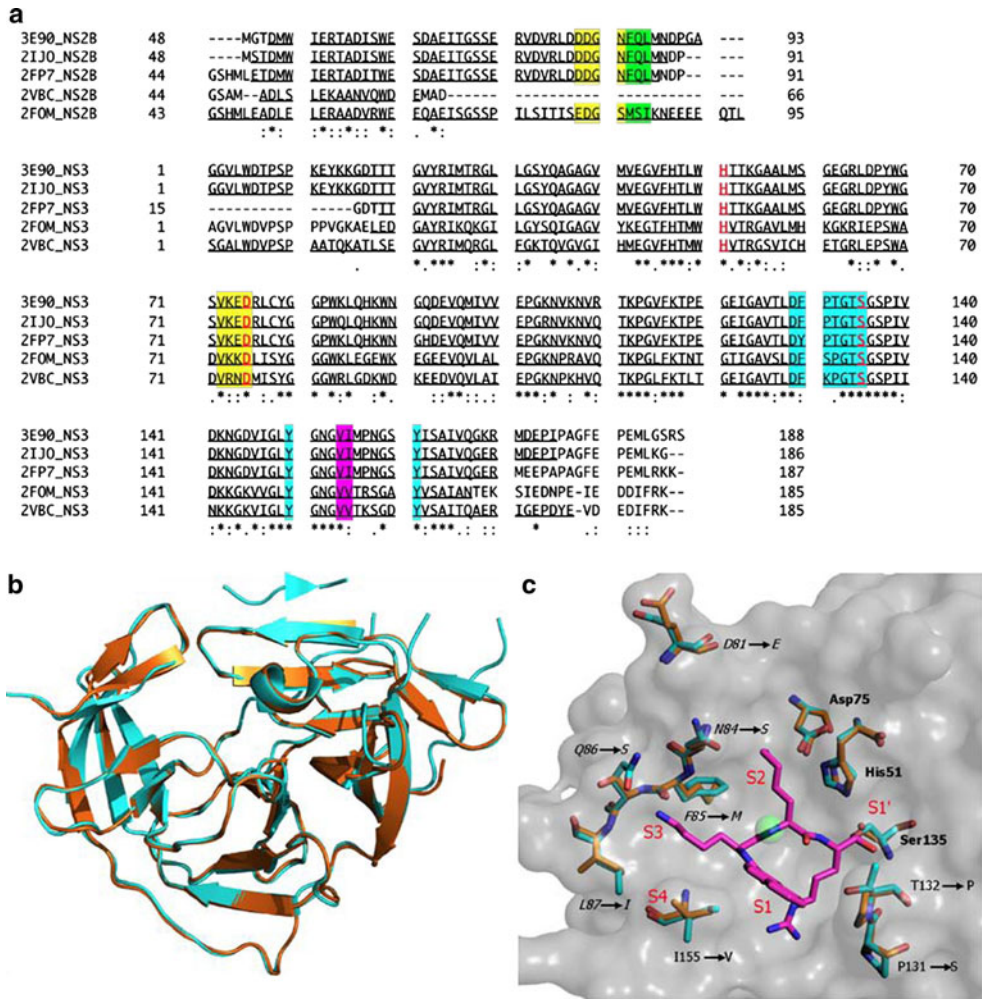

Fig. 1 **a** Multiple sequence alignment of template structures (PDB ids: 2ijo [5]; 2fp7 [6]; 3e90 [7]; all WNV, 2vbc (DEN4) [9]) and the DENV2 target sequence (PDB id: 2fom [6]) used for homology modeling. Underlined residues provide resolved 3D structural information from which restraints could be derived. Residues of the S1, S2, S3 and S4 pocket in cyan, yellow, green and magenta. Stars denote identical residues, dots indicate similarity. Catalytic triad in bold red color. **b** Structural alignment of WNV NS2B-NS3pro (PDB id: 3e90 [7]) in cyan and the DENV2 homology model in orange. **c** Superposition of the substrate binding site of the DENV2 homology model (vdW surface in gray, residues in orange) and the WNV template structure (PDB id: 3e90 [7]) in cyan. NS2B residues are in italics. Arrows indicate residue exchanges in going from the WNV 3e90 sequence to the DENV2 homology model. Residues in bold three-letter code constitute the catalytic triad. The subpockets are labeled S1, S1′, S2, S3 and S4 (in *red letters*). The peptide aldehyde inhibitor Naphtoyl-KKR-H is shown in magenta. A green sphere indicates the center of the binding site selected for docking. Figures generated with PyMol

templates. Based on preliminary analyses the first seven residues at the N-terminus and the last nine residues at the C-terminus of NS2B were removed to avoid unfavorable high energy scores. To conserve the oxyanion hole in the homology modeling, the following distance restraints were derived from WNV protease (PDB id: 3e90); d[N:Gly133:NS3 ↔ OAX:NKK(inhibitor)] = 2.8 ± 0.1 Å; d[N:Thr134:NS3 ↔ OAX:NKK(inhibitor)] = 3.2 ± 0.1 Å; d[OG:Ser135:NS3 ↔ CAW:NKK(inhibitor)] = 1.5 ± 0.1 Å; d[N:Ser135:NS3 ↔ OAX:NKK(inhibitor)] = 2.8 ± 0.1. Five models were generated in MODELLER by employing the variable target function method (VTFM) with conjugate gradients (CG) and were refined using molecular dynamics (MD) combined with simulated annealing (SA) optimizing the objective function until the molecular probability density function (mol pdf) exceeded a value of 10^6 [28].

