Computational Design of Dengue Type-2 NS2B/NS3 Protease Inhibitor: 2D/3D QSAR of Quinoline and Its Molecular Docking

Maywan Hariono[1*], Ezatul E. Kamarulzaman[1], Habibah A. Wahab[1,2]

[1]*Pharmaceutical Design and Simulation Laboratory, School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia.*
[2]*Malaysian Institute of Pharmaceuticals and Nutraceuticals, Ministry of Science, Technology and Innovation, Halaman Bukit Gambir, 11700 Gelugor, Pulau Pinang, Malaysia.*

Received: 1 September 2014 / Accepted: 30 November 2014

Abstract:
The reemergence of dengue outbreak demands an effective and efficient treatment especially for antiviral agent since neither vaccine nor drug is available to overcome this disease. This study designed the new model for Dengue Type-2 (DENV2) NS2B/NS3 protease inhibitor from quinoline scaffold by combining QSAR modeling with molecular docking. Genetic Function Approximation (GFA) method was used to construct the QSAR model showing good correlation coefficient ($r^2 = 0.8996$). The model was validated using Partial Least Square (PLS) analysis ($r^2 = 0.861$), Leave One Out - Cross Validation (LOO-CV with $q^2 = 0.677$) and external validation ($r^2 = 0.854$). The model was then used to design more quinoline derivatives by attaching more electron withdrawing group such as halogen and aldehyde. Ssearching for more diverse structure of protease inhibitors was carried out by screening out NCI database. Results showed that the QSAR model was able to predict the new protease inhibitor model with a nanomolar inhibition. The molecular docking was then performed to predict the binding mode of all protease inhibitor models and revealed that most of essential amino acid residues such as Ile36, His51, Ser135, Asp129, Tyr150, Gly151, Gly153 and Val154 significantly interact with the ligands. Based on the agreement between QSAR modeling and molecular docking results, the designed molecules were reported gaining insight activities as DENV2 NS2B/NS3 protease inhibitor.

Key words: *DENV2 NS2B/NS3 protease, genetic function approximation, molecular docking, qsar, quinoline*

Introduction

Dengue is one of pathogenic viruses which killed more than 22,000 people in a recent year all over the world. During 2013, it was reported that dengue outbreaked in Angola and Kenya while in Laos, Malaysia and Singapore, the cases were found more frequent than the same period in 2012 [1]. Neither vaccine to prevent the infection nor drug is approved by FDA to stop the virus replication [2,3]. Especially in South East Asia, DENV2 is the most prevalence among other serotypes (DENV1, DENV3 and DENV4) detected in the patient biological specimen [4]. This flavivirus is 11 kb positive single strand RNA which is genomically employing three structural proteins (capsid, pre-membrane and envelope) as well as 7 non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [5-8].

DENV2 NS2B/NS3 protease (DENV2 NS2B/NS3pro) is the most common target in the drug design of dengue disease [9, 10]. This non-structural protein is composed by NS3 as the protease and NS2B as the cofactor. As other trypsin like serine protease, NS3 regulates the post-translation of the genome to its protein as well as the virus maturation while the hydrophilic region of NS2B is responsible for holding and promoting the activation of NS3 [11-13]. In addition, the hydrophobic part of NS2B plays around the membrane association upon the cleavage process [14-17].

Up to now, there has been some published compounds exhibiting nano to micro molar inhibition ranges against DENV2 NS2B/NS3pro such as tetrapeptide with electrophilic covalent warhead [12], panduratin derivatives [18], α-ketoamide [19], acrylamide [20], anthracene [21], quinoline [22], aminobenzamide [23], benzothiazole and oxadiazole [24]. Unfortunately, neither peptidomimetic nor small molecule compounds are able to enter the clinical stage as a drug candidate [19]. Therefore, a searching for more potent inhibitors is highly demanded to proceed the new candidate up to the drug that is clinically used for the dengue patient [25].

To date, this present study is aimed to develop the more potent DENV2 NS2B/NS3pro inhibitor by designing new molecules using 2D/3D-QSAR modeling. Beside, the built model is also used to virtually screen the National Cancer Institute (NCI) database in order to explore more alternative structure as DENV2 NS2B/NS3pro hit compounds. The absence of DENV2 NS2B/NS3pro co-crystal structure with its inhibitor, make the ligand-based drug design such as QSAR modeling is more feasible than structure based drug

***Corresponding author:** Maywan Hariono.
E-mail: maywan_har@yahoo.com

design such as molecular docking. However, to predict the binding mode of the designed ligand, the molecular docking was also performed using the DENV2 NS2B/NS3pro model produced by homology modeling [26].

The previous QSAR studies on DENV2 NS2B/NS3pro inhibitor was carried out on panduratin derivatives using Multiple Linear Regression (MLR) combined with Partial Least Square (PLS) method [27]. In the same period, QSAR modeling of DENV2 NS2B/NS3pro inhibitor was also developed on peptidomimetic compounds using LUDI scoring function [28]. The MLR method was reported that it cannot be used in QSAR if the number of descriptors is extremely larger than the number of data set [29]. On the other hand, the peptidomimetic compound frequently stuck on the in vivo pharmacokinetic study due to the instability of peptide compound in the real body system.

In this work, we perform the QSAR modeling using GFA algorithm which beneficially can select the most appropriate descriptors from the large number of descriptors we order in the setting of parameters [30]. Basically, GFA operates iteratively in succession: selection of two parents from the present population due to their fitness of proportional probability, crossover the parent characters to create the new child string and mutation to introduce a measure of diversity into the population [31, 32]. The data set was collected from published paper [22] employing quinoline scaffold with azomethine (Schiff Base) fragment. The consideration of the respective data set collection is due to the drug-like likeness of quinoline compared to that of peptidomimetic compounds. Moreover, in term of a synthetic accessibility in future works, the Schiff base is a kind of simple reaction and the starting material of quinoline scaffold is broadly available.

Experimental

Materials

The 2D structure was constructed using ACD/ Chemsketch Freeware version 12.00 (www.acdlabs.com) and the 3D structure was energetically optimized using Hyperchem Professional version 8.0 (www.hyper.com). The QSAR study was carried out using Discovery Studio Client 2.5 (www.accelrys.com) and the Molecular Docking was performed using AutoDockTools version 1.5.6 for protein-ligand preparation and AutoDock 4 for running the docking job (www.autodock.scrips.edu). The graph was presented using Microsoft Excell 2007.

Data set collection and generation

The data set was collected from Deng et al consists of 26 representative compounds which were tested in vitro using the same method (see Table 1) [22]. The biological activities were calculated as the fifty percent of inhibition concentration, IC_{50} (μM) which was later expressed in negative logarithmic units (pIC_{50}). The 2D structures were constructed using ACD/ Chemsketch and its 3D conversion as well as an energy minimization were calculated using Hyperchem Professional version 8.0. The used force filed was MM+ with Polak-Ribiere (Conjugate Gradient) as the algorithm. The data set was than splitted into two clusters in chemical spaces defined by the principal components of two fingerprints: Extended-connectivity fingerprint (ECFP_6) and Functional-class fingerprints (FCFP_6) using Library Cluster Ligand Analysis embedded in Discovery Studio 2.5, (Accelrys). The splitted clusters were named as training set (20 ligands) and test set (6 ligands), respectively.

Calculation of descriptors

The training set along with its experimental pIC_{50} (pIC_{50} (exp)) was exported to Discovery Studio 2.5. The parameters was set up by ordering 2D descriptors such as AlogP, Molecular Weight, Num_Aromatic Rings, Num_H_Acceptors, Num_H_Donors, Num_Rings, Num_Rotatable Bonds, Molecular_Fractional Surface Area, and Topology Descriptors such as CHI, Kappa, etc. Next, the 3D descriptors were also ordered employing Dipole, Jurs Descriptors, Principle Moments of Inertia, Shadow Indices and Surface Area and Volume. The correlation matrix was set to true and then the job was run to generate all calculated descriptors. The selected descriptor data of the training set are presented in Table 1 (supporting informations).

QSAR modeling

The GFA algorithm parameter was set up in Discovery Studio 2.5 using pIC_{50} as the dependent variable and all calculated descriptors as the independent variables. Six models including Linear (L) and polynomials i.e. Binary Interaction (BI), Simple Quadratic [31], Full Quadratic (FQ), Simple Cubic (SC) and Full Cubic (FC) were built in this calculation. The next parameters were set according to the parentheses as followed: Population Size (100), Maximum Generation (5000), Scoring Method (Single Score) and Score Function (Friedman Lack-Of-Fit). The best statistical values of six models will be selected as the model for further design.

Model Validation

Leave One Out-Cross Validation (LOO-CV)

The LOO-CV was generated using Bayesian Model embedded in Discovery Studio 2.5. The concept of this validation is when one ligand being excluded from the training set; the remained ligands were used as the current training set to predict the activity of excluded ligand as well as the training set it. The LOO-CV is accepted statistically when the $q^2 > 0.6$ [31].

Partial Least Square Analysis

The descriptors from the selected QSAR model were then used for independent variables against the training set with its pIC_{50} (exp) as the dependent variable to validate the respective model using PLS method in Discovery Studio 2.5. The validity of the model is

defined when the r^2 value pf pIC_{50} (exp) vs pIC_{50} (pred) > 0.7 [31] using Microsoft Excell 2007.

External Validation
The test set along with pIC_{50} (exp) was exported to Discovery Studio 2.5 and set up as an input ligand. The same descriptors with the training set were ordered with an additional selected QSAR model. The observed pIC_{50} (pred) was collected and then plotted versus pIC_{50} (exp) using Microsoft Excell 2007. The validity of the model is defined when the r^2 value pf pIC_{50} (exp) vs pIC_{50} (pred) > 0.5 [31] using Microsoft Excell 2007. The selected descriptor data of the test set are presented in Table 2 (supporting informations).

Design of the new models using QSAR model
Five new models were designed based on the most active compound by modifying the functional groups according to the descriptors selected by the best QSAR model. The same protocols with external validation were then used to predict the activity of the corresponding models. The selected descriptor data of the designed ligands are presented in Table 3 (supporting informations).