Model evaluation

The five generated models were evaluated using the discrete optimized protein energy (DOPE) score, DOPE-per-residue score (DPRS) and by visual inspection. DOPE is a statistical potential employed here to assess the quality of a structure model as a whole [30]. The generated models were first ranked by their DOPE score for evaluation. Next,

calculation and comparison of DPRS profiles of the model and the template structures allowed for a more fine-grained evaluation. Finally, structural alignments of models and templates were performed using PyMol (version 0.99; DeLano Scientific, Palo Alto, CA, USA) [31] and structural variances were correlated with DPRS curves. The model showing the most similar DPRS curve to 3e90 was selected. Additionally, a Ramachandran map was generated with the Molecular Operating Environment (MOE, version 2009.10; Chemical Computing Group, Montreal, Canada) [32, 33]. This analysis revealed overall correct backbone φ and ψ angles with the exception of a single amino acid (Leu31) that was projected into the disallowed region of the map. As this amino acid is part of a flexible extended loop region (E1a-D1 loop [6]) located 21.1 Å away from the catalytic Ser135, this minor deviation from optimality is considered to be acceptable. A PDB file of the homology model is provided in Online Resource 2.

Fragment generation

Fragments were derived from the all-purchasable subset of the ZINC database version 9 containing ca. 14 million compounds [34]. The ZINC database compounds were loaded into MOE, hydrogens were added and strong acids or bases were deprotonated or protonated assuming pH 7. Next, the database was filtered by the Lipinski Rule-of-Five (Molecular weight (MW) < 500 Da, number of hydrogen bond (H-bond) donors ≤ 5, number of H-bond acceptors ≤ 10, ClogP ≤ 5) [35]. For fragment generation the retrosynthetic combinatorial analysis procedure (RECAP) as implemented in MOE was applied to all ZINC molecules [36]. Duplicate fragments were removed and a Rule-of-Three filter for fragment-based design (MW < 300, number of H-bond donors ≤ 3, number of H-bond acceptors ≤ 3, ClogP ≤ 3) was applied [37]. The final fragment library contained 149,151 unique molecular fragments with the number of rotatable bonds ranging from 0 to 13, the number of rings from 0 to 6 and within the confines of the filters described above.

Fragment docking

All fragments were exported in the PDB format, non-polar hydrogen atoms were merged and Gasteiger charges [38] were calculated with help of the software AutoDockTools (version 4; Molecular Graphics Laboratory, La Jolla, CA, USA) [39]. To make the homology model suitable for the docking exercise, the model was further refined using MOE as follows: Hydrogen atoms were added and the ionization state of His51 (catalytic triad) was set manually to neutral according to Markley and Westler [40]. The orientation of Asn152 which does not correspond to the reference orientation in either the DENV 2fom or WNV 3e90 template structure was corrected in MOE. Then, the homology model was subjected to a restrained energy minimization in MOE with the AMBER99 force field and a Born solvation term until the RMS gradient fell below 0.05 kcal/(mol Å) [41]. Finally, non-polar hydrogen atoms were merged with their bound carbons and Gasteiger charges were calculated with AutoDockTools 4. The fragment docking was conducted with the software Autodock Vina (version 1.1.0; Molecular Graphics Laboratory, La Jolla, CA, USA) with the following parameters for the selected binding site center $x = 17$, $y = 55$, $z = 4$ (Fig. 1c, green sphere), the dimension parameters $x\text{-}size = 22$, $y\text{-}size = 22$, $z\text{-}size = 24$, a term for thoroughness $exhaustiveness = 8$, and the number of best scored binding poses to store $Number\ of\ modes = 20$ [42]. The coordinates of the binding site center (in Å) refer to the generated homology model provided in Online Resource 2. The docking procedure for 149,151 molecular fragments required 72 h on a HP Xeon two socket Dual-Core-64bit Linux cluster using 10 nodes.

Rescoring

The top-ranked binding pose of each fragment was rescored with the scoring function (Eq. 1):

$$S_k = \sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij} + I_{ij} \quad (1)$$

Where S_k is the score for subpocket k, i is an atom of the residues delineating subpocket k, and j is an atom of a docked fragment. C_{ij} assumes a value of 1 if an hydrophobic interaction was identified between i and j and 0 otherwise, and I_{ij} assumes a value of 1 if an hydrophilic interaction between i and j was identified and 0 otherwise. Potential molecular interactions were calculated with the help of MOE's Protein Ligand Interaction Fingerprint (PLIF). Eq. 1 was applied to subpockets S1, S2, S3, S1' and the oxyanion hole to establish a subpocket assignment for each fragment.

Fragment linking

The linking of fragments was performed implicitly to avoid chemically unfeasible molecules and to restrict the chemical space to the commercially available compounds of the ZINC database. The updated ZINC database version 10 was downloaded in the SMILES format containing 20,878,779 molecules [43]. Next, molecules and fragments were chemically standardized, two dimensional (2D) coordinates were calculated, hydrogens were removed, and a substructure search was performed with the software Pipeline Pilot (version 6.2; Accelrys Inc., San Diego, CA, USA) [44].

Substructure searching for preferred S1 and S2 fragments was based on the SciTegic fingerprints *FCFP_4* and *ECFP_4* [45]. The search hits were required to contain at least one fragment (either S1 or S2). The substructure screening hits were triaged by a number of structural and physicochemical property filters. Briefly, spiro-compounds, compounds with a diameter > 19 bonds, with more than 5 rings and with fused ring systems consisting of > 3 rings were removed. In addition, all molecules were required to contain a terminal cationic or basic moiety expected to be protonated under assay conditions for favorable interactions with the S2 pocket. 18,803 filtered compounds were subsequently clustered into 300 clusters based on maximum fingerprint dissimilarity with help of Pipeline Pilot. Promising clusters were chosen by manual inspection of their cluster representatives selecting for primarily linear molecules containing favorable S1 and S2 fragments at opposite ends. This topology was motivated by molecular docking results. Next, a neighborhood search was conducted based on the *ECFP_4* and *FCFP_4* fingerprints with a Tanimoto coefficient cut-off of 0.7 [46]. Results of the neighborhood search were once again subjected to the structure and property filters leaving 7,915 molecules for re-clustering and selection as described above. Finally, the remaining 815 compounds were docked to the substrate binding site of the NS2B-NS3pro homology model and the final selection of test compounds for experimental validation was manually picked after visual inspection.