Virtual Screening of NCI database using QSAR model
The virtual screening procedure was employed for designing the new model of quinoline based-structure through 1990 NCI compounds at once. The hit compounds were ranked according to the pIC_{50} (pred) of the QSAR model. Five representative ligands with the IC_{50} values at nanoMolar range were selected as the hit compounds, subsequently. The selected descriptor data of the selected NCI ligands are presented in Table 4 (supporting informations).

Molecular docking of the new models
The DENV2 NS2B/NS3pro model complex with tetrapeptide inhibitor (Bz-Nle-Lys-Arg-Arg-H) [26] was used as the protein target. This model was built using WNV NS2B/NS3 crystal structure (PDB 2FP7) as a template and the coordinate was taken from DENV2 NS2B/NS3 crystal structure (PDB 2FOM). The protein was separated from its inhibitor complex using Discovery Studio 2.5 and then both of them were prepared using AutoDockTools 1.5.6. The Kollman charge was added to the protein structure while the ligand was added by Gasteiger charge. The torsion of ligands was set to 17 rotatable bounds out of total 34 bounds, accordingly. Subsequently, the covalent grid map was used for positive control docking since the co-crystallized ligand covalently bound to the corresponding protein. The Grid Box was set in number of points x = 60, y = 60, and z = 60 as well as its spacing i.e. 0.375 Å. The coordinates used in this Grid Box are x = 21.517, y = 43.428 and y = −1.743. Atom type 'Z' was assigned to the carbon atom which forms a covalent bond with the "warhead" Ser135. Docking was run using AutoDock 4 [33] and the re-docked result which gives the RMSD value less than 2.0 Å will be applied as the fix parameter in a further new ligand docking simulation.

Results and Discussion

QSAR
The training set (20 ligands) was simulated its QSAR using 6 models presented in GFA algorithm. Linear is a model containing the original properties on the data stream only, without product terms. Binary interaction contains a (quadratic) product of different properties on the data stream next to linear. The multiplication of N (N-1)/2 should be raised for N properties that may be included in the models. Simple Quadratic may contain the product with the same properties only in addition to linear terms. For N properties, the N additional terms should be raised up. Full quadratic may combine the binary interaction and simple quadratic for additional linear terms. For N properties, this rises the terms up to N (N+1)/2.

Cubic may combine either the squares or cubic of each property for additional linear terms. For N properties this rises up to 2N additional terms. The last model is Full Cubic which is defined as the combination of three properties. For N properties, this gives rise to N(N+1)(N+5)/6 additional terms [34]. The results of QSAR calculation is presented in Table 2.

Based on the statistical results, all models comply the goodness of statistical parameters. As listed, the r^2, r^2 (adj) and r^2 (pred) has a value at the range 0.7 to 0.9 even better than the range required theoretically which is above 0.6. In this study, the correlation coefficient, r^2, measures how strong the relationships between the molecular properties of compounds and their biological activities against protease. Adjusted r^2 is used to penalize equations containing a too diverse terms which qualify their fittings. Predicted r^2, (r^2 (pred)), is the regression coefficient which is predicted from the multiple observation. The RMS residual errors are described by a probability distribution of the values with a constant variance once the model being built. In this experiment, the RMS residual errors show the minimum value in between 0 to 1 [35]. The fitness of GFA is also well measured for the various statistical parameters using Friedman lack-of-fit (LOF). It is formulated as:

$$\mathrm{LOF} = \frac{\mathrm{SSE}}{\left(1-\frac{c+dp}{M}\right)^2} \qquad (1)$$

Where SSE is defined as the sum of squares of errors, c is the number of basic functions (other than the constant term), d is a smoothing parameter, p is the total number of features contained in all the basic functions and M is the number of samples in the training set. This LOF has a better measurement to resist the over fitting than the SSE which ideally having value in between 0 to 1 [34]. As well as the RMS residual error, in this experiment, the Friedman LOF shows a good criterion approximately 0.1. The significant of errors is related with a standard of statistical test for the quality of the variances on two

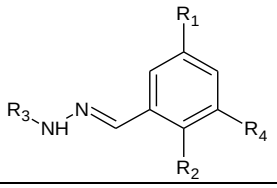
populations which have a normal distribution. When the S.O.R F value is required 95%, the S.O.R p-value must be within 0 to 5%

In general, it can be said that the lower of S.O.R p-value, the better model. Therefore, the observed p -value of S.O.R in this experiment was categorized as a very good model [34].

Among those six models, FC model showed the best statistical results, therefore, this model was proposed to be the best QSAR model for further design. The graph pIC_{50} p vs pIC_{50} e is illustrated in Figure 1. Nevertheless, the validation of the selected model is necessary to evaluate its prediction accuracy. The first validation is carried out using LOO-CV and it shows the $q^2 = 0.677$ which is accepted result reflecting the validity of the training set data.

The variable counts of the descriptors used in FC model with 5000 number of generations (see Figure 2). It is noted that that among 20 counted descriptors, Jurs_DPSA_1, Num_H_Acceptors, AlogP, Dipole_X, V_Dist_Mag and Shadow_Zlength are the most frequent descriptors found giving significant correlation toward the activity of the training set.

Table 1. The data set structures and its biological activities



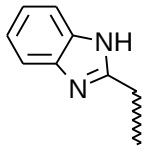
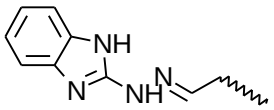
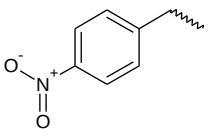
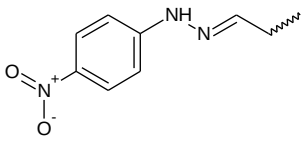
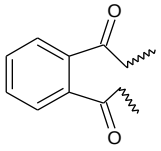
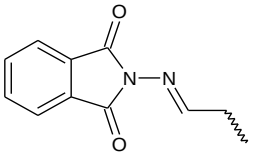
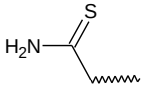
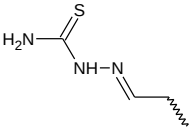
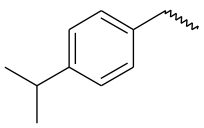
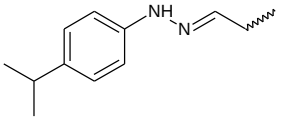
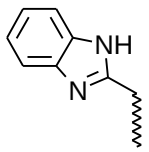
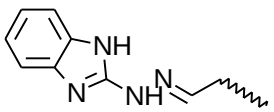
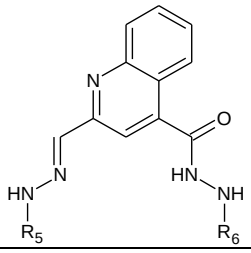
Compound	R_1	R_2	R_3	R_4	IC_{50} (μM)	pIC_{50}
Deng_1*	$-CO_2Et$	-OH			13.12	−1.12
Deng_22a*	$-CO_2Et$	-OH			14.58	−1.16
Deng_22b*	$-CO_2Et$	-OH			>100.00	−2.00
Deng_22h*	$-CO_2Et$	-OH			>100.00	−2.00
Deng_22j*	$-CO_2Et$	-OH			39.46	−1.60
Deng_22n*	-H	-OH			>100.00	−2.00

Table 1. Continued



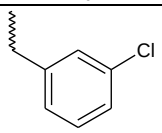
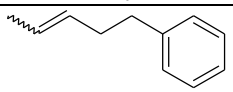
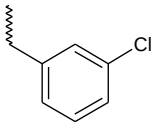
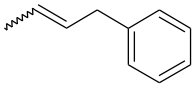
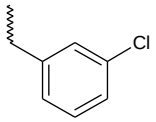
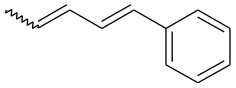
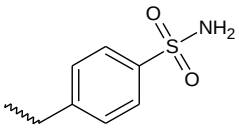
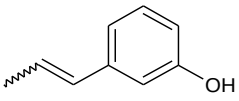
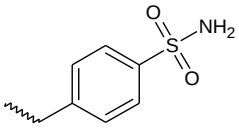
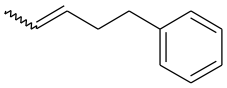
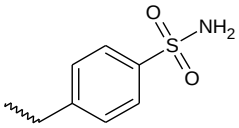
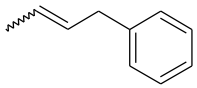
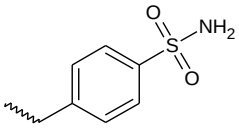
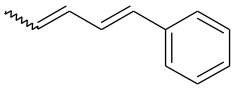
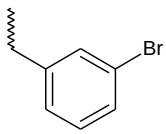
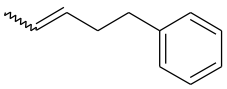
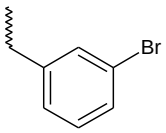
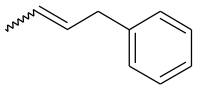
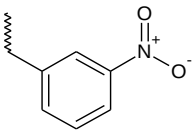
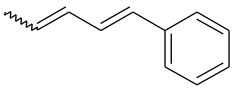
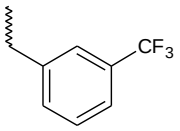
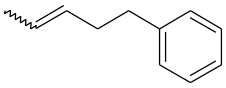
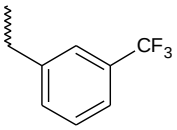
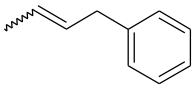
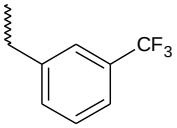
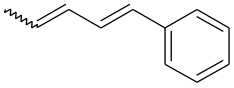
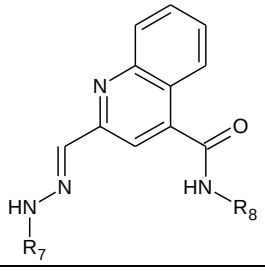
Compound	R_5	R_6	IC_{50} (μM)	pIC_{50}
Deng_23a			29.04	−1.46
Deng_23b			28.12	−1.45
Deng_23c			14.32	−1.16
Deng_23d			7.83	−0.89
Deng_23e			36.02	−1.56
Deng_23f			>100.00	−2.00
Deng_23g			7.46	−0.87
Deng_23h			19.93	−1.30

Table 1. Continued

Deng_23i			9.45	−0.98
Deng_23j			21.96	−1.34
Deng_23k			>100.00	−2.00
Deng_23l			19.80	−1.29
Deng_23m			>100.00	−2.00



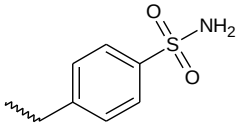
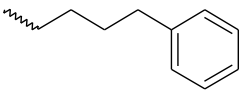
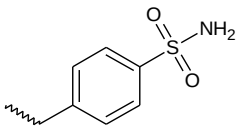
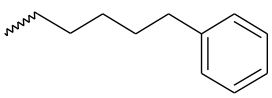
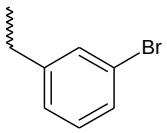
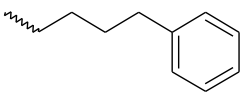
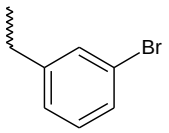
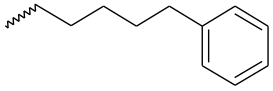
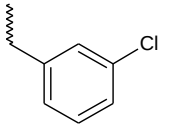
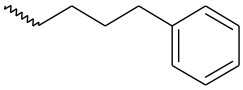
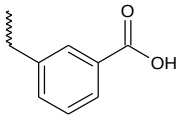
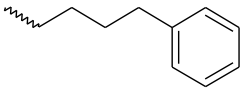
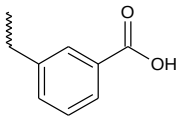
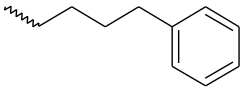
Compound	R_7	R_8	IC_{50} (μM)	pIC_{50}
Deng_24a			>100.00	−2.00
Deng_24b			>100.00	−2.00
Deng_24c			>100.00	−2.00

Table 1. Continued

Deng_24d			>100.00	−2.00
Deng_24e			>100.00	−2.00
Deng_24g			41.24	−1.62
Deng_24h			35.28	−1.55

*Compound of the test set

The IC_{50} value = >100.00 was defined as 100.00 which is converted to pIC_{50} = −2.00.

Table 2. The best QSAR GFA models and their regression statistics.

Model	Equation	r^2	r^2 (adj)	r^2 (pred)	RMS residual errors	Friedman L.O. F	S.O.R p-value
L	20lig_lin_1 = 13.471 − 1.0826 * Dipole_X − 0.018431 * Jurs_DPSA_1 + 5.0919 * Jurs_FNSA_2 − 11.526 * Jurs_RASA + 0.018445 * Molecular_Volume	0.8233	0.7602	0.6735	0.2004	0.1933	7.44×10^{-5}
BI	20lig_BI_1 = 3.7077 + 0.56672 * Num_H_Acceptors − 1.2253 * Dipole_X − 0.13849 * Dipole_Z − 3.1177 * Jurs_FPSA_2 − 0.015312 * Shadow_YZ	0.8322	0.7723	0.7077	0.1953	0.1836	5.26×10^{-5}
SQ	20lig_SQ_1 = 13.471 − 1.0826 * Dipole_X − 0.018431 * Jurs_DPSA_1 + 5.0919 * Jurs_FNSA_2 − 11.526 * Jurs_RASA + 0.018445 * Molecular_Volume	0.8233	0.7602	0.6735	0.2004	0.1933	7.44×10^{-5}
FQ	20lig_FQ_1 = 15.115 + 1.9494 * CIC − 1.5503 * Dipole_X − 0.19382 * Dipole_Z − 3.8411 * Jurs_FPSA_2 − 11.115 * Jurs_RASA	0.8322	0.7722	0.6507	0.1953	0.1837	5.26×10^{-5}
SC	20lig_SC_1 = 2.9116 − 0.68874 * Dipole_X − 0.010288 * Jurs_DPSA_1 − 5.5398 * Jurs_RPCG + 0.97884 * Jurs_RPCS − 0.16656 * Shadow_Zlength	0.7928	0.7188	0.5690	0.2170	0.2267	2.16×10^{-4}
FC	20lig_FC_1 = − 5.1205 − 0.00021184 * V_ADJ_mag + 0.33217 * ALogP * Num_H_Acceptors * Jurs_FPSA_1 − 5.6731×10^{-6} * Molecular_FractionalPolarSurfaceArea * V_DIST_mag * Jurs_DPSA_1 + 0.0024467 * CHI_1 * Jurs_RASA * Shadow_XY − 5.9895×10^{-5} * Dipole_X * Jurs_DPSA_2 * Shadow_Zlength	0.8996	0.8637	0.8113	0.1510	0.1224	1.60×10^{-6}

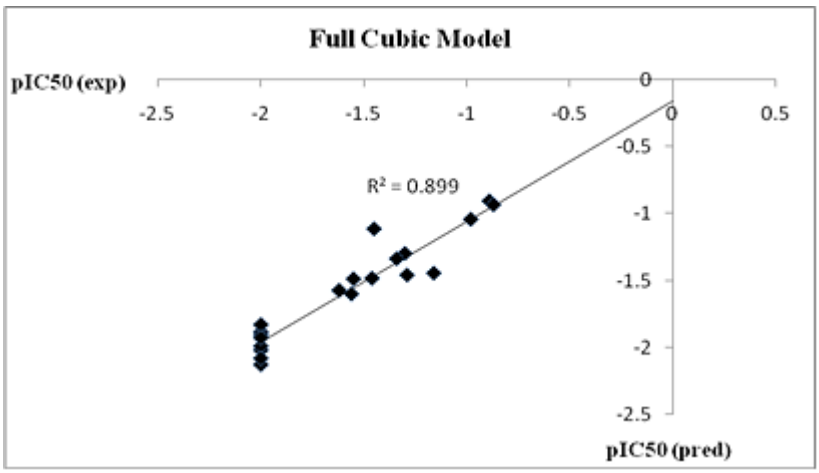


Figure 1. The graph of QSAR calculation using GFA algorithm with full cubic model for training set.

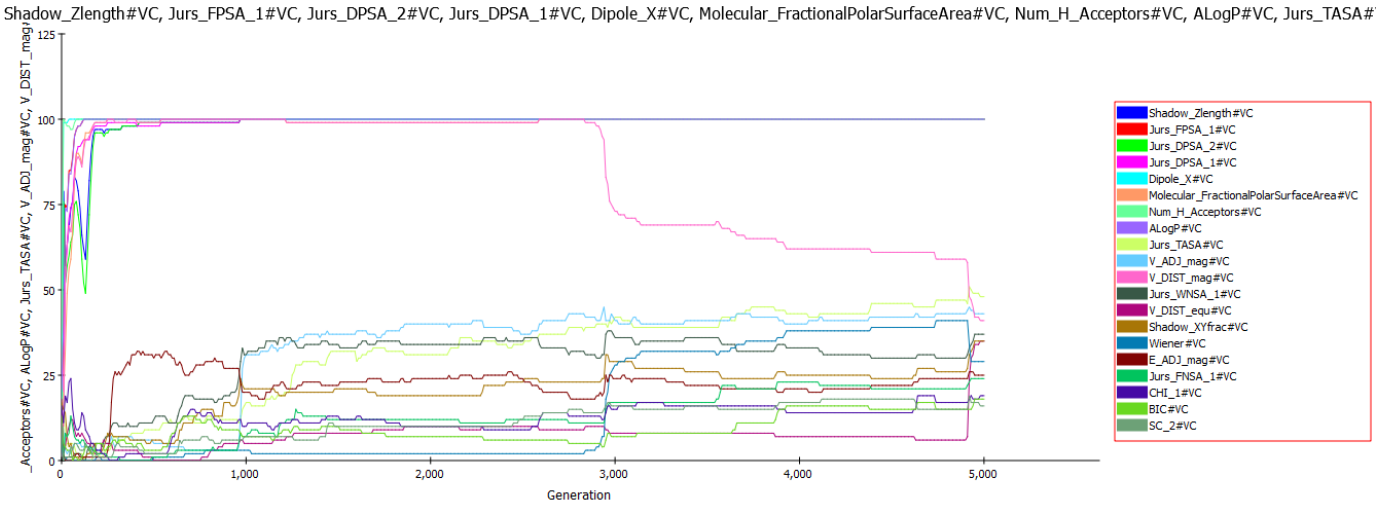


Figure 2. The variable counts of the descriptors used in FC model with 5000 number of generations.

The FC model found that the affinity of compounds toward DENV2 NS2B/NS3pro depends on the partition coefficient (ALogP), the number of hydrogen bond acceptors (Num_H_Acceptor), the molecular fractional surface area, the topology descriptors employ CHI (CHI_1), V_ADJ_mag, V_DIST_mag, the shadow indices (Shadow_XY and Shadow_Zlength) and five 3D descriptors: Jurs_FPSA_1, Jurs_DPSA_1, Jurs_DPSA_2, Jurs_RASA, and Dipole_X. In particular, the ligand affinity is enhanced by ALogP, Num_H_Acceptors, CHI_1, Jurs_FPSA_1, Jurs_RASA and Shadow_XY while it is in contrast with the V_ADJ_mag, molecular fractional surface area, V_DIST_mag, Jurs_DPSA_1, Dipole_X, and Shadow_Zlengt. Although not all mentioned descriptors were found frequently in the variable counts but the combination of them gives the best statistical results.

CHI_1 is the first-order chi index encodes the number of bonds (edges) in the molecular graph. This descriptor describes the ligand connectivity relates to the presence of the series methane through trifluoromethane in compound Deng_23k, 23l and 23m. V_ADJ_mag (Vertex Adjancency Magnitude) and V_DIST_mag (Vertex Adjancency Magnitude) helps to differentiate molecules based on the size, flexibility, degree of branching and entire shape. Particularly, these descriptors view the molecules as a subset structure that is equivalent in some senses [36, 37].

Shadow_XY (area of the molecular shadow in the xy plane) and Shadow_Zlength (length of the molecule in z dimension) are calculated based on the model surface projection for 3 mutually perpendicular planes: xy, yz and xz. Both conformation as well as configuration of the molecules can be calculated by these descriptors by rotating them to align the principle moment of inertia with the x, y and z axes [38]. Dipole is the 3D electronic descriptor which indicates the behavior of the molecules in particular strength and its orientation [39]. This relates to the polarity of the ligand contributed by functional group such as –OH, –NH, C=O, CF_3, SO_2, etc. which make the electropositive atom repulse to interact with the protein.