Molecular docking

Molecular fragments were linked to complete molecules (superstructures), which were docked with the software GOLD (version 4.1; The Cambridge Crystallographic Data Centre, Cambridge, UK) [47]. The protease structure was again refined in MOE for docking as described above omitting the final AutoDockTools step. Ligands were prepared in MOE by adding hydrogens followed by energy minimization (MMFF94x force field, 0.1 kcal/(mol Å) RMS gradient limit) and finally saving them in SD format. GOLDscore and ChemScore fitness functions were selected and early termination was allowed [48]. The genetic algorithm (GA) settings were set to automatic with a search efficiency of 100%. The binding site center of the homology model was set to $x = 17$, $y = 55$, $z = 4$ with a radius of 16 Å. The molecular docking of 815 molecules took 36.6 h on an Intel Core 2 Quad (2.8 GHz) Linux desktop computer. GOLD was selected for docking of linked molecules due to its known reliability. However, Autodock Vina was utilized for fragment docking due to its multi-threading capabilities and free availability allowing for straightforward parallel and distributed docking of a large number of molecules.

In vitro protease inhibition assay

Small molecules were assayed in 384-well microplates in protease buffer (50 mM Tris, pH 8.5, 1 mM Chaps, 20% glycerol) in a final volume of 30 µL as described by Li et al. [49]. Briefly, protease DENV2 $NS2B_{40}$-G_4-S-G_4-NS3pro185 (10 nM) was pre-incubated with various concentrations of test compounds at room temperature for 30 min. The final DMSO concentration was kept constant at 3%. The reaction was then initiated by the addition of competitive, protease specific, fluorophore-tagged substrate Bz-nKRR-AMC at 20 µM. Reaction progress was monitored continuously by following the increase in fluorescence (excitation 380 nm, emission 450 nm) on a Tecan Infinite®M200 microplate reader at room temperature. Relative fluorescence values were determined after 30 min for 8 consecutive compound concentrations and the IC_{50} was derived from the dose–response curve by a non-linear regression using GraphPad prism (version 5.0; GraphPad Software, La Jolla, CA, USA) [50]. All IC_{50} values reported are derived from at least two experiments.

Enzyme kinetics

5 different compound concentrations (0, 10, 20, 30, and 40 µM) and 3 different substrate concentrations (40, 60, and 80 µM) were tested. Enzyme kinetics was performed in 384-well microplates in protease buffer (50 mM Tris, pH 8.5, 1 mM Chaps, 20% glycerol) in a final volume of 30 µL. Protease (10 nM) was pre-incubated with the indicated concentrations of test compounds (3% DMSO) at room temperature (RT) for 30 min. The reaction was then initiated by the addition of substrate Bz-nKRR-AMC at indicated concentrations. Reaction progress was monitored continuously by following the increase in fluorescence (excitation 380 nm, emission 450 nm) on a Tecan Infinite®M200 microplate reader at RT and relative fluorescence values were determined after 60 min. The data was analysed using a Dixon Plot and the K_i was derived from a linear regression of the reciprocal reaction rate (1/v) against inhibitor concentration by taking the average of the pairwise regression line intersection points using GraphPad prism (version 5.0, La Jolla, CA, USA). Experiments were performed in triplicates.

Results and discussion

Homology model

The structure-based design of DENV protease inhibitors is currently impeded by the lack of a closed-form DENV NS2B-NS3pro structure with fully formed substrate binding site despite extensive effort. We therefore generated a

DENV protease homology model (Fig. 1) based on five WNV and DENV protease template structures (Table 1). The model was validated by its DOPE score, DOPE-per-residue profile, and Ramachandran plots, and found to be of acceptable overall accuracy (c.f. Materials and Methods). Figure 1b shows the structural alignment of the homology model with a WNV protease template structure (PDB id: 3e90 [7]). The alignment has a low root mean squared deviation (RMSD) of non-hydrogen atoms of 0.388 Å. Recently, a DENV protease homology model was published by Wichapong et al. (2009) [51] that agrees well with our present work (RMSD 0.350 Å, Table 1). However, a major structural difference of our model from Wichapong's is the conformation of the oxyanion hole (residues Gly133, Thr134, and Ser135). Our dengue protease homology model features a functionally competent oxyanion hole [5–7]. Distance restrains of atoms participating in the oxyanion hole were derived from WNV protease structure 3e90 and explicitly included in our modeling. A close-up inspection of the model published by Wichapong et al. revealed that the oxyanion hole assumed an inactive conformation: Thr134 is "flipped" by 180 degrees causing the backbone nitrogen atom of Gly133 to point away from the oxyanion hole [5, 7]. In our model the backbone nitrogen atoms of Gly133 and Ser135 are oriented to stabilize a potential negative charge on the carbonyl oxygen of the scissile peptide bond by H-bonding. We hypothesized that the oxyanion hole could provide additional interactions for potential protease inhibitors.