The next 3D descriptors are Jurs_FPSA_1 (fractional charge partial surface area), Jurs_DPSA_1 (difference in charge partial surface areas), Jurs_DPSA_2 (difference in total charge weighted surface areas) and Jurs_RASA (relative hydrophobic surface area). These descriptors assigned the atomic partial charge on solvent accessible surface area on individual atoms. The Jurs_FPSA_1 contributes in enhancing ligand-protein affinity describes that some partial charges of the ligand may contact intensively with the solvent in the biological system and probe the nonpolar site of molecule to be accessed by the solvent (entropic effect) [40] .

In this work, the PLS analysis is used to fit and weigh the basic functions relative input in the final model. Figure 3 shows the PLS regression graph plotting pIC_{50} (exp) vs. pIC_{50} (pred) of the training set using the descriptors selected in FC model. The statistical results of PLS analysis shows the good correlation coefficient ($r^2 = 0.861$) between pIC_{50} (e) vs. pIC_{50} p of the training set and the least square error = 0.022 corroborates that the FC model is capable to predict the new model. Furthermore, the external validation using the test set reveals the good correlation between pIC_{50} e vs pIC_{50} p of the test set sustains the high quality of the respective model.

The basic function of QSAR equation is defined by the higher number of hydrogen bond acceptor such as $-NO_2$, $-SO_2$ and -COOH. This is agree with the SAR analysis from Deng et al., 2011 stated that the presence of electron withdrawing group in R_5 and R_7 increases the activity toward the pratease. Other basic function i.e. molecular fractional polar surface area was defined to decrease the ligand activity. Afterward, the design of the new quinoline derivatives is carried underlying the FC Model using the strategy depicted in Figure 4. In the training set, two active compounds: Deng_23g and Deng_23l employs the quinoline scaffold, carbonyl group, hydrazone group, conjugated double bonds, substituted phenyl ring with organohalogen or sulfonamide and unsubstituted phenyl ring. In this present works, while maintaining all of the mentioned fragments, we modify the functional group substituted in both phenyl rings. The pIC_{50}s top ten best new models are presented in Table 3.

As we can see in Table 3 that the modifications in the phenyl rings are predicted to increase the activity of the new model 868 folds higher than the most active compound in the data set (Deng_23g). In these new models, we found that all of them have halogen groups hold on the phenyl rings. These are in agreement with the kind of descriptors suggested by FC model that the number of hydrogen acceptor and CHI_1 topology descriptors have positive correlations with the activity of the training set. Halogen is the electronegative group which is able to strongly accept the electron from the nuclei like hydrogen. In conjunction, the halogen also represents the CHI_1 as the consequence that the ligand connects to the presence of the series of fluorinated methane. The activity of the new models was also predicted by positive correlation with the fractional as well as relative hydrophobic atomic partial charge which can access the solvent. Furthermore, the area of the molecular shadow in the xy plane definitely fulfills the requirement of the FC model.

The virtual screening selected 255 of 1996 of NCI compounds having the IC_{50} (pred) covers from 10 nanomolar to 1 femtomolar. To make it more comparable with the designed ligand above, we select randomly the NCI ligands which have IC_{50} at the range 0 to 10 nM as tabulated in Table 4 (supporting informations). As observed, the four selected NCI ligands (see Figure 5) represents the particular functional group corresponds to that of the new designed ligand such as quinoline, sulfonamide, organohalogen, halogen and carbonyl moieties. Definitely, all of those selected NCI ligands fulfilled the requirements formulated by the FC QSAR model to have the IC_{50} (pred) better than the initial synthesized quinoline derivatives.

Binding Mode Prediction

The control docking shows the similarity pose between the initial pose and the pose of tetrapeptide inhibitor re-docked to DENV2 NS2B-NS3pro. As illustrated in Figure 1 (supporting information), the main chain structure was almost preserved and only at the side chain of the tetrapeptide deviated from the initial structure which is indicated by the RMSD value of 1.4 Å. Roughly, the docked conformation still occupied nicely in the pocket. The major deviation between the docked conformation and the initial structure was at side chain 3 and 4 positions of the peptide. It was because P3 and P4 pockets were shallow that make the flexible ligand rather hard to adopt the favorable conformation.

In the designed ligands, the free energy of binding is in the range of −7 to −10 kcal/mol indicates the sufficient affinity between the ligands and the protein, respectively. Most of them show contact residues in the P1 pocket as well as the P2 pocket while a few of them also shows the interaction with P4 pocket of the protease. As listed down in Table 5 (supporting information), the amino acid residues which commonly present in the complex are Asp129, Ser135, Tyr150 and Tyr161 of NS3 at the P1 pocket, and Asp75 and Asn152 at the P2 pocket. Although P4 pocket is considered shallow but two of ligands are able to adopt their conformations in this cavity by interacting with Val154.

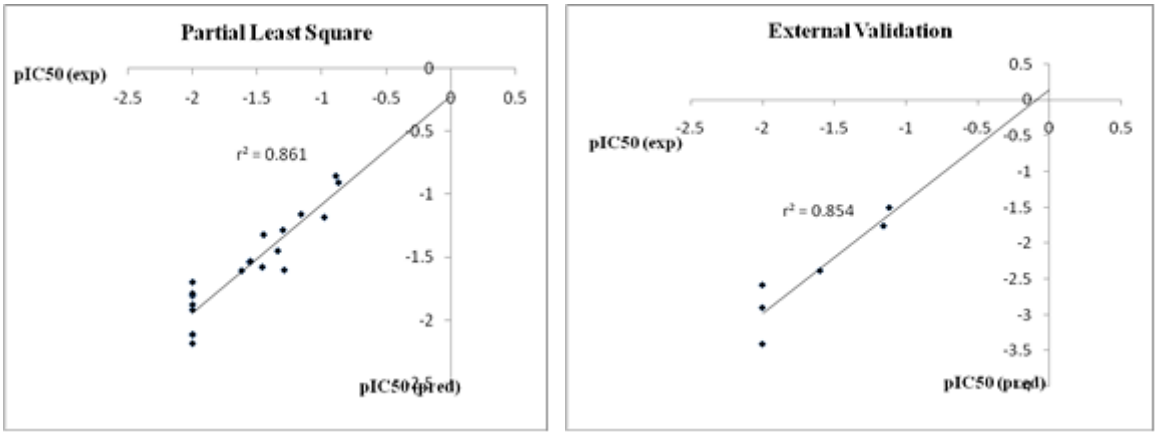


Figure 3. The graph of PLS analysis for training set and external validation for test set.

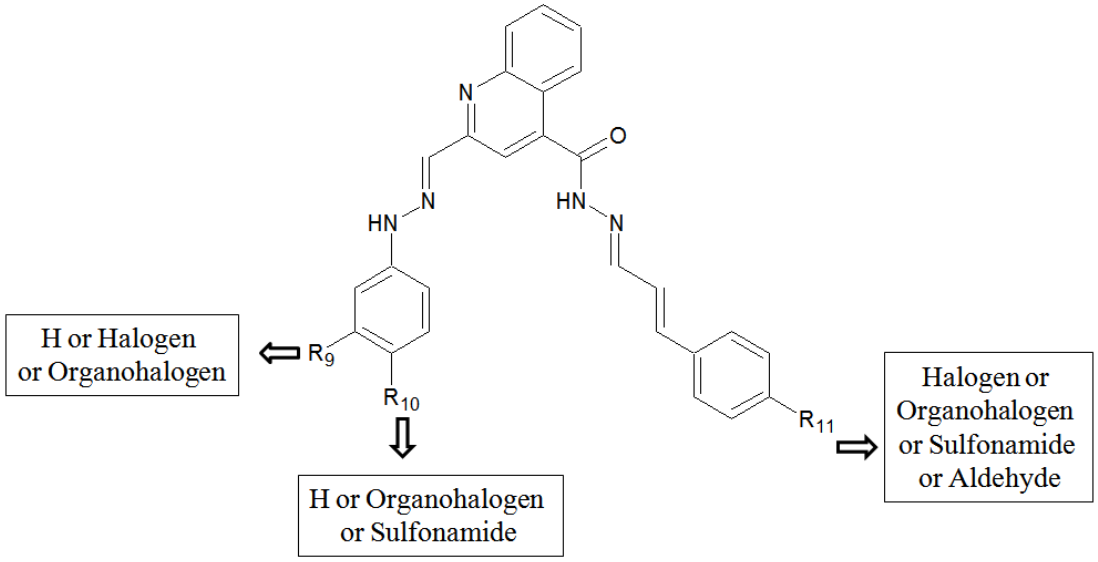
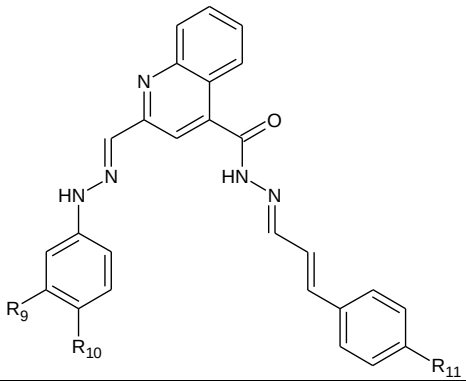


Figure 4. The design strategy of the new quinoline derivatives as DENV2 NS2B/NS3pro inhibitors.

Table 3. The $IC_{50\,(p)}$ data of the new DENV2 NS2B/NS3pro models.



Name	R_9	R_{10}	R_{11}	pIC50	IC_{50} (nM)
Model_1	CCl_3	H	COH	2.2861	5.1755
Model_2	Br	H	COH	1.8721	13.4246
Model_3	H	CF_3	SO_2NH_2	1.5903	25.6850
Model_4	H	CH_2Cl	SO_2NH_2	1.2070	62.0926
Model_5	CF_3	H	COH	1.1663	68.1930

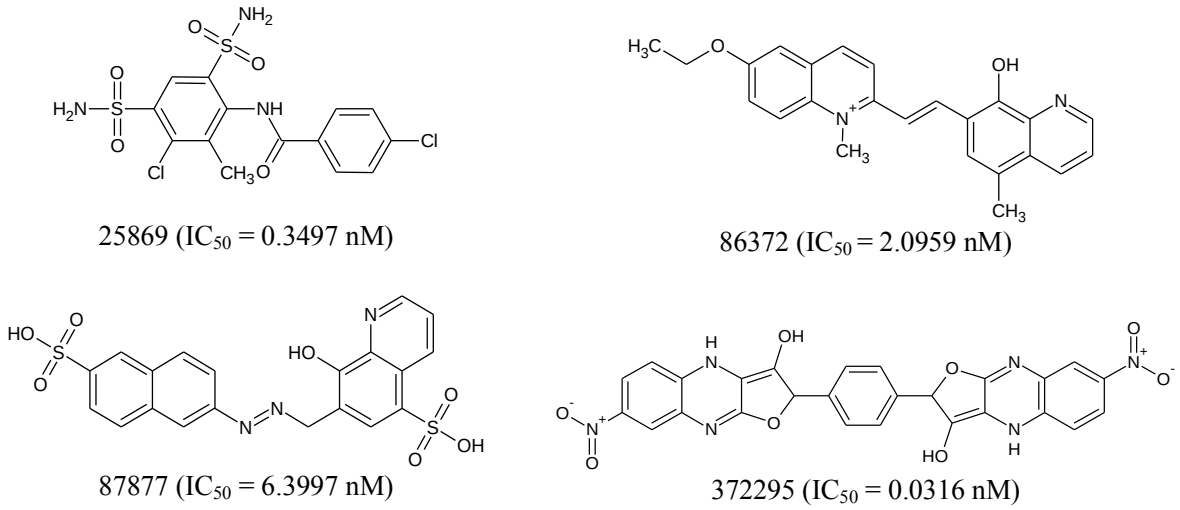


Figure 5. The structure of four selected NCI ligands and its prediction of IC_{50} screened out by the FC model.

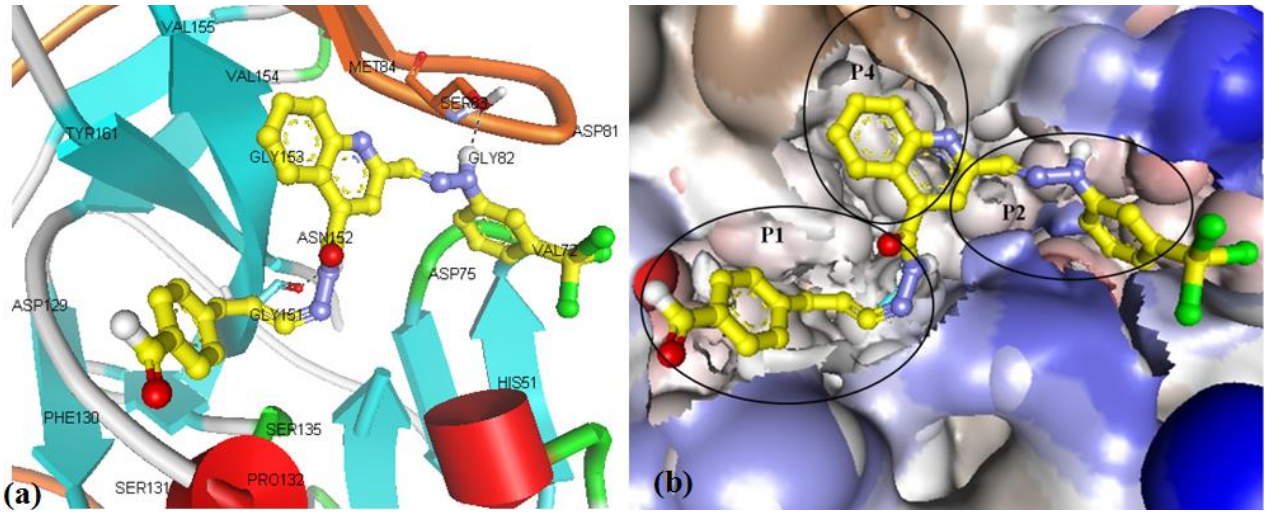


Figure 6. The intermolecular interaction between Model_1 with the protease in (a) schematic form and (b) in a surface form. In surface form, the blue area represents the hydrophobic side chain; brown areas are the hydrophilic side chain while white area represents the neutral side chain accordingly. The picture was visualized using Discovery Studio Client 2.5.

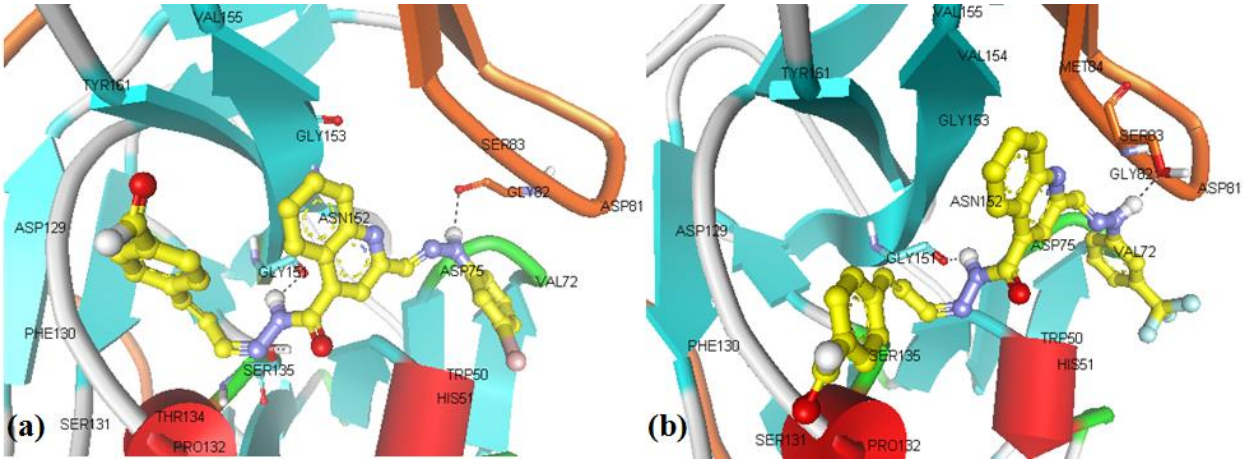


Figure 7. The intermolecular interaction between (a) Model_2 and (b) Model_5 with the protease in schematic form

Model_1 was found having hydrogen bond interaction between the H of amide with the carbonyl oxygen of Gly151 of NS3 and the H of hydrazone linker with hydroxyl oxygen of Gly82 at NS2B (see Figure 6). The most van der Waals interaction was observed at the imine bond while the hydrophobic interaction occurs when the quinoline moiety occupy the P4 pocket by interacting with Val154. On the other hand, the trichlorobenzene is projected toward hydrophobic region of P2 pocket by interacting with Val72. In the schematic form, the NS3 is presented in blue blade except Pro132 and His51 which are in red tubes while NS2B is in orange color. The ligands are presented in a ball and stick form employing carbon is yellow, hydrogen is white, nitrogen is steelblue and sulfur is darkorange. The black dotted line represents the hydrogen bond interactions. The picture was visualized using Discovery Studio 2.5.

As observed, Model_2 and Model_5 (see Figure 7) have the canonical binding modes with Model_1 but the predicted IC50 of Model_1 is approximately 14 times lower than Model_5 while it is 2 times lower than Model_2, respectively. It could be impacted by the more steric shape of trichloromethane in Model_1 seems like carrying a hydrophobic contact more effective than trifluoromethane in Model_5 as well as bromine in Model_2. However, those three ligands are still predicted having IC_{50} in nanoMolar ranges.

Either Model_3 or Model_4 shows more H-bonds interaction with the protease than the previous ligands (see Figure 8). In a similar way, most of van der Waals interaction are possessed by aliphatic unsaturated bond linker while the hydrophobic interactions are possessed by quinoline moieties at P4 pocket as well as the halobenzene in the hydrophobic side chain of Val72 in P2 pocket. Model_3 posses H-bonds interactions i.e. the H of sulfonamide with Phe130 and the O sulphonyl with Tyr150 of NS3 at P1 pocket as well as N of imine with Asn152 of NS3 and H of hydrazone with Gly82 of NS2B at P2 pocket. Interestingly, Model_4 shows H-bond interaction between H of sulfonamide with Asp129. This is kindly agreed with the hypothesis suggested by Frecer and co-worker in 2010, that Asp129 in P1 pocket was the key residue to reveal the substrate recognition. Furthermore, the interaction between N of sulfonamide and H of Ser135 supports that the predicted IC_{50} of Model_4 was also contributed by the interaction of the ligand with this one of catalytic site of the protease. To be pointed out, even though the protein used in molecular docking is different with the published work by Deng et al., 2011 but some molecular interactions were equally performed to Asn152 and Val154, respectively. The substitution of EWG at the para position of phenylbutadiene moieties of the parent molecule especially in Model_3 and Model_4 increase the H-bond interaction significantly with the protease which is agree with the QSAR model. Overall, all the selected designed ligands exhibit the similar poses in P1, P2 and P4 pocket which gain the insight activity of the corresponding ligands towards DENV2 NS2B/NS3pro.

In a similar way, five selected NCI ligands exhibit the docked poses as well as representative molecular interactions toward DENV2 NS2B/NS3pro. As elucidated, Diversity25869 was docked to P1 pocket interacting with Tyr150, Phe130, Asp129, His51 and Ser135 via hydrogen bonding (see Figure 9a). Three particular functional groups which commonly present in the previous designed ligand such as sulfonamide, amide and chlorine were also presented in this ligand. The sulfonamide group in this ligand shows the same pattern with the Model_3 and Model_4 which interacts with Asp129 and Phe130, respectively. The amide group intensively interacts with two catalytic sites i.e. His51 and Ser135 whereas the chlorine group contacts with the surrounding nonpolar residues. In these complexes, the protein is also presented in a schematic form. The NS3 is presented in maroon blade while NS2B is in steelblue color. The ligands are presented in a ball and stick form employing carbon is yellow, hydrogen is white, nitrogen is steelblue and sulfur is darkorange. The black dotted line represents the hydrogen bond interactions. The picture was visualized using Discovery Studio 2.5.

Diversity86372 has a quinoline scaffold, thus, the activity of this ligand was also in conjunction with the previous designed ligand (see Figure 9b). The hydrogen bonding interaction of this ligand was possessed by the hydroxyl group by contacting with His51, Ser135 and Gly151. Interestingly, the oxygen of the methoxy group strongly interact with Val155 via hydrogen bonding with a distance 1.75 Å which makes the hydrophobic part of the methoxybenzene closer to interact with hydrophobic side chain in P4 pocket, accordingly.