Comparing the binding sites of DENV2 and WNV

A detailed comparison of the modeled DENV2 substrate binding site with the WNV binding site revealed overall high similarity (Fig. 1a). Eight amino acids differences at the substrate binding site distinguish WNV from DENV2 (Fig. 1c). These are D81E, N84S, F85 M, Q86S, and L87I in NS2B (WNV amino acid numbering); and P131S, T132P, and I155 V in NS3. D81E is a conservative change that is not likely to have a large impact on the substrate binding site. N84S replaces the polar amide moiety of the side chain by a polar and smaller hydroxyl group. Asn84 is within hydrogen bonding distance of the P2 residue and was shown to be responsible for S2 pocket specificity differences between WNV and DENV (WNV protease prefers a P2-lysine while DENV prefers a P2-arginine) [52, 53]. For F85 M the hydrophobic side chains of both amino acids are buried within the protease and their exchange is unlikely to impact the binding properties. At position 86 the Gln in WNV is engaged in hydrogen bonding with the P3 residue whilst the shorter side-chained Ser residue in DENV might influence the binding properties of the S3 pocket. However, mutagenesis and peptide substrate/inhibitor data indicate that binding to the S3 pocket is less specific [13, 14, 49, 53]. Lastly, L87I is a conservative hydrophobic exchange unlikely to impact the binding behavior significantly. In NS3pro the amino acid substitutions P131S and T132P occur in a loop connecting the oxyanion hole with the S1 pocket. Interestingly, the position of proline is shifted by one amino acid position without altering the loop conformation ("proline shift"). T132P increases the van der Waals (vdW) surface of the S1 pocket by 18 $Å^2$ and removes the hydrogen bond donor side chain of Thr132. Figure 1c and our docking results indicate that Pro132 increases the hydrophobicity of the DENV2 S1 pocket compared to WNV. This is supported by peptide inhibition studies where hydrophobic aromatic residues in the S1 pocket are more readily tolerated by DENV2 than by WNV protease [13, 14]. The proline shift from position 131 to 132 was further implicated in modifying the binding specificity of the S1' pocket of DENV and WNV protease [54]. Finally, I155 V is a conservative substitution occurring in the S4 pocket that is not expected to have a significant impact on the binding properties of this pocket. In summary, the substrate binding sites of WNV protease and our DENV2 protease model are overall similar with a number of mostly conservative sequence differences. This is in line with the very similar substrate specificity of both proteases [14, 49].

Fragment generation and docking

The DENV protease model subsequently served as the basis of our fragment-based drug design approach. Molecular fragments were generated by retrosynthetically cleaving 14 million compounds of the ZINC screening database version 9. The ZINC compounds were filtered by the Lipinski Rule-of-Five to provide molecules from a drug-like chemical space. 149,151 fragments were generated by application of the RECAP rules [36] and subsequent filtering by the Rule-of-Three for fragment-based drug design (Fig. 2) [37]. In the next step, fragments displaying high *in silico* binding affinity for the DENV protease binding site were identified by parallel high-throughput docking with the software Autodock Vina (see Materials and Methods for details). Scoring of binding poses proved to be difficult as the AutoDock docking score correlated with the molecular weight of the fragments, that is, larger fragments were scored higher on average. We therefore devised our own scoring function which provided a separate score for the S1-S3, S1' subpockets and the oxyanion hole by summarizing polar and hydrophobic interactions. 41,902 fragments (29%) were docked into the S1 and S2 pocket and 1,446 fragments (1%) were docked exclusively into the S2 pocket. Preferred fragment scaffolds of the S1 pocket were indole and other bicyclic

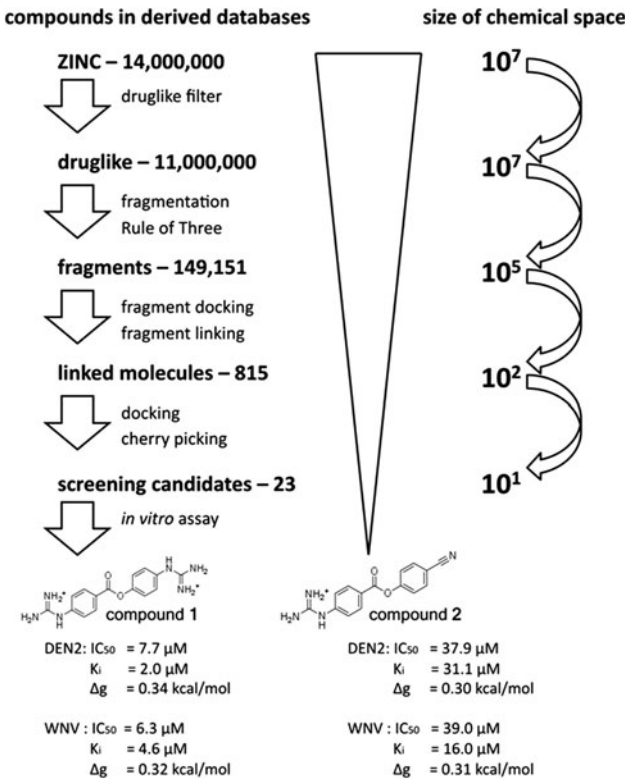

Fig. 2 Schematic of the applied virtual screening cascade. (1) The ZINC database contained 14 million commercially available screening compounds and was filtered for drug-like molecules. (2) 149,151 fragments were generated by a RECAP analysis and filtering by the Rule-of-Three. (3) The fragments were docked to the modeled DENV2 substrate binding site and 220 high-scoring fragments were identified and subsequently linked by our implicit linking approach. (4) 815 linked molecules were docked and 23 compounds were selected for experimental evaluation by visual inspection. (5) Experimental validation by an in vitro protease inhibition assay determined two bioactive molecules with IC_{50} values in the micromolar range

systems. This is supported by data from a tetrapeptide inhibitor study indicating that the indole-containing amino acid tryptophan is tolerated at the P1 position [13]. Preferred fragments of the S2 pocket were exclusively positively charged molecular fragments (pH 7). Two sets of high-scoring fragments, each containing 110 fragments, were chosen based on the S1 and S2 binding scores.

Implicit fragment linking

High-scoring fragments were linked to generate potential DENV protease inhibitors. In order to restrict the chemical space of designed molecules to commercially available compounds and to avoid chemically infeasible molecules fragment linking was performed implicitly. The general concept is outlined as follows: (i) Compounds containing high-scoring fragments were identified in the ZINC database by substructure searches. (ii) Structural and physicochemical property filters were applied forcing potentially favorable interactions with the S1 and S2 pockets. (iii) The compounds were clustered and promising clusters were selected by manual inspection. (iv) A neighborhood search based on chemical similarity was performed around promising candidate molecules.