Diversity87877 show the more hydrogen bond character to interact with the protease (see Figure 10a). Sulfanilic acid group is a strong electron withdrawing group which interacts with Arg54 but on the other hand, this kind of functional group also acts as electron donating group by interacting with Asp129. The diazo group (-N=N-) strongly binds the two catalytic sites of the enzyme, i.e. His51 and Ser135, respectively. Again, the hydroxyl group in the quinoline ring interacts with Ile36 which causes the hydrophobic part of this particular quinoline ring next to the hydrophobic chain of Ile36.

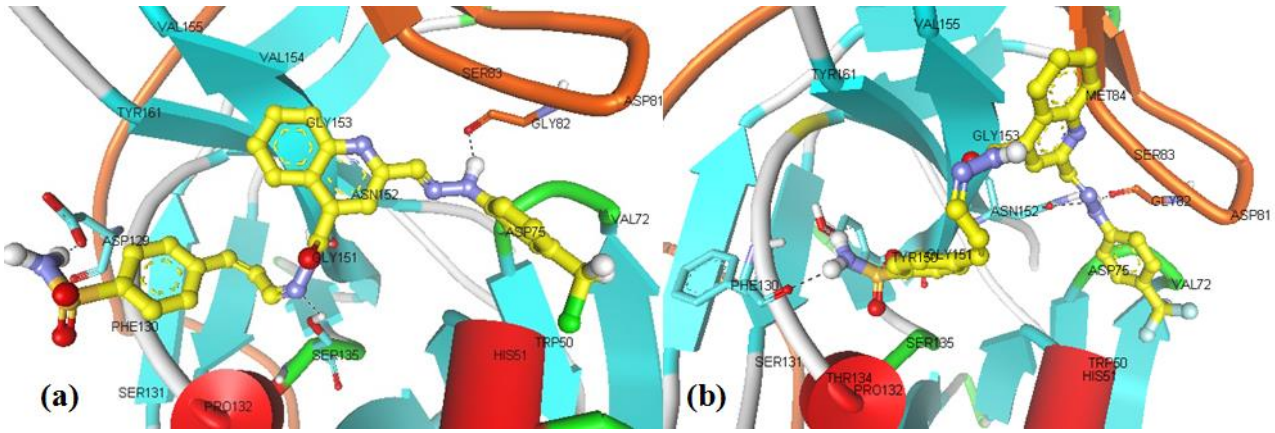


Figure 8. The intermolecular interaction between (a) Model_4 and (b) Model_3 with the protease in schematic form.

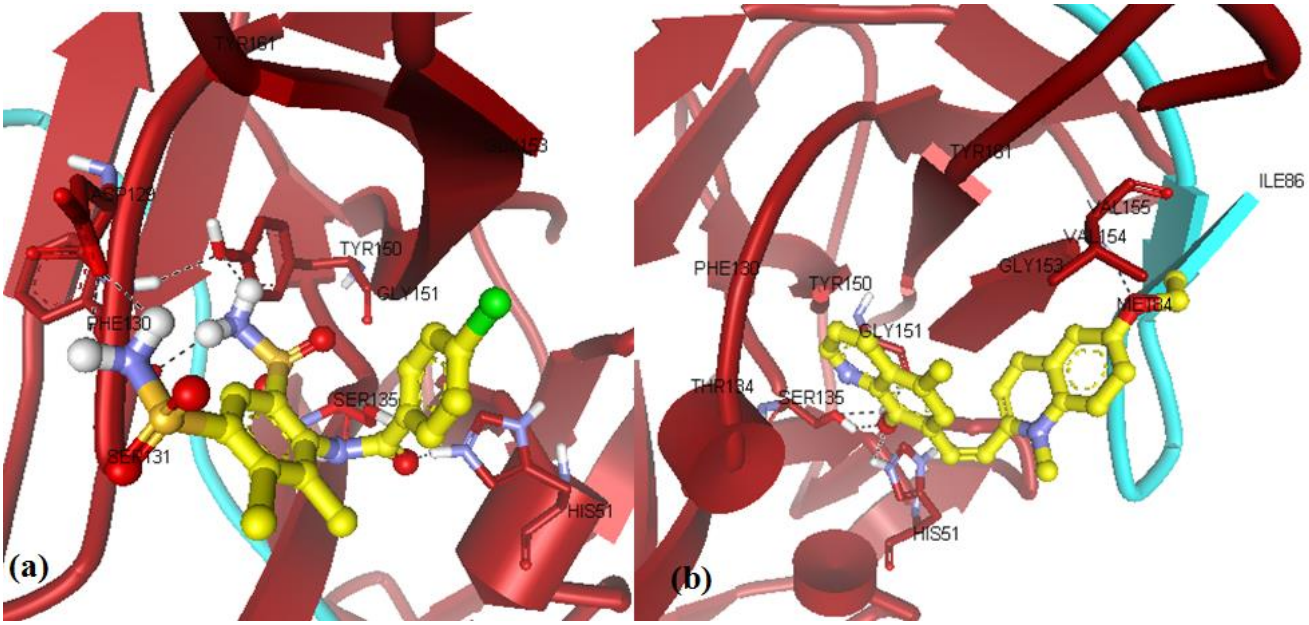


Figure 9. The intermolecular interaction between (a) Diversity25869 and (b) Diversity8672 with the protease in schematic form.

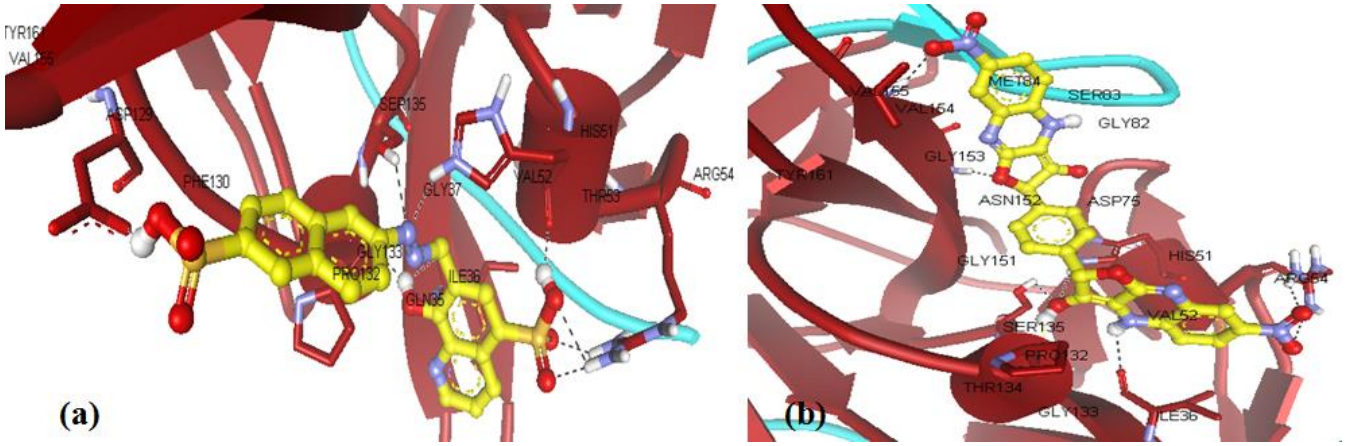


Figure 10. The intermolecular interaction between (a) Diversity87877 and (b) Diversity372295 with the protease in schematic form.

Diversity372295 employs two nitro groups which strongly interact with Arg54 via ionic bonding and Val154, respectively (see Figure 10b). Beside the hydrogen bond interaction constructed by Oxygen of nitro group from the ligand and the hydrogen of the amine group from Val155, this forms an extra hydrophobic interaction between phenyl ring in the P4 pocket. In a proper way, the OH group of the ligand forms the H-bond interaction with two catalytic sites, i.e. His51 and Ser135. Moreover, the heterocyclic oxygen forms H-bond interaction with Gly153 as well as the secondary amine in forming H-bond interaction with Ile36 push the benzene ring next to the hydrophobic side chain of this particular amino acid residue.

Conclusions

The QSAR modeling of quinoline based scaffold for designing the model for DENV2 NS2B/NS3pro inhibitor has been carried out. The QSAR equation generated by GFA algorithm was constructed in goodness and a predictive capability. ALogP, Num_H_Acceptors, Jurs_FPSA_1, CHI_1, Jurs_RASA and Shadow_XY were observed having positive contribution in the model activity toward the protease. Therefore, the attachment of more electron withdrawing group such as halogen and aldehyde were predicted to increase the model activities up to nanomolar concentrations. Beside, the constructed QSAR model was capable to screen out the NCI diversities in such a way that enriches the diverse model for the protease inhibitor. Furthermore, the binding mode of all models were observed exhibiting the ligand-amino acid residues as commonly studied such as Ile36, His51, Ser135, Asp129, Tyr150, Gly151, Gly153 and Val154. Thus, these models could be explored as the further DENV2 NS2B/NS3pro inhibitors with better activities.

Acknowledgement

A great acknowledgement is addressed to Universiti Sains Malaysia for fully facilitating the research as well as the conference publication under Research University Grant for Team (RUT) No. 1001/PKIMIA/855006 with Assoc. Prof. Hasnah Osman as the grant holder. MH specially acknowledges Universiti Sains Malaysia for awarding USM Fellowship (2011-2012) and Malaysian Ministry of Science and Technology Innovation (MoSTI) under IPHARM Grant No. 311/IFN/6926012 for hiring position as research associate.

References

[1] WHO, Dengue and severe dengue, 2013 ; Available from: http://www.who.int/mediacentre/factsheets/fs117/en/index.html.

[2] V. Nair, G. Chi, J. Julander, and D.F. Smee, A heterocyclic molecule with significant activity against dengue virus, *Bioorg. Med. Chem. Lett.*, **19**, 2009, 1425-1427.

[3] U.S.F. Tambunan, N. Apriyanti, A.A. Parikesit, W. Chua, and K. Wuryani, Computational design of disulfide cyclic peptide as potential inhibitor of complex NS2B NS3 dengue virus protease, *Afr. J. Biotechnol*., **10**, 2011, 12281-12290.