A substructure search of 220 high-scoring fragments in the ZINC database version 10 (20,878,779 molecules) retrieved 352,758 hits (1.7%). The search hits were required to contain at least one preferred fragment (either S1 or S2). The number of hits was reduced to 18,803 by applying a number of structural and physicochemical property filters (see Materials and Methods). Filtered molecules were required to contain a terminal cationic or basic moiety expected to be protonated under assay conditions to meet our observations for preferred S2 pocket fragments. This requirement effectively linked preferred S1 fragments with an S2 pocket feature. The filtered molecules were clustered and promising clusters were selected by manual inspection of their cluster representatives. This manual selection step ensured that the (largely linear) compounds contained both a favorable S1 and S2 binding fragment at opposite ends. Then, 554 promising compounds were docked with GOLD to the DENV protease model for further analysis [47].

These molecules were subjected to a neighborhood search (similarity search) to explore the chemical space around the selected molecules. A similar filtering cascade as described above was applied to focus the set of molecules resulting in 261 additional compounds. Finally, 23 compounds were selected based on their docking pose ("cherry picking") for experimental evaluation. Of these, 17 originated from the substructure search and 6 from the neighborhood search. The concept of implicit linking is illustrated in Fig. 3 for a preferred indole amide scaffold.

Experimental evaluation

20 small molecules were purchased from ENAMINE Ltd. (Kiev, Ukraine) and 3 compounds were provided free of charge by the National Cancer Institute (NCI, Developmental Therapeutics Program, Bethesda, MD, USA). Of these, compounds **1** and **2** were found to inhibit both DENV2 and WNV protease in the low to moderate micromolar range (Fig. 4). In an in vitro protease inhibition assay, compound **1** exhibited an IC_{50} of 7.5 µM for DENV2 and 6.3 µM for WNV while compound **2** displayed an IC_{50} of 37.9 µM for DENV and 39.0 µM for WNV. In vitro analysis of the remaining 21 compounds can be found in Online Resource 1, Fig. S1. Compounds **1** and **2** are close analogs of a small molecule recently published by Caflisch and colleagues which was shown to inhibit WNV protease ($IC_{50} = 34.2$ µM, $K_d = 16$ µM) [17].

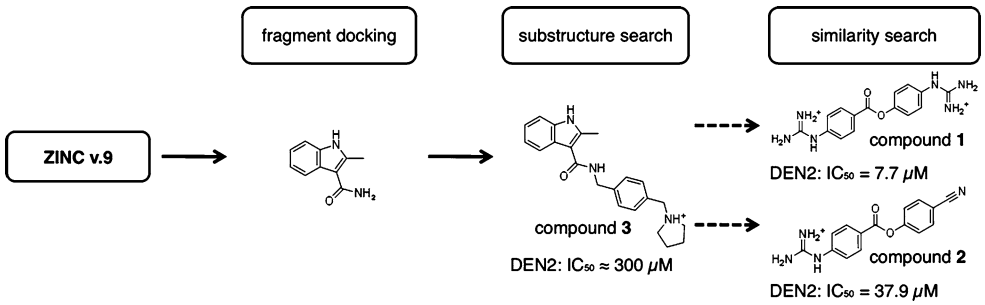

Fig. 3 Compound evolution of compound **1** and compound **2**. **a** The indole amide was identified as a preferred binding fragment of the S1 pocket. **b** Application of an implicit linking approach identified compound **3** ($IC_{50} \approx 300$ μM) containing an indole amide moiety (preferred S1 fragment) and a pyrrolidine fragment (preferred S2 fragment). **c** Hit compounds **1** and **2** were discovered by a neighborhood search (chemical similarity search) and subsequent chemical focusing (clustering, docking, visual inspection)

Caflisch's group devised a structure-guided fragment-based *in silico* drug discovery strategy which is similar to our approach. In their studies fragments were derived either from the ZINC database (Sept. 2006) or the iResearch database (ChemNavigator Inc., 2006) with the software DAIM, docked and scored with the software SEED, followed by docking of screening molecules using the position and orientation of their fragments as anchors by the program FFLD, and finally followed by pose filtering [55–57]. Despite some differences in the implementation of FBDD and the computer programs used, the principle difference to the present study is the choice of biological target molecule. Caflisch's studies were based on a WNV protease structure (PBD id: 2fp7 [6]) and molecular dynamics (MD) snapshots thereof, whereas we used a model of closed-form DENV2 protease derived by a multi-template homology modeling approach. Interestingly, their WNV-targeted study and our DENV-targeted study resulted in very similar compounds. In our protease assay compounds **1** and **2** exhibited almost identical IC_{50} values against DENV2 and WNV protease. This could provide another indication that the substrate binding sites of both proteases are in fact very similar. However, it should be considered that compounds **1** and **2** are carboxyl esters. The ester bond is readily cleaved by proteases (protease acylation reaction), however, the deacylation reaction is much slower [49]. Thus, compounds **1** and **2** may act as covalent inhibitors. The ester bond might also react with nucleophilic groups outside the active site opening the possibility for unspecific or allosteric binding. However, enzyme kinetics analysis of compounds **1** and **2** indicated a competitive mode of inhibition as expected by active site-directed inhibitors (Fig. 4b). In this analysis compound **1** exhibited an inhibition constant K_i of 2.0 μM for DENV2 and 4.6 μM for WNV, and compound **2** showed a K_i of 31.1 μM for DENV2 and 16.0 μM for WNV.