[4] K. Irie, P.M. Mohan, Y. Sasaguri, and R. Putnak, Sequence analysis of cloned dengue virus type 2 genome (new guinea-c starin), *Gene*, **75**, 1989, 197-211.

[5] D.A.S. Ray, Recent advances in flavivirus antiviral drug discovery and vaccine development, *Recent Pat Antiinfect Drug Discov*., **1**, 2006, 45-55.

[6] S.L. Alcaraz-Estrada, M. Yocupicio-Monroy, and R.M. de Angel, Insights into dengue virus genome replication, *Future Virol.*, **5**, 2010, 575-592.

[7] B.B. Folly, A.M. Weffort-Santos, C.G. Fathman, and L.B.R. Soares, Dengue-2 structural proteins associate with human proteins to produce a coagulation and innate immune response biased interactome, *BMC Infect. Dis*., **11**, 2010, 1-12.

[8] S.A.P. Noisakran, Alternate hypothesis on the pathogenesis of dengue hemorrhagic fever (DHF)/ dengue shock syndrome (DSS) in dengue virus infection, *Exp. Biol. Med*., **233**, 2008, 401-408.

[9] D. Luo, T. Xu, C. Hunke, G. Gruber, S.G. Vasudevan, and J. Lescar, Crystal structure of the NS3 protease-helicase from dengue virus, *J. Virol.*, **82**, 2008, 173-183.

[10] N.H. Mueller, *et. al.*, Identification and biochemical characterization of small-molecule inhibitors of west nile virus serine protease by a high-throughput screen, *Antimicrob. Agents Chemother.*, **52**, 2008, 3385-3393.

[11] N. Frimayanti, S.M. Zain, V.S. Lee, H.A. Wahab, R. Yusof, and N.A. Rahman, Fragment-based molecular design of new competitive dengue Den2 Ns2b/ Ns3 inhibitors from the component of fingerroot (Boesenbergia rotunda), *In Silico Biol,* **11**, 2012, 29-37.

[12] Z. Yin, S.J. Patel, W.L. Wang, G. Wang, W.L. Chan, K.R.R Rao, J. Alam, D.A. Jeyaraj, X. Ngew, V. Patel, D. Beer, S.P. Lim, S.G. Vasudevan, and T.H. Keller, Peptide inhibitors of dengue virus NS3 protease, Part 1: Warhead, *Bioorg. Med. Chem. Lett*., **16**, 2006, 36-39.

[13] Z. Yin, S.J. Patel, W.L. Wang, W.L. Chan, K.R.R. Rao, G. Wang, X. Ngew, V. Patel, D. Beer, J.E. Knox, N.L. Ma, C. Ehrhardt, S.P. Lim, S.G. Vasudevan, and T.H. Keller, Peptide inhibitors of dengue virus NS3 protease, Part 2: SAR study of tetrapeptide aldehyde inhibitors, *Bioorg. Med. Chem. Lett*., **16**, 2006, 40-43.

[14] S. Chanprapaph, P. Saparpakorn, C. Sangma, P. Niyomrattanakit, S. Hannongbua, C. Angsuthanasombat, and G. Katzenmeier, Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites, *Biochem. Biophys. Res. Commun*., **330**, 2005, 1237-1246.

[15] L.Y. Kee, T.S. Kiat, H.A. Wahab, R. Yusof, and N.A. Rahman, Nonsubstrate based inhibitors of dengue virus serine protease: a molecular docking approach to study binding interactions between protease and inhibitors, *AsPas J. Mol. Biol. Biotechnol.*, **15**, 2007, 53-59.

[16] P. Niyomrattanakit, P. Winoyanuwattikun, S. Chanprapaph, C. Angsuthanasombat, S. Panyim, and G. Katzenmeier, Identification of residues in the dengue virus type 2 NS2B cofactor that are critical for NS3 protease activation, *J. Virol*, **78**, 2004, 13708-13716.

[17] C. Steuer, K.H. Heinonen, L. Kattner, and C.D. Klein, Optimization of assay conditions for dengue virus protease: effect of various polyols and nonionic detergents, *J. Biomol. Screen*., **14**, 2009, 1102-1108.

[18] T.S. Kiat, R. Pippen, R. Yusof, H. Ibrahim, N. Khalid, N.A. Rahman, Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoides of fingerroot, Boesenbergia rotunda (L.) towards dengue-2 virus NS3 protease, *Bioorg. Med. Chem. Lett*., **16**, 2006, 3337-3340.

[19] C. Steuer, C. Gege, W. Fischl, K.H. Heinonen, R. Bartenschlager, and C.D. Klein, Synthesis and biological evaluation of α-ketoamides as inhibitors of dengue virus

protease with antiviral activity in cell culture, *Bioorg. Med. Chem.*, **19**, 2011, 4067-4074.
[20] C. Nitsche, C. Steuer, and C.D. Klein, Arylcyanoacrylamides as inhibitors of the Dengue and West Nile virus proteases. *Bioorganic & Medicinal Chemistry*, **19**(24), 2011, 7318-7337.
[21] S.M. Tomlinson and S.J. Watowich, Anthracene-based inhibitors of dengue virus NS2B-NS3 protease, *Antiviral Res*, **89**, 2011, 127-135.
[22] J. Deng, N. Li, H. Liu, Z. Zuo, O.W. Liew, W. Xu, G. Chen, X. Tong, W. Tang, J. Zhu, J. Zuo, H. Jiang, C.G. Yang, J. Li, and W. Zhu, Discovery of novel small molecule inhibitors of dengue viral NS2B-NS3 protease using virtual screening and scaffold hopping, *J. Med. Chem*., 2011.
[23] S. Aravapalli, H. Lai, T. Teramoto, K.R. Alliston, G.H. Lushington, E.L. Ferguson, R. Padmanabhan, and W.C. Groutas, Inhibitors of dengue virus and west nile virus proteases based on the aminobenzamide scaffold, *Bioorg. Med. Chem*., **20**, 2012, 4140-4148.
[24] H. Lai, G.S. Prasad, and R. Padmanabhan, Characterization of 8-hydroxyquinoline derivatives containing aminobenzo-thiazole as inhibitors of dengue virus type 2 protease in vitro. *Antiviral Res*., **97**, 2013, 74-80.
[25] J.R. Tisoncik, Y. Guo, K.S. Cordero, J. Yu, J. Wang, Y. Cao, and L. Rong, Identification of critical residues of influenza neuraminidase in viral particle release, *Virol. J*., **8**, 2011, 1-14.
[26] K. Wichapong, S. Pianwanit, W. Sippl, and S. Kokpol, Homology modeling and molecular dynamics simulations of dengue virus NS2B-NS3 protease: insight into molecular interaction, *J. Mol. Recognit*., **23**, 2010, 283-300.
[27] N. Frimayanti, S.M. Zain, and N.A. Rahman, Discovering new competitive dengue DEN2 NS2B/NS3 inhibitors using similarity searching, in International Conference on Chemistry and Chemical Engineering (ICCCE) 2010: Kyoto, Japan, 30-34.
[28] V. Frecer and S. Miertus., Design, structure based focusing and in silico screening of combinatorial library of peptidomimetic inhibitors of dengue viruses NS2B-NS3 protease, *J. Comput. Aided Mol. Des*., **24**, 2010, 195-212.
[29] J.M. Luco and F.H. Ferreti, QSAR based on multiple linear regression and PLS methods for the anti-HIV of a large group of HEPT derivatives, *J. Chem. Inf. Comput. Sci.*, **37**, 1997, 392-401.
[30] P.M. Sivakumar, S.K.G. Babu, and D. Mukesh, QSAR studies on chalcones and flavonoids as anti-tuberculosis agents using genetic function approximation (GFA) method, *Chem. Pharm. Bull.*, **55**, 2007, 44-49.
[31] K.F.M. Pasqualoto and M.M.C. Fereira, Molecular modeling and receptor-dependent (RD) 3D-QSAR approach to a set of antituberculosis derivatives, *QSAR Comb. Sci*., **28**, 2009, 1455-1464.
[32] B.O. Al-Najjar, H.A. Wahab, T.S.T. Muhammad, A.C. Shu-Chien, N.A.A. Noruddin, and M.O. Taha, Discovery of new nanomolar peroxisome proliferator activated receptor γ activators via elaborate ligand-based modeling, *Eur. J. Med. Chem.*, **46**, 2011, 2513-2529.
[33] R. Huey and G.M. Morris, Using AutoDock 4 with AutoDockTools: a tutorial., 2008, California, United State of America: The Scripps Research Institute Molecular Graphics Laboratory.
[34] D. Rogers, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, *J. Chem. Inf. Comput. Sci*., **34**, 1994, 854-866.
[35] Discovery Studio 2.5.5: Help, 2010, Accelrys.
[36] S. Thangapandian, S. John, and K.W. Lee, Genetic function approximation and bayesian models for the discovery of future HDAC8 inhibitors. *IBC*, **3**, 2011, 1-11.
[37] L.M. Shi, Y. Fan, T.G. Myers, P.M. O'Connor, K.D. Paull, S.H. Friend, and J.N. Weinstein, Mining the NCI anticancer drug discovery database: genetic function approximation for the QSAR study of anticancer ellipticine analogues, *J. Chem. Inf. Comput. Sci*., **38**, 1998, 189-199.
[38] A. Beheshti, E. Pourbasheer, M. Nekoei, and S. Vahdani, QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm-multiple linear regressions, *J. Saud. Chem. Soc*., 2012.
[39] D.H. Cho, S.K. Lee, B.T. Kim, and K.T. No, Quantitative structure-activity relationship (QSAR) study of new fluorovinyloxyactemides, *Bull. Korean. Chem. Soc*., **22**, 2001, 388-394.
[40] R. Sharma and S.K. Mahmood, An approach towards quantitative structure-activity relationship study for the affinity of diphenyl-pyridines toward CBI receptor, *J. Pharm. Sci. Tech*., **3**, 2011, 477-493.