A binding mode of compound **1** was predicted by molecular docking (Fig. 5). The compound was predicted to interact with Asp129 of the S1 pocket via hydrogen bonding and electrostatic interactions. The phenylguanidine moiety might be engaged in a stacked arene-arene interaction with Tyr161. This π-π stacking could be further supported by hydrophobic interactions with Pro132 where Tyr161 and Pro132 form a "hydrophobic clamp" in the S1 pocket. The opposite phenyl moiety of compound **1** was predicted to interact with His51 of the catalytic triad by arene-arene interactions and the S2 residue Asn152 was predicted to be engaged in hydrogen bonding with the opposite guanidinium group. Its positive charge is likely stabilized by the overall negatively charged environment of the S2 pocket.

Conclusions

We demonstrated the successful application of a structure-guided fragment-based *in silico* drug design approach of DENV protease inhibitors. Due to the lack of a closed-form DENV protease structure with fully formed substrate binding site we generated a model of such a structure by homology modeling. FBDD was driven by fragment docking and implicit linking (fragment-focused virtual screening cascade). The concept of implicit linking allowed for the straightforward discovery of commercially available compounds containing preferred fragments. Scoring of docking poses was a general issue which was circumvented by the development of a target-specific scoring function based on counts of potential interactions per protease subpocket. The development of accurate scoring functions should be a focus of future research in this area. Fragments were generated by performing a RECAP analysis on drug-like ZINC compounds. Our

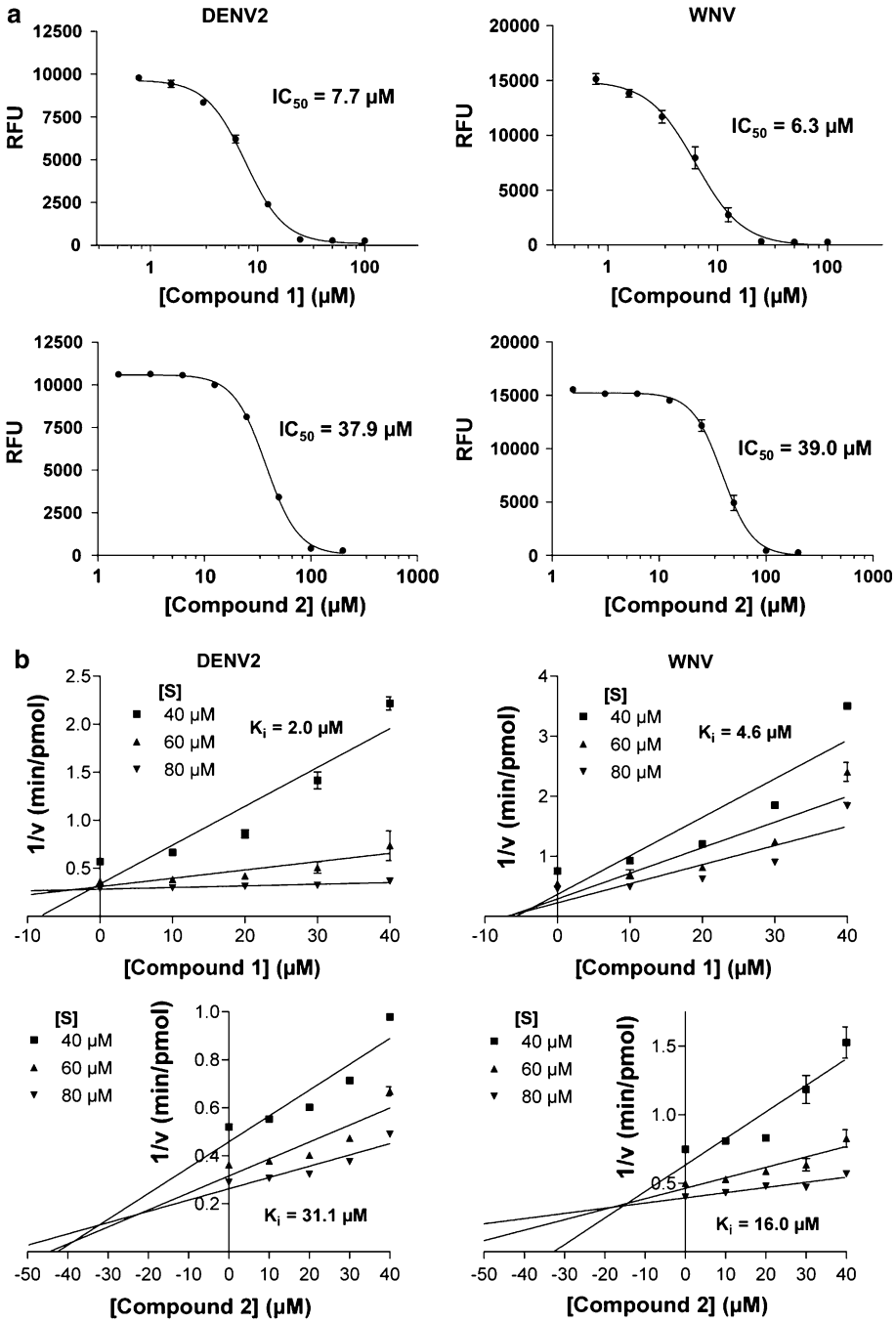

Fig. 4 **a** Dose–response curves of compounds **1** and **2** generated by a fluorescence-based in vitro protease inhibition assay. The compounds were tested for DENV2 and WNV protease inhibition. Relative fluorescence units (RFU) were plotted against inhibitor concentration ([Compound 1/2]). Reported IC_{50} values are the result of at least two experiments. **b** Dixon plots of enzyme inhibition kinetics experiments. The reciprocal reaction rate (1/v) was plotted against inhibitor concentration ([Compound 1/2]) for three different substrate concentrations ([S]). The results indicate a competitive mode of inhibition of compounds **1** and **2**. Kinetics experiments were performed in triplicates

results show that the ZINC database generally provides a diverse pool of molecular fragments useful for FBDD; however, fragmentation by RECAP was incomplete. Alternative approaches have been suggested in the literature and should be explored in the future [58–61]. Two compounds with micromolar activity against DENV2 and WNV protease were discovered. The compounds have favorable ligand efficiency $\geq$ 0.30 kcal/mol but are likely to be positively charged under physiological pH [62]. As carboxyl esters compounds **1** and **2** potentially observe suboptimal properties such as aqueous instability and unspecific reactivity restricting their usefulness. Replacement of the ester moiety should be the target of further research. The presented approach has several advantages over regular high-throughput docking: (i) It is faster. Instead of docking millions of compounds only a much smaller set of fragments has to be docked. (ii) Additional chemical knowledge is derived during the process. This

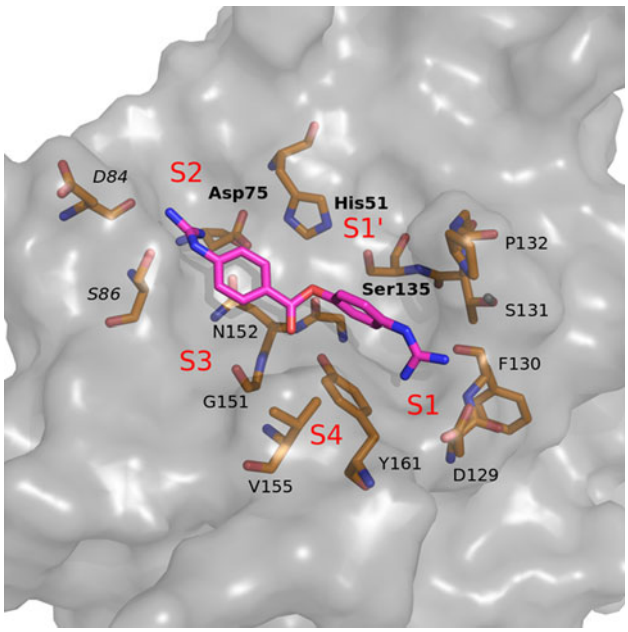

Fig. 5 Predicted binding pose of compound **1**. The compound (shown in *magenta sticks*) was docked into the modeled DENV2 substrate binding site (vdW surface in gray, residues orange sticks) by GOLD version 4.1 yielding a favorable Goldscore of 50.3. Hydrogen bonding and electrostatic interactions were predicted between the guanidinium groups and Asp129, F130 (S1 pocket) and Asn152 (S2 pocket). The binding pose suggested the presence of a stacked arene-arene interaction of one phenylguanidine moiety with Tyr161 possibly supported by hydrophobic interactions with Pro132 in a "hydrophobic clamp" conformation. The docking pose of compound **2** was similar but two possible binding modes rotated by 180 degrees were observed (data not shown). NS2B residues in italics. Figure generated with PyMol

includes, for instance, information about preferred fragments, binding modes, and interaction "hotspots" in the binding pocket. Nonetheless, high-throughput docking could have potentially identified similar hit compounds. In summary, our combined approach of homology modeling, fragment docking and implicit linking offers a promising avenue for structure-based drug design of DENV protease inhibitors.

Acknowledgments We are grateful to Mr. Kok Soon Lai and Ms. Si Fang Wang for expert technical assistance. The National Cancer Institute is thanked for providing three compounds free of charge. Funding was provided by Duke-NUS Signature Research Program (funded by the Agency for Science, Technology and Research, Singapore and the Ministry of Health, Singapore) as a startup grant to SV.

References

1. Division of Vector-Borne Infectious Disease. Centers for Disease Control and Prevention, Atlanta, USA (2010) http://www.cdc.gov/Dengue/ Accessed 23 Sept 2010
2. World Health Organization - Dengue (2010) Geneva, Switzerland. http://www.who.int/topics/dengue/en/ Accessed 23 Sept 2010
3. Lescar J, Luo D, Xu T, Sampath A, Lim SP, Canard B, Vasudevan SG (2008) Towards the design of antiviral inhibitors against flaviviruses: The case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res 80:94–101
4. Luo D, Wei N, Doan DN, Paradkar PN, Chong Y, Davidson AD, Kotaka M, Lescar J, Vasudevan SG (2010) Flexibility between the Protease and Helicase Domains of the Dengue Virus NS3 Protein Conferred by the Linker Region and Its Functional Implications. J Biol Chem 285:18817–18827
5. Aleshin AE, Shiryaev SA, Strongin AY, Liddington RC (2007) Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. Prot Sci 16:795–806
6. 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
7. Robin G, Chappell K, Stoermer MJ, Hu S, Young PR, Fairlie DP, Martin JL (2009) Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. J Mol Biol 385:1568–1577
8. Chandramouli S, Joseph JS, Daudenarde S, Gatchalian J, Cornillez-Ty C, Kuhn P (2010) Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. J Virol 84:3059–3067
9. 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
10. Su X, Ozawa K, Qi R, Vasudevan SG, Lim SP, Otting G (2009) NMR analysis of the dynamic exchange of the NS2B cofactor between open and closed conformations of the West Nile virus NS2B-NS3 protease. PLoS Negl Trop Dis 3:e561
11. Tomlinson SM, Malmstrom RD, Russo A, Mueller N, Pang Y, Watowich SJ (2009) Structure-based discovery of dengue virus protease inhibitors. Antiviral Res 82:110–114
12. Yin Z, Patel SJ, Wang W, Wang G, Chan W, 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
13. Yin Z, Patel SJ, Wang W, Chan W, Ranga Rao K, 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
14. Knox JE, Ma NL, Yin Z, Patel SJ, Wang W, Chan W, Ranga Rao KR, Wang G, Ngew X, Patel V, Beer D, Lim SP, Vasudevan SG, Keller TH (2006) Peptide inhibitors of West Nile NS3 protease: SAR study of tetrapeptide aldehyde inhibitors. J Med Chem 49:6585–6590
15. Stoermer MJ, Chappell KJ, Liebscher S, Jensen CM, Gan CH, Gupta PK, Xu W, Young PR, Fairlie DP (2008) Potent cationic inhibitors of West Nile virus NS2B/NS3 protease with serum stability, cell permeability and antiviral activity. J Med Chem 51:5714–5721
16. Sidique S, Shiryaev SA, Ratnikov BI, Herath A, Su Y, Strongin AY, Cosford ND (2009) Structure-activity relationship and improved hydrolytic stability of pyrazole derivatives that are allosteric inhibitors of West Nile virus NS2B-NS3 proteinase. Bioorg Med Chem Lett 19:5773–5777
17. Ekonomiuk D, Su X, Ozawa K, Bodenreider C, Lim SP, Otting G, Huang D, Caflisch A (2009) Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. J Med Chem 52:4860–4868