SUPPORTING INFORMATIONS

Table 1. The selected descriptor data of the training set

No	Name	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Deng_23a	5.96	5	16.18	0.60	0.91	125.47	0.184	444.23	10641.20	152.38	1,633.31	8.14	2.32
2	Deng_23b	5.50	5	15.68	0.59	0.91	117.46	0.192	429.05	9920.87	122.06	1,389.88	7.15	2.44
3	Deng_23c	5.94	5	16.18	0.57	0.91	112.31	0.186	444.23	10641.20	103.02	1,416.02	7.84	2.46
4	Deng_23d	3.38	8	16.79	0.56	0.72	131.66	0.377	474.84	12179.10	97.96	2,920.36	6.43	0.95
5	Deng_23e	4.00	7	17.39	0.60	0.81	129.78	0.318	490.26	12980.80	174.47	3,007.47	8.08	0.60
6	Deng_23f	3.54	7	16.89	0.59	0.79	124.81	0.329	474.84	12174.70	135.77	2,626.25	8.11	1.11
7	Deng_23g	3.98	7	17.39	0.58	0.78	119.88	0.321	490.26	12980.80	115.03	2,558.41	7.70	0.74
8	Deng_23h	6.04	5	16.18	0.59	0.92	126.80	0.18	444.23	10641.20	146.65	1,642.39	7.83	2.27
9	Deng_23i	5.59	5	15.68	0.58	0.91	119.07	0.187	429.05	9920.87	109.49	1,391.78	7.35	2.47
10	Deng_23j	5.17	7	17.09	0.57	0.77	118.71	0.284	474.84	12178.70	94.050	1,921.33	7.71	2.69
11	Deng_23k	6.24	5	17.39	0.56	0.92	128.83	0.177	490.26	12990.00	105.13	2,032.86	8.23	2.42
12	Deng_23l	5.78	5	16.89	0.55	0.91	120.82	0.184	474.84	12184.70	69.11	1,738.72	7.63	2.47
13	Deng_23m	6.21	5	17.39	0.53	0.91	115.56	0.179	490.26	12990.00	49.93	1,770.84	7.89	2.59
14	Deng_24a	4.33	6	16.89	0.62	0.81	137.71	0.297	474.84	12174.70	194.50	2,800.39	6.61	1.09
15	Deng_24b	4.79	6	17.39	0.64	0.83	139.63	0.287	490.26	12980.80	230.98	3,007.60	7.68	0.49
16	Deng_24c	6.38	4	15.68	0.62	0.93	132.43	0.155	429.05	9920.87	182.14	1,538.46	6.97	2.27
17	Deng_24d	6.83	4	16.18	0.63	0.93	135.17	0.149	444.23	10641.20	201.04	1,654.50	6.62	2.25
18	Deng_24e	6.29	4	15.68	0.63	0.93	130.07	0.159	429.05	9920.87	193.11	1,548.39	6.87	2.26
19	Deng_24g	5.26	6	16.59	0.64	0.80	135.14	0.24	459.50	11401.70	215.83	1,974.27	6.73	1.97
20	Deng_24h	5.71	6	17.09	0.65	0.81	144.12	0.231	474.84	12178.70	235.54	2,102.08	6.47	2.00

A = AlogP, B = Num_H_Acceptors, C = CHI_1, D = Jurs_FPSA_1, E = Jurs_RASA, F = Shadow_XY, G = Molecular_FractionalPolarSurfaceArea, H = V_ADJ_mag, I = V_Dist_mag, J = Jurs_DPSA_1, K = Jurs_DPSA_2, L = Shadow_Zlength, M = Dipole_X

Table 2. The selected descriptor data of the test set

No	Name	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Deng_1	3.52	9	17.10	0.65	0.81	149.01	0.35	444.23	4,958.02	239.57	2,314.38	4.24	4.59
2	Deng_22a	3.66	7	15.25	0.65	0.86	133.73	0.40	459.5	5,062.94	– 6.91	2,587.29	4.25	34.76
3	Deng_22b	3.76	8	16.67	0.65	0.85	147.42	0.37	521.31	5,330.55	80.80	2,360.06	4.25	2.61
4	Deng_22h	7.06	7	16.77	0.76	0.88	137.20	0.60	254.08	1,849.99	22.38	1,441.54	4.25	2.77
5	Deng_22j	1.38	7	10.84	0.51	0.80	109.84	0.19	459.5	5,062.94	445.80	2,166.13	8.07	2.88
6	Deng_22n	3.32	11	16.77	0.49	0.74	149.81	0.33	429.05	3,907.48	210.30	1,746.18	3.41	3.88

Table 3. The selected descriptor data of the new quinoline derivative models

No	Name	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Model_1	6.568	6	18.33	0.5077	0.89	119.64	0.18	521.31	14,690	11.70	1,958.07	8.83	– 0.88
2	Model_2	5.784	6	17.11	0.58	0.88	112.13	0.20	474.84	12,169	112.99	1,669.32	8.84	– 1.29
3	Model_3	4.924	7	19.00	0.48	0.82	119.89	0.29	552.65	16,499	– 19.01	2,937.63	9.27	0.70
4	Model_4	4.582	7	18.32	0.55	0.82	118.33	0.29	521.31	14,670	77.61	2,716.20	9.39	– 0.16
5	Model_5	5.978	6	18.32	0.54	0.88	114.63	0.20	521.31	14,690	59.96	2,054.02	8.77	– 0.23

Table 4. The selected descriptor data of the NCI diversities models.

No	Name	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Div25869	1.84	5	11.81	0.39	0.70	101.98	0.49	310.76	15.120	−121.95	2,438.37	6.42	−3.86
2	Div86372	5.674	3	13.57	0.74	0.89	116.10	0.12	369.16	15.028	327.22	1,4929.21	6.56	−7.69
3	Div87877	3.717	10	14.53	0.43	0.61	110.38	0.47	413.94	15.354	−84.90	3,174.36	6.94	1.03
4	Div372295	3.077	12	19.26	0.48	0.53	102.76	0.43	600.16	15.876	−29.37	3,447.78	7.29	−0.76

A = AlogP, B = Num_H_Acceptors, C = CHI_1, D = Jurs_FPSA_1, E = Jurs_RASA, F = Shadow_XY, G = Molecular_FractionalPolarSurfaceArea, H = V_ADJ_mag, I = V_Dist_mag, J = Jurs_DPSA_1, K = Jurs_DPSA_2, L = Shadow_Zlength, M = Dipole_X

Table 5. The ligands and their corresponding FEB, amino acid residue-ligand H-bond interaction as well as its distances produced by docking against DENV2 NS2B-NS3pro model, JMR_977_sm.

Ligand	FEB	H-Bond	H-Bond distance
Model_1	– 9.29	Gly82	1.98
		Gly151	2.16
Model_2	– 9.44	Gly82	2.05
		Gly151	2.36
Model_3	– 8.64	Phe130	1.85
		Tyr150	1.95
		Asn152	2.19
		Gly82	1.78
Model_4	– 8.73	Gly82	2.02
		Ser135	2.35
		Gly151	2.01
		Asp129	1.81
Model_5	– 8.39	Gly151	2.12
		Ser83	1.90
25869	– 9.44	Tyr150	1.97; 2.14
		Phe130	1.84
		Asp129	2.03; 2.30
		Ser135	2.05
86372	– 6.71	His51	2.28
		Ser135	1.78; 2.30
		Val155	1.75
		Gly151	1.96
87877	– 8.02	Arg54	2.48; 1.95; 1.82
		His51	1.96; 2.11
		Ile36	2.22
		Pro132	2.36
		Ser135	2.36
372295	– 10.38	His51	2.45
		Arg54	2.20: 1.94; 1.76
		Ser135	2.09
		Gly153	2.14
		Val155	1.98
		Ile36	2.45
		Pro132	1.70

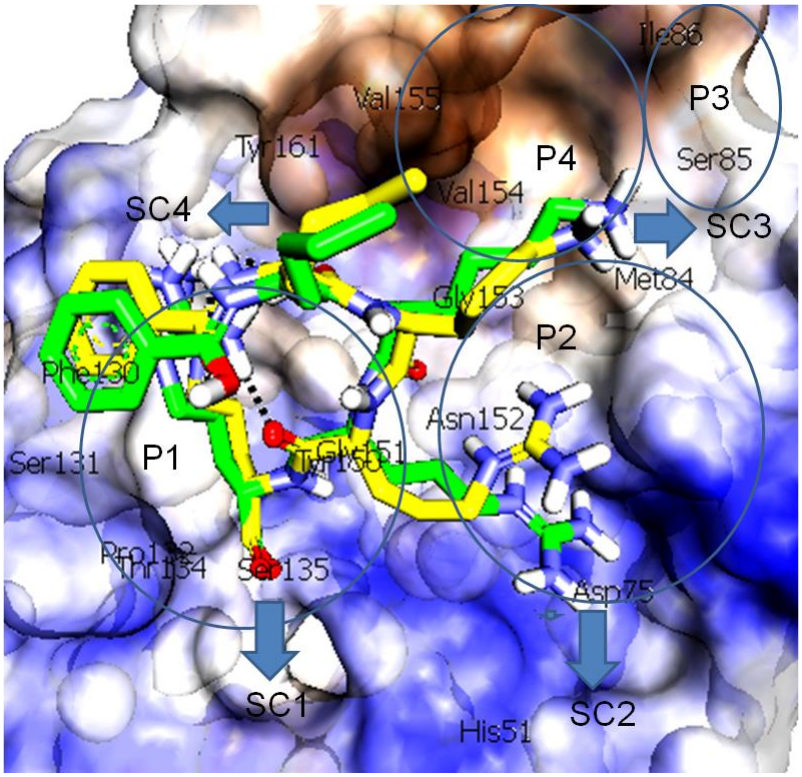


Figure 1. The overlay of an initial pose and a control docking of tetrapeptide inhibitor (NDL1001) to the DENV2 NS2B-NS3pro from the Wichapong model. The NS2B-NS3pro is presented in a surface form. The ligands are presented in a stick form employing carbons are yellow (for initial pose) and green (for control docking pose), hydrogen is white, nitrogen is steelblue and sulfur is dark orange. In surface form, the blue area represents the negative ionizable side chain, brown area the positive ionizable side chain while white area represents the neutral side chain accordingly. P1 = pocket 1, P2 = pocket 2, P3 = pocket 3, P4 = pocket 4, SC1 = side chain 1, SC2 = side chain 2, SC3 = side chain 3, SC4 = side chain 4. The picture was visualized using Discovery Studio Client 2.5.