18. Ekonomiuk D, Su X, 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
19. Ganesh VK, Muller N, Judge K, Luan C, Padmanabhan R, Murthy KHM (2005) Identification and characterization of non-substrate based inhibitors of the essential dengue and West Nile virus proteases. Bioorg Med Chem 13:257–264
20. Leung D, Schroder K, White H, Fang N, 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
21. Nall TA, Chappell KJ, Stoermer MJ, Fang N, Tyndall JDA, Young PR, Fairlie DP (2004) Enzymatic characterization and homology model of a catalytically active recombinant West Nile virus NS3 protease. J Biol Chem 279:48535–48542
22. Johnston PA, Phillips J, Shun TY, Shinde S, Lazo JS, Huryn DM, Myers MC, Ratnikov BI, Smith JW, Su Y, Dahl R, Cosford NDP, Shiryaev SA, Strongin AY (2007) HTS identifies novel and specific uncompetitive inhibitors of the two-component NS2B-NS3 proteinase of West Nile virus. Assay Drug Dev Technol 5:737–750
23. 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
24. Mueller NH, Pattabiraman N, Ansarah-Sobrinho C, Viswanathan P, Pierson TC, Padmanabhan R (2008) Identification and biochemical characterization of small-molecule inhibitors of west nile virus serine protease by a high-throughput screen. Antimicrob Agents Chemother 52:3385–3393
25. Shiryaev S, Ratnikov B, Chekanov A, Sikora S, Rozanov D, Godzik A, Wang J, Smith J, Huang Z, Lindberg I, Samuel M, Diamond M, Strongin A (2006) Cleavage targets and the D-arginine-based inhibitors of the West Nile virus NS3 processing proteinase. Biochem J 393:503
26. Tomlinson SM, Watowich SJ (2008) Substrate inhibition kinetic model for West Nile virus NS2B-NS3 protease. Biochemistry 47:11763–11770
27. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27:157–162
28. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
29. Marti-Renom MA, Madhusudhan MS, Sali A (2004) Alignment of protein sequences by their profiles. Protein Sci 13:1071–1087
30. Shen M, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524
31. DeLano W (2009) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto
32. Chemical Computing Group (2009) MOE—the molecular operating environment. Montreal, Canada
33. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99
34. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model 45:177–182
35. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
36. Lewell XQ, Judd DB, Watson SP, Hann MM (1998) RECAP—retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Comput Sci 38:511–522
37. Congreve M, Carr R, Murray C, Jhoti H (2003) A 'Rule of Three' for fragment-based lead discovery? Drug Discov Today 8:876–877
38. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. Tetrahedron 36:3219–3228
39. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791
40. Markley JL, Westler WM (1996) Protonation-state dependence of hydrogen bond strengths and exchange rates in a serine protease catalytic triad: bovine chymotrypsinogen A. Biochem 35:11092–11097
41. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21:1049–1074
42. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461
43. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36
44. Accelrys Inc. (2007) Pipeline Pilot. Accelrys Inc., San Diego
45. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50:742–754
46. Tanimoto T (1957) IBM internal report. 1957
47. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
48. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 11:425–445
49. 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
50. GraphPad Software Inc. (2009) GraphPad prism. GraphPad Software Inc., La Jolla
51. Wichapong K, Pianwanit S, Sippl W, Kokpol S (2010) Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction. J Mol Recognit 23:283–300
52. Chappell KJ (2007) Structure-function relationships of the West Nile virus protease NS3 and its cofactor NS2B. PhD thesis, University of Queensland, Australia
53. Chappell KJ, Stoermer MJ, Fairlie DP, Young PR (2006) Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies. J Biol Chem 281:38448–38458
54. Shiryaev SA, Ratnikov BI, Aleshin AE, Kozlov IA, Nelson NA, Lebl M, Smith JW, Liddington RC, Strongin AY (2007) Switching the substrate specificity of the two-component NS2B-NS3 flavivirus proteinase by structure-based mutagenesis. J Virol 81:4501–4509
55. Kolb P, Caflisch A (2006) Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. J Med Chem 49:7384–7392

56. Majeux N, Scarsi M, Caflisch A (2001) Efficient electrostatic solvation model for protein-fragment docking. Proteins 42: 256–268
57. Budin N, Majeux N, Caflisch A (2001) Fragment-based flexible ligand docking by evolutionary optimization. Biol Chem 382:1365–1372
58. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–2893
59. Xu Y, Johnson M (2001) Algorithm for naming molecular equivalence classes represented by labeled pseudographs. J Chem Inf Comput Sci 41:181–185
60. Schuffenhauer A, Ertl P, Roggo S, Wetzel S, Koch MA, Waldmann H (2007) The scaffold tree—visualization of the scaffold universe by hierarchical scaffold classification. J Chem Inf Model 47:47–58
61. Xu Y, Johnson M (2002) Using molecular equivalence numbers to visually explore structural features that distinguish chemical libraries. J Chem Inf Comput Sci 42:912–926
62. Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431